LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

TRINITY COLLEGE
0 5 AUG 2007
LIBRARY DUBLIN
THESIS
8170

Exploration of a Therapeutic Strategy for Osteogenesis Imperfecta.

A Thesis submitted to the University of Dublin for the Degree of Doctor of Philosophy, by,

Helena McMahon MSc

Department of Genetics

University of Dublin

Trinity College Dublin

October 2006

Declaration

I declare that this thesis is entirely my own work. This thesis has not previously been submitted for a degree in this or any other university. I agree that this thesis may be lent or copied at the discretion of the librarian, Trinity College Dublin.

Helena McMahon

Helena McMahon 31/10/06

Acknowledgements

I would like to thank Dr. Jane Farrar for her support, guidance and patience through out this entire process, from the bench work and throughout the writing of this thesis right through to the tears and upset. Thank you. Sophia, thank you for always being there, to calm the storm, highlight the good results and nudge me in the right direction. I would also like to thank Professor Humphries for his support and encouragement.

I want to thank everyone in the lab, I won't mention you all by name as I'm sure everyone has contributed in some way. Thank you. In addition I would like to thank everyone else in the department in particular Dave and Brenda.

My family have been amazing through out this entire process. Mom, Teresa the phone bills may become a bit more normal now, thank you for everything not just through the time of my PhD studies but for the support and encouragement throughout my education. Also to my brother Sean and Paul, thank you for being there. To all my friends, for not judging, for knowing when to and not ask how the work was going and most importantly for keeping me in touch with the lighter side of life!!!!

One person who gets a special mention is Mandy, for always being there to listen and thanks especially for all the laughs. Finally, keeping the best for last I want to thank my amazing boyfriend, Dave, for his understanding, support, patience, guidance and advice. You were the light in my life when things got difficult, always bringing a smile and a laugh when most needed. I know without you I couldn't have stuck it out and definitely wouldn't be writing this acknowledgement.

Table of Contents

Title page i

Declaration ii

Acknowledgements iii

Table of Contents iv

List of Figures xii

List of Tables xv

Abbreviations xvii

Summary xviii

Chapter 1: General Introduction

1.1 Bone 1

1.1.1 Micro-structure 1

1.1.2 Bone macro-structure 2

1.1.3. Cellular Composition of bone 3

1.1.4 Bone Formation 7

1.2 Mesenchymal Stem Cells 8

1.2.1 The Immunological properties of Mesenchymal stem cells 11

1.2.2 Homing capacity of mesenchymal stem cells 13

1.2.3 Therapeutic applications of Mesenchymal stem cells 14

1.2.4 Mesenchymal stem cell bone therapeutics 15

1.3 Osteogenesis Imperfecta 17

1.3.1 Classification of Osteogenesis Imperfecta 17

1.3.2 Type I Collagen 17

1.3.3 Molecular Pathology of Osteogenesis Imperfecta 20

1.3.4 Management of Osteogenesis Imperfecta 23

1.3.5 Cell based therapies 24

1.3.6 Gene supplementation 26

1.3.7 Gene Targeting 27

1.3.8. Mutation independent gene therapies 28

1.4 RNA Interference 30

1.4.1 Discovery 30

1.4.2 RNAi silencing mechanism 31

1.4.3 Specificity of RNAi based mRNA targeting and cleavage 32

1.4.4 Micro RNAs (miRNAs) 33

1.4.5 miRNA mediated Translational repression or targeted 34

1.4.6 miRNA associated disorders 35

1.4.7 Chromatin effects 35

1.4.9 Applications of RNAi 36

Chapter 2: Material and Methods

2.1 Reagents 39

2.2 Solutions 50

2.3 Primers 52

2.4 DNA Manipulations 55

2.4.1 Oligonucleotide design and synthesis 55

2.4.2 Polymerase Chain Reaction (PCR) 55

2.4.3 Agarose Gel Electrophoresis 55

2.4.4 Gel Extraction using QIAquick Gel Extraction kit (Qiagen) 55

2.4.5 Phenol/Chloroform Extraction of DNA and RNA. 56

2.4.6 Ethanol precipitation 56

2.4.7 Restriction Enzyme Digests 56

2.4.8 DNA Ligations 57

2.4.9 End filling of 5' DNA overhangs using Klenow 57

2.4.10 End filling of 3' DNA overhangs using T4 polymerase 57

2.4.11 Dephosphorylation of DNA fragments using Antartic 57

2. 5 Microbiology 58

2.5.1 Competent Cells 58

2.5.2 Transformation of XL1-Blue E.Coli Cells 58

2.5.3 Transformation of XL-10 Gold ultracompetent cells 58
2.5.4 Plasmid PCR from individual bacterial colonies 59
2.6 Preparation of Plasmid DNA 59
2.6.1 Plasmid Miniprep (M-Stet Boiling Prep) 59
2.6.2 HiYield Plasmid Miniprep (Real Biotech Corporation) 60
2.6.3 Plasmid Midi-Prep (Qiagen High Speed Plasmid Midi -Kit) 60
2.6.4 Maxi preps. (Qiagen High Speed Plasmid Maxi Kit) 61
2.6.5 DNA/RNA quantification 62
2.7 DNA Sequencing 62
2.7.1 ABI Prism Capillary Sequencing 62
2.8 Tissue Culture Techniques 63
2.8.1 Revival of Frozen cell stocks 63
2.8.2 Sub-culturing of mammalian cells 63
2.8.3 Counting mammalian cells 64
2.8.4 Freezing down cell stocks 64
2.8.5 Transfection of HeLa Cells using Lipofectamine 2000 64
2.8.6 Genomic DNA extraction from cultured cells (Gentra) 64
2.9 RNA Techniques 65
2.9.1 RNA Extraction. 65
2.9.2 DNase Treatment 66
2.9.3 RNA extraction using RNeasy RNA extraction kit (Qiagen) 66
2.9.4 Real-Time Reverse Transcription reaction (RT-PCR) 66

Chapter 3: Evaluation of Col1a1 targeting siRNAs in HeLa cells.

3.1 Introduction 69
3.2. Materials and Methods 72
3.2.1 Design of Col1a1 and Non-targeting siRNAs. 72
3.2.2 HeLa cell culture 72
3.2.3 RNA extraction from HeLa cells. 73
3.2.4 Real time RT-PCR analysis endogenous levels of Col1a1 73
3.2.5 Lipofectamine 2000 based siRNA transfection of HeLa cells 74

3.2.6 Quantitative Real time RT-PCR analysis of Col1a1 expression 75

3.3 Results 81

3.3.1 Evaluation of endogenous Col1a1 81

3.3.2 Real time RT-PCR evaluation of Col1a1 expression levels 81

3.4 Discussion 93

Chapter 4: Evaluation of Col1a1 targeting shRNAs in HeLa Cells.

4.1 Introduction. 98

4.2 Materials and Methods 103

4.2.1 shRNA oligonucleotides design. 104

4.2.2 Generation of H1 Bluescript-shRNA constructs. 105

4.2.3 Transfection of HeLa cells with shRNA plasmids. 106

4.2.4 Real time RT-PCR evaluation of Col1a1 expression levels in 106

4.3 Results 110

4.3.1 shRNA construct generation 111

4.3.2 Real time RT-PCR evaluation of the potency of shRNA 111

4.3 Discussion 123

Chapter 5: Evaluation of RNAi in Mesenchymal Stem Cells.

5.1 Introduction 128

5.2 Materials and Methods 133

5.2.1 Mesenchymal stem cell culturing conditions 134

5.2.2 Mesenchymal stem cell transfection using the Amaxa 134

5.2.2.1 Optimisation of Mesenchymal stem cell nucleofection 134

5.2.2.2. Optimised MSC nucleofection protocol 135

5.2.3 RNA extraction from siRNA and shRNA nucleofected 135

5.2.4 Real time RT-PCR evaluation of Col1a1 mRNA expression levels 135

5.2.8 Evaluation of Type I Collagen protein production in MSCs 137

5.3 Results 144

5.3.1 Optimization of Mesenchymal Stem Cell 144
5.3.2 Nucleofection of Mesenchymal Stem Cells with 144
5.3.3 Real time RT-PCR evaluation of the potency of Col1a1 145
5.3.4 Nucleofection of MSCs with Col1a1 targeting 152
5.3.5 Metra CICP ELISA evaluation of de novo Type I collagen 160
5.3.6 Metra CICP ELISA evaluation of d*e novo* type I collagen 161
5.5 Discussion 170

Chapter 6: Exploration of ΦC31 Integrase system for site specific genomic integration of therapeutic nucleic acids.

6.1 Introduction 178
6.2 Material and Methods 187
6.2.1 Mesenchymal Stem Cell Differentiation 187
6.2.1.1 Osteogenic differentiation. 187
6.2.1.2. Adipogenic Differentiation 187
6.2.2 Histological staining of differentiating MSCs 188
6.2.2.1 Alkaline Phosphatase (Sigma Kit No. 85) 188
6.2.2.2 Oil red O Staining 188
6.2.2.3 Microscopic evaluation of MSC differentiation 189
6.2.3 Generation of pEGFP1.attB plasmid. 189
6.2.4 Generation of pEGFP1.CMV.attB plasmid 189
6.2.5 Generation of pEGFP1.AttB.Col1a1 promoter plasmid. 190
6.2.6 Generation of pcDNA.CMV⁻.AttB 191
6.2.7 Generation of shRNA.AttB plasmids. 191
6.2.8 Transfection of HeLa cells with pEGFP1.CMV.attB 191
6.2.9 Selection of HeLa cells which had integrated the 192
6.2.10. Evaluation of plasmid integration sites in HeLa cells. 192
6.2.11 Bacterial plasmid PCR of *E.Coli* colonies generated from 193
6.2.12 Restriction enzyme digestion of plasmids isolated from HeLa 194
6.2.13 Nucleofection of MSCs with pEGFP.AttB.Col1a1 promoter 195
6.2.14 Microscopic Analysis of MSC eGFP expression 195

6.2.15 Selection of MSCs which had integrated the pEGFP.AttB 197
6.2.16 MSC nucleofection with shRNA.AttB and pCMVInt. 197
6.2.17 Selection of MSCs which had integrated the shRNA.AttB 197
6.2.18. Four week shRNA suppression study in MSCs stably expressing 197

6.3 Results 199
6.3.1 Evaluation of the multipotency of a Mesenchymal Stem Cell 199
6.3.1.1 Evaluation of Osteogenic differentiation of MSCs 200
6.3.1.2 Evaluation of Adipogenic differentiation of MSCs 201
6.3.2 Generation of *AttB* containing plasmids 210
6.3.3 Generation of stable HeLa cell lines to assess ΦC31 integrase 210
6.3.4 Retrieval of plasmids integrated into HeLa cell genome using 210
6.3.5 Bacterial plasmid PCR screening of rescued plasmids. 217
6.3.6 Restriction enzyme digestion of rescued plasmids. 217
6.3.7 Sequencing of plasmids selected based upon restriction digest 223
6.3.8 Generation of pEGFP1.AttB.Col1a1 promoter plasmid. 239
6.3.9 Nucleofection of MSC with pEGFP1.AttB.Col1a1 promoter 242
6.3.10 Cloning of shRNA.AttB plasmids 247
6.3.11 Nucleofection of MSCS with shRNA.AttB and pCMVInt 251
6.3.12 Evaluation of shRNA.AttB plasmid integration in MSCs. 255
6.3.13 Evaluation of ΦC31 integrase expression in HeLa cells 256
6.4 Discussion 258

Chapter 7: Generation and testing of Col1a1 replacement gene

7.1 Introduction 267
7.2 Materials and Methods 275
7.2.1 Generation of Col1a1 genomic/cDNA replacement gene 275
7.2.1.1 Sequencing of the Col1a1 Minigene 5' UTR & 3' UTR 275
7.2.1.2 Removal of KpnI restriction site from multiple cloning 276
7.2.1.3. Removal of Col1a1 cDNA from pHUC & generation of 276
7.2.1.4 Addition of Col1a1 3'Untranslated Region (3'UTR) 277

7.2.1.5 Addition of Col1a1 5'Untranslated Region (5 'UTR) 278
7.2.1.6 Removal of CMV promoter from pcdCol1a1 282
7.2.1.7 Sequencing of the entire Col1a1 genomic/cDNA 282
7.2.1.8. Identification of mutations in Col1a1 genomic/cDNA 282
7.2.1.9 Correction of the mutations identified in the Col1a1 283
7.2.2 Culturing of Mov13 cells. 284
7.2.3 Nucleofection of Mov13 cells. 287
7.2.3.1 Optimisation of Mov13 nucleofection protocol 287
7.2.3.2 Fluorescence microscopy of Mov13 nucleofected cells 287
7.2.3.3 Optimised Mov13 nucleofection protocol 287
7.2.4 Generation of Mov13.Col1a1 stable cell lines 288
7.2.5 Real time RT-PCR analysis of Col1a1 expression in Mov13 288
7.2.6 Culturing of Type I OI cell lines 289
7.2.7 Nucleofection of Type I OI cell lines 290
7.2.8 Generation of stable Type I OI cell lines 291
7.2.9 Real time RT-PCR analysis of Col1a1 expression in Type I OI 291
7.2.10 Metra CICP ELISA analysis of type I collagen expression 291
7.2.11 Western Blot analysis of type I collagen expression. 292
7.2.11.1 Protein Extraction from Type I OI fibroblasts 293
7.2.11.2 Bradford Assay 294
7.2.11.3 SDS PAGE electrophoresis 295
7.2.11.4 Antibody probing 296
7.2.11.5 Chemiluminiscent detection 296
7.2.12 Type I collagen Immunofluorescence analysis in Type I OI cells 297
7.2.13 Microscopic analysis of Type I collagen Immunofluorescence in Type I 297
7.3. Results **298**
7.3.1 Generation of Col1a1 genomic/cDNA replacement gene. 299
7.3.2 Sequencing of the Col1a1 genomic/cDNA replacement gene. 299
7.3.3 Evaluation of the timing of the Col1a1 genomic/DNA replace 303
7.3.4 Correction of the mutations which had arisen during generation 303
7.3.5 Optimisation of Mov13 nucleofection. 311

7.3.6 Generation of Mov13 cells stably expressing the Full Length 316
7.3.7 Evaluation of Col1a1 genomic/cDNA replacement gene 316
7.3.8 RT-PCR analysis of Col1a1 expression in Type I OI fibroblasts. 322
7.3.9 Metra CICP ELISA of de novo Col1a1 expression in Type I OI 327
7.3.10. Western Blot analysis of Col1a1 expression in Type I OI 331
7.3.6 Immunofluorescent analysis of Col1a1 protein production 332
7.4 Discussion 336

Discussion and concluding remarks 345
Appendix 1 354
Appendix 2 360
References 373

List of Figures

Figure 1.1 Illustration of compact and spongy bone tissue 5

Figure 1.2 Illustration of a long bone 6

Figure 1.3 Multipotency of Mesenchymal Stem Cells. 10

Figure 1.4 Schematic of Col1a1 and Col1a2 genes 19

Figure 1.5 Assembly of wild type and mutant type I collagen triple helix 21

Figure 1.6 Mechanisms of RNAi mediated gene silencing 37

Figure 2.1 DNA Ladders 41

Figure 2.2 BioRad Kaleilidoscope Protein marker 42

Figure 2.3 Cloning vector pcDNA3.1- 43

Figure 2.4 Cloning vector pcDNA3.1- CMV 44

Figure 2.5 H1-Bluescript Cloning vector 45

Figure 2.6 EGFP1 vector 46

Figure 2.7 pEGFP1-CMV 47

Figure 2.8 pTOPO vector 48

Figure 2.9 pCMVInt vector 49

Figure 3.1 Col1a1 siRNA target sequences, oligonucleotides & duplexes 77

Figure 3.2 siRNA target sequences, oligonucleotides & duplexes 78

Figure 3.3 Col1a1 siRNA target sequences, oligonucleotides & duplexes 79

Figure 3.4 siNon -Targeting sequence oligonucleotides & duplex 80

Figure 3.5 Graph of Col1a1 expression levels in HeLa cells 91

Figure 4.1 Three principle RNAi mechanisms 102

Figure 4.2. ShRNA Oligonucleotide design 105

Figure 4.3 shRNA target sequence and annealed sense and antisense 108

Figure 4.4 shRNA target sequence and annealed sense and antisense 109

Figure 4.5 shRNA bacterial plasmids PCR 114

Figure 4.6. Schematic of H1-shRNA plasmid. 114

Figure 4.7 Chromatogram of Col1a1 targeting shRNA, shCol2 115

Figure 4.8 Chromatogram of Col1a1 targeting shRNA, shGFP 115

Figure 4.9 Chromatogram of Col1a1 targeting shRNA, shCol4 116

Figure 4.10 Graph of Col1a1 expression data in cells transfected with Col1a1 120

Figure 5.1 Type I collagen triple helix. 138

Figure 5.2 Col1a1 siRNA target sequences, sense and antisense oligonucleotides 140

Figure 5.3 Col1a1 siRNA target sequences, sense and antisense oligonucleotides 141

Figure 5.4 shRNA target sequence and annealed sense and antisense 142

Figure 5.5 shGFP target sequence, annealed sense and antisense 143

Figure 5.6. Evaluation of the efficacy of plasmid delivery to MSCs 148

Figure 5.7 Relative Col1a1 mRNA expression levels in MSCs following siRNA 151

Figure 5.8 Relative Col1a1 expression levels in shRNA nucleofected 158

Figure 5.9. De novo Type I collagen protein production in MSCs following 168

Figure 6.1 ΦC31 integrase mediated plasmid integration into the human 182

Figure 6.2 Plasmid rescue technique 195

Figure 6.3 Evaluation of Osteogenic differentiation of MSCs Day 7 202

Figure 6.4 Evaluation of Osteogenic differentiation of MSCs day 14 203

Figure 6.5 Evaluation of Osteogenic differentiation of MSCs day 21 204

Figure 6.6 Evaluation of Osteogenic differentiation of MSCs day 28 205

Figure 6.7 Evaluation of adipogenic differentiation of MSCs day 7 206

Figure 6.8 Evaluation of adipogenic differentiation of MSCs day 14 207

Figure 6.9 Evaluation of adipogenic differentiation of MSCs day 21 208

Figure 6.10 Evaluation of adipogenic differentiation of MSCs day 28 209

Figure 6.11 *AttB* sequence in the sense orientation. 212

Figure 6.12 *AttB* sequence in the antisense orientation 212

Figure 6.13 pEGFP1.AttB plasmid 213

Figure 6.14 Sequence chromatogram generated from pEGFP1.AttB 213

Figure 6.15 Schematic of pEGFP1-CMV.AttB plasmid 215

Figure 6.16 Sequence chromatogram generated from pEGFP1.CMV.AttB 215

Figure 6.17 pcDNA.AttB plasmid 216

Figure 6.18 Sequence chromatogram generated from pcDNA.AttB plasmid 216

Figure 6.19 Bacterial Plasmid PCR of colonies generated from plasmid rescue 219

Figure 6.20 Restriction enzyme digest of plasmids rescue from genomes of 221

Figure 6.21 Restriction enzyme digest of plasmids rescue from genomes of 222
Figure 6.22 Restriction enzyme digest of plasmids rescue from genomes of 223
Figure 6.23 Sequence of AttB site generated from sequencing plasmids 224
Figure 6.24 ClustalW alignment of DNA sequence detected in each of the 225
Figure 6.25 Alignment of consensus HeLa *AttB* DNA sequence insertion 226
Figure 6.26 pEGFP1.AttB.Col1a1 promoter plasmid. 240
Figure 6.27 pEGFP1.AttB.Col1a1 promoter digest with HindIII:BamH1. 241
Figure 6.28 Assessment of Nucleofection efficiencies and eGFP from Col1a1 243
Figure 6.29. Fluorescence microscope analysis of GFP expression in MSCs 244
Figure 6.30. Fluorescence microscope analysis of GFP expression in MSCs 245
Figure 6.31. Fluorescence microscope analysis of GFP expression in MSCs 246
Figure 6.32 Schematic of shRNA.AttB plasmid. 248
Figure 6.33 shCol2.AttB sequence chromatogram 249
Figure 6.34 shCol4.AttB sequence chromatogram 250
Figure 6.35 Sequence of AttB site generated from sequencing plasmids 254
Figure 6.36 Evaluation of ΦC31 Integrase expression in HeLa cells 256
Figure 7.1.1 Col1a1 genomic/cDNA replacement gene cloning step 1: 279
Figure 7.1.2 Col1a1 genomic/cDNA replacement gene cloning step 2: 280
Figure 7.1.3 Col1a1 genomic/cDNA replacement gene cloning step 3: 281
Figure 7.1.4 Cloning design for generation of the Col1a1 genomic/cDNA 300
Figure 7.2 Chromatogram of sequence derived from the Col1a1 genomic 304
Figure 7.3 Chromatogram of sequence derived from pHUC Col1a1 cDNA 304
Figure 7.4 Chromatogram of sequence derived from the corrected Col1a1 305
Figure 7.5 Chromatogram of sequence derived from the Col1a1 genome 305
Figure 7.6 Chromatogram of sequence derived from pHUC Col1a1 cDNA 306
Figure 7.7 Chromatogram of sequence derived from Col1a1 genomic/cDNA 306
Figure 7.8 Sequence derived from the Col1a1 genomic/cDNA replacement 307
Figure 7.9 Sequence derived from pHUC Col1a1 cDNA construct using the 307
Figure 7.10 Mutated Col1a1 genomic/cDNA replacement gene and repair 308
Figure 7.11 Evaluation of Mov13 nucleofection efficiencies 311
Figure 7.12 Evaluation of Mov13 nucleofection efficiencies. 312

Figure 7.13 Evaluation of Mov13 nucleofection efficiencies 313

Figure 7.14 Evaluation of Mov13 nucleofection efficiencies 314

Figure 7.15 Graph of Col1a1 RNA expression levels in Mov13 cells and 318

Figure 7.16 Gel electrophoresis of Col1a1 Real Time RT-PCR cDNA products 319

Figure 7.17 Gel electrophoresis of EEF2 Real Time RT-PCR cDNA products 320

Figure 7.18 Real time RT-PCR evaluation of Col1a1 mRNA expression data 325

Figure 7.19 Metra CICP ELSIA analysis of *de novo* Type I collagen protein 329

Figure 7.20 Western Blot of protein extracted from Type I OI Col1a1 stable 333

Figure 7.21 Evaluation of Co1a1 production by type I OI fibroblasts 334

List of Tables

Table 1.1 Col1a1 and Col1a2 mutations characterised to date in OI 22

Table 3.1 Evaluation of Col1a1 expression levels in MSCs, 293 cells 82

Table 3.2 Sample of relative quantification of RT-PCR data. 86

Table 3.3 GAPDH normalized relative Col1a1 expression values in HeLa 88

Table 3.4 β-Actin normalized relative Col1a1 expression values in HeLa 89

Table 3.5 Table of results generated from Dunns multiple comparisons 92

Table 4.1 Relative Co1a1 expression levels in HeLa cells transfected 117

Table 4.2 Relative Co1a1 expression data from HeLa cells transfected 118

Table 4.3 P values obtained from Dunns multiple comparison 121

Table 4.4 P values obtained from two tailed student t-test 122

Table 5.1 Summary of GAPDH normalised relative Co1a1 mRNA expression 149

Table 5.2 Summary of β-Actin normalized relative Co1a1 mRNA expression 150

Table 5.3 P values generated from comparison of relative Col1a1 expression 156

Table 5.4 Summary of relative Co1a1 mRNA expression levels in MSCs 157

Table 5.5 T-test analysis to determine if the Col1a1 targeting siRNAs 159

Table 5.6 ELISA data based calculation of type I collagen expression in MSC 163

Table 5.7 Type I collagen protein expression levels following nucleofection 165

Table 5.8 Type I collagen protein expression levels following nucleofection 167

Table 5.9. P values generated from t-test comparison of the relative levels 169

Table 6.1 Col1a1 expression levels in MSCs nucleofected with shCol2.AttB 173

Table 7.1 Summary of Real Time RT-PCR expression data from Mov13 317

Table 7.2 Summary of Col1a1 RNA expression data from type I OI 324

Table 7.3 Metra CICP ELSIA analysis of *de novo* Type I collagen protein 328

Abbreviations

AttB	Bacterial attachment site
AttP	Bacteriophage attachment site
CMV	Cytomegalovirus
DNA	Deoxyribonucleic acid
cDNA	Complementary deoxyribonucleic acid
eGFP	Enhanced green fluorescent protein
ELISA	Enzyme-linked immuoscorbent assay
FCS	Fetal calf serum
RT-PCR	Reverse transcription polymerase chain reaction
GAPDH	Glyceraldehyde – 3- phosphate
MSCs	Mesenchymal stem cells
OI	Osteogenesis Imperfecta
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PolII	Polymerase two promoter
PolIII	Polymerase three promoter
RdDM	RNA dependent DNA methylation
RISC	RNA induced silencing complex
RNAi	RNA interference
dsRNA	Double stranded RNA
mRNA	Messenger RNA
miRNA	Micro RNA
shRNA	Short hairpin RNA
siRNA	Small interfering RNA

Summary

The studies described in this thesis focus upon an RNAi-base therapeutic approach for the treatment of Col1A1 linked Osteogenesis Imperfecta (OI). OI is an autosomal dominant negative disorder in which the presence of mutated α1 (I) or α1 (2) chains in the type I collagen triple results in the formation of weak and brittle bones. It is genetically heterogeneous with over 350 mutations in the Col1a1 and Col1a2 genes known to cause OI. As it would not be feasible to design effectors of RNAi to target every OI linked mutation, a mutation independent approach termed suppression and replacement was undertaken, this involves targeting the wild type coding sequence of the Col1a1 gene. This should result in suppression of both Col1a1 alleles; consequently it is likely that a replacement Col1a1 gene would be required to restore the levels of Col1a1 expression to close to wild type levels. This replacement gene would have to be engineered such that it is protected from the Col1a1 targeting RNAi effectors.

Initially 5 Col1a1 targeting siRNAs were evaluated in HeLa cells, each of the 5 siRNAs potently reduced Col1a1 expression at the mRNA level. For a Col1a1 suppression strategy to be effective life long suppression of the mutant transcripts would likely be necessary, therefore selection of siRNA expression plasmids were generated. The design adopted was the expression of short hairpin RNAs (shRNAs) under the control of the H1 promoter. The Col1a1 targeting shRNAs potently reduced Col1a1 mRNA expression in HeLa cells. At this point it was necessary to proceed with the evaluation of the Col1a1 targeting siRNAs and shRNAs in a target cell line for OI, mesenchymal stem cells (MSCs). In this regard it is notable that MSCs can differentiate in a variety of cell types including osteoblasts the bone producing cells of the body. Analysis of Col1a1 expression levels in MSCs at both the mRNA and protein levels following delivery of the siRNAs or shRNAs revealed that Co11a1 expression had been reduced by up to 98%.

Whilst identification of potent suppressors of Col1a1 expression represents an important step in the development of a therapeutic strategy for Col1a1 linked OI, clearly a vector enabling delivery of such nucleic acid based therapies will also be required. Hence, the potential application of the ΦC31 integrase system to a Col1a1 linked OI therapeutic approach was explored, this system is reported to effect site specific integration of exogenous DNA sequences into the human genome. The system was explored in HeLa cells and MSCs and is described in chapter 6 of this thesis. The multipotency of an MSC population was demonstrated with MSC differentiation into both osteogenic and adipogenic lineages. A 4 week Col1a1 targeting shRNA suppression study in MSCs is also described.

The replacement aspect of the therapeutic approach is detailed in chapter 7. It is essential that the Col1a1 replacement gene is expressed in a tissue specific manner at as close to endogenous levels as possible. Thus a wild type col1a1 replacement gene which is a hybrid of genomic and cDNA sequences was successfully generated which expresses efficiently at the mRNA level. This construct was found to restore Col1a1 expression levels to within normal range in a Type I OI cells, both at the mRNA and protein level and the α1(I) chains expressed from the Col1a1 genomic/cDNA construct appeared to be processed in an identical manner to the endogenous α1(I) chains.

Thus the topics explored during my PhD studies and described in the chapters of this doctoral thesis represent a number of elements which may contribute toward the development of potential nucleic acid based therapeutics for Col1a1 linked Osteogenesis Imperfecta. In summary the studies involved the initial validation of Col1a1 targeting siRNAs and shRNAs in HeLa cells, evaluation of the efficacy of these effectors of Col1a1 RNAi in mesenchymal stem cells a potential target cell type for OI therapeutics, the exploration of the ΦC31 integrase system non viral gene delivery system as a potential delivery tool and finally the generation and assessment of a wild type Col1a1 replacement gene in Type I OI fibroblasts.

Abstract

Osteogenesis Imperfecta (OI) is an autosomal dominant negative disorder in which the presence of mutated α1(I) or α2(I) chains in type I collagen results in the formation of weak and brittle bones. Due to the extensive mutational heterogeneity associated with OI a mutation independent RNAi based suppression and replacement strategy for Col1a1 linked OI was explored. A selection of Col1a1 targeting siRNAs and shRNAs were assessed at the RNA and protein level in Mesenchymal stem cells which are a therapeutically relevant cell type for OI, following infusion an autologous corrected MSC population, the MSCs should in theory provide the OI recipient with a life long supply of normal osteoblasts. The Col1a1 targeting suppressors were extremely potent with Col1a1 expression levels reduced by up to 98%. The

ΦC31 integrase system which is reported to effect site specific integration of exogenous DNA into the human genome was explored in MSCs to evaluate its potential application for use in the delivery of Col1a1 linked OI therapeutics Despite the generation of a selection of stable cell lines in the studies employing the ΦC31 integrase system site specific integration was not demonstrated. In relation to replacement a wild type Col1a1 replacement gene was successfully generated, evaluation of the construct revealed that it expressed efficiently at the RNA level, restoring Col1a1 expression to close to wild type levels in a Type I OI cell line. Protein analysis revealed that the Co1la1 RNA was efficiently translated into α1(I) chains, which in turn appeared to be processed in an identical manner to endogenously expressed wild type Col1a1, In conclusion a selection of potent effectors of Col1a1 RNAi in MCSs were identified and a Col1a1 replacement gene which expresses within the endogenous Col1a1 expression range was generated which together lay the foundations upon which further explorations toward a Col1a1 linked OI therapeutic strategy.

Chapter 1

General Introduction

1.1 Bone

1.1.1 Micro-structure

Bones are a complex multifunctional tissue providing movement, mechanical support, organ protection in addition to a site for ion storage and haematopoiesis. Bony tissue consists of two distinct layers an outer layer of hard dense cortical bone and an inner layer of porous cancellous bone. Cortical bone also known as compact bone comprises 80% of the mature skeleton. It consists of precise arrangements of cylindrical structures known as Osteons or Haversian Systems. These units are comprised of calcified matrix arranged in concentric layers known as lamellae. The lamellae surround a central canal (haversian canal) which contains an arteriole, venule and nerve. Lacunae, small fluid filled cavities within which the bone producing osteoblasts reside occur at regular intervals in the lamellae. Small channels called canaliculi radiate from the lacunae to the haversian canal creating a network through the hard matrix. Cancellous or Spongy bone is located deep within bone, it is comprised of trabeculae which have a mesh like arrangement giving the spongy appearance. The spaces are filled with bone marrow and the canaliculi connect to adjacent cavities instead of haversian canals to receive their blood supply (Marieb, 2003) (See Figure 1.1)

1.1.2 Bone macro-structure

The bones of the body come in a variety of shapes and sizes, the four principle types are long (humerous), flat (sternum), short (metatarsals) and irregular (vertebrae). In describing the general structure of a bone the long bone will be used as an example.

The main shaft of the bone is called the diaphysis, it surrounds the medullary cavity which contains the bone marrow and is lined with a thin layer of connective tissue, the

endosteum. On each end of the diaphysis is the epiphysis consisting of spongy bone surrounded by a layer of compact bone. At the junction between the diaphysis and the epiphysis is the cartilginous growth plate, a region of mitotic activity responsible for longitudinal bone growth. A layer of dense connective tissue, the periosteum covers the entire bone except the ends of the epiphysis which are covered in a layer of articular cartilage. The collagenous periosteum is highly vascularised and responsible for appositional growth (Allen, et al., 2004; Sommerfeldt and Rubin, 2001). Figure 1.2

1.1.3. Cellular Composition of bone

Four cell types are found in bone, three of which arise from mesenchymal lineage; osteoblasts, osteocytes and bone forming cells with the final type osteoclasts arising from the haemopoietic lineage. The location of these cells also differs, osteoblasts, osteoclasts and bone lining cells are found on the surface of bone whilst osteocytes are embedded within the interior of the bone (Downey, 2006).

Osteoblasts

Osteoblasts are derived from pluripotent mesenchymal stem cells (MSCs). Osteogenic differentiation proceeds through a number of increasingly committed cell stages to the mature osteoblast. This is regulated by a wide variety of transcription factors such members of the transcription growth factor β (TGF β) and bone morphogenic protein (BMP) families (Aubin, 2001). The primary function of osteoblasts is the production and subsequent mineralization of the bone matrix, which consists primarily of type I collagen (94%) and a small amounts of osteocalcin, osteopontin, bone sialoprotein and matrix gla protein. Signalling molecules such as the BMPs, cytokines and growth factors are also produced and embedded within the matrix (Ross and Christiano, 2006; Sommerfeldt and

Rubin, 2001). This activity is reflected by the presence of a large number of biosynthetic organelles within the osteoblasts (Kartsogiannis and Ng, 2004).

Osteoblasts are polarised cells, secreting the bone matrix unidirectionally toward the bone surface on which they lie, which is subsequently mineralized by the addition of hydroxyapatite like crystals (Downey, 2006; Sommerfeldt and Rubin, 2001). This process is poorly understood, it is thought that mineral laden vesicles are released from the osteoblasts, when saturation is reached the minerals begin to crystallise initiating a mineralisation cascasde (Sommerfeldt and Rubin, 2001). Mature osteoblasts follow one of 3 pathways, 1) remain as an active osteoblast, 2) become surrounded by bone matrix forming an osteocyte or 3) become a relatively inactive bone lining cell. Osteoblasts are anchorage dependent and rely on extensive cell-cell contacts to maintain their responsiveness to metabolic and mechanical stimuli (Sommerfeldt and Rubin, 2001)

Osteocytes

Osteocytes account for 90% of bone cells in the adult skeleton, they are terminally differentiated osteoblasts encased by the bone matrix. Transformation from a motile active osteoblast to an entrapped osteocyte takes about three days. During this time the cell produces a volume three times that of its own cellular volume, as more matrix gets laid down the cells become located deeper within the bone tissue. They eventually reside within lacunae and communicate with each other and with bone lining cells on the bone surface through long intercellular processes which extend through the canaliculi. This cell network allows exchanges with the vasculature, facilitates the detection of mechanical stress and communication with other cells. It is now thought that osteocytes may coordinate the recruitment of bone forming and resorbing cells. (Downey, 2006; Kartsogiannis and Ng, 2004; Knothe Tate, *et al.*, 2004; Sommerfeldt and Rubin, 2001)

Bone lining cells.

These are thin elongated cells which cover most surfaces within the mature skeleton which are neither undergoing resorption nor formation, cytoplasmic extensions often link bone lining cells to each other or osteocytes (Downey, 2006; Kartsogiannis and Ng, 2004). The function of bone lining cells was largely unknown and subject to much speculation but an exciting discovery was made with back to back reports in Nature detailing the involvement of these mesenchymally derived cells in the regulation of proliferation and differentiation of the Haemopoeitic Stem cells (Calvi, et al., 2003; Zhang, et al., 2003).

Osteoclasts

Osteoclasts are terminally differentiated, multinucleated cells formed via the fusion of monocyte/macrophage precursor cells, differentiation into mature osteocytes is thought to be directly influence by osteoblasts (Blair, 2002; Ross and Christiano, 2006). These motile cells are generally located on the bone surface (Downey, 2006). A pivotal player in bone growth, remodelling and healing their functional role is the resorption of the mineralised bone matrix. The plasma membrane of an osteocyte has an infolded appearance known as a ruffled border, these infolds wrap around bone or lie along the surface forming an enclosed area (Karsenty and Wagner, 2002). The osteoclasts pump protons and proteases into these specialized zones creating an acidic milieu which demineralises bone matrix and creates an optimum environment for the enzymatic degradation of the demineralised bone matrix (Blair, 2002; Karsenty and Wagner, 2002). The co-operative action of osteoblasts and osteoclasts in maintaining skeletal architecture is evident from their opposing functionality.

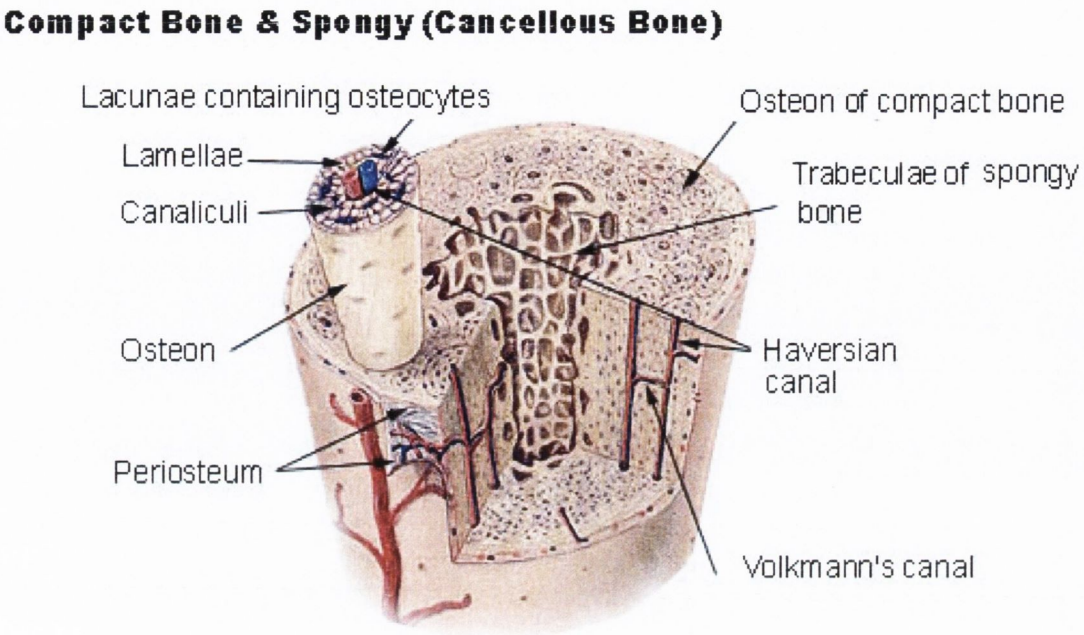


Figure 1.1 Illustration of compact and spongy bone tissue. Compact bone is comprised of osteons, concentric rings of calcified tissue (lamellae) which surround the Haversian canal which contains an arteriole and venule. Haversian canals are linked by smaller Volkmann canals and small channels called canaliculi which radiate outward from the haversian canals creating a network through the hard matrix. Small cavities, lacunae, within which osteoblasts reside, occur at regular intervals within the osteons. Spongy tissue is located deep within bone and has a different structural arrangement. The mineralized bony matrix has a mesh like arrangement - trabeculae which give the spongy appearance. The spaces are filled with bone marrow and the canaliculi connect to adjacent cavities instead of haversian canals to receive their blood supply. The periosteum a highly vascularised layer responsible for appositional growth surrounds the outer surface of the bone (except articulating surfaces). Illustration from *http://en.wikibooks.org/wiki/Image:Illu_compact_spongy_bone.jpg*

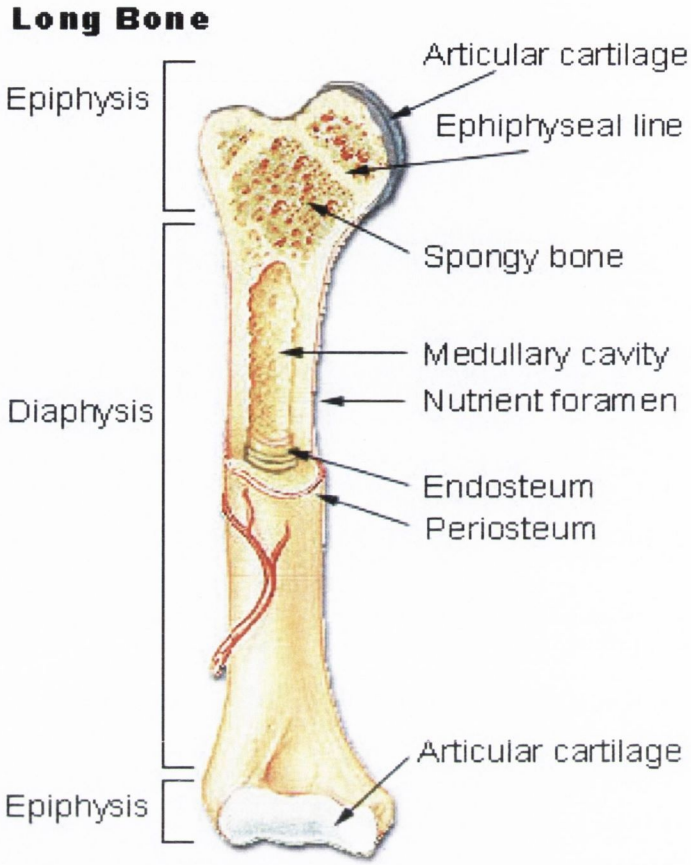


Figure 1.2 Illustration of a long bone. The central shaft of a long bone the diaphysis surrounds the medullary cavity which is lined by the endosteum and contains the bone marrow. Epiphysis cap each end of the diaphysis, they are capped by articulate cartilage and contain spongy bone. The ephiphseal line, also known as the growth plate is comprised of cartilage and is the site of longitudinal bone growth. The nutrient foramen is the site of entry of a blood vessel into the diaphysis.

http://en.wikibooks.org/wiki/Image:Illu_long_bone.jpg

1.1.4 Bone Formation.

There are two distinct mechanisms of bone formation, *Endochondral Ossification* and *Intramembrous Ossification.*

Endochondral Ossification.

In this process a cartilaginous model of the skeleton is formed and replaced by bone. At the onset of skeletal development the mesenchyme (collection of undifferentiated cells derived from the embryonic mesoderm) condense in the locations that will eventually be occupied by the skeleton. These cells then proceed into the chondrocytic pathway, laying down cartilaginous matrix in the shape of the future skeleton (Wagner and Karsenty, 2001). A layer of undifferentiated mesenchymal stem cells remains at the periphery of the cartilaginous template forming the perichondrium (Allen, *et al.*, 2004). The innermost chondrocytes further differentiate into hypertrophic chondrocytes, whilst matrix deposition and mineralization continues (Blair, 2002). The chondrocytes then secrete metalloproteinases which induces angiogenesis (Karsenty and Wagner, 2002). Vascularistion facilitates the introduction of osteoclasts and mesenchymal precursors to the interior of the mineralised cartilage. The osteoclasts remove the mineralised cartilaginous matrix, the mesenchymal stem cells differentiate into osteoblasts which replace the osteoclast degraded matrix with bone matrix which consists primarily of type I collagen. Following vascularisation the hypertrophied chondrocytes die via apoptosis becoming restricted to the growth plate of the bone which together with osteoblasts orchestrates longitudinal bone growth which continues through childhood and adolescence to maturity. A layer of fibrous connective tissue, the periosteum forms and surrounds the outer surface of the bone and which providies a source of osteoblasts to lay down new bone throughout growth and development (Olsen, *et al.*, 2000).

Intramembembrous Ossification

All flat bones form in this manner, in this process the mesenchyme differentiates directly into osteoblasts without a chondrocyte intermediary. The mesenchyme clusters and condenses at the location of the flat bones of the skeleton, this connective tissue becomes highly vascularised which promotes proliferation of the mesenchymal stem cells many of which proceed down the osteogenic lineage and begin to lay down the bone matrix. This matrix is mineralised by the osteoblasts, many of which become entrapped within the hard matrix forming osteocytes. A periosteum also develops during intramembrous ossification, it surrounds the outer surface of the flat bones providing a source of osteoblasts throughout growth and development (Eames, 2003; Karsenty and Wagner, 2002; Olsen, *et al.*, 2000).

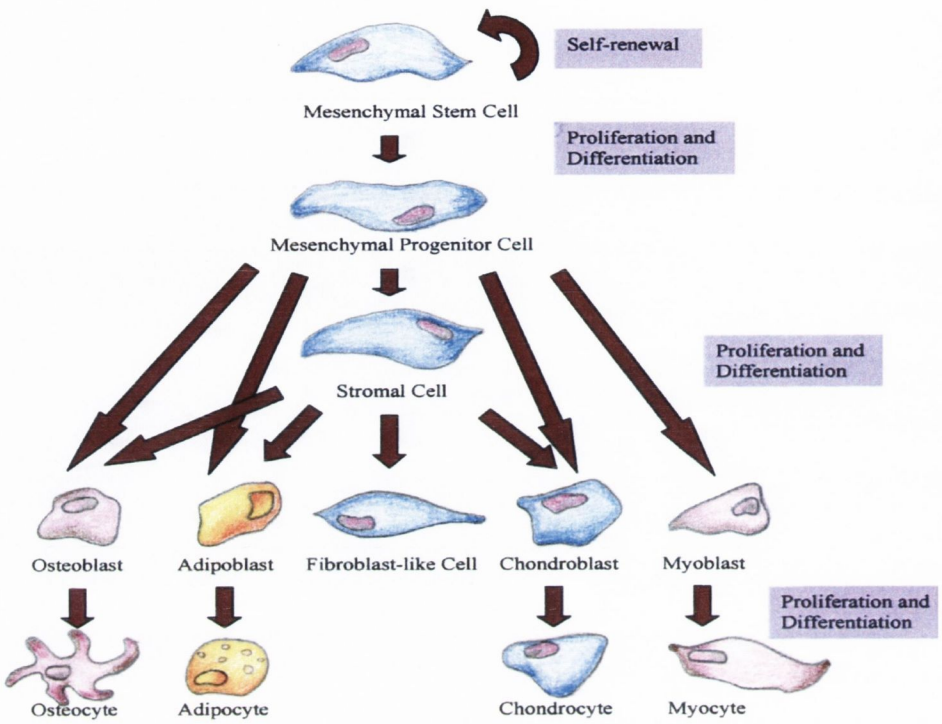
1.2 Mesenchymal Stem Cells

First described 30 years ago by Friedenstein, mesenchymal stem cells (MSCs) have received considerable attention in recent years (Friedenstein, 1966). This is in no doubt due to the realisation of the therapeutic potential of embryonic stem cells and the restrictions put upon them. MSCs are an excellent alternative to embryonic stem (ES) cells, in fact their multipotency and scope of application was initially underestimated. The potential application of MSCs for therapeutic purposes is supported by the fact that they do not appear to possess the tumorigenic potential associated with embryonic stem cells. In addition, the restricted differentiation repertoire of MSCs has further enhanced their utility as they are easier to direct into a specific cell lineage than embryonic stem cells, facilitating more controlled concise tissue engineering (Gafni, 2004).

MSCs are multipotent, self renewing adult stem cells with the capacity to differentiate into a variety of different cell types of non-haemopoietic lineages in response to biochemical stimuli (Figure 1.3). The hallmark of MSCs is their tri-lineage differentiation

potential *in vitro,* that is; the osteogenic, adipogenic and chondrogenic (Pittenger, *et al.*, 1999) lineages. The capacity of these cells to differentiate into neurogenic (Long, *et al.*, 2005), angiogenic (Nagaya, *et al.*, 2004), skeletal (De Bari, *et al.*, 2003) and cardiac muscle cells (Nagaya, *et al.*, 2004; Orlic, 2003) has expanded the scope of the potential clinical applications for MSCs. Traditionally isolated from the bone marrow of a variety of bones, in humans this tends to be the superior iliac crest of the pelvis, MSCs are thought to represent a very small fraction (0.001 - 0.01%) of the total population of nucleated cells in the bone marrow (Pittenger, *et al.*, 1999). However, cultures enriched for MSCs can be generated and expanded with high efficiency. Frequently the cells are fractionated on a Percoll gradient, plated for 2 weeks at which point the MSC population will have adhered to the tissue culture plates, the non-adherent haemopoietic population is then removed leaving behind the adherent MSCs (Gregory, *et al.*, 2005a; Sekiya, *et al.*, 2002).

New insights into stem cells biology have revealed that MSCs reside in a vast number of adult tissues outside of the bone marrow. MSCs with potent differentiation capacities have been isolated from adipose tissue (Dicker, *et al.*, 2005), peripheral blood (Kassis, *et al.*, 2006), muscle (Henry E. Young, 2001), placenta (Zhang, *et al.*, 2006), skin (Henry E. Young, 2001) and synovium (De Bari, *et al.*, 2003). A recent study proposes that all organs contain a mesenchymal stem cell reservoir, (frequently referred to as a niche) as cultures derived from brain, spleen, liver, kidney, lung, bone marrow, muscle, thymus, pancreas have all yielded MSC populations (Meirelles, *et al.*, 2006). It is now widely accepted that adult stem cell reservoirs exist throughout adult life and are central to the functionality of tissues and organs in which there are high levels of cell turnover (Moore and Lemischka, 2006).



(Millington-Ward, *et al.*, 2005)

Figure 1.3 Multipotency of Mesenchymal Stem Cells.

Uncommitted MSCs undergo either self-renewal, maintaining multipotency, or proliferation and differentiation. Multipotent MSCs can differentiate into a variety of precursor cells such as osteoblasts, stromal cells, adipoblasts, chrondroblasts and myoblasts. These cells may differentiate into osteocytes (bone cells), adipocytes (fat cells), chondrocytes (cartilage cells) and myoctes (muscle cells). (Millington-Ward, *et al.*, 2005)

1.2.1 The Immunological properties of Mesenchymal Stem Cells.

MSCs are thought to be hypo- Imunogenic cells as the normal process of allogenic cell rejection does not appear to apply to MSCs in either human or animal models (Ryan, *et al.*, 2005). For example, administration of allogenic MSC prolonged allogenic skin graft survival in a primate model (Bartholomew, *et al.*, 2002). In another study intra-cardiac injection of human MSCs in a porcine model of myocardial infarction resulted in engraftment and differentiation of the cells with preservation of cardiac function. There was no apparent evidence of an immune response to the human MSCs (Makkar, 2005). Horwtiz *et al* reported an improvement in bone quality in patients with Osteogenesis Imperfecta following allogenic MSC transplantation once again without evidence of an immune response (Horwitz, *et al.*, 2002), it is notable that the patients who received the MSC infusion had previously received bone marrow transplants from the same donors (Horwitz, *et al.*, 1999; Horwitz, *et al.*, 2001) . The apparent ability of MSCs to avoid induction of an immune response in recipients is believed to be due to three factors, (1) multipotent MSC do not express HLA class II cell surface markers (Le Blanc and Ringden, 2005), (2) MSCs do not induce T cell proliferation (Bartholomew, *et al.*, 2002; Krampera, *et al.*, 2003), (3) MSCs release cytokines such as IL-10 and prostaglandins which are thought to dampen the immune response (Ryan, *et al.*, 2005).

The unique immunological properties of allogenic MSCs transplantation have sparked huge interest in the potential therapeutic application of MSCs for immune system associated dysfunction. It appears that not only can allogenic MSCs evade the recipients' immune system but they can apparently modulate the behaviour of the hosts' immune system. In one small trial 8 patients with steroid refractory Graft verus host disease (GvHD) received an allogenic MSC infusion. Two patients died and with no evidence of an MSC response, in the 6 remaining patients there was no evidence of toxicity or ectopic tissue formation with complete resolution of the GvHD (Le Blanc and Ringden, 2006). Whilst this is a small study, requiring further exploration and validation, the results are encouraging.

Despite this apparent success there have also been contradicting reports. In one study allogenic MSCs were found to suppress T cell proliferation *in vitro* but the results from the study suggests that subseqeuent to delivery the allogenic MSC populations did not display similar immunosuppressive properties *in vivo* (Inoue, 2006; Sudres, *et al.*, 2006). Another study found that allogenic MSCs were rejected following transplantation in major histocompatability complex (MHC) mismatched mice (Eliopoulos, *et al.*, 2005). This indicates that more rigorous investigations are required to determine the true immuogenicity of MSCs before the development of allogenic MSC based therapeutics.

An inherent difficulty in MSC based studies is that as of yet MSCs have not been definitively characterised. For example, the cell surface marker repertoire of MSCs is subject to conflicting reports and has yet to be fully delineated (Baksh, 2004; Barry and Murphy, 2004; Dazzi, *et al.*, 2006). The techniques employed in MSC isolation and culturing often vary, consequently the cells are heterogeneous in their cell surface markers, homing potential and self renewal capacity. In addition, as the cellular population will most likely vary in their levels of commitment to particular cell lineages, their differentiation capacity will also be heterogeneous. This level of potential diversity between MSC populations can result in conflicting experimental results. Furthermore, the behaviour and activity of the transplanted MSCs will in part be influenced by the biochemical milieu into which they are delivered and engraft.

Without a clear definition of what constitutes an MSC, researchers will have to contend with the variability in reported research results to date. Movreover, once this has been resolved the information will likely only be a definition of the *in vitro* characteristics of the MSCs. These cells undoubtedly behave in a totally different manner *in vivo*. The *in vivo* stem cell milieu consists of a complex network of cells, extracellular matrix and cell to cell signalling environment (Le Blanc and Ringden, 2006). Whilst MSCs are stimulated to divide and proliferate in the artificial *in vitro* tissue culture environment, it is known that MSCs may be quiescent *in vivo* (Baksh, 2004). This is most likely one of the mechanisms by which life long resources of MSCs with extensive self renewal and

differentiation capacity are maintained. It has been shown that MSC which have been extensively cultured loose their differentiation and homing capacities (Javazon, et al., 2004). In fact, the reduction in their self renewal capacity is thought to be central to the process of aging as the regenerative capacity of tissues and organs is thought to diminish with the reduction in vital MSC populations (Sethe, *et al.,* 2006). Given this scenario it would be important to use MSCs of low passage number when developing potential MSC based cell therapies.

1.2.2 Homing capacity of mesenchymal stem cells

MSCs appear to have an innate capacity to populate a variety of tissues and organs following systemic injection (Devine, *et al.*, 2003; Liechty, *et al.*, 2000). MSCs are thought to sense the location of injury, preferentially migrate to that site, engraft, differentiate into the damaged tissue type and contribute to tissue repair and regeneration at that site. MSCs are known to preferentially migrate to the site of injury, for example, to sites of neuronal (Coronel, *et al.*, 2006; Satake, 2003), cardiac (Jiang, *et al.*, 2006; Nagaya, *et al.*, 2004) and bone (Devine, 2002; Shirley, 2005; Taguchi, *et al.*, 2005) damage. The attraction of MSCs to these regions is mediated by the presence of specific signals and chemoattractive molecules in the area of damage (Frank, 2003; Kraus and Kirker-Head, 2006). This homing ability may eliminate the need for invasive delivery of MSCs to target tissues and the potential risks associated with such procedures as systemic infusion may potentially be a sufficient means of MSC delivery. There appears to be no significant difference in the homing capacity of allogenic or autologous MSCs, for example a primate study showed identical tissue distribution 21 months after systemic administration (Devine, *et al.*, 2003).

1.2.3 Therapeutic applications of Mesenchymal Stem Cells.

The range of diseases and clinical conditions that may potentially be treated using MSCs is vast, due to the multitude of positive attributes of these cells. MSCs are a self-renewing cell population readily isolated from a variety of accessible tissues types which are easily expanded and genetically manipulated in culture. Following systemic administration, the cells migrate and engraft in a variety of tissues, with sites of injury being a preferential destination. Once established, the cells should in principle provide a life long source of normal functional cells through their capacity for self renewal and differentiation. Results from studies have shown that MSCs contribute to repair and moreover the transplanted MSCs appear to recruit additional MSCs to sites of injury further potentiating the regenerative response (Barry and Murphy, 2004; De Bari, *et al.*, 2003; Pittenger and Martin, 2004; Rojas, *et al.,* 2005). Finally, the unique immunological properties of the cells have presented the possibility of perhaps using allogenic MSCs based therapeutics. This may have most benefit in acute settings where early intervention may limit tissue damage hence providing the opportunity for optimal recovery, for example in the case of stroke or cardiac dysfunction. The regenerative capacity of MSC has been used in the treatment of cardiac dysfunction (Orlic, 2003; Pittenger and Martin, 2004), neurological diseases (Dezawa, *et al.*, 2005; Kurozumi, *et al.*, 2005), skeletal muscle disorders (Goncalves, et al., 2006), lung injury and disease (Rojas, *et al.,* 2005; Yen, *et al.*, 2006) and renal damage (Morigi, *et al.*, 2004). In a murine model of multi-organ failure systemically administered MSCs were found to migrate to and engraft into a variety of damaged tissues such as intestine, liver, kidney, spleen and contribute to their repair and regeneration (Chapel, 2003)

1.2.4 Mesenchymal stem cell bone therapeutics

One of the most prolific areas of MSC research has been in the field of bone repair and tissue engineering. Whilst osteoblasts are the bone forming cells, these cells have a finite life span, in principle, millions of osteoblasts would be required to repopulate defective bone and secrete the required normal extracellular matrix and repeat dosing would most likely be required. For this reason MSCs with their capacity for self-renewal, proliferation and differentiation into the osteogenic lineage are considered the ideal cell source for treatment of orthopaedic and osteogenic defects (Kofron and Laurencin, 2006). Further impetus toward the use of MSCs has come from the extensive investigation and delineation of the molecular mechanisms of bone repair. It has been revealed that the process of fracture healing and bone repair recapitulates intramembrous and endochondrial bone formation (Dimitriou, *et al.*, 2005; Ferguson, *et al.*, 1999).

Whether systemically or locally administered, MSCs localise primarily on the edges of the osteogenic defect in an organised rather than random manner. Therefore, MSC participation in bone regeneration appears to be directed and co-ordinated by the regenerative biochemical milieu which developed in response to the bone tissue damage (Gafni, 2004; Kraus and Kirker-Head, 2006). Whilst administration of MSCs has itself been shown to be an effective means of enhancing bone repair and regeneration (Pereira, *et al.*, 1998), one of the primary focuses of research to date has been upon the application of MSCs for enhanced bone repair at the sites of non-union factures, this type of fracture is usually unable to heal without surgical intervention (Panagiotis, 2005). In these studies MSCs were delivered impregnated on a scaffold to the fracture site, the MSCs self-renew, many differentiate into osteoblasts laying down the new bone required at the fracture site for repair (Jorgensen, *et al.*, 2004; Meinel, *et al.*, 2005; Yoshikawa, 2005).

In many studies MSCs have been genetically modified to express osteogenic genes (Cheng, *et al.*, 2001; Dragoo, 2003; Gysin, 2004; Tsuda, 2005; Zhang, 2004). The bone morphogenic protein (BMP) family has been extensively investigated as these proteins are potent inducers of osteogenesis (Abe, 2006). Due to its highly osteoinductive properties BMP-2 has been extensively used and shown to contribute substantially to bone formation *in vivo* (Chang, *et al.*, 2003; Cheng, *et al.*, 2001; Cowan, *et al.*, 2005; Dragoo, 2003; Tsuda, 2005). Other BMPs such as BMP 4 and BMP9 have also been used (Aslan, *et al.*, 2006; Dayoub, *et al.*, 2003; Dumont, 2002; Gysin R, 2004; Zhang, 2004) for the *in vivo* promotion of bone formation and regeneration. It has been shown that the use of MSCs to express BMPs is more effective than use of many other cell types in promoting efficient and organised bone regeneration. It appears that expression of the osteogenic BMP transgenes by MSCs affects not only the donor cell and its progeny (autocrine effect) but the BMP proteins expressed by transplanted MSCs also stimulate resident host MSCs (paracine effect) (Gazit, 1999).

1.3 Osteogenesis Imperfecta

Osteogenesis Imperfecta (OMIM 166210) is a congenital brittle bone disorder which affects 1 in 15:000 to 1:20 000 births. Caused by mutations in either the Col1a1 or Col1a2 genes, it is a phenotypically heterogeneous disorder in which severity of bony deformations, height, susceptibility to fracture, deafness and the involvement of other tissue types such as skin, dentin, tendon and ligament varies (Primorac, *et al.*, 2001; Rauch and Glorieux, 2004; Roughley, 2003).

1.3.1 Classification of Osteogenesis Imperfecta.

Sillence classified the disease into four Types I – IV based upon clinical, genetic and radiographic findings, all of which are associated with mutations in either the Col1a1 or Col1a2 genes (Sillence, *et al.*, 1979). Type I, the mildest form of the disease, is characterised by normal height and lack of deformity, Type II in contrast is associated with perinatal lethality and severe deformities. Type III is the most severe surviving form, presenting with severe deformities short stature and high fracture rate and Type IV is moderately severe with moderated deformity, stature and fracture rate. Since this initial classification three new types of OI have been reported, Types V – VII for which no direct association with any type I collagen mutations have been detected (Glorieux, *et al.*, 2000; Glorieux, *et al.*, 2002; Ward, *et al.*, 2002).

1.3.2 Type I collagen.

Of the 27 different types of collagens identified to date, Type I collagen is the most abundant collagen in the body. It accounts for 90% of the organic mass of bone forming a template for mineral deposition and providing the tensile and elastic strength characteristic of healthy bone (Di Lullo, *et al.*, 2002; Niyibizi, *et al.*, 2004; Pace, *et al.*, 2002). Type I collagen is a triple helical heterotrimer consisting of two α1(I) chains and a single α2(I) chain. The Col1a1 gene (AF017178) which encodes the α1(I) chains is located

on chromosome 17q21.3. It contains 51 introns and 52 exons and is approximately 18kb in length. Two mRNA transcripts of differing lengths are transcribed from the Col1a1 gene due to the presence of two poly adenylation sites; both transcripts encode a α1 pro-collagen protein of 1,442 amino acids. The Col1a2 gene (AF004877) which encodes α2(I) chain is located on chromosome 7q22.1 the gene is also comprised of 51 introns and 52 exons and is approximately 38.7kb in length. The Col1a2 gene has 6 different poly adenlyation sites, each of the 6 transcripts expressed from the Col1a2 gene encode a α2 pro-collagen of protein of 1,366 amino acids (Figure 1.4).

Type I collagen is synthesised in a procollagen form which consists of a central $(Gly - X - Y)_n$ amino acid repeat motif flanked by N and C globular propeptides (Figure 1.4).The presence of Glycine, the smallest amino acid at every third amino acid position is a prerequisite for fibril forming collagens (Pace, *et al.*, 2001; Persikov, *et al.*, 2002). A distinctive $(Gly\text{-}Pro\text{-}Hyp)_4$ -Gly-Pro-Pro sequence is found at the C-terminus of the triple-helix domain of the α1(I) and α2(I) chains of type I collagen. This is thought to be a nucleation site at which the three chains come together to initiate formation of the type I collagen heterotrimer. This is followed by propagation which proceeds like a "zipper" from the C to N terminus. The procollagen is secreted from the cell and processed into a mature fibril via propeptide cleavage following which the type I collagen heterotrimers associate to form the type I collagen fibril (Beck, *et al.*, 2000; Persikov, *et al.*, 2002; Prockop, 1998). This process is termed spontaneous self assembly (Prockop, 1998) which requires that all subunits are perfect, consequently mutations are poorly tolerated and the presence of such mutations invariably compromises the functional integrity of the collagen fibril.

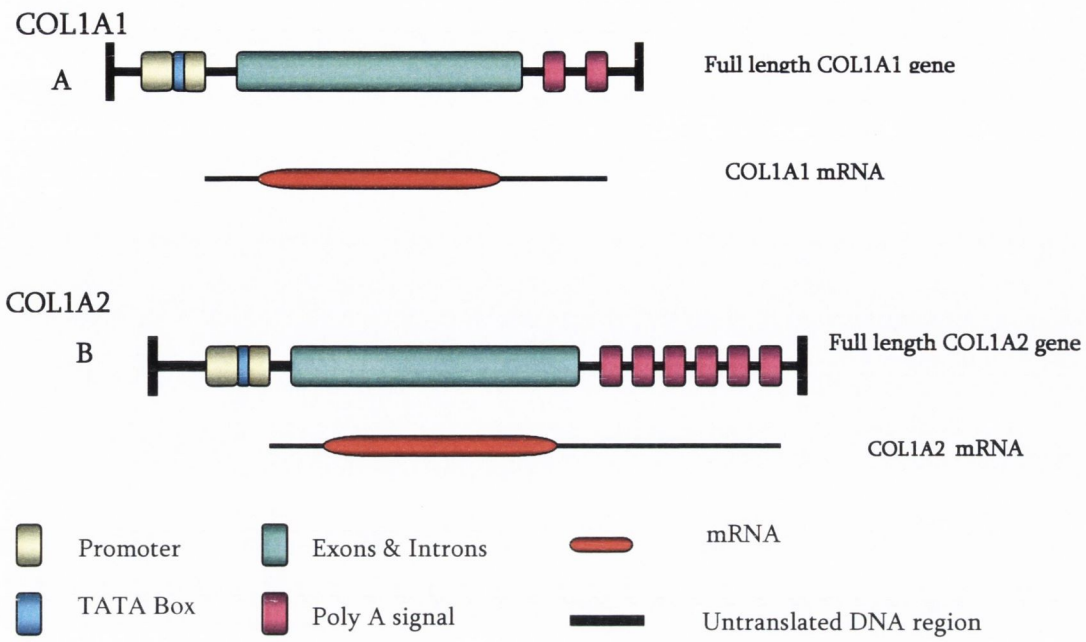


Figure 1.4. Schematic of Col1a1 and Col1a2 genes. A) The Col1a1 gene consists of promoter sequence, 52 exons and 51 introns and 2 Poly A signals in the 3' UTR **B)** The Col1a2 gene consists of promoter sequence, 52 exons and introns and 6 Poly A signals in the 3' UTR. The length of the 3'UTR depends upon the Poly A site used. Col1a1 and Col1a2 can generate 2 or 6 alternative mRNA transcript lengths respectively.

1.3.3 Molecular Pathology of Osteogenesis Imperfecta.

The majority of mutations are due to substitution of the essential Glycine residues in the Gly – X – Y triplet with bulkier amino acids which sterically hinder triple helix formation (Figure 1.5). A variety of other mutation have been identified including deletions, insertions, duplications and exon skipping all of which are detailed in the Collagen mutation database http://www.le.ac.uk/genetics/collagen/, this comprehensive resource is frequently updated and contains all reported mutations, of which over 350 have been characterised and reported to date (Table 1.1). The severity of the OI phenotype depends upon a number of factors; (1) the proximity of the mutation to the C-terminus, most perinatal lethal cases are due to mutations in the C-terminus due to the negative affect on triple helix assembly, (2) the identity of the mutated chain (α1(I) or α2(I) chain), due to the presence of two Col1a1 chains in the type I collagen triple helix, it would be expected that 25% of molecules will be normal and 75% will contain either one or two mutant Col1a1 chains whilst if the Col1a2 allele is carrying the mutation 50% of molecules will be normal and 50% mutated (provided that the half life of the mutant and wild type chains do not differ) and finally (3) in the case of a glycine substitution the identity of the replacing amino acid, directly impacts on the severity of the phenotype, the larger the residue the more disruptive its impact on the structural integrity of the type I collagen fibrils and consequently the more severe the phenotype (Baum and Brodsky, 1999). The majority of OI Col1a1 mutations act in a dominant negative manner; the cells produce both normal and abnormal α1 (I) chains both of which will be incorporated into newly formed triple helices. The abnormal helices are then incorporated into the collagen fibrils.

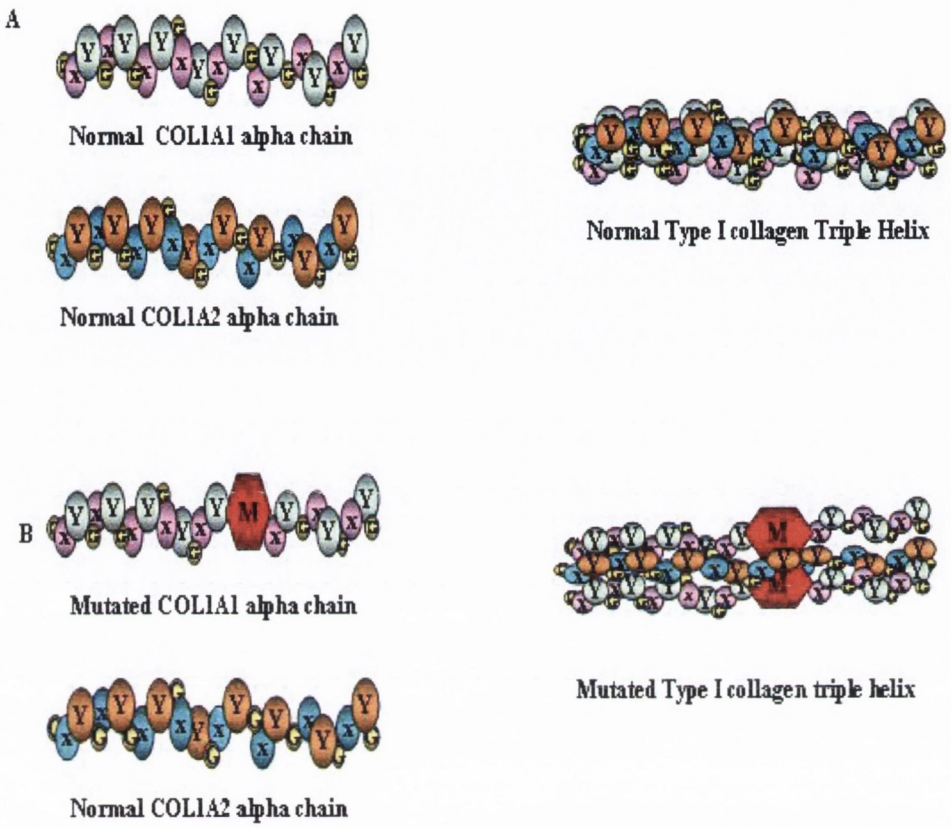


Figure 1.5 Assembly of wild type and mutant type I collagen triple helix. Type I collagen triple helix heterotrimer is comprised of two α1(I) and a single α1(I) chains. Triple helix formation and stability is dependent upon the Gly-X-Y amino acid sequence motif present in the chains, (in which G is glycine, X is frequently proline and Y is hydroxyproline) **A)** Wild type, type I collagen triple helix **B)** The presence of a classic glycine substitution mutation in the α1(I) chain, indicted by the red hexamer, disrupt the formation of the triple helix, which in turn disrupts and weakens the type I collagen fibrils into which it is incorporated leading most likely to the formation of weak bone tissue, resulting in an OI phenotype.

Mutation Type	Number characterized in Col1a1	% OI Types generated	Number characterized in Col1a2	% OI Types generated
Amino Acid Substitutions	137	Type I 22 % Type II 36% Type III 25% Type IV 17%	79	Type I 11 % Type II 22 % Type III 31 % Type IV 36 %
Splice mutations	50	Type I 60 % Type II 11 % Type III 18 % Type IV 11 %	18	Type I 26 % Type II 47 % Type III 11 % Type IV 16 %
Duplications	2	Type III 100%	2	Type III 50 % Type IV 50 %
Frame Shifts	47	Type I 96% Type IV 4 %	1	Type III 100%
Deletions (non-frameshift)	9	Type II 66 % Type III 22 % Type IV 12 %	9	Type I 22 % Type II 33 % Type III 11 % Type IV 33 %
Insertions (non-frameshift)	2	Type II 100 %	None reported	

Table 1.1. Col1a1 and Col1a2 mutations characterised to date in patients with Osteogenesis Imperfecta. Adapted from http://www.le.ac.uk/genetics/collagen.

Biochemical abnormalities also arise as a consequence of the incorporation of the mutant chain in the type I collagen triple helix, these include, increased post-translational modification, decreased thermal stability and disruption of fibrillar architecture, all of which negatively impacts bone flexibility and strength (Di Lullo, *et al.*, 2002; Gajko-Galicka, 2002; Liat Ries-Levavi, *et al.*, 2004). Alternatively, the presence of one or more mutated chains in the triple helix can result in it being targeted for intracellular degradation reducing the amount of secreted protein by 75% (Jones, *et al.*, 2002; Pochampally, *et al.*, 2005). In contrast mutations which generate a premature stop codon most frequently result in Type I OI (Willing, *et al.*, 1996) the truncated mRNA is unstable and most likely targeted for non-sense mediated decay (Byers, 2002). Consequently, only the full length normal transcripts are translated and normal collagen triple helices are produced albeit in a reduced amount.

1.3.4 Management of Osteogenesis Imperfecta.

OI is associated with remarkable clinical diversity, there is little correlation between genotype and phenotype and it is immensely difficult to treat. Traditionally treatment involves a combination of physiotherapy, rehabilitation and orthopaedic therapy (Jones, *et al.*, 2002; Millington-Ward, *et al.*, 2005; Primorac, *et al.*, 2001; Rauch and Glorieux, 2004). These approaches have had limited success, currently the only pharmacological agents that have shown any beneficial effects are the bisphosphonates. In addition to the presence of reduced or abnormal type I collagen, in individuals with OI bones are smaller, thinner and fragile, this is due to the fact that the balance between bone resorption and formation, which normally favours formation is tipped toward resorption (Jones, *et al.*, 2002; Lindsay, 2002). Bisphosphonates are pharmacological inhibitors of bone resorption, they bind to the hydroxyapatite crystal of bone preventing osteoclast action (Roughley, 2003). Extensive studies using Pamidronate have been carried out in children, these agents have resulted in a dramatic reduction in bone pain and fracture number (Frank Rauch, 2005). Despite these successes the optimal treatment regimen and

the consequences of long-term bisphosphonate treatment in children are currently unknown. In addition this therapeutic approach does not address the underlying genetic defect and life long pharmacological intervention will be required.

1.3.5 Cell based therapies.

The basic principle behind cell based therapies is that delivery of a population of normal cells to the affected individual should lead to phenotypic benefit through the expression of normal alleles of the mutated protein in question. The cell population delivered ideally would be stem cell based as these cells should establish a reservoir of normal cells for the life of the patient. Moreover, these normal cells may possibly have a growth advantage over the endogenous mutant cells further potentiating the treatment.

Mosaic OI patients are a naturally occurring OI cell therapy model, these patients present with the very mild disease or lack of symptoms which suggests a degree of type I collagen mutational burden that the body can tolerate (Constantinou, *et al.*, 1990; Edwards, *et al.*, 1992; Wallis, *et al.*, 1990). This suggested that a relatively small increase in the amount of normal collagen available may lead to a significant improvement in bone quality. In fact a recent study has shown that 40% to 75% of osteoblasts may be heterozygous for a Col1a1 mutation without negatively impacting skeletal growth, density or histology (Cabral and Marini, 2004). Whilst osteoblasts are the major type I collagen producers in bone, they are terminally differentiated cells which have arisen from mesenchymal stem cells (MSC). In order for an OI targeted therapy to be long lasting it is MSCs and not the terminally differentiated osteoblasts or even osteoblast precursor cells that would need to be delivered. Initial studies into the feasibility of such an approach to treat OI were carried out in mice. Marrow cells were isolated from Col1a1 minigene transgenic mice which express a truncated form of the human Col1a1 gene, these cells (enriched for MSC's) were infused into the femurs of irradiated mice (Pereira, *et al.*, 1995). The cells were found in the spleen, lungs, brain, cartilage and bone and Col1a1 expression from the

cells was found occur in a tissue specific manner (Pereira, 1995). In a follow up study irradiated minigene mice were the recipients of a population of normal MSC's once again cells were found to have engrafted in various tissues and a small increase in both type I collagen and bone mineral content was observed one month after infusion (Pereira, *et al.,* 1998). Hou *et al* reported similar results in irradiated mice with widespread engraftment of donor cells, differentiation and tissue specific expression, similar results have also been reported to occur in sheep (Mackenzie and Flake, 2001) and non-human primates (Devine, *et al.,* 2001).

MSC's appeared immensely suited to this type of therapy, they are easily isolated, can migrate and engraft into a variety of tissues including bone, express in an appropriate tissue specific manner and retain their differentiation capacities (Barry and Murphy, 2004; Devine, *et al.*, 2001; Mackenzie and Flake, 2001; Pereira, *et al.*, 1998; Pereira, 1995). In addition these cells migrate to sites of injury (Prockop, 2003) and have the unique feature of being non immunogenic, thereby opening up the possibility of allogenic MSC transplantation (Ryan, *et al.,* 2005).

A human clinical trial of allogenic bone marrow transplantation for the treatment of OI has been carried out. Following ablative conditioning therapy three children with Type III OI were intravenously infused with bone marrow cells. Engraftment occurred in all three children with 1.5% – 2.0% of osteoblasts thought to be of donor origin after three months, despite the low level of engraftment increases in bone mineral content, growth and a decrease in fracture incidence was reported (Horwitz, *et al.*, 1999). The short follow up time of 6 months and the lack of proper controls (Horwitz, *et al.*, 2001) somewhat weakens these results. A second study was carried out in which whole marrow cells were administered to 5 children, in the three children with engraftment there was an increase in bone formation and growth rates but this dissipated over time (Horwitz, *et al.,* 2001). A third study was carried out in which 6 children were infused with MSC's, engraftment occurred in 5 of the children, three showed an increase in growth velocity and only one patient had an increase in mineralization (Horwitz, *et al.*, 2002). Despite the difficulty in

both assessing and interpreting the actual clinical benefit, the merit in this approach is clear and warrants further investigation.

1.3.6 Gene supplementation.

Based on the premise that provision of additional functional Col1a1 and Col1a2 alleles may possibly result in phenotype alleviation, gene supplementation strategies have been investigated. The first study was carried out in 1987 by Schienke *et al*, cells derived from the Col1a1 null Mov13 mice were transfected with a functional human Col1a1 transgene resulting in the formation of a cross species heterotrimer consisting of 2 human α1(I) and a single murine α2(I) chain. Gene supplementation has also been explored in the *oim* mouse, which do not express Col1a2 (Chipman, *et al.*, 1993). *Oim* bone marrow cells transduced with an adenoviral Col1a2 construct secreted normal type I collagen heterotrimer. The *in vivo* efficacy of the adenoviral Col1a2 construct was also demonstrated with synthesis of normal murine type I collagen at the site of injection in the *oim* mouse following interdermal injection (Niyibizi, 2001). In a follow up study MSC migration, engraftment and differentiation in bone was observed four weeks after systemic injection of an MSC population into 2 day old *oim* mice (Wang, *et al.*, 2006a). Initially gene supplementation was thought to only be suitable for the treatment of Type I OI a recent study demonstrated that over expression of a wild type Col1a1 allele in Type III OI MSCs increased the expression of wild type Col1a1 at the mRNA levels and protein level. The correction of the wild type protein deficit was accompanied with restoration of the differentiation capacity of the cells which prior to transfection had been defective (Pochampally, *et al.*, 2005).

1.3.7 Gene Targeting

When treating a dominant negative disorder the key to relieving the phenotype is to prevent the expression of the mutant protein, a variety of tools are available for the task. Antisense olignucleotides, ribozymes and siRNA have been used to suppress Col1a1 or Col1a2 expression. Antisense olignucleotides (ASO) are DNA sequences which bind to RNA transcripts in a sequence specific manner, the DNA:RNA hybrid is targeted by RNase H for degradation. ASO molecules targeting human Col1a1 were introduced into NIH 3T3 cells which stably expressed the Col1a1 minigene and down regulated the human transcripts by 37 – 80% (Colige, *et al.*, 1993). Efficient down regulation of the Col1a1 minigene was achieved *in vivo,* mice which expressed the minigene in the antisense orientation were crossed with mice which expressed it in the correct sense orientation, the presence of the antisense allele reduced the occurrence of embryonic lethality from 93% to 27% (Khillan, *et al.*, 1994). ASO's designed to target Col1a2 allele with a point mutation have also been explored, efficient down regulation of expression was achieved, but the specificity was poor with transcripts from both alleles targeted for degradation (Wang and Marini, 1996).

Ribozymes are catalytic RNA molecules which bind to complementary mRNA molecules and effect cleavage of the mRNA. Ribozymes have superior allele specificity than ASO's and their inherent catalytic activity in principle confers a higher level of efficiency. (Grassi, *et al*., 1997) achieved down regulation in the region of 50% in fibroblasts isolated from and OI patient using a Col1a1 targeting ribozmye (Grassi, *et al.*, 1997). The Col1a1 minigene has also been efficiently suppressed by ribozymes (Peace, *et al.*, 2005; Smicun, 2003; Toudjarska I, 2001). RNAi is the most recent addition to the molecular toolbox, it is effected by double stranded RNA molecules which induce sequence specific RNA degradation, the only report of RNAi for OI so far is a mutation independent strategy (Millington-Ward, 2004). With the enormous genetic heterogeneity associated with OI it would be prohibitively expensive and time consuming to design either ASO's or

ribozymes for each individual mutation. To overcome this, a variety of mutation independent strategies have been developed.

1.3.8. Mutation independent gene therapies

Antisense oligonucleotides, ribozymes and siRNA identify their target mRNA through sequence specific complementary base pairing. Utilising this feature three novel mutation independent suppression approaches have been developed. Three mutation independent suppression strategies have been developed by members of the research team in TCD (Millington-Ward, *et al.*, 1997) in which either 3' or 5' untranslated (UTR) sequences, the coding sequence or a polymorphism is targeted by a suppression agent e.g. siRNA. In the first approach either the 3' or 5' UTR is targeted by the suppression agent resulting in the suppression of transcripts derived from both the mutant and wild type alleles of the target gene. Consequently a wild type copy of the gene has to be provided in parallel with the suppression agent to restore expression of the target gene within the cell; this approach has been described as suppression and replacement. The replacement transgene provided will be modified at the nucleotide level at the target sequence for suppression such that transcripts from the gene are refractory to suppression. However, as the modified nucleotides are located within the 5' and 3' UTRs the replacement gene codes for the wild type α1(I) protein.

In the second approach the suppression agent is targeted toward the coding sequence of the RNA, once again both the wild type and mutant alleles are suppressed. In this case the replacement gene is engineered to contain sequence alterations at the wobble base(s) of each codon located within the target sequence of the suppression agent. In principle transcripts from the replacement gene will encode the wild type protein yet be refractory to suppression. The final approach is based upon designing the suppressor to target a polymorphism known to be present within the mutated allele. For this approach to be effective the suppressor needs to be highly sequence specific such that it can discriminate

between the polymorphic alleles down to a single nucleotide difference. If this sensitivity can be achieved then in principle only the mutated allele should be targeted for down regulation, a recent publication has indicated that indeed it is possible to achieve this levels of specificity; the report detailed specific suppression of the mutated alleles of the superoxidase dismutase and Huntington genes were achieved (Schwarz, *et al.*, 2006).

Targeting of a mutation linked polymorphic allele using a ribzoyme based approach has been successfully carried out *in vitro* (Millington-Ward. S, 1999) and in MSC's achieving a 50% down regulation at the RNA and protein level with allele specificity (Millington-Ward, *et al.*, 2002). The same polymorphism was targeted in MSC's using siRNA and 85% down regulation was achieved, however allele specificity was not demonstrated in this case (Millington-Ward, 2004). An innovative strategy was recently reported in which a severe form of OI was converted to a less severe form by inactivating the mutated allele i.e. converting the severe phenotype to the mild phenotype of Type I OI. This was achieved through the use of an adenoviral gene targeting vector designed to insert into exon 1 of the Col1a1 gene. Adenoviral transduction of MSC's isolated from patients with OI resulted in 31 to 90% of cells undergoing gene targeting at either the wild type or mutated allele. When the mutated allele is targeted normal type I collagen synthesis is restored, in contrast when the normal allele is targeted there is no collagen synthesis. The cells in which the mutant allele was inactivated retained their capacity to differentiate *in vivo* (Chamberlain, *et al.*, 2004).

1.4 RNA Interference

1.4.1 Discovery

RNA Interference was first described in 1998 by Fire & Mello who discovered that injection of double stranded RNA (dsRNA) into *C.elegans* resulted in degradation of homologous mRNA transcripts. RNAi was soon reported to occur in a variety of organisms including Hydra (Lohmann, *et al.*, 1999), Trypanosomes (Ngo, *et al.*, 1998) and Drosophila (Caplen, *et al.*, 2000; Kennerdell and Carthew, 2000). With the observation of RNAi in plants and fungi (Cogoni and Macino, 1999) and the isolation of short dsRNA molecules of 21 – 25ntds from plants (Hamilton, 1999) and Drosophila (Elbashir, 2001b) into which long dsRNA had been introduced, it became clear that the dsRNA triggered silencing phenomena previously observed in plants, (post transcriptional gene silencing (PTGS)) and fungi (quelling) (Romano, 1992; van der Krol, *et al.*, 1990) were functionally related to RNAi.

RNA interference is an evolutionarily conserved gene silencing mechanism, which is believed to have developed as a primitive form of an immune system targeting dsRNA molecules generated during viral replication (Raoul, *et al.*, 2005). The utility of RNAi in mammalian cells was thought to be limited due to the induction of an interferon response by dsRNA > 30ntd in length, interferon triggers global down- regulation of RNA expression and ultimately cell death (Stark, *et al.*, 1998). A crucial discovery was made at that time, the introduction of chemically synthesised small double stranded RNA molecules of 21 – 22ntds (siRNA) in mammalian cells mediated degradation of complementary mRNAs sequences without induction of the interferon response (Elbashir, 2001a; Elbashir, 2001b). Therefore, use of these short interfering RNAs (siRNAs) 21mer dsRNA duplexes facilitated the exploitation of the RNAi pathway in mammalian cells.

1.4.2 RNAi silencing mechanism

dsRNA is cleaved by Dicer a dsRNA specific RNAase III endonuclease (Hammond, et al., 2000) generating siRNAs, which are 21 – 23mer dsRNA duplexes with symmetric 2 ntd 3' overhangs with 5'phosphate and 3'hydroxyl groups (Elbashir, 2001a). The siRNA is then incorporated into a multi-enzyme ribonucleic complex, the RNA Induced Silencing Complex (RISC) (Hammond, *et al.*, 2000). Despite the huge amount of research focused upon elucidation of RISC structure no consensus has been reached (Sontheimer, 2005). The one common element of all RISC complexes is the presence of a member of the Argonaute family which have PAZ and PIWI domains (Hall, 2005). Some groups have reported the unwinding of the siRNA duplex with retention of a single strand within the RISC complex bound by the PAZ domains, forming siRISC (Liu, *et al.*, 2004; Martinez, *et al.*, 2002; Matranga, *et al.*, 2005). Other reports state that it is not unwinding but cleavage that occurs to release one of the duplex strands whilst the other is retained as described (Leuschner, 2006; Matranga, *et al.*, 2005). The bound strand (ideally the antisense strand) guides the RISC to the homologous mRNA, the PIWI RNAase H like domain of the RISC complex, Slicer executes cleavage of the mRNA (Liu, *et al.*, 2004). The RISC complex is subsequently released and participates in repeated rounds of mRNA cleavage (Martinez, *et al.*, 2002; Martinez and Tuschl, 2004). The synthetic siRNA 21mers bypass the Dicer step and are incorporated directly into the RISC complex (Elbashir, 2002; Elbashir, 2001a). Whilst siRNA molecules which bypass the Dicer step are effective mediators of RNAi, it has been reported that 27mer siRNA which are processed by Dicer are more potent than their corresponding 21mer siRNA which suggests that siRNA processing can enhance RISC entry and consequently the potency of the RNAi effect (Silva, *et al.*, 2005). The cleaved mRNA strands are then processed by the general eukaryotic machinery for mRNA degradation (Valencia-Sanchez, *et al.*, 2006).

1.4.3 Specificity of RNAi based mRNA targeting and cleavage.

The siRNA unwinding or cleavage step is central to determining the specificity of siRNA triggered RNAi. The strand that is incorporated into the RISC complex is determined by the thermal stability of the siRNA duplex. The strand with its 5' terminus at the thermodynamically less stable end of the duplex is preferentially incorporated into RISC (Khvorova, *et al.*, 2003; Schwarz, *et al.*, 2003). This 'thermodynamic asymmetry" is exploited when designing siRNA molecules to ensure that the intended anti-sense strand is preferentially incorporated into RISC thereby minimising off target effects (Mittal, 2004; Reynolds, *et al.*, 2004). For a given siRNA the site of RNA cleavage is highly specific, the phosphodiester bond between nucleotide 10 & 11 (Elbashir, 2001b) . This accuracy is maintained even when there are up to 5 mismatches on the 5' siRNA terminus, but it is notable that this number of mismatches significantly reduces cleavage efficiencies (Haley and Zamore, 2004; Martinez and Tuschl, 2004).

The power of siRNA and its appeal as a functional genomics tool lies within its sequence specificity, initial reports indicated that allelic specificity was achievable with only a single siRNA:mRNA mismatch required to prevent gene silencing (Abdelgany, 2003; Elbashir, 2001b; Pedro Gonzalez-Alegre, 2003). This unfortunately was not to last, reports detailing a variety of off target and non-specific suppression events came to light, such as mRNA degradation induced by partial complementarities (Jackson, *et al.*, 2003) and miRNA like translation repression mediated by incomplete homology or complementarity's with 3' UTR sequences (Birmingham, *et al.*, 2006; Doench, *et al.*, 2003; Jackson, *et al.*, 2003; Jackson, *et al.*, 2006b; Saxena, *et al.*, 2003; Zeng, *et al.*, 2003). The use of high concentrations of siRNA has been shown to contribute significantly to off target effects (Persengiev, *et al.*, 2004).

Chemical modification of siRNA molecules is one avenue investigators are exploring as a means of reducing these off target effects, the use of a 2' O- methyl-ribose substitution at nucleotide 2 of the siRNA duplex was shown to reduce non-specific effects without reducing target suppression efficiency regardless of the sequence composition of the siRNA (Jackson, *et al.*, 2006a). In addition increasingly more complex algorithms are coming on stream to aid investigators in siRNA design which will contribute further to minimising off target effects (Arziman, *et al.*, 2005; Naito, *et al.*, 2005; Yamada and Morishita, 2005). Additionally the levels of allele discrimation achievable appear to be down to a single base pair with the recent demonstration of spective suppression of mutant over wild type alleles (Schwarz, 2006)

1.4.4 Micro RNAs (miRNAs).

siRNA resemble endogenously occurring non-coding small RNA molecules micro-RNA's (miRNA's). These molecules direct targeted mRNA degradation or translation repression. Functionally they have been shown to be involved in a diverse number of regulatory pathways including apoptosis, haemopoietic cell differentiation, developmental timing control and organ development (Bartel, 2004). It has been suggested that up to 30% of human genes may be regulated by miRNA's (Lewis, *et al.*, 2005). miRNA genes were thought to be located in intergenic chromosomal regions, but recent analysis has shown that 70% of mammalian miRNA are located within defined transcription units (Rodriguez, *et al.*, 2004). Their genomic locations can be categorised as follows; intronic miRNA in protein coding regions, intronic or exonic in non-coding genes and finally those generated by splicing events (Kim and Nam, 2006; Rodriguez, *et al.*, 2004; Sevignani, *et al.*, 2006).

miRNA are initially transcribed as long transcripts termed pre-miRNAs which are initially cleaved in the nucleus by the Microprocessor complex which contains an

RNAase III like enzyme Drosha, which in human's is associated with DCGR8, generating 60 – 70 ntd hairpin structures termed Pre-miRNA's (Gregory, et al., 2004; Han, et al., 2006). Pre-miRNA's are exported to the cytoplasm facilitated by Exportin 5, (Bohnsack, *et al.*, 2004; Lund, *et al.*, 2004) where further processing by Dicer takes place generating 21mer dsRNA molecules (Kim and Nam, 2006; Zamore and Haley, 2005). In this final step Dicer is coupled to the RISC complex into which the guide miRNA strand is efficiently incorporated. In fact the RISC complex has 10 fold greater activity when using pre-miRNA Dicer substrates than when using siRNA duplexes (Gregory, *et al.*, 2005b).

1.4.5 miRNA mediated translational repression or targeted mRNA degradation.

In order for targeted mRNA degradation to occur there needs to be a high level of homology between the miRNA and its cognate mRNA, especially at the Splicer cleavage site located between nucleotides 10 & 11 (Elbashir, *et al.*, 2001b; Haley and Zamore, 2004). If the required level of homology is present and the correct Argonaute protein present in the RISC, mRNA cleavage will occur (Valencia-Sanchez, *et al.*, 2006; Zamore and Haley, 2005). miRNA duplexes tend to have incomplete homology with their target, this is insufficient to induce mRNA cleavage and induces translational repression (Doench, *et al.*, 2003). Binding of a single miRNA is usually insufficient to exert translational repression and several miRNAs may bind at different sites along the transcript. This opens up the whole area of multi-miRNA regulatory pathways where different numbers or combinations of miRNA molecule may have distinct effects on gene expression. In fact mammalian genes can have more than one target site in their 3'UTR's and a given miRNA can target multiple mRNA's (Bartel, 2004; Sevignani, *et al.*, 2006; Zamore and Haley, 2005).

1.4.6 miRNA associated disorders.

miRNAs have been shown to have a role in human disease Many miRNA's have roles in homeostasis with alterations in the expression profiles of miRNA genes contributing to the pathogenesis of a variety of cancers, with aberrant expression of *mir-15* & *mir-16* associated with Chronic lymphocytic leukaemia (CLL) (Calin and Croce, 2006), *mir-155* with B-cell lymphoma (Eis, *et al.,* 2005), *mir-143 & mir-145* with colorectal cancer (Michael, et al., 2003) and *mir-21* with Gliobastoma (Chan, *et al.*, 2005). Other diseases associated with perturbed miRNA expression include Fragile X and spinal muscular atrophy (Sevignani, *et al.*, 2006). In this regard it is notable that the DCGR8 gene which associates with Drosha is located in the DiGeorge syndrome chromosomal region (Landthaler, *et al.*, 2004).

1.4.7 Chromatin effects

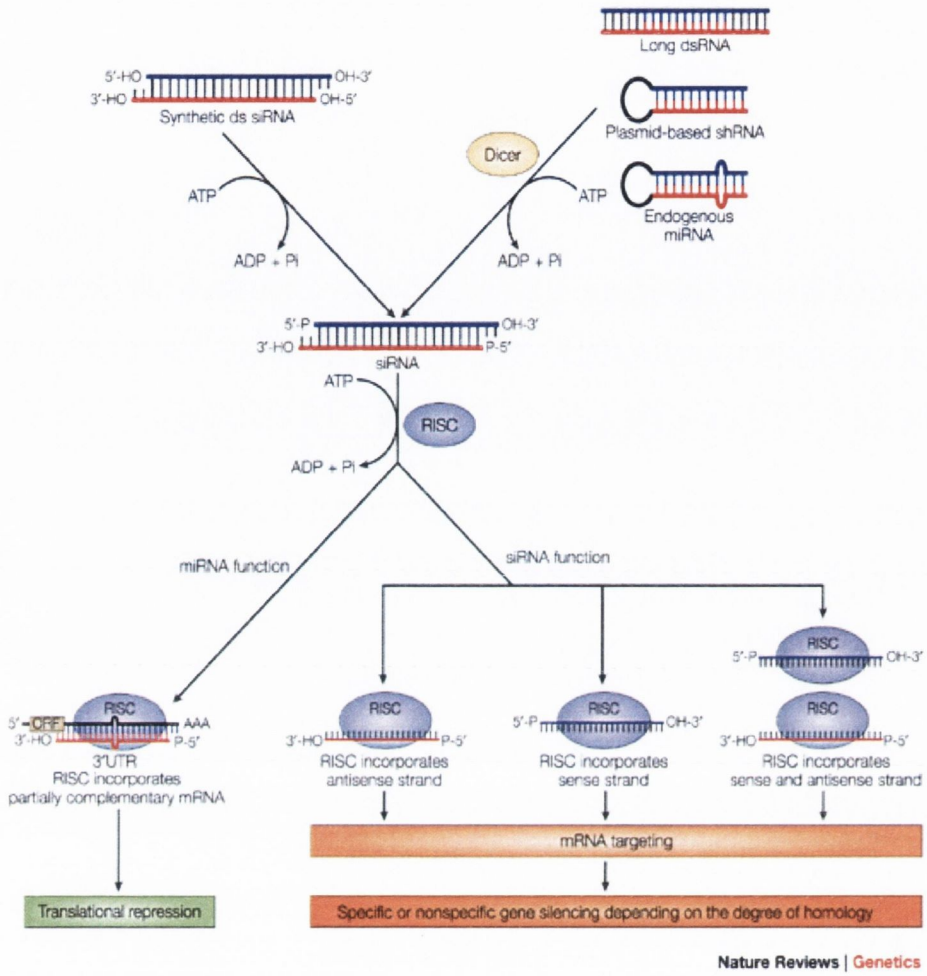
RNAi acts mainly at the post transcriptional level either directing mRNA degradation or translational repression, but miRNA and siRNA are now thought to effect gene silencing through genomic modifications which occur at the DNA level via RNA directed DNA methylation (RdDM) and RNAi mediated heterochromatin formation. Both are natural epigenetic gene regulation mechanisms controlling heritable changes in gene expression that are not caused by mutations but are epigenetic processes (Wassenegger, 2005). RdDM has been observed extensively in plants, where dsRNA homologous to promoter regions can trigger promoter methylation and gene silencing, cytosine's are methylated and the methylation is largely restricted to the regions of homology (Bayne and Allshire, 2005; Matzk, 2005).

The occurrence of RdDM in mammalian cells is an area of conflicting evidence. Morris et al 2004 demonstrated RdDM in human cells where siRNA targeting of the elongation factor 1 alpha (EF1A) promoter which was driving expression of a GFP reporter gene in human 293FT cells induced DNA methylation of the promoter sequences and down

regulation of GFP expression (Morris, *et al.*, 2004). In another report an siRNA targeting the promoter of RASSF1A a tumour suppressor gene did appear to induce DNA methylation albeit at a low level (Castanotto, *et al.*, 2005). Park and colleagues reported that they were unable to induce DNA methylation of the Huntington gene perhaps this is due to the fact that the siRNAs targeted the coding sequence of the gene rather than the promoter (one siRNA spanned the ATG start site and the second was completely located within the fourth exon) and this appears to be a requirement of RdRM (Park, *et al.*, 2004).

1.4.8. Applications of RNAi

RNAi is an extremley versatile genomics tool, it lends itself to a variety of applications; gene discovery, signal transduction pathway elucidation, target identification, target validation and generation of transgenic animals (Cullen and Arndt, 2005; Echeverri and Perrimon, 2006; Ito, *et al.*, 2005; Leung and Whittaker, 2005). With the completion of a large number of genome sequencing projects, including the human genome, a huge number of genes were identified to which no function has yet been assigned. RNAi provides a rapid, sensitive and inexpensive means by which loss of function studies can be carried out in order to characterise unknown coding sequences. The most efficacious way to do this is do a whole genome screen, a number of these studies have been carried out in *C. elegans,* Drosophila and mammalian cells, resulting in the identification of novel genes involved in processes such as longevity, genome stability, fat storage, cell morphology, cell growth and apoptosis, reviewed in (Cullen and Arndt, 2005; Kuttenkeuler and Boutros, 2004; Sugimoto, 2004). Since the realisation that the RNAi pathway could be exploited in mammalian cells, there has been a flurry of investigation into the direct application RNAi based therapeutics. The attractiveness of RNAi based therapeutics is due to 1) the ease at which accessible target sites can be identified, 2) the very low concentrations of siRNA required for gene silencing, 3) RNAi effector molecules are very stable, 4) custom synthesis can be carried out rapidly and 5) it is a relatively inexpensive technology.



(Mittal, 2004)

Fig. 1.4 Mechanisms of RNAi mediated gene silencing. Long (> 30 bp) dsRNA, vector-based shRNAs and endogenous miRNAs are processed by Dicer to form 21–23 bp siRNA duplexes with symmetric 3′-overhangs and 5′-phosphate groups. The duplexes associate with RISC, which contains a helicase that unwinds the duplex. Each siRISC associates with its complementary mRNA which depending on its complementarity, results in either targeted mRNA cleavage or translational arrest. In theory, the antisense siRNA strand will target the desirable mRNA for destruction, whereas the RISC that incorporates the sense strand can potentially associate with other unintended mRNAs and cause off-target activities. (Mittal, 2004).

The entire spectrum of diseases have been targeted (for review see (Dallas A, 2006)), incuding, viral infections such as HIV (Morris and Rossi, 2006b), Hepatitis B (Peng, *et al.*, 2005), Hepatitis C (Hamazaki, *et al.*, 2006), influenza (Bennink and Palmore, 2004) and west nile virus (Ong, *et al,.* 2006). Neurodegenerative disease have also been explored as therapeutic candidates, targeting of genes involved in the pathogenesis of diseases such as Huntington's, Alzheimer's, Spinocerebral Ataxia and dementia have all shown promising results (Miller, *et al.,* 2005; Ralph, *et al.*, 2005). Investigation into the potential therapeutic application of RNAi for various forms of cancer has been equally promising, with genes involved in Chronic Myeloid Leukaemia (Yang, *et al.*, 2005), cervical cancer (Gu, *et al.*, 2006), pancreatic carcinoma (Tsutsumida, *et al.*, 2006) and lung cancer (Ko, *et al.*, 2006) targeted successfully. Although the development of siRNA based therapeutics was initiated only a few years ago, siRNA drug candidates have already entered Phase I clinical trials. A potential therapeutic for age related macular degeneration (AMD), the anti-VEGF siRNA Cand5 (Acuity Pharmaceutical) is now recruiting for Phase II trials. Another anti-VEGF siRNA-027 (Sirna Therapeutics) which is also directed toward treating AMD and a siRNA targeting Respiratory Syncytial Virus (Alnylam pharmaceuticals) are also in clinial trials (www.rnaiweb.com).

With the depth of knowledge regarding the molecular pathogenesis of Osteogenesis imperfecta it is has become clear that if it were possible to reduce the amount of mutant Col1a1 or Col1a2 chains incorporated into the type I collagen triple helix alleviation of the pathology associated with the disorder may occur. RNAi provides a tool by which this may potentially be achieved, based upon this fact and the realisation of the potential therapeutic application of MSCs for the treatment of bone disorders such as OI an RNAi based suppression and replacement strategy for the treatment of Osteogenesis Imperfecta was devised and was explored in the variety of studies described in this thesis.

Chapter 2

Materials and Methods

2.1 Reagents

(I) **Chemicals:** All reagents, growth media and antibiotics unless otherwise stated were obtained from Sigma Chemical Corporation, BDH Laboratory Supplies or Gibco/BRL.

(II) **Enzymes and Buffers**: All restriction enzymes and buffers were obtained from New England Biolabs (NEB), Promega or Roche. T4 DNA Ligase and RNaseIT cocktail were obtained from Stratagene. RQ1 DNase was purchased from Promega. Alkaline Phosphatase Enzyme was obtained from New England Biolabs (NEB).

(III) **Markers and ladders.** 1kb and 100bp DNA ladders were both obtained from NEB (Figure 2.1). Kaleidoscope pre stained protein ladder was purchased from BioRad (Figure 2.2).

(IV) **Kits**: Hi-Speed Plasmid Maxi-kits, Hi-Speed Plasmid Midi-kit RNeasy Mini Prep Kits, QIAquick Gel Extraction kits and the QuantitTech SYBR Green RT-PCR kits were all purchased from Qiagen. Metra CICP EIA kits were purchased from the Quidel Corporation. HiYield Plasmid Mini Kits were purchased from Real Biotech Corporation. Big Dye Terminator v3.1 Cycle Sequencing kits were purchased from Applied Biosystems. All procedures were carried out as described by the manufacturer.

(V) **Cloning Vectors**: Cloning vectors used were pcDNA3.1- purchased from Invitrogen, pcDNA3.1 without the CMV promoter was generated by Dr. M.'O' Reilly. EGFP-1 was purchased from Clontech. EGFP1 with a CMV

promoter was generated by Dr. N. Chadderton and Dr. S. Kiang generated a H1-Bluescript vector. All vectors were illustrated in Figures 2.3 -2.7.

(VI) **Cell Lines**: HeLa cells were a kindly provided by Prof. Seamus Martin Dept of Genetics Trinity College Dublin and Mesenchymal Stem Cells by Dr. Grant Jordan, Department of Trauma & Orthopaedic Surgery, Musgrave Park Hospital Belfast Ireland.

(VII) **Tisssue Culture**: Dulbecco's Modified Eagle Medium (DMEM), α-MEM Medium, L-Glutamine and Sodium Pyruvate which were all purchased from Cambrex. Standard Fetal Bovine serum and Fetal Bovine Serum of South American origin, Trypsin-EDTA, Genetecin and Phosphate Buffered Saline were all obtained from Gibco. DMSO was obtained from Sigma.

(VIII) **Differentiation Regents**. L-Ascorbic Acid, β-glycerophosphate, Inodomethacin, dexamethasone, 3 Wasobutyl-1-methylxanthine and insulin all purchased from Sigma.

(IX) **Histochemical Stains.** Oil Red O and Alkaline Phosphatase kit 85C were purchased from Sigma.

(X) **Protease Inhibitors.** Complete Mini Protease Inhibitor Cocktail was purchased from Roche.

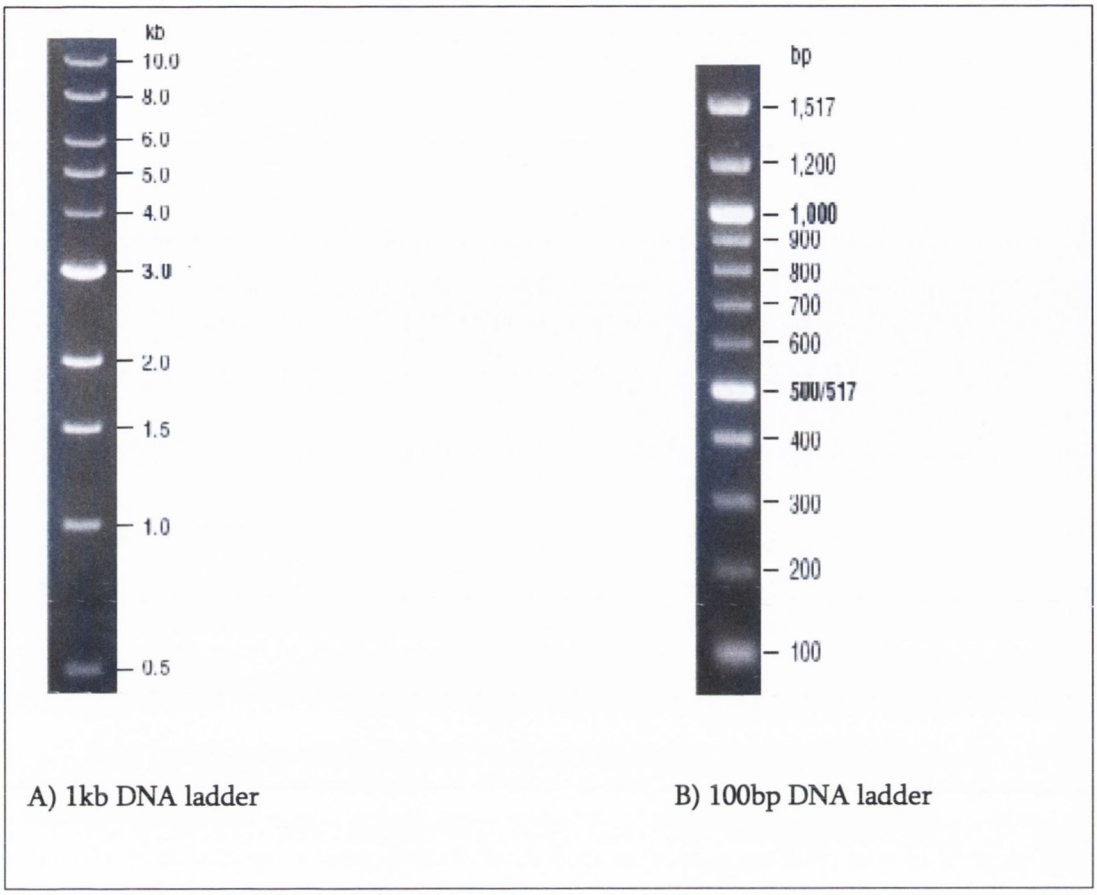


Figure 2.1 DNA Ladders. A) 1kb ladder which ranges from 500bp to 10.0kb and **B)** 100bp ladder which ranges from 100bp to 1517bp.

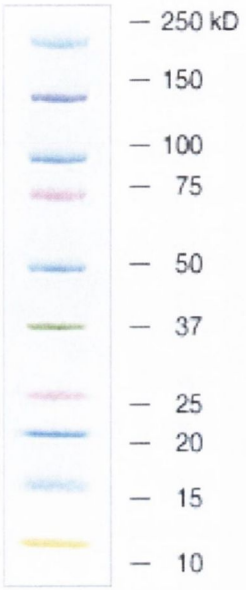
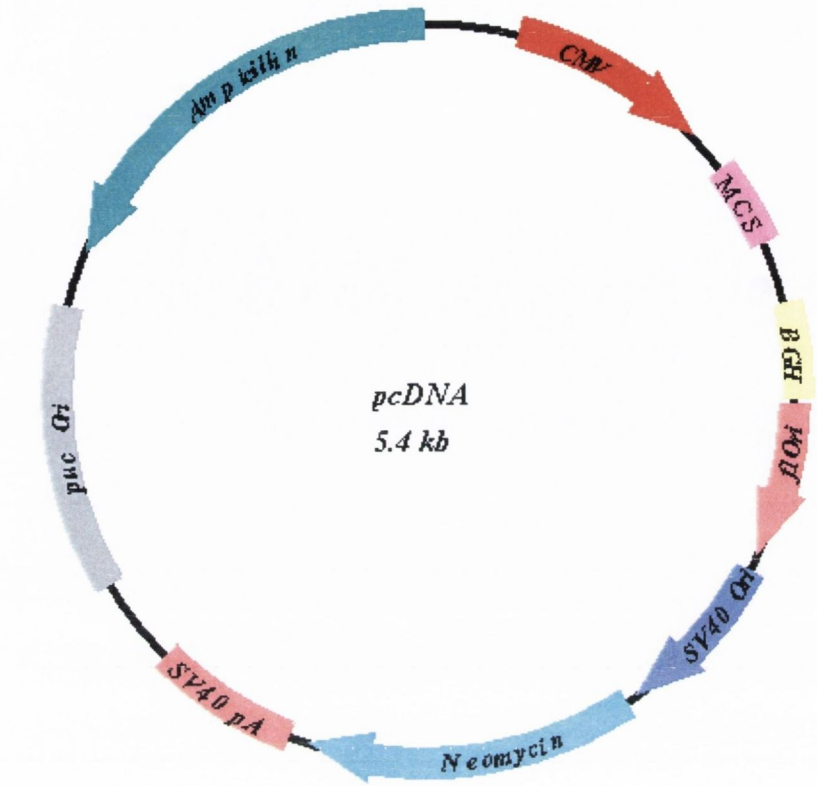


Figure 2.2. BioRad Kaleidoscope Protein marker. The protein marker which ranges from 10 - 250kDa.



T7 promoter/primer binding site — *Nhe* I — *Pme* I — *Afl* II — *Hind* III — *Asp*718 I — *Kpn* I

869 GACTCACTAT AGGGAGACCC AAGCTGGCTA GCGTTTAAAC TTAAGCTTGG TACCGAGCTC

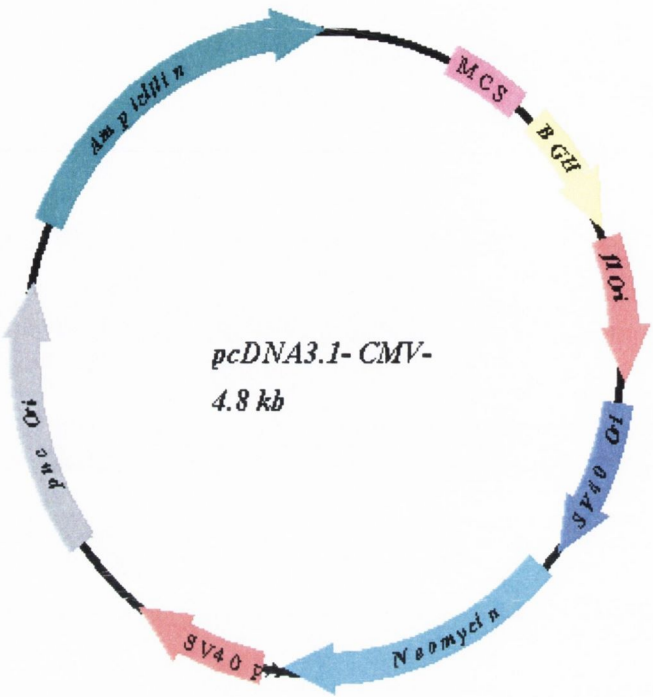
*Bam*H I — *Bst*X I* — *Eco*R I — *Eco*R V — *Bst*X I* — *Not* I — *Xho* I

929 GGATCCACTA GTCCAGTGTG GTGGAATTCT GCAGATATCC AGCACAGTGG CGGCCGCTCG

Xba I — *Apa* I — *Pme* I — pcDNA3.1/BGH reverse priming site

989 AGTCTAGAGG GCCCGTTTAA ACCCGCTGAT CAGCCTCGAC TGTGCCTTCT AGTTGCCAGC

Figure 2.3 Cloning vector pcDNA3.1- (Invitrogen) with Multiple Cloning Site (MCS) detailed.



T7 promoter/primer binding site — *Nhe* I — *Pme* I *Afl* II *Hind* III *Asp*718 I *Kpn* I

869 GACTCACTAT AGGGAGACCC AAGCTGGCTA GCGTTTAAAC TTAAGCTTGG TACCGAGCTC

*Bam*H I — *Bst*X I* *Eco*R I — *Eco*R V — *Bst*X I* *Not* I *Xho* I

929 GGATCCACTA GTCCAGTGTG GTGGAATTCT GCAGATATCC AGCACAGTGG CGGCCGCTCG

Xba I — *Apa* I *Pme* I — pcDNA3.1/BGH reverse priming site

989 AGTCTAGAGG GCCCGTTTAA ACCCGCTGAT CAGCCTCGAC TGTGCCTTCT AGTTGCCAGC

Figure 2.4. Cloning vector pcDNA3.1- CMV- (generated by Dr. M O' Reilly Trinity College Dublin) with Multiple Cloning Site (MCS) detailed. The CMV promoter has been removed from this plasmid.

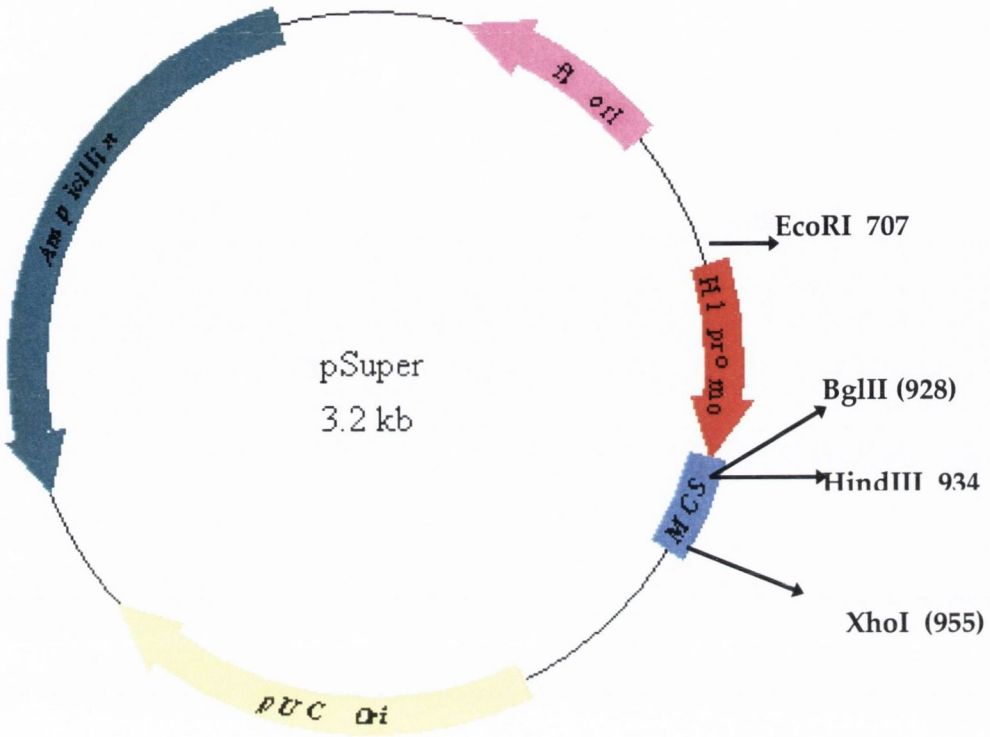
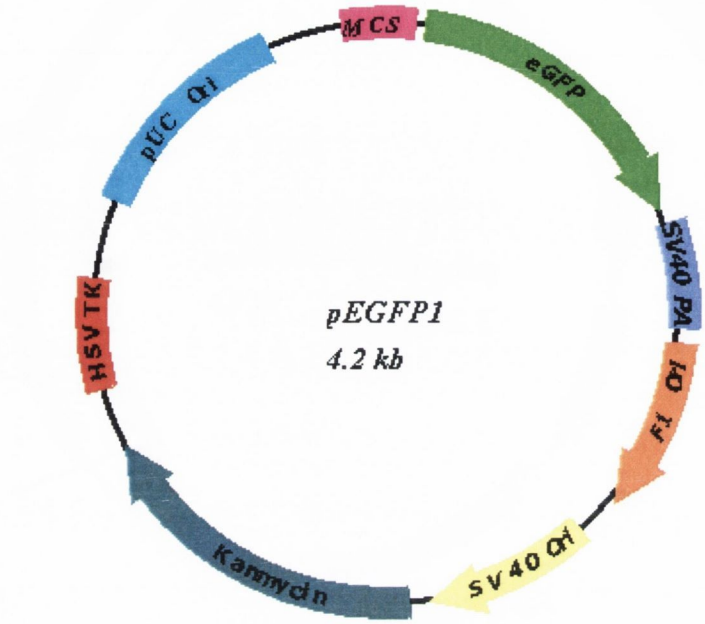


Figure 2.5. H1-Bluescript Cloning vector (Generated by Dr. S. Kiang) with key restriction enzyme sites and H1 promoter highlighted.



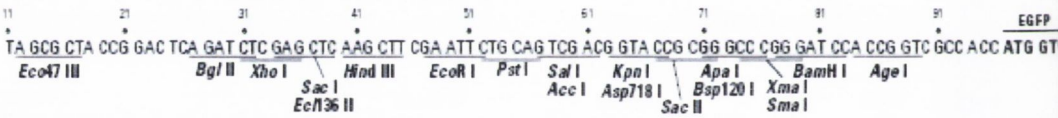


Figure 2.6. EGFP1 vector (Clontech) with multiple cloning site (MCS) illustrated.

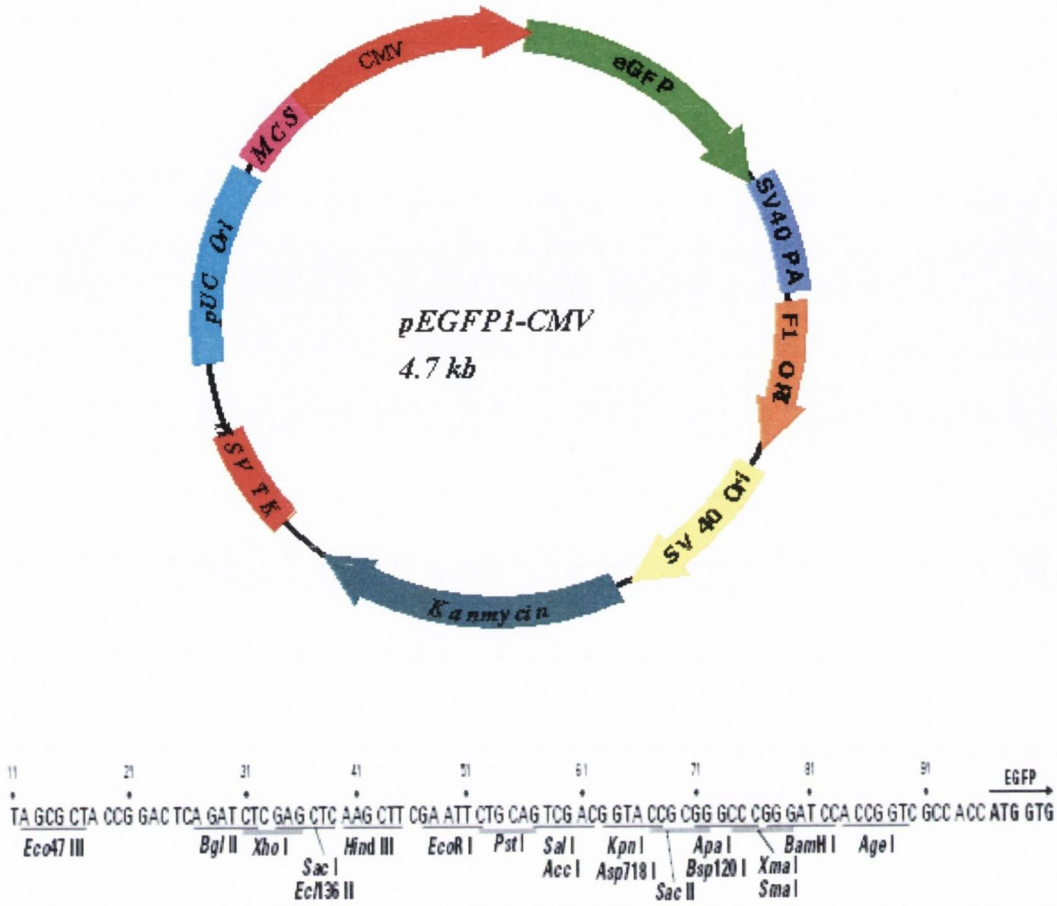


Figure 2.7. pEGFP1-CMV which has CMV promoter cloned into the SmaI site was generated by Dr. N. Chadderton, Trinity College Dublin. The multiple cloning site (MCS) is detailed.

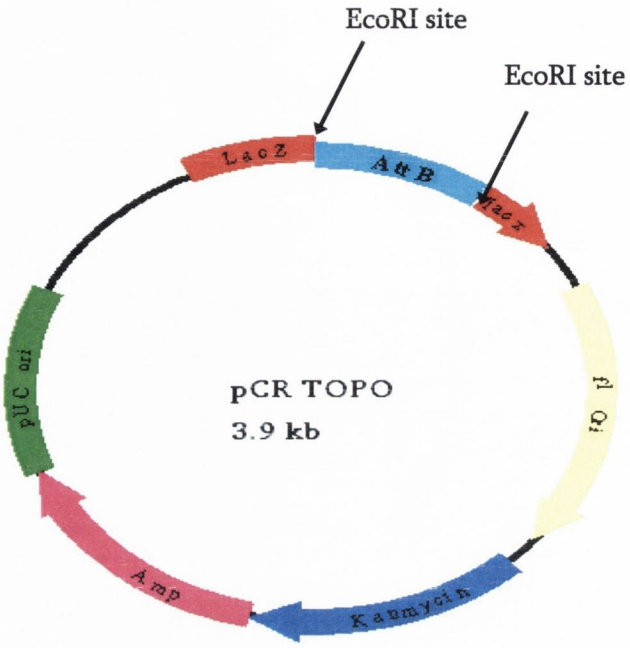


Figure 2.8 pCR TOPO cloning vector in which the *AttB* sequence was provided, the EcoRI sites used to remove the 300bp *AttB* sequence are indicated by the black arrows. This *AttB* containing plasmid was a kind gift of Dr.M.Calos, Dept of Genetics, Stanford University School of Medicine, Stanford California, USA.

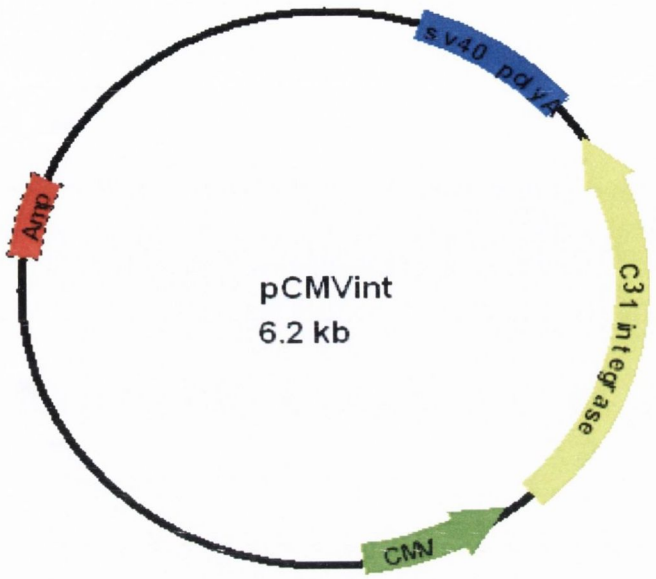


Figure 2.9. pCMVInt, this plasmid contains the gene encoding the ΦC31 integrase enzyme under the control of the CMV promoter. This plasmid was a kind gift of Dr.M.Calos, Dept of Genetics, Stanford University School of Medicine, Stanford California, USA.

2.2. Solutions

(I) **1 X TAE**: 0.04M Tris-acetate, 0.0001 EDTA (ethylenediaminetetra-acetic disodium salt).

(II) **Agarose gels:** generated at the required percentages by dissolving 1g of agarose in 100ml 1XTAE per percentage required. Loading Buffer consisted of 0.25% Bromophenol blue, 0.25% Xylene cyanol and 30% Glycerol. Ethidium Bromine was added to a final concentration of 0.5ug/ml

(III) **Luria Bertani Medium** (LB medium): 1% (w/v) NaCl, 1% (w/v) tryptone, 0.5% (w/v) yeast, 0.2% (w/v) maltose and 10Mm $MgSO_4$. Media was autoclaved at 15lbs pressure psi for 20 minutes. For solid LB plates 1.5% (w/v) bacto-agar was added prior to autoclaving. Selection of desired bacteria was achieved using antibiotic selection with 50ug/ml ampicillin and 25μg/ml tetracycline or kanamycin

(IV) **Tissue Culture**: *DMEM complete* - 500ml DMEM, 10% FCS, 0.05mg/ml penicillin/-streptomycin, 1mM sodium pyruvate and 2mM L-glutamine. 2x *Freezing Medium*- 1ml DMSO, 20% FCS and 2ml DMEM complete. *α MEM 20* - 500ml α MEM , 20% FCS (South American origin), 0.05mg/ml penicillin/-streptomycin, 1mM sodium pyruvate and 2mM L-glutamine. *Selective Medium* – 600μg/ml G418 in Complete medium.

(V) **PCR:** *10X PCR Buffers* - 500mM KCL, 100mM TrwasCl (pH 8.8), 0.1% gelatine & 1% Triton (pH 9.0) supplemented with either 5mM, 10mM, 15mM or 20mM $MgCl_2$. dNTP mix (Roche) 10ul of dATP, dTTP, dCTP and dGTP in dH_20 to a final concentration of 1.25mM

(VI) **Mini Preps solutions:** *Lysis Solution 1*- 50mM Glucose, 25mM Trwas pH 8.0 & 10mM ETDA. *Lysis Solution 2*- 0.2M NaOH & 1% SDS. *Lysis Solution 3*- 3M NaOAc pH 5.2.

(VII) **Western Blot Analysis:** *RIPA Buffer*- 50mM Trwas.Cl pH 7.5, 150mM 1% Non-Idet P-40, 0.5% Sodium deoxycholate and 0.1% SDS. Heat to 68^{0}C. Prior to use 10ηg leupeptin, 20ng apoprontin and 100ng PMSF was added to 10mls of RIPA buffer. *1x SDS PAGE Running Buffer*- 25mM TrwasCl (pH 8.3), 250mM Glycine, 0.1%SDS. *Transfer Buffer* - 39mM glycine, 48mM Trizma base, 0.037% SDS, 20% Methanol. *Blocking Solution*- 3% BSA-TBST, 20mM Tris.Cl (pH7.5), 500mM NaCl, 0.05% (v/v) Tween 20, 3% BSA (bovine serum albumin). *Washing Solutions* – TBS Buffer - 10mM Tris.Cl (pH 7.5) 150mM NaCl. TBS-Tween/Triton buffer: 20mM Tris.Cl (pH 7.5), 500mM NaCl, 0.05% (v/v) Tween 20, 0.2% Triton.

2.3 Primers

2.3.1 Plasmid PCR and sequencing primers – described in 5' – 3' orientation

pcDNA3.1F	CTCTGGCTAATAGAGAAC
pcDNA3.1R	CCAGGGTCAAGGAAGGCACG
pBluescript 1201	AACAGCTATGACCATG
pBluescript 1224	CGCCAGGGTTTTCCCAGTCACGAC
pcDNA 82F	TTGTGTGTTGGAGGTCG

2.3.2 Col1a1 PCR and sequencing Primers described in 5' – 3' orientation

3' insert detect	AATCTGTCTCCCAGAAGAC
Col1a1 cmv-detectR	ACGTGGAGTTCTAGAGCCATCTTC
KpnIF	GAACGGCCTCAGGTACCATGACCGAG
AccIII	GAGAACATCCGGACCCAGAGGGAAGCC
NotIF	TAGTCGCGGCCGCAGCAGACGGTTTCTCCT
ClaI F	TCCGCCTGCCCATCATCGATGTGGCCCCCTTGGACGTT
AflIIR	GATTACAGTATGAACCTTAAGGCCCATCACCCCAAGGC

2.3.4 Col1a1 mutation detection sequencing primers described in 5' – 3' orientation

Deletion F	CAAGGCTGGAGAGCGAGGTG
Deletion R	GTCTCCAGGAACAACGTGTTC
Gly – AsnF	CTGGACGAGACGGTTCTCC
Gly – AsnR	GCCTGTCTCACCCTTGTCAC
Stop Signal F	TCCTGGTCCCAAGGTTAACA
Stop Signal R	ATCGCGCCAGGGAAACCAC

2.3.5 Minigene sequencing primers described in 5' – 3' orientation

102 F	GGTTCAAGAACTGCCATCTCAG
103 F	CCGCTTCCACGTTTGCAGCTC
104 F	CACTGAGTCCAGGTACAACT
105 F	GGGTTTGGAGGAAGGCATTTG
106 F	GCCCGGTACCAGCTTTTGTT
107 F	GAAGAGAGCAGAGGGGCACC
101 3' F	CGGAGCCAGGGAGCAGCGA
102 3' F	GGAATCAGTGTAGAGGAT
103 3' F	ACCATGTTAGCCAGCCTGGT

2.3.6 RT-PCR Primers described in 5' – 3' orientation

GAPDH F	CAGCCTCAAGATCATCAGCA
GAPDH R	CATGAGTCCTTCCACGATAC
Col1a1F	CCTGGAAAGAATGGAGATGATGG
Col1a1R	AGCACCAGCAATACCAGGAGCACC
Col1a1 Duo F	GAGAGCATGACCGATGGATTCC
Col1a1 Duo R	TTAGGTGATGTTATGGGAGGC
Actin F	TCACCCACACTGTGCCCATCTACGA
Actin R	CAGCGGAACCGCTCATTGCCAATGG
EEF2 F	CATCGCTGAACGCATCAAC
EEF2 R	TGGAAGGTCTAGAGC

2.3.7 Integrase sequencing primers described in 5' – 3' orientation

AttB F 1201	TCGACGATGTAGGTCACGGTCTCGAAG
AttB R1200	CCGTGCTCACCGTGACCACCGCGC

2.4 DNA Manipulations

2.4.1 Oligonucleotide design and synthesis

Primers were usually designed to between 18 and 21 nucleotides in length, with the GC: AT ratios as close to 1:1 as possible. Primer sequences were checked to ensure low secondary structure and to minimise primer dimer formation. All oligos were ordered from Sigma-genosysis. All oligonucleotides were resuspended in nuclease free sterile water at concentration of 100pmoles/ul or 100μM.

2.4.2 Polymerase Chain Reaction (PCR)

Standard PCR reactions were carried out in a 50μl volume. Each 50μl reaction consisted of 10pM of each primer (forward and reverse), 200μM dNTPs, 1X PCR buffer and 1U Taq Polymerase. Optimal $MgCl_2$ concentrations ranging from 5mM, 10mM, 15mM or 20mM were determined for each specific reaction. The standard PCR program was as follows; 94°C for 5 minutes followed by 35 cycles of 94°C for 1 minute, 55°C for 1 minute and 72°C for 1.2 minutes. Finally reactions were incubated for 10 minutes at 72°C.

2.4.3 Agarose Gel Electrophoresis

Agarose gels of the required percentage were prepared in TAE buffer, 0.5ug per ml Ethidium Bromide was added to facilitate DNA visualisation under UV light.

2.4.4 Gel Extraction using QIAquick Gel Extraction kit (Qiagen).

DNA fragments generated in PCR or restriction enzyme digests were excised from gels with a sterile scalpel. The equivalent of 3 gel volumes of buffer GQ was added to each gel slice containing the required DNA (100mg = 100ul). Gel slices were incubated at 50°C for 10 minutes and vortexed every 2 to 3 minutes to aid solution. 1 gel volume of Isopropanol was added to the sample. Samples were added to the QIAquick spin column and centrifuged at 18,000*g* for 1 minute to bind the DNA to the column. 0.75ml of Buffer PE was added to the column and samples were centrifuged for 1 minute at 18,000*g*. The flow-through was discarded and the QIAquick column centrifuged for a further minute.

The DNA was eluted from the column with 50ul of Buffer EB (10mM TrisCl, pH 8.5) or dH_20.

2.4.5 Phenol/Chloroform Extraction of DNA and RNA.

Nucleic Acids in aqueous solution were mixed with an equal volume of phenol, vortexed for 30 seconds and centrifuged at 13,000 rpm (IEC Micromax bench top centrifuge) for 5 minutes. The top aqueous phase containing the nucleic acid was removed and transferred to a fresh eppendorf. An equal volume of a 1:1 phenol/chloroform mix was added to the sample, vortexed and centrifuged as above. The top layer was removed and transferred to a fresh eppendorf and an equal volume of chloroform added, vortexed and centrifuged as above. The top phase was then removed transferred to a new eppendorf and the nucleic acids were ethanol precipitated (section 2.4.6).

2.4.6 Ethanol precipitation.

Nucleic acids were precipitated by adding a 10% volume 3M NaOAc (pH5.2) and 2.5 volumes of 100% ice cold ethanol. The sample was vortexed for 10 seconds and left to precipitate at –20°C for 30min.Nucleic acids were pelleted by centrifugation at 13,000 rpm for 15 minutes. The liquid was carefully removed. 400ul of 75% EtOH was added to the tube and it was centrifuged for 15min at 18,000*g* and carefully removed. Pellets were air dried and resuspended in dH_20.

2.4.7 Restriction Enzyme Digests.

Typically DNA was digested with the required restriction enzyme using 2U of enzyme per µg DNA diluted in the appropriate reaction buffer. 0.1% BSA was added when appropriate. The mix was then incubated at the optimal temperature for ~ 1 ½ hours. Digestion was verified by agarose gel electrophoresis.

2.4.8 DNA Ligations.

Ligations were carried out in 15µl reaction volume. Reaction buffer was used at a 1X concentration supplemented with rATP to a final concentration of 1mM. One unit of ligase was used to ligate 1µg of DNA. Multiple reactions were carried out for each ligation in which the ratio of vector: insert was varied, typically from 1:1 to 1:10. In addition the appropriate digest, insert and ligase controls were performed.

2.4.9 End filling of 5' DNA overhangs using Klenow.

Reactions were carried out in 50µl volume. 1U of Klenow enzyme was used per µg of DNA. 33µM of dNTPs were included in the reaction mix. Reactions were incubated at 37^{0}C for 30 minutes. Reactions were terminated by heating samples to 65^{0}C for 5 minutes. Samples were purified as described in 2.4.5

2.4.10 End filling of 3' DNA overhangs using T4 polymerase

Reactions were carried out in 50µl volumes. Reactions contained 1-3 Units of T4 polymerase per µg of DNA, 100uM of dNTPs and 1X BSA. Reactions were carried out at 12^{0}C for 10 minutes. Reactions were terminated by heating the samples to 65^{0}C for 5 minutes. Samples were purified as described in 2.4.5

2.4.11 Dephosphorylation of DNA fragments using Antartic Phosphatase enzyme.

Antarctic Phophatase (AP) enzyme reactions were carried out in 70µl volumes, containing 1X AP specific buffer, 2 Units of Antartic Phosphatase enzyme per µg of DNA was used. This enzyme required stringent deactivation (5mM EDTA, 10 minutes at 65^{0}C followed by 1 hour at 65^{0}C). A phenol/chloroform extraction of every desphosphorlyated sample was carried out (2.4.5) following deactivation.

2. 5 Microbiology

2.5.1 Competent Cells

A 10ml culture of XL1 Blue *E.Coli* cells were grown in an orbital shaker overnight at 37°C in LB supplemented with 25μg/ml of tetracycline. 100ml of LB broth supplemented with 25μg/ml tetracycline was inoculated with 1ml of the overnight culture and grown shaking in an orbital shaker a 37°C until an OD600 of 0.3 – 0.5 was reached (~2.5 hours). The culture flask was placed on ice for at least 30 minutes. Cells were pelleted by centrifuging at 3K/min for 15 min. Pellets were gently re-suspended in 50ml of ice-cold 50mM CaCl2 and placed on ice for 30 minutes. Cells were pelleted again by centrifugation at 3K/min for 15 min. The pellet was gently re-suspended in 10ml of ice cold $CaCl_2$. Competent cells were stored on ice and could be used within 24 hours. Alternatively cells were stored at –80°C in 15% glycerol for up to 3 months.

2.5.2 Transformation of XL1-Blue E.Coli Cells

5μl of ligated DNA was added to 200μl of competent cells. The cells were placed on ice for 30min. Cells were heat shocked by incubating at 42°C for 5min, cells were then placed on ice immediately and 1ml of LB broth added. Cells were shaken gently at 37°C for an hour. 200μl of transformed cells were plated on 10cm Petri dishes containing selective LB agar. Plates were placed upside-down in a 37°C oven and grown overnight at 37°C.

2.5.3 Transformation of XL-10 Gold ultracompetent cells

Once thawed on ice the cells were mixed gently and 100μl added to a pre-chilled polypropylene tubes. 4 μl of β-Mercaptoethanol was added to each tube, mixed gently and incubated on ice for 10 minutes with gentle mixing every 2 minutes. 0.1 – 50ng of DNA was added to each aliquot, mixed gently and incubated on ice for 30 minutes. Following a 30 second heat shock at 42°C the cells were incubated on ice for a further 2 minutes. 0.9mls of pre-heated NYZ^+ broth was added and the cell suspension incubated at

37°C for an hour. 100ul of the transformation mix was plated on to LB agar plates containing the appropriate antibiotic and the plates incubated overnight at 37°C.

2.5.4 Plasmid PCR from individual bacterial colonies.

Using primers which flanked the multiple cloning site of the vector colony PCRs were carried out as follows. Single colonies were picked using a sterile toothpick. The colony was first streaked onto a selective LB agar plate and the remainder of the colony added to 5µl of Tris-Triton lysis buffer in a sterile 500µl tube. In addition each 50µl reaction consisted of 10pM of each primer (forward and reverse), 200µM dNTPs, 1X PCR Buffer (20mM $MgCl_2$) and 1U Taq Polymerase. The PCR program runs as follows; 95°C for 5 minutes, 94°C for 5 minutes followed by 35 cycles of 94°C for 1 minute, 55°C for 1 minute and 72°C for 1.2 minutes ending with a final 10 minutes at 72°C. The PCR products were analysed on a 1% or 2% agarose gel.

2.6 Preparation of Plasmid DNA

2.6.1. Plasmid Miniprep (M-Stet Boiling Prep)

Single colonies from the selective agar plates were used to inoculate 10ml of selective LB media and streaked on a selective agar plate. The cultures were grown shaking in an orbital shaker at 37°C overnight and the plates were placed in a 37°C incubator overnight. The cultured cells were pelleted by centrifugation at 18,000*g* for 15 minutes and re-suspended in 400µl of Lysis Solution I. After incubation at room temperature for 5 minutes 800µl of Lysis solution II was added and the samples placed on ice for 5 minutes. 600µl of ice cold 3M NaOAc (pH5.2) was added and samples placed on ice for a further 5 minutes. Samples were then centrifuged at 13,000rpm for 15 minutes. 600 µl of the supernatant was transferred to a fresh eppendorf and a phenol extraction and ethanol precipitation carried out (section 2.4.5). Pellets were re-suspended in 45µl of dH_20. 1µl of RnaseIT RNase enzyme was added and samples incubated at 37°C. The presence of the

desired DNA insert was determined via restriction enzyme digest analysis and sequencing.

2.6.2 HiYield Plasmid Miniprep (Real Biotech Corporation)

A 5ml overnight culture was grown in the presence of the appropriate selective antibiotic. 1.5mls of this culture was centrifuged at 18,000*g* for 1 minute, the supernatant discarded and the pellet resuspended in 200µl of buffer PD1. 200µl of buffer PD2 was then added and mixed gently by inverting the tube. After a 2 minute incubation at room temperature 300µl of buffer PD3 was added, gently mixed and the tubes were centrifuged immediately at 18,000*g* for 3 minutes. The cleared lysate was then applied to a spin column and centrifuged at 11,000*g* for 30 seconds. The flow was discarded; the spin column returned to the collection tube. 400µl of buffer W1 was added. The tubes were then centrifuged at 11,000*g* for 30 seconds, the flow through discarded and the spin column returned to the collection tube. 600µl of wash buffer was added and the tubes were centrifuged at 8,000rpm for 30 seconds. The flow through was discarded once again and the spin column returned to the collection tube which was centrifuged at 18,000*g* for 2 minutes to dry the column matrix. The dried spin column was placed in a clean tube; 50µl of water added on the centre of the matrix, following 2 minutes incubation the tube was centrifuged for 2 minutes at 18,000*g* to elute the purified DNA.

2.6.3 Plasmid Midi-Prep (Qiagen High Speed Plasmid Midi -Kit)

A single colony containing the plasmid of interest was inoculated into 5ml of selective LB and grown for 8 hours shaking at 37°C. 1ml of this 5ml culture was added to 50ml of selective LB and grown for an additional 16 hours shaking at 37°C. Bacterial cells were harvested by centrifugation at 4°C for 15min at 6,000*g* (Sorvall RC-5B refrigerated super centrifuge – rotor GSA). The pellet was gently resuspended in 6ml Buffer P1. 6ml of Buffer P2 was added and the sample was mixed gently by inverting. Whilst the mixture was incubating for 5 minutes, the QIA cartridges were prepared by adding 6ml of chilled Buffer P3. The High Speed Maxi Tips were equilibrated by allowing of 4ml of Buffer QBT

flow through the filter via gravity flow. The cell lysate was then poured onto the Maxi Tip, the bound DNA was then washed with 20ml of Buffer QC and eluted with 5ml Buffer QF. DNA was precipitated with 3.5mls of isopropanol at room temperature. A QIAprecipitator maxi module was attached to the 20ml syringe and the precipitate was flushed through the module under constant pressure. The DNA which bound to the module was washed with 2ml 70% ethanol. The module membrane was dried by forcing air through it using the syringe plunger. Once dry the DNA was eluted with 1ml of TE Buffer. DNA concentration was determined spectrophotometrically at 260nm (section 2.6.5) and visualised on a 1% agarose gel.

2.6.4 Maxi preps. (Qiagen High Speed Plasmid Maxi Kit)

A single colony containing the plasmid of interest was innoculated into 5ml of selective LB and grown for 8 hours shaking at 37^{0}C. 1ml of this 5ml culture was added to 500ml of selective LB and grown for an additional 16 hours shaking at 37^{0}C. Bacterial cells were harvested by centrifugation at 4^{0}C for 15min at 6,000g (Sorvall RC-5B refrigerated super centrifuge – rotor GSA). The pellet was gently resuspended in 10ml Buffer P1. 10ml of Buffer P2 was added and the sample was mixed gently by inverting. Whilst the mixture was incubating for 5 minutes, the QIA cartridges were prepared by adding 10ml of chilled Buffer P3. The High Speed Maxi Tips were equilibrated by allowing of 10ml of Buffer QBT flow through the filter via gravity flow. The cell lysate was then poured onto the Maxi Tip, the bound DNA was then washed with 60ml of Buffer QC and eluted with 15ml Buffer QF. DNA was precipitated with 10.5mls of isopropanol at room temperature. A QIA precipitator maxi module was attached to the 30ml syringe and the precipitate was flushed through the module under constant pressure. The DNA bound to the module was washed with 2ml 70% ethanol. The module membrane was dried by forcing air through it using the syringe plunger. Once dry the DNA was eluted with 1ml of TE Buffer. DNA concentration was determined spectrophotometrically at 260nm (section 2.6.5) and visualised on a 1% agarose gel.

2.6.5 DNA/RNA quantification

DNA and RNA can be quantified based upon it absorbance at 260nm using the following formulae;

OD 260nm X dilution factor X 50 = μg/μl DNA

OD 260nm X dilution factor X 40 = μg/μl RNA

2.7 DNA Sequencing

2.7.1 ABI Prism Capillary Sequencing

Each reaction contained 2μl Terminator ready reaction mix, 0.4μl of 5X sequencing buffer, 3.2pM of the forward or reverse primer and 200pM – 400pM of DNA. The DNA was amplified using the following sequencing program: 96°C for 2 minutes followed by 25 cycles of 96°C for 10 seconds, 50°C for 30 seconds, 60°C for 4 minutes. PCR products were precipitated by adding 14.0μl dH_20, 3.0μl 3M NaOAc, pH 4.6 and 62.5μl 95% non-denatured ethanol. Tubes were vortexed and incubated at -20°C for a minimum of 30minutes Tubes were then centrifuged for 20 minutes at 18,000*g*, the supernatant aspirated carefully and the pellet was washed with 75% ethanol. Once air-dry the pellet was re-suspended in 25ul of Template Suppressor Reagent, denatured at 95°C for 5 minutes and immediately placed on ice. Samples were then loaded onto the ABI Sequencer for analysis.

2.8 Tissue Culture Techniques

2.8.1 Revival of Frozen cell stocks

Cells were defrosted on ice. Once thawed, 10ml of pre-warmed medium was added to the cell suspension and the cells were then centrifuged for 5 minutes at 1000rpm (IEC Centra-3c bench top centrifuge). The supernatant was then removed; cells were gently re-suspended in 10ml of pre-warmed medium and plated into a $60cm^2$ tissue culture dish. The cells were grown at 37°C in the presence of 5% CO_2.

2.8.2 Sub-culturing of mammalian cells

The medium was pipetted off and the cells rinsed with 10ml PBS to remove traces of medium. 2mls of Trypsin-EDTA was added and the plate incubated at 37°C for 5 minutes. 10ml of warm DMEM (FCS inactivates Trypsin-EDTA) was added and the remaining cells detached by pipetting the medium against the monolayer. The cell suspension was centrifuged at 1000rpm (IEC Centra-3c bench top centrifuge) for 5 minutes. The supernatant was discarded and the cells were re-suspended in 10mls of medium 3mls was added to each new plate and brought up to 10ml. The cells were grown at 37°C in the presence of 6% CO_2.

2.8.3 Counting mammalian cells

Cells were harvested as described 2.8.2, once re-suspended an aliquot of the cell suspension was taken and a 1:5 dilution with trypan blue was prepared. 20µl of this suspension was placed on a haemocytometer. The haemocytometer was transferred to a microscope and viewed under the 10x objective. The cells were then counted. This involves counting the number of cells in each 4x4 grid and and calculating the average number of cells. This was the number of cells by $1x10^4$, the dilution factor was adjusted (5X) for and the number of cells per ml calculated.

2.8.4 Freezing down cell stocks

Freezing ampules were labelled with the cell name, number of cells/ml, passage number, date and pre-cooled in the specialised "Mr. Frosty" rapid freeze alcohol container. Once again the cells were harvested and counted as described 2.8.2 and 2.8.3. Dilutions were made such that each 500µl contained an average of $2x10^6$ cells. An equal volume of freezing media (20% DMSO, 60% FCS and 20% complete DMEM) was added and the 1ml then placed in the pre-cooled ampules and place in the –80°C. Cells can be stored in the –80°C for up to 3 months thereafter being transferred to liquid nitrogen for long-term storage.

2.8.5 Transfection of HeLa Cells using Lipofectamine 2000

Cells were harvested, counted and plated in antibiotic free DMEM so that the cells would have become 90% – 95% confluent at the time of transfection. All transfections unless otherwise stated were carried out in 6 well plates (Nunclon). $5x10^5$ HeLa cells were seeded per well of the 6 well plate. Using the 6 well format a total of 4.0µg of DNA or 40pm of siRNA was diluted in 250µl of Optimem (specialised serum free medium) and 5µl of Lipofectamine 2000 was diluted in 250µl of Optimem per well. After 5 minutes incubation the diluted DNA or siRNA and the diluted lipofectamine were combined and complexes allowed form for 30 minutes. The 500µl mix was then added to the well, swirled gently and placed in the 37°C incubator. 24 hours later the media was changed and assayed at 24 -48 hours post transfection.

2.8.6 Genomic DNA extraction from cultured cells (Gentra Genomic DNA Purification kit), Extraction from 10-20 million cells.

The cells were harvested via trypsinisation and centrifuged at 500g for 3 minutes. The supernatant was removed leaving behind 200µl of residual liquid in which the cells were then resuspended. 3mls of Lysis solution was added and the cells were vortexed for 10

seconds, 15µl of RNase A solution was added and mixed by inverting the tube 25 times. Following incubation at 37°C for 1 hour the sample(s) were placed on ice for 3 minutes after which 1ml of protein precipitation solution was added. The cell lysate was vortexed for 20 seconds to uniformly mix the protein precipitation solution. The solution was centrifuged at 2000g to pellet the precipitated protein, the supernatant containing the DNA was transferred to a clean tube containing 3mls of 100% Isopropanol. The sample was centrifuged at 2,000g for 3 minutes, the supernatant poured off and the pellet washed in 70% EtOH. The sample was centrifuged at 2000g for 1 minute the supernatant discarded and the pellet allowed to air dry. Once dry 250µl of DNA hydration solution was added, heated to 65 °C for an hour and allowed to resuspend overnight at room temperature.

2.9 RNA Techniques

2.9.1 RNA Extraction.

RNA was extracted from cells in 6 well plates. The cells were rinsed in PBS and 1ml of Trizol Reagent (MRC) added to each well. The cells were suspended in the Trizol by repeatedly pipetting up and down and left to incubate at room temperature for 10 minutes. The cell suspension was then transferred to a 1ml eppendorf, 200µl of chloroform added and vortexed for 15 seconds. The tubes were centrifuged at 16,000*g* for 15 minutes at 4°C. 400µl of the upper phase was transferred to a new eppendorf. 333µl of isopropanol was added and incubated at room temperature for 15 minutes to allow the RNA to precipitate. The suspension was centrifuged at16, 000g for 15 minutes at 4°C. The pellet was washed with 75% EtOH, air dried and resuspended in 50µl nuclease free dH_20. Trace DNA was removed using DNase enzyme.

2.9.2 DNase Treatment

1 U of RQ1 RNase free DNase (Promega) enzyme was added to 50µl RNA in the presence of 1X RQ1 RNase free DNase buffer. After the 30-minute incubation at 37^{0}C the enzyme was removed via phenol/chloroform extraction (section 2.4.5).

2.9.3 RNA extraction using RNeasy RNA extraction kit (Qiagen)

Trypsin harvested cells were lysed in 350 – 600µl of buffer RLT supplemented with 10µl/ml β-mercaptoethanol. One volume of 70% EtOH was then added and mixed gently. 700µl of the solution was applied to a spin column and centrifuged at 14 000*g* for 15 seconds and the flow through discarded and repeated if required. 350µl of buffer RW1 was added and centrifuged at 14 000*g* for 15 seconds and the flow through discarded. Per column 10µl of DNaseI was added to 70µl of buffer RDD, mixed gently, this 80µl was then applied to each column and incubated for 15 minutes to 2 hours. The column was centrifuged 10 seconds at 14, 000*g*. 350µl of buffer RW1 was added and the column centrifuged at 14 000*g* for 15 seconds. The column was then placed into a fresh 2ml collection tube and 500µl buffer RPE added. The column was centrifuged at 14 000*g* for 15 seconds and the flow through discarded. 500µl of RPE was added the tube centrifuged and the tube centrifuged at 14 000*g* for 2 minutes. The flow through was discarded. The column was then centrifuged at 14 000*g* for 1 minute. The column was inserted into a 1.5ml collection tube, 50ul of RNase free H_20 added and the RNA eluted by centrifuging at 14,000*g* for 1 minute

2.9.4 Real-Time Reverse Transcription reaction (RT-PCR).

All RT-PCR reactions were carried out using the QunatiTech™ SYBR Green RT-PCR kits and the ABI Prism 7700. This system can facilitate reverse transcription and gene specific amplification in a one-tube format. The fluorescent dye SYBR green 1 binds to double stranded DNA molecules emitting a fluorescent signal on binding which is the basis of the quantification. All primers used were designed to span an intron yielding a

product of 100 – 150bp which facilitates the identification of amplification from RNA or contaminating genomic DNA.

Each individual reaction contains 1x of QunatiTech™ SYBR Green RT-PCR, 10μm of each primer, 0.2μl of QuantiTech RT mix and between 1pg-1ug of RNA with the volume brought to 20μl with nuclease free water. A master mix containing all reagents except the RNA was prepared for each primer set. The samples were analysed in duplicate or triplicate. The cycling parameters were as follows:

.

Stage 1 - Reverse Transcritption : 20 minutes at 50 °C, ***Stage 2 – Activation:*** 15 minutes at 95°C, ***Stage 3 - Amplification:*** 40 cycles of; 30 seconds at 57°C & 30 seconds at 72°C, ***Stage 4 – Dissociation:***15 seconds at 95°C, 30 seconds at 60°C & 15 seconds at 95°C.

The fluorescence data collected during the RT-PCR was presented as a sigmoidal graph of fluorescence v cycle number. No increase in fluorescence was usually detected between cycles 3 and 15 (unless RNA was very concentrated), this is the baseline. The threshold is the point at which the fluorescence emitted goes above the baseline; this is always in the log linear phase of the RT-PCR reaction. The cycle at which the fluorescence emitted by a sample crosses the threshold is called the Ct value (Threshold Cycle), the more abundant the target transcript is in a given sample lower the Ct number.

A standard curve was included in each run, in order to generate the curve 3 to 4 different concentrations of the standard were measured (usually $1x10^{-2}$ to $1x10^{-5}$) The amount of your target gene transcript must fall within the range of the standard curve to enable valid analysis. The curve is illustrated as a plot of the Ct value for each standard v log of the concentration. The standard curve allows for the concentration of the initial amount of target RNA in the sample to be calculated from the crossing point of its Ct value on the curve. The standard curve is also used to determine the efficiency of the RT-

PCR reaction. The slope of the line will have a value of -3.323 if the efficiency of the PCR reaction is 100%. It is essential that the PCR efficiency fall between 90 and 110% to facilitate meaningful analysis. The concentration of the test samples was then normalised against the expression levels of the selected housekeeping gene(s). Normalisation is achieved by dividing the quantity of the target/test RNA in a sample by the quantity of the housekeeping RNA in the same sample. Expression levels were evaluated using the relative quantification method. In the case of the work carried out in this thesis the normalised Col1a1 expression values were then weighted against Col1a1 expression levels in samples which were transfected with control non targeting siRNA duplexes or shRNA plasmids. Cola1a1 expression values were set at 100% in those cells transfected with control nucleic acids and the relative Col1a1 expression values in the test samples calculated as a percentage of this.

Chapter 3

Evaluation of Col1a1 targeting siRNAs in HeLa Cells

3.1 Introduction

Osteogenesis Imperfecta is caused by mutations in the Col1a1 (α1) or Col1a2 (α2) genes which encode for the α1(I) and α2(I) collagen proteins. The α1 mutations which give rise to OI Types II – IV (section 1.3) act in a dominant negative manner. As described previously (section 1.3), the type I collagen triple helix consists of 2 α1(I) chains and a single α2(I) chain. In the case of Col1a1 linked OI, cells will produce both a wild type and mutated α1(I) chain, consequently 75% of the pro-collagen triple helices will have one or two mutated chains and the remaining 25% will contain two wild type α1(I) chains. Depending upon the type of mutation the mutated heterotrimers may be targeted for destruction prior to secretion and incorporation into the collagen fibrils, this potentially could result in a 75% reduction in available type I collagen (Pochampally, *et al.*, 2005). Alternatively, the abnormal collagen triple helix may be incorporated into the collagen fibril, the presence of the α1(I) mutant chain interrupts and disorganises the type I collagen fibril organisation. The abnormal and irregular fibril arrangement severely impairs the structural integrity of the bone. Due to the difficulties associated with treating OI it appears that the most effective approach to relieving the pathology associated with the disorder may be to prevent incorporation of the mutant α1(I) chain into the triple helix.

There are a variety of molecular tools available for the sequence specific targeting of mRNA for degradation including antisense oligonucleotides (ASO), ribozymes and RNAi technology (section 1.3). All of these have successfully been applied in the sequence specific down regulation of Col1a1 gene expression(Colige A, 1993; Grassi, *et al.*, 1997; Millington-Ward, 2004; Millington-Ward, *et al.*, 1999; Millington-Ward, *et al.*, 1997; Millington-Ward. S, 1999; Peace, *et al.*, 2005; Pereira, *et al.*, 1998; Smicun, 2003; Toudjarska I, 2001). Whilst effective suppression reagents, antisense oligonucleotides are limited by their short half life and toxicity associated with chemical modifications employed to extend their half life. Ribozymes may be limited by their specific target

sequence requirements and in addition it has also been difficult to replicate *in vitro* ribozymes cleavage efficiencies *in vivo* (Millington-Ward, *et al.*, 2005).

For this reason it was decided to explore the use of RNAi (section 1.4) for down regulation of Col1a1 gene expression. RNAi is an endogenous sequence-specific mRNA silencing pathway, induced by the presence of long double stranded RNA. Initially the use of this pathway to induce suppression of mRNA in mammalian cells was thought to be precluded, as long dsRNA induces the interferon response resulting ultimately in death of the cells (Stark, *et al.*, 1998). Fortuitously it was subsequently discovered that synthesised small interfering RNAs (siRNA) of approximately 15 -27 nucleotides in length could induce the RNAi response without an immunological response (Elbashir, *et al.*, 2001a). This observation led to the application of RNAi technology in many areas of human genetics. More specifically, in relation to Col1a1 linked OI, RNAi was deemed as a potentially powerful technology to enable silencing of the mutated form of the gene (Millington-Ward, 2004).

Whilst the availability of potent suppression agents is a prerequisite for the therapeutic approach under evaluation during the course of this study, OI presents additional challenges when considering the design of effective therapeutics. As discussed in the introduction OI is a genetically heterogeneous disorder, thus far over 350 OI linked mutations in the Col1a1 and Col1a2 genes have been identified http://www.le.ac.uk/genetics/collagen/. Therefore, the design of siRNAs to target each of the individual Col1a1 mutations may be precluded by the prohibitively high timelines and financial costs and in particular the technical difficulties associated with the targeting of single nucleotide based mutations. To overcome this a mutation independent strategy for the suppression of Col1a1 expression has been employed in the current study. As discussed in the introduction (section 1.3.) three mutation independent suppression strategies have been developed by members of the research team in TCD (Millington-Ward, *et al.*, 1997) in which either 3' or 5' untranslated (UTR) sequences, the coding

sequence or a polymorphism is targeted by a suppression agent. The suppression studies described thorough out this thesis are focused upon targeting the coding sequence of the Col1a1 genes such that expression from both Col1a1 alleles should be reduced if the suppressors are sufficiently potent.

The work described in this chapter of the thesis is focused upon a mutation independent approach to suppression in which 5 different siRNAs were designed to target the coding sequence of the Col1a1 gene. In order to assess the potency of each of the siRNA duplexes for the suppression of the target Col1a1 mRNA, a cell line which endogenously expressed human Col1a1 was required. HeLa cells, a human cervical cancer cell line, represent an ideal *in vitro* cell model due to their ease of culturing, transfection, 24 hour doubling time and expression of Col1a1 at sufficiently high levels to facilitate meaningful analysis of the potency of the Col1a1 targeting siRNAs. The following chapter describes the design and generation of the siRNA duplexes, the culturing and transfection of the HeLa cells. Col1a1 expression levels in cells transfected with both the Col1a1 targeting and control siRNAs were evaluated using real-time RT-PCR. The Col1a1 mRNA expression data was analysed for statistical significance and is presented and discussed in this chapter.

3.2. Materials and Methods

The specific methods and materials employed to carry out the research presented in this chapter are detailed below, all general methods and primer sequences are described in full in chapter 2 of this thesis.

3.2.1 Design of Col1a1 and non-targeting siRNAs.

The four for silencing siRNA package was obtained from Qiagen. These siRNA duplexes are high performance guaranteed such that at least two will knockdown expression by 70% or more. The Hi-Performance Algorithm employed in the design of the Col1a1 targeting siRNA duplexes searched for the most potent siRNA molecules with minimal off target effects. The duplexes (siCol1, siCol2, siCol3 & siCol5) were provided annealed and were resupsended in 1ml of the siRNA Suspension Buffer (100mM potassium acetate 30mM HEPES-KOH and 2m magnesium acetate at a concentration of 20pm/µl. An additional siRNA (siCol4) molecule initially evaluated by Dr. D. Allen Trinity College Dublin and found to suppress Col1a1 expression by 85% in COS 7 cells was also included in the study. In addition two control siRNAs were also purchased from Qiagen, a GFP targeting siRNA (siGFP) and a non targeting siRNA which does not target any known human genes (siNon-Tgt). siRNA target sequences, oligonucleotides and annealed duplexes are outlined in Figure 3.1 to Figure 3.3.

3.2.2 HeLa cell culture

HeLa cells were grown at 37°C and 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 5% L-Glutamine, 5% Sodium Pyruvate and 10% fetal calf serum. The cells were cultured until about 85% confluent at which point they were sub-cultured as described in section 2.8.2. Briefly, the medium was pipetted off and the cells rinsed with 10ml PBS to remove any remaining traces of medium. 2mls of Trypsin-EDTA was added and the plate incubated at 37°C for 5 minutes. 10ml of warm DMEM was

added (the fetal calf serum in the medium inactivates the Trypsin-EDTA) and the remaining cells detached by pipetting the medium against the monolayer. The cell suspension was centrifuged at 1400*g* for 5 minutes and the supernatant discarded. If the cells were to be plated into a 6 well plate and the cells were re-suspended in 2mls of medium, counted (section 2.8.3) and $5x10^5$ HeLa cells plated per well. Alternatively, the cells were resuspended in 6mls, split 1:3 and replated into 60mm tissue culture dishes.

3.2.3 RNA extraction from HeLa cells.

In all cases RNA was extracted from HeLa cells which had been plated in 6 well plates. The cells were rinsed in PBS and 1ml of Trizol Reagent (MRC) added to each well, pipetted up and down and left to incubate at room temperature for 10 minutes. The cell suspension was then transferred to a 1ml eppendorf and 200µl of chloroform added and then vortexed for 15 seconds. The tubes were centrifuged at 17,000*g* for 15 minutes at 4°C. 400µl of the upper phase was transferred to a new eppendorf. 333µl of isopropanol was added and incubated at room temperature for 15 minutes to allow the RNA to precipitate. The suspension was centrifuged at 17,000*g* for 15 minutes at 4°C, washed with 75% EtOH, air dried and resuspended in 50µl nuclease free dH_20. DNA contamination was removed using the DNase enzyme RQDNaseI as described in section 2.9.2.

3.2.4 Real time RT-PCR analysis endogenous levels of Col1a1 expression in a selection of Human Cell lines

Two human cell lines, 293 and HeLa cells were available for use as the *in vitro* model for evaluation of the potency of the 5 Col1a1 targeting siRNAs. It was necessary to determine if either cell line expressed Col1a1 at sufficiently high levels to facilitate meaningful analysis of the Col1a1 targeting siRNAs. As the Col1a1 expression levels of the cells was unknown mesenchymal stem cell RNA was used to prepare the standard curve, it was

also used a positive control against which the levels of Col1a1 expression in the 293 and HeLa cells could be compared. $5x10^5$ cells per were plated in duplicate per well of a 6 well plate for each cell line. i.e. 293, HeLa and MSCs. RNA was trizol extracted (section 2.9.1) from each of the cell lines and a single real time RT-PCR run was carried out, with the reactions set up as described in 2.9.4.

3.2.5 Lipofectamine 2000 based siRNA transfection of HeLa cells

As described in the introduction (section 1.4) off target siRNA effects can be induced by the introduction of excessively high concentrations of siRNA. A series of optimisation studies carried out by members of our group had previously established that 40pm of siRNA per transfection i.e. per single well of a 6 well plate was more than sufficient to elicit an effective RNAi response. Hence this concentration was used in all transfection experiments. All HeLa cell transfections were carried out as described in section 2.8.5. Cells were harvested, counted and $5x10^5$ HeLa cells seeded per well of a 6 well plate in antibiotic free DMEM such that cells were 85% – 90% confluent at the time of transfection. 40pm (2µl) of each siRNA, siCol1 -siCol5, siGFP & siNon-Tgt were diluted in 250µl of Optimem, 5µl of Lipofectamine 2000 was diluted in 250µl of Optimem per well. After 5 minutes incubation the diluted siRNA and the diluted lipofectamine were combined and siRNA:liposome complexes allowed form for 30 minutes. The 507µl mix (2µl siRNA, 5µl lipofectamine 2000 and 500µl Optimem) was then added per well of HeLa cells, swirled gently and placed in the 37°C incubator. Each Col1a1 targeting siRNA (siCol1 – siCol5) was transfected in triplicate (i.e. 3 separate wells of HeLa cells) and the control siRNAs, (siGFP & siNon-Tgt) were transfected in duplicate (i.e. 2 separate wells of HeLa cells) in a single transfection experiment. RNA was extracted 24 hours later. Each transfection experiment was carried out three times. This generated 9 independent RNA samples for each Col1a1 targeting siRNA (siCol1 – siCol5) and 6 independent RNA samples for each of the control siRNA (siGFP & siNon-Tgt). The 3 RNAs generated per Col1a1 targeting siRNA in a single transfection experiment were pooled, generating a

single pooled RNA sample representing 3 separate transfections. This was carried out for all 3 experiments generating 3 pooled RNA samples per Col1a1 targeting siRNA, which represents a total of 9 transfections for each Col1a1 targeting siRNA. The RNAs generated for each control siRNA siGFP and siNon-Tgt were also pooled generating 3 pooled RNA samples each representing 6 transfections for siGFP and siNon-Tgt. (Note in the case of siCol4 four transfection experiments were carried out generating a total of 4 pooled RNA samples).

3.2.6 Quantitative Real time RT-PCR analysis of Col1a1 expression

Col1a1 expression levels were evaluated using real time RT-PCR and the relative quantification method. All real time RT-PCR reactions were set up as described in section 2.9.4. Each pooled RNA sample was analysed in triplicate (i.e. 3 wells per RNA) per primer set in a single 96 well plate RT-PCR assay. Each 96 well plate real time RT-PCR assay was run 3 – 5 times. In relative quantification method the Col1a1 expression data generated from all runs was normalized, this is achieved by dividing the Col1a1 expression value in a given sample by the housekeeper expression value in the same sample. In this case Col1a1 is divided by GAPDH or β-Actin value from the same sample. The relative Col1a1 expression value is determined by comparing the normalized Col1a1 expression values from HeLa cells transfected with Col1a1 targeting siRNAs (siCol1 – siCol5) to the normalized Col1a1 expression values from HeLa cells transfected with the non-targeting control siRNAs, siGFP and siNon-Tgt. Statistical analysis of the relative expression values was then carried out using GraphPad InStat3. The mean and standard deviation of the data generated for each sample was determined. Non parametric ANOVA anlaysis was carried out to determine if the Col1a1 expression observed in cells transfected with the Col1a1 targeting siRNAs, siCol1 – siCol5, siGFP and siNon-tgt differed to any degree of significance. ANOVA analysis was first carried out comparing siCol1 – siCol5 and siGFP Col1a1 expression data sets. A P value of $p<0.05$ indicates that the Col1a1 expression data sets analysed in the ANOVA differ significantly. In order to

specifically determine which of the data sets differ post test analysis was carried out. The Dunn's multiple comapriosn test was employed, (this is also called the Friedman test but will be refered to as Dunn's multiple comparison test through out this thesis).
This test carried out pairwise analysis of every data set in the ANOVA, i.e the siRNA controls (siGFP or siNon-Tgt) Col1a1 expression data to all Col1a1 targeting siRNA (siCol1 – siCol5) data, and each of the Col1a1 targeting Col1a1 expression data (siCol1 – siCol5) to each other. A p value of $p<0.05$ indicates that the two data sets compared differ significantly whilst a p value of $p>0.05$ indicates that the two data sets compared are not significantly different. The ANOVA and Dunn's multiple comparison analysis was carried for both GAPDH and β-Actin normalised data.

A) siCol1

Target RNA *Col1a1 NM_000088*

Nucleotides 613- 633 5' ATGGCTATGATGAGAAATCAA 3'

Sense 5' r(GGCUAUGAUGAGAAAUCAA)d(TT) 3'

Antisense 5' r(UUGAUUUCUCAUCAUAGCC)d(AT) 3'

Duplex 5' GGCTATGATGAGAAATCAATT 3'

3' ATCCGAUACUACUCUUUAGUU 5'

B) siCol2

Target RNA *Col1a1 NM_000088*

Nucleotides 1404-1404 5' CCAAGGGTAACAGCGGTGAA 3'

Sense 5' r(CAAGGGUAACAGCGGUGAA)d(TT) 3'

Antiense 5' r(UUCACCGCUGUUACCCUUG)d(GG) 3'

Duplex 5' CAAGGGUAACAGCGGUGAATT 3'

3' GGGUUCCCAUUGUCGCCACUU 5'

Figure 3.1 Col1a1 siRNA target sequences, oligonucleotides & duplexes. A) siCol1 and B) siCol2.

C) siCol3

Target DNA *Col1a1 NM_000088*

Nucleotides 5572 -5592 5' AAGGTGCTATTTAACATGGGA 3'

Sense 5' r(GGUGCUAUUUAACAUGGGA)d(TT) 3'

Antisense 5' r(UCCCAUGUUAAAUAGCACC)d(TT) 3'

Annealed 5' GGUGCUAUUUAACAUGGGATT 3'

3' TTCCACGAUAAAUUGUACCCU 5'

D) siCol4

Target RNA *Col1a1 NM_000088*

Nucleotides *227-247* 5' ACCAATCACCTGCGTACAGA 3'

Sense 5' r(CAAUCACCUGCGUACAGAA)d(TT) 3'

Antiense 5' r(UUCUGUACGCAGGUGAUUG)d(GT) 3'

Annealed 5' CAAUCACCUGCGUACAGAATT 3'

3' GTGUUAGUGGACGCAUGUCUU 5'

Figure 3.2 siRNA target sequences, oligonucleotides & duplexes. C) siCol3 and D) siCol 4.

E) siCol5

Target RNA *Col1a1 NM_000088*

Nucleotides 293 -313 5' CCGATCTGCGTCTGCGACAA 3'

Sense 5' r(CAAUCACCUGCGUACAGAA)d(TT) 3'

Antisense 5' r(UUCUGUACGCAGGUGAUUG)d(GT) 3'

Annealed 5 ' CAAUCACCUGCGUACAGAATT 3'

3' GTGUUAGUGGACGCAUGUCUU 5'

C) siGFP Qiagen Control

Target DNA

Nucleotides 120 -143 *5'* CGGCAAGCTGACCCTGAAGTTCAT 3'

Sense 5' r(GCAAGCUGACCCUGAAGUUCAU) 3'

Antisense 5' r(GAACUUCAGGGUCAGCUUGCCG) 3'

Duplex 5' GCAAGCUGACCCUGAAGUUCAU 3'

3' GCCGUUCGACUGGGACUUCAAG 5'

Figure 3.3 Col1a1 siRNA target sequences, oligonucleotides & duplexes. E) siCol5, B) siGFP

D) siNon Targeting

Sense	AAUUCUCCGAACGUGUCACGT
Anti-Sense	UUAAGAGGCUUGCACAGUGCA

Duplex	AAUUCUCCGAACGUGUCACG
	UUAAGAGGCUUGCACAGUGCA

Figure 3.4 siNon -Targeting sequence oligonucleotides & duplex.

3.3 Results

3.3.1 Evaluation of endogenous Col1a1 expression levels in a selection of cell lines

The human cell lines 293T and HeLa were available for use as the *in vitro* model for evaluation of the potency of the 5 Col1a1 targeting siRNAs. It was necessary to determine if the cells expressed Col1a1 at sufficiently high levels to facilitate meaningful analysis of the Col1a1 targeting siRNAs. RNA was extracted from each of the cell lines and a single real time RT-PCR run was carried out. As the Col1a1 expression levels of the cells was unknown, mesenchymal stem cell RNA was used to prepare the standard curve and as a positive control against which the Col1a1 expression levels in 293 and HeLa cells could be compared. Col1a1 expression levels in MSCs were set at 100% for comparison purposes and Col1a1 expression was calculated to be in the region of 22.11% and 67.30% in 293 and HeLa cells respectively. As MSCs are known to highly express Col1a1 the level of expression in HeLa cells was considered to be sufficient to facilitate evaluation of the efficacy of the Col1a1 targeting siRNAs and the cells were selected for use in the Col1a1 suppression studies described in this chapter of the thesis.

3.3.2 Real time RT-PCR evaluation of Col1a1 expression levels in HeLa cell transfected with siCol1 – siCol5.

As HeLa cells proved to express Col1a1 at sufficiently high levels to jusifty further use, these cells were used in the preliminary siRNA Col1a1 suppression studies in which the efficacy of five Col1a1 targeting siRNA duplexes was evaluated. As described in 3.2.5 each Col1a1 targeting siRNA (siCol1 – siCol5) was transfected in triplicate per 6 well plate (i.e. 3 separate wells of HeLa cells) and the control siRNAs, (siGFP & siNon-Tgt) were transfected in duplicate on a 6 well plate (i.e. 2 separate wells of HeLa cells) in a single transfection experiment. Each transfection experiment was carried out three times. This generated 9 independent RNA samples for each Col1a1 targeting siRNA (siCol1 – siCol5) and 6 independent RNA samples for each of the control siRNA (siGFP & siNon-Tgt).

	MSC	*293 Cells*	*HeLa Cells*
Raw Data	0.00076	0.0003	0.0007
	0.00135	0.0002	0.0006
	0.00101	0.0002	0.0008
Average	0.00104	0.0002	0.0007
%Col1a1 expression	100%	22.11%	67.30%

Table 3.1 Evaluation of Col1a1 expression levels in MSCs, 293 cells and HeLa cells. A cell line was required which expressed Col1a1 at sufficiently high levels to facilitate meaningful analysis of the Col1a1 targeting siRNAs. The Col1a1 mRNA expression levels were evaluated in 293 and HeLa cells and compared to Col1a1 expression levels in mesenchymal stem cells (MSCs). $5x10^5$ cells were plated in duplicate per cell line (MSC, 293 and HeLa cells) in a 6 well plate. RNA was extracted from each well 48 hours later. A single real time RT-PCR run was carried out to determine Col1a1 expression levels. Col1a1 expression in MSCs was set at 100% and Col1a1 expression levels in 293 (22.11%) and in HeLa cells (67.30%) estimated by comparison.

The RNA samples generated per siRNA per transfection experiment were pooled, generating 3 RNA samples from the 9 transfections for siCol1 to siCol5 and 3 samples for the controls siGFP to siNon-Tgt from 6 transfections. (Note in the case of siCol4 four transfection experiments were carried out generating a total of 4 pooled RNA samples).

Col1a1 expression levels in all RNA samples were evaluated using real time RT-PCR and the analysed using the relative quantification method. As described in section 3.2.6 each RNA sample was analysed in triplicate (i.e. three wells per 96 well plate) per primer set (Col1a1, GAPDH and β-Actin) in a single RT-PCR run. The Col1a1 expression values were then normalized to each housekeeping gene, GAPDH and β-Actin. This is achieved by dividing the Col1a1 expression value by either the GAPDH expression value or the β-Actin expression value in the same sample. This accounts for any concentration differences in the RNA sample or pipetting error. The next step is to determine the relative Col1a1 expression values, i.e. levels of Col1a1 expression in HeLa cells transfected with Col1a1 targeting siRNAs relative to the Col1a1 expression levels in cells transfected with negative control, non-targeting siRNAs. The non targeting siRNAs used siGFP and siNon-tgt are thought to be processed in an identical manner in the cell as the Col1a1 targeting siRNAs, therefore comparative molecular and biochemical events should, in principle, be occurring in the siCol1 – siCol5 transfected HeLa cells as in the siGFP and siNon-tgt transfected cells, up to the point of RNA cleavage as siGFP and siNon-tgt have no known target genes in mammalian cells. Therefore Col1a1 expression in these cells should in principle be close to the normal endogenous expression levels and use of such controls should enable one to determine if non specific suppression of the target gene under evaluation is occurring.

The relative Col1a1 expression data calculated for each sample in every run was compiled and analysed using GraphPad InStat3. The mean and standard deviation of each RNA data set (Col1a1 expression relative to both siGFP and siNon-Tgt following normalisation to both GAPDH and β-Actin) was determined. The relative expression values are

presented as percentage (%) expression, with Col1a1 expression in cells transfected with siGFP and siNon-Tgt set at 100%. A sample calculation is presented in Table 3.2. The Col1a1 expression data calculated for each of siRNAs is presented in Table 3.3, Table 3.4 and Figure 3.7.

Each of the five siRNAs siCol1 – siCol5 reduced endogenous Co1a1 expression levels significantly. When normalized to GAPDH, the relative Col1a1 expression levels in cells transfected with siCol1– siCol5 ranged from 38.04% to 47.34% (±6.65 – 24.63%). More specifically the relative range of Col1a1 expression was 38.04 % to 42.89% (±6.65 – 20.23%) and 42.55% to 47.34% (± 7.38 – 24.63%) when compared to Col1a1 expression levels in HeLa cells transfected with siGFP and siNon-Tgt respectively. Non parametric ANOVA analysis was carried out in which levels of Col1a1 expression in HeLa cells following transfection with siCol1-siCol5, siGFP or siNon-Tg were compared. A p value of $p<0.0001$was generated from each ANOVA indicating that Col1a1 expression data sets compared in each ANOVA, i.e. siCol1 – siCol5 and siGFP or siCol1 – siCol5 and siNon-tgt differed significantly (Table 3.3 and Figure 3.5).

When β-Actin was used as the internal control housekeeping gene the relative levels of Col1a1 expression in HeLa cells transfected with siCol1 – siCol5 expression ranged from 28.75 to 39.46% (± 5.90% - 21.21%) when compared to the levels of Col1a1 expression levels in HeLa cells transfected with the control siRNAs, siGFP or siNon-Tgt. More specifically, the relative range of Col1a1 expression was 28.75% to 37.13% (± 13.77% - 19.95%) and 30.55% to 39.56% (± 5.78% - 21.21%) when compared to Col1a1 levels in HeLa cells transfected with siGFP and siNon-Tgt respectively. Non parametric ANOVA analysis was carried in which the Col1a1 expression data for each Col1a1 targeting siRNA (siCol1 – siCol5) siGFP or siNon-Tgt Col1a1 was comapred. A p value of $p<0.0001$ was generated from each ANOVA indicating that Col1a1 expression data sets compared in each ANOVA, i.e. siCol1 – siCol5 and siGFP or siCol1 – siCol5 and siNon-tgt differed significantly (Table 3.4 and Figure 3.5).

Following ANOVA analysis of the GAPDH or β-Actin normalized Col1a1 expression data, post test analysis was carried out to determine which of the Col1a1 data sets differed, and to determine which, if any, of the Col1a1 targeting siRNAs (siCol1 – siCol5) effected more potent suppression of Col1a1 expression in HeLa cells.The post test employed was the Dunns multiple comparison test, the test carried out pairwise comaprison of each of the 6 data sets originally compared in the ANOVA. In the case of the pairwise comparison of siGFP or siNon-tgt Col1a1 expression data to each of the siCol1 – siCol5 Co1a1 expression data sets a p value of $p<0.001$ was generated from each comparison, indicating that Col1a1 expression levels in HeLa cells transfected with each of the Col1a1 targeting siRNAs differed significantly from Col1a1 expression levels in HeLa cells transfected with siGFP or siNon-tgt (Table 3.5)

The pairwise comparisons of each of Col1a1 expression data sets generated following transfection of the 5 Col1a1 targeting siRNAs (siCol1 – siCol5) was also carried out, each pariwise comparison generated p value of $p>0.05$ indicating that the Col1a1 expression data sets did not differ significantly and therefore no single Col1a1 targeting siRNA appeared to more potently reduced Col1a1 expression levels in HeLa cells (Table 3.5)

Column	1	2	3	4	5
	GAPDH	Col1a1	Col1a1/ GAPDH	Relative expression%	
siCol4	670	261.99	0.39	36.44	
	458	187.3	0.4	37.38	
A)	343.17	83.99	0.24	22.42	
	675.87	206.88	0.3	28.03	
	3.61	1.65	0.45	42.05	
	1.5	0.819	0.54	50.46	
	3.43	1.88	0.54	50.46	
	2	0.779	0.39	36.44	
	336.15	153.56	0.45	42.05	
	608.3	307.51	0.5	46.72	
	268.12	180.79	0.67	62.6	
	288.26	114.08	0.39	36.38	
	192.68	72.55	0.37	34.51	
	29.79	11.38	0.38	35.44	
	23.06	7.23	0.31	28.91	
	14.02	4.6	0.32	29.85	
	608.3	307.51	0.5	46.72	**39.23%**
siGFP	7.12	8	1.12	104	
	11.01	10.99	1	93	
B)	683.47	571.99	0.84	78	
	142.73	195.3	1.36	127	
	2.62	2.31	0.88	82.2	
	3.65	4.2	1.15	107	
	11.54	11.25	0.97	90	
	60.06	64.18	1.06	99	
	4.53	5.24	1.15	107	
	1.16	1.38	1.19	111	***99.82%***
			1.072		

Table 3.2 Sample of relative quantification of RT-PCR data. Column 1 contains the GAPDH expression value from each RNA sample analysed; **A)** data generated from analysis of the RNA extracted from HeLa cells transfected with the Col1a1 targeting siRNA siCol4, **B)** data generated from analysis of the RNA extracted from HeLa cells transfected with non targeting control siGFP, Column 2 contains the Col1a1 expression value from each RNA sample analysed, Column 3 contains the normalised Col1a1

expression values for each sample, normalisation is achieved by dividing the Col1a1 in a sample by the GAPDH value in a sample, i.e. dividing the value in Column 2 by the value in Column 1. Column 4 contains the relative Col1a1 expression values expressed as a percentage. The relative expression values are determined by dividing each sample by the average of the normalised Col1a1 expression values of the HeLa cells transfected with the non-targeting siRNA, in this case siGFP, the average value calculated for this data set was 1.072. The normalised Col1a1 expression value of each sample is divided by 1.702, the relative expression value is multiplied by 100 so that it can be represented as a percentage. Column 5 contains the average relative Col1a1 expression value in HeLa cells transfected with siCol4, the value highlighted in blue is 39.23%. The value highlighted in green is the relative Col1a1 expression value in HeLa cells transfected with siGFP, in this data set the relative Col1a1 expression value is 99.82%, this would be represented as 100% on the graph for clarity. The RT-PCR results generated in each mRNA quantification based study detailed in this thesis were calculated in an identical manner. Note the Col1a1 expression value calculated for siCol4 does not match the result presented in the chapter all of the siGFP RT-PCR expression data is not presented in the table, the purpose of this table is the illustrate how the actual calculations were carried out.

		siGFP	*siNon-tgt*	*siCol1*	*siCol2*	*siCol3*	*siCol4*	*siCol5*	*ANOVA*
GAPDH	*siGFP*	100%		40.36%	38.65%	42.89%	38.93%	38.04%	p<0.0001
	StdDev	23.13		20.23	8.27	6.65	9.50	14.00	
	siNon-tgt		100%	43.72%	42.65%	47.34%	42.97%	42.55%	p<0.0001
	stdDev		20.90	24.63	9.10	7.38	10.47	16.65	

Table 3.3 GAPDH normalized relative Col1a1 expression values in HeLa Cells transfected with siCol1-siCol5, siGFP or siNon-tgt. Col1a1 targeting siRNAs (siCol1 – siCol5) and control siRNAs (siGFP & siNon-tgt) were transfected into HeLa cells and Col1a1 expression levels evaluated 24 hours later by RT-PCR. Col1a1 expression was normalized to GAPDH for all samples. Col1a1 expression in control siRNAs, siGFP & siNon-Tgt was set at 100% and normalized Col1a1 expression values for siCol1 – siCol5 compared to determine the relative Col1a1 expression values. Col1a1 expression ranged from 38.04 – 42.89% when compared to siGFP (Green bar) and 42.55% – 47.34% when compared to siNon-Tgt (Blue bar). Data was analysed using GraphPad InStat 3, Non-paramteric ANOVA analysis was carried out in which the relative Col1a1 expression data for each Col1a1 targeting siRNA, siCol 1- siCol 5 and siGFP or siNon-tgt was compared, a p value of $p<0.0001$ was generated in both cases indicating that the Col1a1 expression data sets compared were significantly different. In order to determine which of the data sets differed post test analysis was carried out using the Dunns multiple comparison test which carries out pairwise comparison of each of the data sets evaluated in the ANOVA, for details see Table 3.5.

		siGFP	*siNon-tgt*	*siCol1*	*siCol2*	*siCol3*	*siCol4*	*siCol5*	*ANOVA*
β-Actin	*siGFP*	100%		30.66%	29.32%	36.47%	28.75%	37.13%	p<0.0001
	StdDev	16.81		12.63	8.79	6.06	13.77	19.95	
	siNon-tgt		100%	31.18%	32.55%	39.56%	30.55%	39.46%	p<0.0001
	StdDev		16.32	9.35	12.90	5.90	14.64	21.21	

Table 3.4 β-Actin normalized relative Col1a1 expression values in HeLa Cells transfected with siCol1-siCol5, siGFP or siNon-tgt Col1a1 targeting siRNAs (siCol1 – siCol5) and control siRNAs (siGFP & siNon-Tgt) were transfected into HeLa cells and Col1a1 expression levels evaluated 24 hours later by RT-PCR. Col1a1 expression was normalized to β-Actin for all samples. Col1a1 expression in control siRNA, siGFP & siNon-Tgt was set 100% and normalized Col1a1 expression values for siCol1 – siCol5 compared to determine the relative Col1a1 expression values.Col1a1 expression ranged from 28.75% - 37.13% when compared to siGFP (Green bar) and from 30.55% to 39.56% when compare to siNon-Tgt (Blue bar). Data was analysed using GraphPad InStat 3, Non-parametric ANOVA analysis was carried out in which the relative Col1a1 expression data for each Col1a1 targeting siRNA, siCol 1- siCol 5 and siGFP or siNon-tgt was compared, a p value of $p<0.0001$ was generated in both cases indicating that the Col1a1 expression data sets compared were significantly different. In order to determine which of the data sets differed post test analysis was carried out using the Dunns multiple comparison test which carries out pairwise comparison of each of the data sets evaluated in the ANOVA, for details see Table 3.5.

Figure 3.5. Graph of Col1a1 expression levels in HeLa cells following transfection with each of the 5 Col1a1 targeting siRNAs. Five Col1a1 targeting siRNAs (siCol1 – siCol5) and two control siRNAs (siGFP & siNon-Tgt) were transfected into HeLa cells and Col1a1 expression levels evaluated 24 hours later using RT-PCR. Col1a1 expression was normalized to GAPDH or β-Actin for all samples. Col1a1 expression in control siRNA, siGFP & siNon-Tgt was set 100% and normalized Col1a1 expression values for siCol1 – siCol5 compared to determine the relative expression values. In the case of GAPDH normalised data, Col1a1 expression ranged from 38.04 – 42.89% when compared to siGFP (Green bar) and 42.55% – 47.34% when compared to siNon-Tgt (Blue bar). In the case of β-Actin normalised data, Col1a1 expression ranged from 28.75% - 37.13% when compared to siGFP (Green hatched bar) and from 30.55% to 39.56% when compare to siNon-Tgt (Blue hatched bar). All data was analysed using GraphPad InStat 3, the mean Col1a1 expression levels in HeLa transfected with siCol1 – siCol5, siGFP and siNon-Tgt and standard deviations were calculated. Non-parametric ANOVA analysis (of both the GAPDH and β-Actin normalised data sets) was carried out in which siCol1 – siCol5, siGFP or siNon-Tgt Col1a1 expression data sets were compared a p value of $p<0.0001$, indicating that the Col1a1 expression data sets differed significantly. In order to determine which of the data sets differed post test analysis was carried out using the Dunns multiple comparison test which carries out pairwise comparison of each of the data sets evaluated in the ANOVA, for details see Table 3.5. The error bars represent the standard deviation of each data set, and the ANOVA p value is indicated.

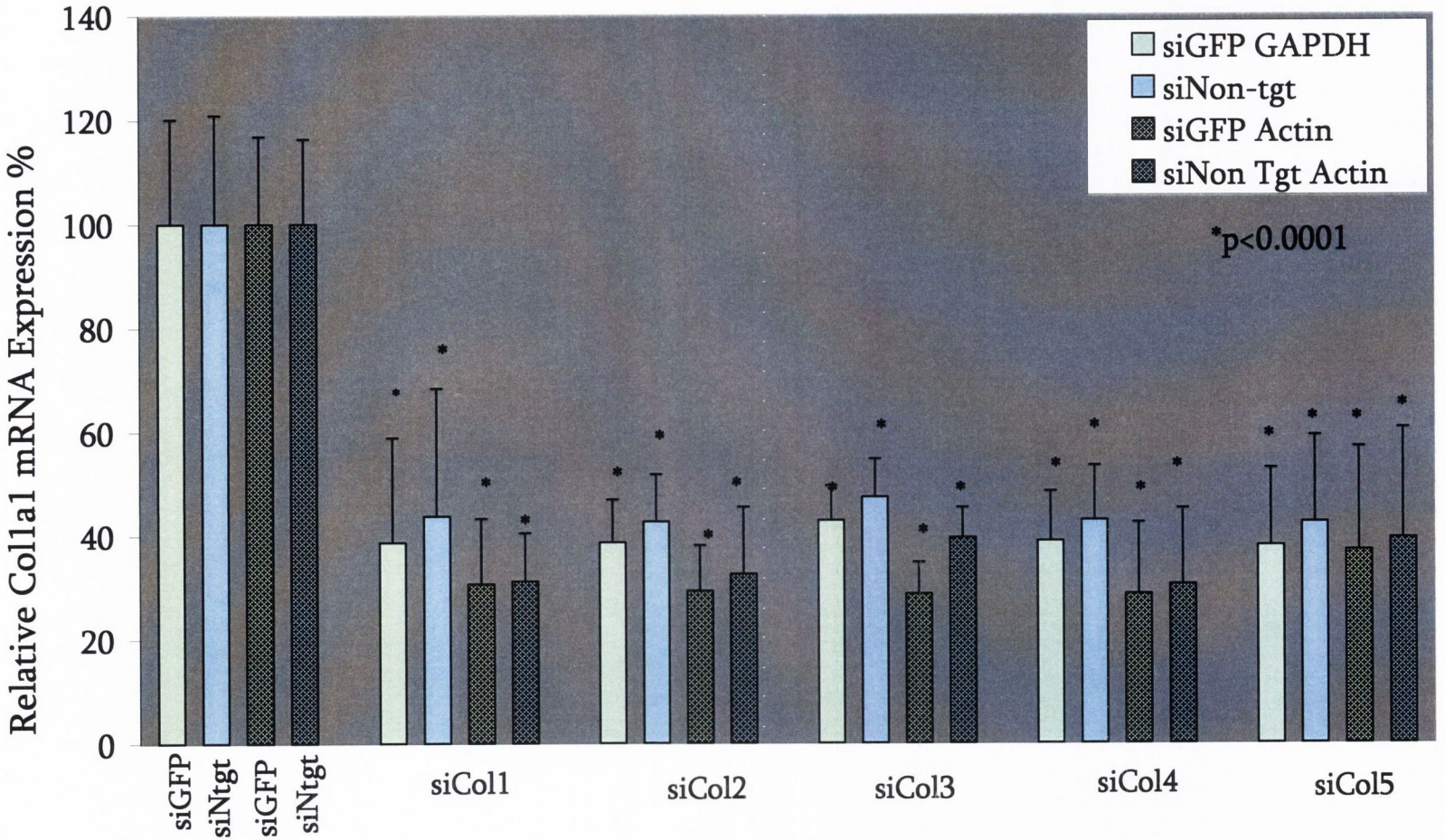


Figure 3.5. Graph of Col1a1 expression levels in HeLa cells following transfection with each of the 5 Col1a1 targeting siRNAs

	siCol1	*siCol2*	*siCol3*	*siCol4*	*siCol5*
siGFP	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
siNon-tgt	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
siCol1	p>0.05	p>0.05	p>0.05	p>0.05	p>0.05
siCol2			p>0.05	p>0.05	p>0.05
siCol3				p>0.05	p>0.05
siCol4					p>0.05

Table 3.5 Table of results generated from Dunn's multiple comparisons test.
Following ANOVA analysis in which the levels of Col1a1 expression in HeLa cells following transfection with siCol1 – siCol5 and siGFP or siNon-Tgt were compared, Dunn's multiple comparison post test analysis was carried out, this analysis carried out pairwise analysis of each of the data sets. As the results generated from ANOVA and Dunns multiple comparison analysis of GAPDH and β-Actin normalised data were the same this table represents both data sets. Pairwise comparison of Col1a1 expression data from HeLa cells transfected with each of the 5 Col1a1 targeting siRNAs to Col1a1 expression in HeLa cells transfected with siGFP or siNon-tgt generated a p value of p<0.001 from each of the comparisons, indicating the the Col1a1 expression levels in HeLa cells transfected with any of the 5 Col1a1 targeting siRNAs differed significantly from Col1a1 expression levels in HeLa cells transfected with siGFP or siNon-tgt. Comparison of Col1a1 expression data sets generated for each of the 5 Col1a1 targeting siRNAs (siCol1 – siCol5) to one another generated p values of p>0.05 indicating that the Col1a1 expression levels in HeLa cells transfected with each of the Col1a1 targeting siRNAs did not differ, therefore it appears that no single Col1a1 targeting siRNA more potently reduces Col1a1 expression in HeLa cells. All data was evaluated using Graph Pad Instat 3.

3.4 Discussion

As it had previously been demonstrated that RNAi is an effective means by which endogenous Col1a1 expression levels can be reduced (Millington-Ward, 2004), the main aim of the work described in this chapter of the thesis was to design and evaluate a selection of Col1a1 targeting siRNA duplexes. To this end five Col1a1 targeting siRNAs were evaluated in HeLa cells.

It was decided that a suitable *in vitro* model for this work would be a cell line which was of human origin, plentiful, easy to transfect and endogenously expressed Col1a1 at sufficiently high levels to undertake subsequent analysis by RT-PCR. HeLa cells, one of the most extensively used cell lines, meet most of these criteria but the Col1a1 expression profile of the cells was unknown. Real time RT-PCR quickly revealed that these cells did indeed express Col1a1, expression levels were found to be approximately 60% of that in mesenchymal stem cells which express Col1a1 at high levels. This level of expression was considered to be sufficiently high and HeLa cells a suitable *in vitro* system for the testing of the Col1a1 targeting siRNAs (Table 3.1).

Five Col1a1 targeting siRNAs siCol1 – siCol5 were evaluated in the study described in this chapter of the thesis. The potency of the Col1a1 targeting siRNAs was evaluated by real time RT-PCR using the relative quantification method. The Col1a1 expression levels were normalized against expression of housekeeping gene(s). To achieve normalisation the quantity of the target gene mRNA in a sample is divided by the quantity of the housekeeping mRNA in the same sample (in this study GAPDH and β-Actin), this is to reduce potential qualitative and quantitative differences caused by possible variation in the initial sample amount, possible RNA degradation, differences in sample quality, pipetting errors and potential variations in cDNA synthesis efficiency. The normalized Col1a1 expression values in cells transfected with Col1a1 targeting siRNAs were then divided by the normalized Col1a1 expression values from cells transfected with the non targeting control siRNAs (in this study siGFP and siNon-tgt) to reduce any possible variations caused by the presence of siRNAs on the

cell machinery which may manifest at the mRNA level. The value obtained is called the relative expression value, i.e. the expression of Col1a1 in cells transfected with a Col1a1 targeting siRNA relative to Col1a1 expression values in cells transfected with control siRNAs, siGFP or siNon-Tgt.

The evaluation of the efficacy of the Col1a1 targeting siRNAs was quite extensive. In each transfection experiment each of the 5 Col1a1 targeting siRNAs, siCol1 – siCol5 was transfected in triplicate. This generated 3 RNA samples per Col1a1 targeting siRNA, the samples were pooled generating a single sample which represented three transfections. A total of 3 transfection experiments were carried out, therefore a total of 3 pooled samples was generated for each of the 5 Col1a1 targeting siRNAs, siCol1-siCol5, which represented a total of 9 transfections per siRNA, Additionally two non-targeting controls siGFP and siNon-Tgt were included in the analysis. Each was transfected in duplicate in each of transfection experiment; the two RNA samples were pooled generating 3 RNA samples each for siGFP and siNon-Tgt representing 6 transfections each.

The Col1a1 expression levels in each sample were evaluated using real time RT-PCR. Two housekeeping genes were included in the assay, each sample was analysed in triplicate per primer set in a single RT-PCR run and each run carried out 3 – 5 times. Col1a1 expression data was normalised to both GAPDH and β-Actin, and the relative Col1a1 expression levels determined by comparing the GAPDH or β-Actin normalised Col1a1 expression levels determined for each Col1a1 targeting siRNA to the GAPDH and β-Actin normalised Col1a1 expression levels in HeLa cells transfected with siGFP or siNon-Tgt. The non targeting controls are thought to be processed in an identical manner as the Col1a1 targeting siRNAs, it is thought that both siGFP and siNon-tgt are incorporated into the RISC complex. Therefore, in principle, one would expect that similar molecular and biochemical processes are occurring in the HeLa cells transfected with the Col1a1 targeting siRNAs and the non targeting siRNAs, except for the final step of mRNA cleavage, as the non targeting siRNAs are not known to have homology to any known genes. Therefore, the use of

such controls should in theory enable one to determine if suppression of the target gene under evaluation is due to sequence specific suppression events.

It is clear form the rigorous testing of the 5 Col1a1 duplexes that siRNA mediated RNAi is indeed an effective approach for mutation independent suppression of endogenous Col1a1 expression levels. The reduction in Col1a1 expression in HeLa cells transfected with each of the 5 Col1a1 targeting siRNAs ranged from 28.75% – 47.34% (± 5.90% - 21.21%) compared to Col1a1 expression levels in HeLa cells transfected with the non targeting control siRNAs Table 3.1, Table 3.2 and Figure 3.1. In order to determine if these reduced levels of Col1a1 expression were significant non parametric ANOVA analysis was carried out in which the Col1a1 expression data generated from HeLa cells following transfection with the Col1a1 targeting siRNAs siCol1 – siCol5 and the non targeting siRNAs, siGFP or siNon-tgt were compared. It was found the Col1a1 expression data sets differed significantly, a p value of $p<0.0001$ was generated when the GAPDH or Actin data was analysed, for details refer to Table 3.1, Table 3.2 and Figure 3.1. Dunns multiple comparison post test analysis was carried out in order to specifically determine which of the data sets differed. This test carried out pairwise analysis of each of the data sets present in the ANOVA. Comparison of each of the siCol1 – siCol5 Col1a1 expression data sets with the siGFP or siNon-tgt Col1a1 expression data generated a p value of $p<0.001$ indicating that Col1a1 expression levels in HeLa cells transfected with each of the 5 Col1a1 targeting siRNAs differed significantly from Col1a1 expression levels in HeLa cells transfected with siGFP or siNon-tgt (Table 3.5). Additionally, pairwise analysis of each of the siCol1 – siCol5 Col1a1 expression data sets was carried out, each pairwise comparison generated a p value of $p>0.05$ in each case, indicating that there was no significant difference between the data sets. Thus it appears that none of the 5 Col1a1 targeting siRNAs more potently reduced Col1a1 expression in HeLa cells (Table 3.3). As the siRNAs which were purchased from Qiagen they came with a guarantee that each would reduce target suppression by at least 70%. Obviously this 70% suppression is achievable under Qiagen optimized experimental conditions, nevertheless suppression rates in our hands ranged from 52.66% to 71.25% which indeed demonstrates the efficacy of this group of 5 Col1a1 suppressors.

Whilst the amount of mutant type I collagen down regulation required to effect therapeutic benefit has not been defined, one could postulate that any reduction in the mutant protein burden may possibly provide an improvement in the OI phenotype. One must note that the results obtained in the initial study are preliminary and not in the target cell type (mesenchymal stem cells) for OI. In this regard additional work has to be carried out to evaluate the potency of the Col1a1 targeting siRNAs both at the RNA and protein level in MSCs, described in chapter 5 of this thesis. In addition, it is likely that using this therapeutic approach for Col1a1 suppression that wild type Col1a1 expression would have to be supplemented through the provision of Col1a1 replacement gene refractory to the Col1a1 targeting siRNAs, this aspect of the current study will be discussed in further detail in chapter 7 of this thesis.

It is of interest to note that patients mosaic for a Col1a1 mutation associated with OI are frequently asymptomatic, frequently only discovering that they are carriers of a mutation when it presents in their children (Cabral and Marini, 2004; Cohen-Solal, et al., 1996; Edwards, 1992; Wallis, 1990). The proportion of cells which carry the mutation will vary from tissue to tissue and between individuals. In one particular study the mutational burden of two mosaic carriers of Osteogenesis Imperfecta was evaluated in osteoblasts the primary producers of Col1a1 and effectors of bone mineralization. Both patients had normal height and bone histology. In the case of the first patient whose sons had Type IV OI, osteoblasts were cultured from bone chips harvested from a left patellar repair and right iliac bone biopsy. PCR analysis of the derived osteoblasts cultures revealed that 73% of osteoblasts obtained from the left patellar repair and 50% of osteoblasts obtained from the right illac bone biopsy were mutated. The second patient had two children with Type III OI. Histology of autopsy samples from rib and calvarium was normal; PCR analysis of DNA extracted from autopsy clavarial tissue revealed that 40% of cells carried the mutation. It was highlighted that DNA was primarily isolated from osteoblasts and osteocytes although presence of other cells could not be excluded (Cabral and Marini, 2004). Nevertheless, it appears that the skeleton may have the potential to bear substantial mutational burden (40% - 70%) with only mild phenotypic affects. This suggests that the body

may have a tolerance or a threshold for the amount of mutant type I collagen that can be present before pathology will manifest. Therefore, one could hypothesise, that there may possibly be potential therapeutic benefit derived from the application of siRNAs that reduce the amount of mutated Col1a1 mRNA incorporated into type I collagen heterotrimers. In addition Type I OI, the mildest form of the disease is due to mutations which result in a null allele, mostly as a result of the introduction of a premature stop codon (Willing, et al., 1996). Patients have a 50% reduction in the amount of type I collagen biosynthesized yet the disease is mild with a low rate of fracture incidence and high mobility. Therefore, if a patient with severely deforming OI were able to reduce their mutational burden by a significant percentage and restore levels of wild type collagen biosynthesis to even 50% of normal levels one could hypothesise that there may be potential for possible clinical benefit.

A mutation independent approach to gene suppression was undertaken in the design and execution of the suppression experiments described in this chapter of the thesis. The flexibility afforded by a mutation independent therapeutic approach is pivotal to the development of a single therapeutic which can be used to treat the majority of patients with Col1a1 linked OI. The 5 siRNAs which target the wild type Col1a1 mRNA evaluated were found to be potent suppressors of endogenous Col1a1 expression *in vitro*, it is notable that whilst siRNA are excellent mediators of RNAi they do not provide long term target suppression, as they are diluted out of the target cell population as the cells divide and are also subject to nuclease degradation. To overcome this transience, vector based RNAi expression systems which enable long term suppression of the target gene will be required. With this goal in mind the two Col1a1 targeting siRNAs, siCol2 and siCol4 were used for vector based RNAi studies which are described in chapter 4 of this thesis.

Chapter 4

Evaluation of Col1a1 targeting shRNAs in HeLa Cells

4.1 Introduction.

RNAi is a powerful tool for sequence specific gene silencing (Elbashir, et al., 2001a) (section 1.4). Whilst siRNAs can potently reduce target transcript levels their application in mammalian cells is limited by the fact that the siRNA duplexes are diluted over successive cell divisions. Suppression with synthesised siRNAs is transient lasting on average approximately 3 – 5 days subsequent to transfection of siRNAs in cell culture (Amarzguioui, *et al.*, 2005). Long term *in vitro* suppression studies or investigation of proteins with long half lives may not be feasible utilising synthesised siRNAs. The *in vivo* applications of siRNA are also limited by the transient nature of siRNA based suppression. For slowly dividing or non-dividing cells such as neurons this may be somewhat less of an issue but for highly proliferative cells such as stem cells this factor is likely to be particularly relevant. It is notable that chemical modifications have been incorporated into siRNA designs to increase siRNA duplex stability and to increase the longevity of suppression (Manoharan, 2004). Whilst such modifications do significantly improve the duration of mRNA suppression in many situations stable almost permanent suppression, may be required. To address the transient nature of siRNA mediated gene suppression, siRNA expression plasmids have been developed which help overcome some of the deficiencies associated with synthesised siRNAs.

The first generation of siRNA expression plasmids generated have been driven predominantly by the U6 or H1 Pol III promoters (Brummelkamp, *et al.*, 2002a; Miyagishi, 2002; Yu, *et al.*, 2002). These promoters are ideally suited for use in siRNA expression plasmids as they drive high levels of expression of small RNAs, they are small in size with no additional sequence requirements and they have a well defined start and stop site which ensures that no additional bases are incorporated into the siRNA (Myslinski, 2001; Paule, 2000). This was considered to be of particular importance due to the strict sequence requirements of siRNA activity, whilst siRNAs can tolerate a few 3' and 5' nucleotide overhangs the incorporation of a large number of additional bases may reduce their potency, in addition it is possible that the incorporation of additional bases may result in off target suppression effects

Chapter 5

Evaluation of RNAi in Mesenchymal Stem Cells

(Amarzguiou, *et al.,* 2005; Brummelkamp, *et al.,* 2002b; Paddison, 2002; Vermeulen, *et al.*, 2005)

Two principle vector designs have emerged over the past few years termed tandem promoter and short hairpin RNA (shRNA) constructs. In tandem promoter constructs polymerase III promoters such as the U6 promoter drive expression of the sense and antisense strands from separate transcription units which anneal intracellularly generating a functional siRNA which enters the RISC complex in the same manner as synthesised siRNA (Miyagishi, 2002). These constructs have been shown to mediate down regulation of gene expression *in vitro* but have not been used extensively due to the superior silencing capacity of the shRNA based expression systems (Miyagishi, 2004).

In short hairpin RNA (shRNA) expression systems, typically U6 or H1 promoters drive expression of a single transcript in which the sense and antisense sequences are separated by a short loop (Brummelkamp, *et al.*, 2002b). The sense and antisense sequences of the single transcript self anneal forming a hairpin structure which is processed by Dicer into a functional siRNA and proceeds through the RISC based RNAi pathway (Figure 4.1) (Amarzguioui, *et al.*, 2005; Brummelkamp, *et al.*, 2002b; Paddison, 2002; Paul, *et al.*, 2002). As mentioned previously, these vector systems have a variety of advantages over synthesised siRNAs, the plasmids can be readily generated, can facilitate long term gene silencing and can permit the inclusion of selectable markers or reporter genes for the identification of transfected cells and generation of cell lines stably expressing shRNAs. The use of the basic shRNA plasmid itself may be restricted as many cell lines can be difficult to transfect with plasmids, the ability to move the H1/U6 shRNA transcription unit potentially into any viral vector is extremely advantageous and can, in principle, overcome the difficulties associated with plasmid based transfection. Given that the promoter and shRNA sequences together are typically less than 300bp, they should be readily accommodated into the majority of viral vectors. Indeed, in many cases additional stuffer sequences or reporter genes may have to be included to attain an optimal size for viral vector generation. As cloning capacity should not be a limiting factor for

shRNAs, viral vector selection may be based upon cell tropism, the mitotic status of the target cell or tissue type and if the absence or presence of genomic integration is required. Due to the presence of existing viral vector technologies the application of viral vectors to RNA occurred in parallel with the emergence of plasmid based RNAi technology (Brummelkamp, *et al.*, 2002a). Adenoviruses (Chen, *et al.*, 2005; Shen, *et al.*, 2003; Xia, *et al.*, 2002; Zhao, *et al.*, 2003), lentiviruses (Morris and Rossi, 2006a), adeno-associated viruses (Ezquer, *et al.*, 2005; Machida, *et al.*, 2006; Wang, *et al.*, 2006b) to name a few have been extensively used for delivery target cells *in vitro* and *in vivo*. These viral systems have been extremely effective at achieving potent target gene suppression in both *in vitro* and *in vivo* studies focused towards the development of treatments for cancer (Brummelkamp, *et al.*, 2002a; Chen, *et al.*, 2005; Czauderna, *et al.*, 2003; Shen, *et al.*, 2003; Takeshita and Ochiya, 2006), neurological diseases (Machida, *et al.*, 2006; Ralph, *et al.*, 2005; Raoul, *et al.*, 2005) and viral infections (Chen, *et al.*, 2006; Grimm and Kay, 2006; Morris and Rossi, 2006a).

Analysis of stable loss of function phenotypes requires a population of cells which stably and continuously express the shRNA of interest. In some instances it may be desirable to have more control over the initiation and duration of shRNA expression and consequent gene silencing. Inducible vector systems have been developed to facilitate this temporal control of gene silencing. A selection of doxycycline responsive H1 promoter systems have been shown to work effectively, inducible suppression of β-catenin in colorectal cancer cells (van de Wetering, 2003) and p53 in breast carcinoma cells (Wiznerowicz and Trono, 2003) has been reported. Dose-dependent inducible p53 and c-Myc suppression has also been achieved in human hepatoma and lung carcinoma cells (Hosono, *et al.*, 2004).

Doxcycline inducible U6 promoter shRNA constructs have also been successfully applied. *In vitro* inducible suppression of endogenous cytokine gene expression has been achieved in a CML cell line K562 (Yang, *et al.*, 2005). Inducible suppression of DNA methyltransferase 1 (DNMT1) in a colorectal cancer cell line resulted in growth arrest (Matsukura, *et al.*, 2003) and the i*n vivo* induction of shRNAs targeting PI3 kinase transcripts significantly reduced the formation of metastasis (Czauderna, *et al.*,

2003). Doxcycline inducible expression of p53 targeting shRNA from a minimal U6 promoter suppressed endogenous p53 expression in HEK 293T, MCF-7 and A549 cells. Potent RNAi was induced with up to 90% suppression of endogenous p53 expression achieved in a dose and time-dependent manner and was reversible in these cell lines upon doxcycline withdrawal (Amar, 2006).

Clearly the use of inducible shRNA constructs provides an additional level of control to the ubiquitously expressing Pol III promoters as discussed above. However in many instances it may be preferable to use tissue specific promoters (typically Polymerase II promoters in mammals) to drive expression of shRNA constructs. Initially thought to be unsuitable for expression of shRNAs the possibility of using tissue specific PolII promoters opened up following demonstration of efficient reduction in expression of both endogenous and exogenous genes *in vivo* by CMV driven shRNA expression (Xia, *et al.*, 2002; Yuan, *et al.*, 2006). A hybrid CMV/H1 promoter has been developed which both increases the efficacy and duration of silencing beyond that of the H1 promoter alone (Ong, *et al.*, 2005). Tissue specific expression has been achieved in rat alveolar cells both *in vivo* and *in vitro* using the SP-C promoter (Gou, *et al.*, 2004). Prostate specific shRNA expression under the control of the pol II prostate specific antigen promoter has also been reported (Song, *et al.*, 2004a)

Whilst no single Col1a1 targeting siRNA of the five evaluated siCol1-siCol5 was statistically a more potent effector of Col1a1 mRNA suppression in HeLa cells, siCol2 and siCol4 were determined to have reduced Col1a1 mRNA levels by the highest percentage. siCol2 and siCol4 reduced Col1a1 expression levels to an average of 35.79% and 34.85% (average of all GAPDH and β-Actin normalised data). Therefore, these two siRNAs were selected and taken forward for incorporation into the design of the siRNA expression plasmids generated in the study and described in this chapter of the thesis. This chapter provides a description of the design and generation of the shRNA expression plasmids, the culturing and transfection of the HeLa cells. In addition analysis of Col1a1 expression levels in cells transfected with both the Col1a1 targeting and control shRNAs were evaluated using real-time RT-PCR.

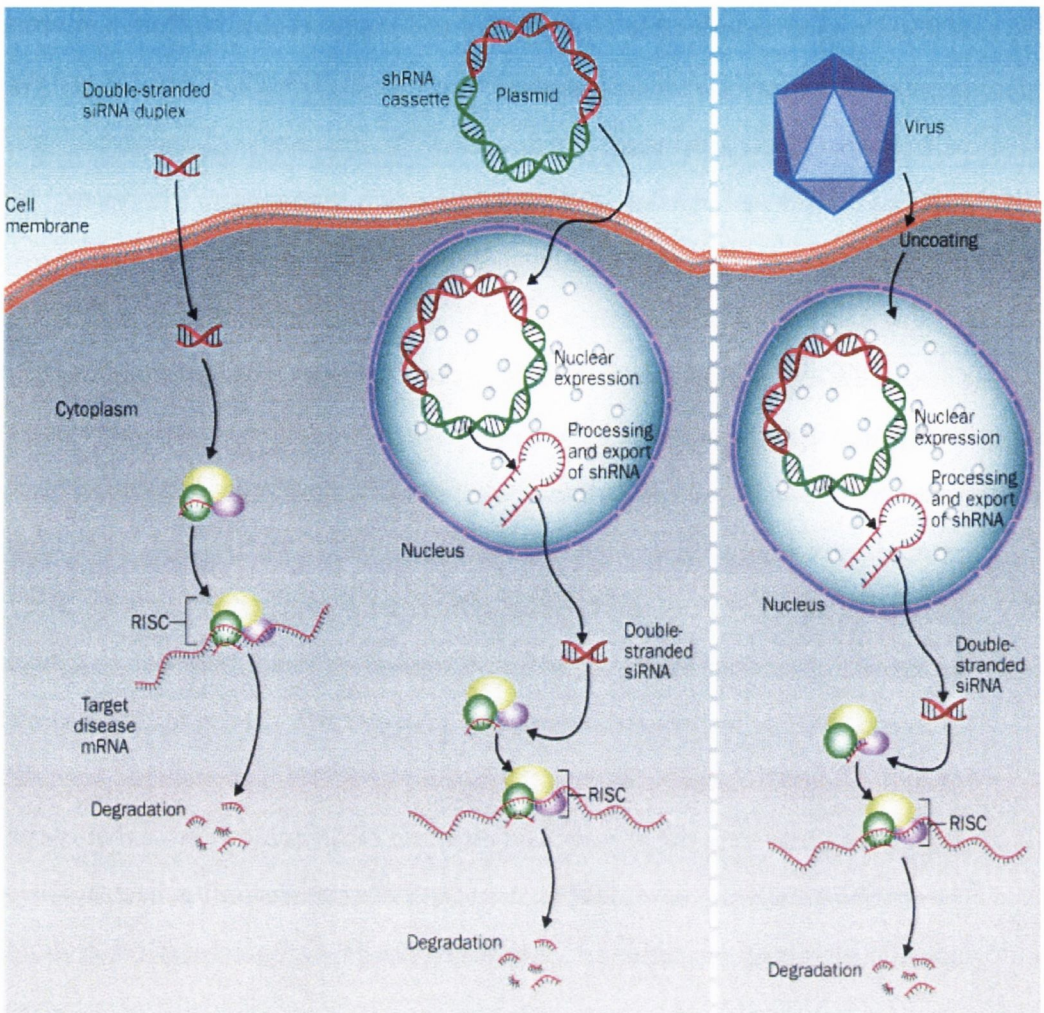


Image obtained from Davidson and Paulson 2004.

Figure 4.1 Three principle RNAi mechanisms. Left: cells are transfected with preformed siRNA duplexes which are incorporated into the RISC following cell entry, leading to degradation of target gene expression. Middle: transfected plasmid vectors must reach the nucleus for transcription of shRNA. The shRNA is processed and exported to the cytoplasm where it is incorporated into the RISC for directed gene silencing. Right: viral vectors encoding shRNA enter target cells via receptor-mediated endocytosis viral genomes enter the nucleus where they either remain episomal or integrate into the genome and express the shRNA which is then processed and incorporated into RISC and effects gene silencing. (Davidson and Paulson, 2004).

4.2 Materials and Methods

All methods and materials particular to this chapter are described below, all general methods and primer sequences are described in chapter two of this thesis.

4.2.1 shRNA oligonucleotides design.

The short hairpin siRNA (shRNA) design used is that originally described by Brummelkamp *et al.,* 2002. This system places the shRNA sequence under the control of the H1 RNA polymerase promoter. These promoters are ideally suited for use in siRNA expression plasmid as they drive high level expression of small RNAs, they are small in size with no additional sequence requirements and they have defined start and stop sites which ensures that no additional bases are incorporated into the siRNA (Myslinski, 2001; Paule, 2000). This is of particular importance due to the strict sequence requirements of siRNA activity, whilst they tolerate a few 3' and 5' overhangs the incorporation of a large number of additional bases reduces their suppressional potency (Amarzguiou, *et al.*, 2005; Paddison, 2002; Vermeulen, et al., 2005)

The shRNA oligonucleotides were ordered from a commercial company (sigma-genosys) as 64mers, which includes the forward and reverse 19 nucleotide target sequence, a 9 nucleotide hairpin sequence and the H1 termination signal which is a stretch of 4 or more thiamine's. Finally as the constructs are always cloned into the BglII and HindIII sites of the H1 Bluescript plasmid, the restriction site overhangs are also included in the oligonucleotides (Figure 4.1). In addition to the two Col1a1 targeting shRNA's, shCol2 and shCol4 control shRNA constructs were required. A GFP targeting shRNA, shGFP was also generated with an identical sequence to the siGFP used in chapter 3. An additional non targeting control shRNA targeting Luciferase (shLuciferase) was kindly provided by Dr A. Palfi, Trinity College Dublin.

4.2.2 Generation of H1 Bluescript-shRNA constructs.

3μg of each sense and antisense oligonucleotide was added to 48μl of annealing buffer (100mM potassium acetate 30mM HEPES-KOH and 2m magnesium acetate). The sample was incubated at 95^{0}C for 4 minutes, 70^{0}C for 10 minutes and slowly cooled to 4^{0}C. H1-Bluescript was digested with BglII and HindIII restriction enzymes (section 2.4.7). The digested plasmid was phenol chloroform extracted (section 2.4.5) and dephosphorylated (section 2.4.11).

The annealed shRNA oligos were ligated into H1-Bluescript (digested with BglII and HindIII restriction enzymes and dephosphorylated) and transformed into XL1 blue *E.coli* cells (section 2.5.2). The bacterial colonies were initially screened via PCR (section 2.5.4) to determine which colonies had plasmids in which the shRNA had successfully been cloned. Primers pBluesscript 1201 and pBluesscript 1224 were used; these primers flank the multiple cloning site of H1-Bluescript. Positive colonies yielded a PCR product of 504bp whilst empty vectors generated 440bp fragment. A selection of the positive colonies identified in the colony PCR were minipreped (section 2.6.1) and sequenced (section 2.7.1) with both the forward and reverse primers pBluescript 1201 and pBluescript 1224 to confirm presence of shRNA inserts with the correct sequence. Plasmids were then maxipreped (section 2.6.3) and the concentrations determined spectrophotometrically (section 2.6.5).

4.2.3 Transfection of HeLa cells with shRNA plasmids.

5x10^5 HeLa cells were seeded per well of a 6 well plate such that cells were 80% - 90% confluent upon transfection. 4μg of each Col1a1 targeting shRNA shCol2 and shCol4 was transfected per well (section 2.8.5). In addition two control shRNA plasmids shGFP and ShLuciferase were transfected in each set of experiments. In each individual transfection experiment each shCol2 plasmid and shCol4 plasmid transfection was carried out in triplicate (i.e. three separate wells of a 6 well plate) and the control plasmids shGFP and shLuicferase were transfected in duplicate (i.e. two separate wells of a 6 well plate). RNA was trizol RNA extracted 24 hours post transfection (section 2.9.1). Three independent transfection experiments were carried

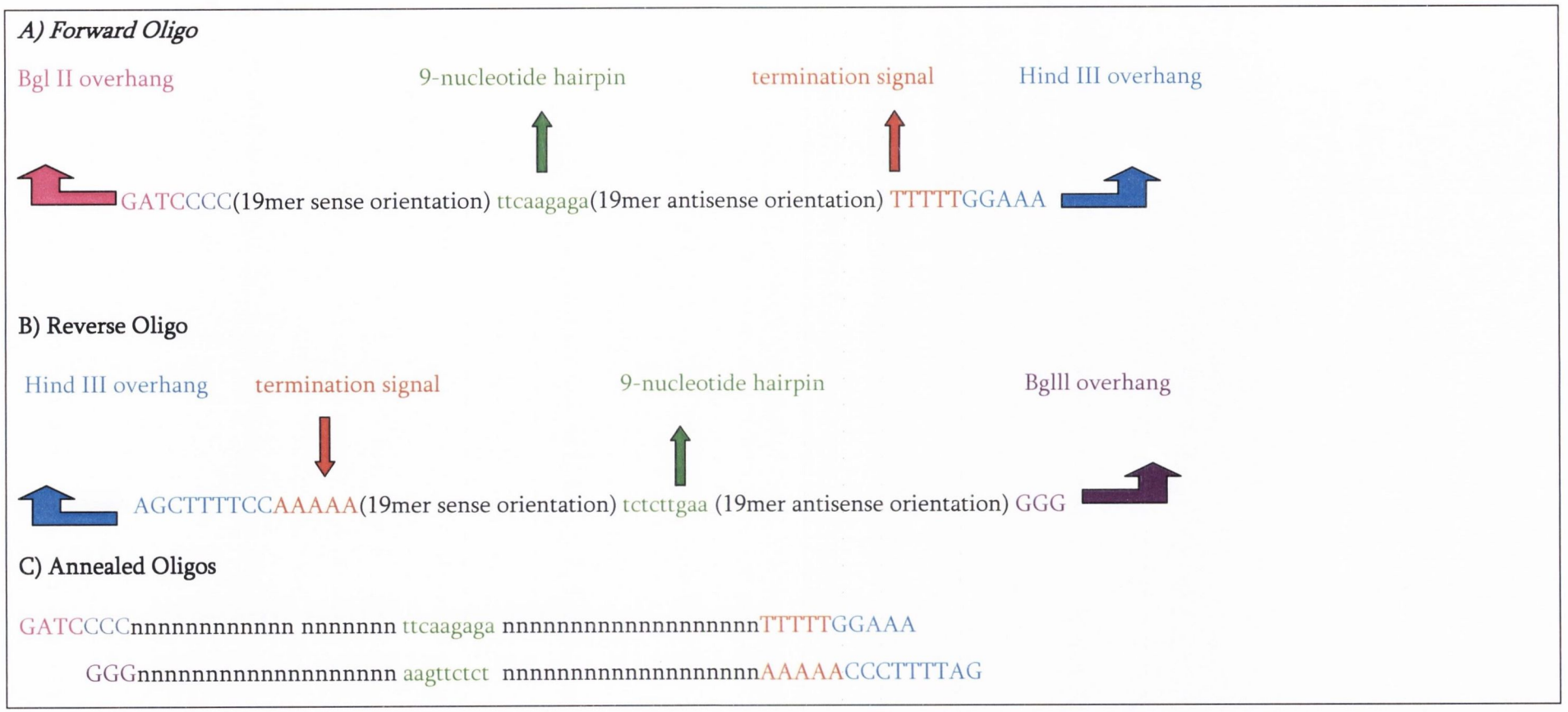


Figure 4.2. ShRNA Oligonucleotide design. Each 64mer A) sense and B) antisense oligonucleotide consists of a 19mer of sense and antisense orientation separated by a 9ntd hairpin sequence. The oligos are annealed and cloned into pSuper using the BglII and HindIII overhangs which are included in the 64mer. This design is based upon that of Brummelkamp et al 2002.(Brummelkamp, et al., 2002b)

out generating a total of 9 RNA samples for shCol2 and shCol4 with 6 RNA samples generated for both shGFP and shLuciferase

4.2.4 Real time RT-PCR evaluation of Col1a1 expression levels in HeLa cells transfected with Col1a1 targeting shRNAs

Col1a1 expression levels in HeLa cells transfected with Col1a1 targeting shRNAs was evaluated using real time RT-PCR and the relative quantification method. As described (section 4.2.3) each Col1a1 targeting shRNA shCol2 and shCol4 was transfected in triplicate per transfection, this generated 3 RNA samples for shCol2 and shCol4 per transfection experiment. For the real time RT-PCR analysis the three RNA samples generated per transfection experiment were pooled, this generated 3 pooled RNA samples for shCol2 and shCol4 representative of 9 independent transfections. For the non-targeting controls the two RNAs generated per transfection experiment were pooled, generating 3 pooled RNA samples representing 6 independent transfections.

All RT-PCR reactions were set up as described in section 2.9.4. The housekeeping genes GAPDH and β-Actin were used for this series of experiments as per chapter 3 of this thesis. Each RNA sample was analysed in triplicate (i.e. 3 wells per RNA) per primer (Col1a1, GAPDH and β-Actin) set in a single 96 well plate RT-PCR assay. Each 96 well plate real time RT-PCR assay was run 3 times. The Col1a1 expression levels determined were normalised, this is achieved by dividing the target expression value in a sample by the housekeeper expression value in the same sample. In this case Col1a1 is divided by GAPDH or β-Actin value from the same sample. The relative Col1a1 expression levels were then determined by comparing the normalised Col1a1 expression values for cells transfected with Col1a1 targeting siRNAs to normalised Col1a1 expression values in cells transfected with the non-targeting control shRNAs siGFP and shLuciferase in which the Col1a1 expression levels were set as 100%. (Calculations were carried as illustrated in Table 3.1). The relative Col1a1 expression data for shCol2, shCol4, siGFP and shLuciferase was compiled and

evaluated using GraphPad InStat3. The mean and standard deviation of the data generated for each sample was determined. Non parametric ANOVA analysis was carried out in which Col1a1 mRNA expression levels in Hela cells transfected with shCol2, shCol4 and shGFP or shLuciferase were compared. A p value of $p<0.05$ indicated that the Col1a1 expression data sets differed significantly, and a value of $p>0.05$ indicated that the data sets compared did not significantly differ. Post test analysis was also carried out in which pairwise comparisons of each of the data sets was carried out, e.g shCol2 to shGFP, shCol4 to shGFP, shCol2 to shCol4 the data was considered to differ if $p<0.05$ and to not differ if $p>0.05$. This analysis was carried out for both GAPDH and β-Actin normalised data, all data was evaluated using GraphPad Instat3 and all graphs generated using Excel.



Figure 4.3 shRNA target sequence and annealed sense and antisense oligonucleotides. A) shCol2, B) shCol4. Each 64mer sense and antisense oligo consists of a 19mer of sense and antisense orientation separated by a 9ntd hairpin sequence (highlighted in green). The oligos are annealed and cloned into pSuper using the BglII (highlighted in pink) and HindIII (highlighted in blue) overhangs which are included in the 64mer. The design is based upon that of Brummelkamp et al 2002 (Brummelkamp, et al., 2002b).

D) *ShGFP*

Target mRNA *CGGCAAGCTGACCCTGAAGTTCAT*

ShGFP oligos

*GATCCCC*CGGCAAGCTGACCCTGAAGTTCAT*ttcaagaga* ATGAACTTCAGGGTCAGCTTGCCG*TTTTTGGAAA*

*GGG*GCCGTTCGACTGGGACTTCAAGTA*aagttctct* TACTTGAAGTCCCAGTCGAACGGC*AAAAACCTT*

E) *ShLuciferase*

Target mRNA *GCGTACGGGGAAAACTTCGA*

ShLuciferase oligos

*GATCCCC*GCGTACGGGGAAAACTTCGA *ttcaagaga* TGGTTAGTGGACGCATGTCT*TTTTTGGAAA*

GGG CGCATGCCCCTTTTGAAGCT *aagttctct* AGCTTCATAAGGCGCATCG*AAAAACCT*

Figure 4.4 shRNA target sequence and annealed sense and antisense oligonucleotides. A) shGFP, B) shLuciferase. Each 63mer sense and antisense oligo consists of a 19mer of sense and antisense orientation separated by a 9ntd hairpin sequence (highlighted in green). The oligos are annealed and cloned into pSuper using the BglII (highlighted in pink) and HindIII (highlighted in blue) overhangs which are included in the 64mer. The design is based upon that of Brummelkamp et al 2002 (Brummelkamp, et al., 2002b)

4.3 Results

Evaluation of the potency of Col1a1 targeting shRNAs in HeLa cells

The key objective of the work carried out in this chapter of the thesis was to determine the efficacy of the Col1a1 targeting shRNAs, shCol2 and shCol4. This initial validation study was carried out in HeLa cells. In addition the potency of shCol2 and shCol4 was compared to that of their siRNA counterparts described in chapter 3 of this thesis.

4.3.1 shRNA construct generation.

As described in the introduction to this chapter due to the transient gene suppression mediated by synthesised siRNAs an alternative system which facilitates long term gene suppression was required. siRNA expression plasmids, short hairpin RNAs (shRNAs) have been developed in parallel with the development of RNAi technology, these plasmids facilitate long term RNAi. The shRNA design of Brummelkamp, *et al.*, (2002b) was adopted for the generation of shRNA counterparts of siCol2 and siCol4, two of the five siRNAs evaluated in chapter 3 of this thesis. Three shRNA expression plasmids, shCol2, shCol4 and shGFP were generated as described in section 4.2.2 for the work carried out in this chapter. The second control plasmid shLuciferase was a kind gift of Dr. A. Palfi, Trinity College Dublin.

The sense and antisense oligonucleotides for shCol2, shCol4 and shGFP were successfully annealed, ligated into H1-Bluescript and transformed into XL1 blue *E.Coli* cells. The screening of bacterial colonies is a very rapid means of detecting a selection of plasmids in which the shRNA insert has been successfully cloned. Plasmids which were positive for shRNA inserts generated a PCR product of 504bp whilst empty vectors generated a PCR product of 440bp (Figure 4.5). Once the positive plasmids were identified for each of the shRNAs, plasmids shCol2, shCol4 and shGFP a selection of colonies for each plasmid was minipreped and sequenced. Frequently mutations arise during the cloning process and the constructs can be difficult to sequence due to the ability of the sequencing PCR product to form very tight hairpin structures. The sequence data obtained for each minprep for each construct was carefully analysed and three constructs in which error free sequence

was verified were selected and maxi-preped. The maxi-preps were then sequenced to confirm that the sequence of each shCol2 (Figure 4.7), shGFP (Figure 4.8) and shCol4 (Figure 4.9) was correct.

4.3.2 Real time RT-PCR evaluation of the potency of shRNA mediated Col1a1 suppression in HeLa cells.

Based upon the efficacy of the HeLa cells as the *in vitro* model for the evaluation of the potency of the 5 Col1a1 targeting siRNAs evaluated in chapter 3 of this thesis, HeLa cells were also used as the *in vitro* model for the evaluation of the potency of the shRNA counterparts of siCol2 and siCol4, i.e. shCol2 and shCol4. For each transfection experiment $5x10^5$ HeLa cells were seeded per well of a 6 well plate such that the cells were 80% - 90% confluent at the time of transfection. In each experiment each Col1a1 targeting shRNA, shCol2 and shCol4 was transfected in triplicate per 6 well plate (i.e. 3 individual wells of a 6 well plate) and each control shRNA, shGFP and shLuciferase, were transfected in duplicate (i.e. 2 individual wells of a 6 well plate). The three RNA samples generated per transfection experiment were pooled, this generated 3 pooled RNA samples for shCol2 and shCol4 representative of 9 independent transfections. For the non-targeting controls the two RNAs generated per transfection experiment were pooled, generating 3 pooled RNA samples representing 6 independent transfections.

The efficacy of the Col1a1 targeting shRNA constructs was evaluated using real time RT-PCR and the relative quantification method. The housekeeping genes used for this series of experiments was GAPDH and β-Actin as per chapter 3 of this thesis. Each sample was plated in triplicate per real time RT-PCR run (i.e. 3 individual wells of the 96 well plate). Each RT-PCR run was carried out three times. Col1a1 expression in each sample was then normalised to each housekeeping gene. Normalisation is achieved by dividing the Col1a1 expression value of a sample by the GAPDH or β-Actin expression value of the same sample. The relative Col1a1 expression is then determined by comparing the normalised Col1a1 expression values for cells transfected with Col1a1 targeting shRNAs to normalised Col1a1 expression values in cells transfected with the non-targeting control

shRNAs, shGFP and shLuciferase. This robust analysis revealed that the Col1a1 shRNA targeting constructs, shCol2 and shCol4 were very effective at reducing the levels of Col1a1 expression in HeLa cells. When GAPDH was used as the housekeeping control gene, Col1a1 expression levels in the cells transfected with shCol2 were 29.11% (± 18.32%) and 27.79% (± 17.41%) when compared to Col1a1 expression levels in HeLa cells transfected with shGFP and shLuciferase respectively. shCol4 also mediated a potent RNAi response in HeLa cells when normalised to GAPDH, Col1a1 expression levels were reduced to 16.78% (± 6.04%) and 15.85% (± 5.43%) in cells transfected with shCol4 when compared to Col1a1 expression levels in cells transfected with the control plasmids shGFP and shLuciferase respectively (Table 4.1).

The data was also normalised using β-Actin as the housekeeping gene. shCol2 again resulted in significant suppression of Col1a1 with expression levels reduced to 17.28% (± 10.46%) when compared to Col1a1 expression levels in cells transfected with the control plasmid shGFP and reduced to 16.71% (± 10.39%) when compared to Col1a1 expression levels in cells transfected with shLuciferase. shCol4 also potently reduced Col1a1 expression levels to 16.23% (± 10.76%) when compared to Col1a1 expression levels in HeLa cells transfected with shGFP and 18.50% (± 11.89%) when compared to Col1a1 expression levels in HeLa cells transfected with shLuciferase (Table 4.2). Therefore it appears that shCol2 and shCol4 elicit significant suppression of the target gene Col1a1 in HeLa cells (Table 4.1, Table 4.2 and Figure 4.10)

In relation to the standard deviations, error and variation can be introduced due to a variety of reasons in studies such as this involves multiple transfections, RNA extractions and RT-PCR evaluations. As each RNA sample analysed is a pool of three RNA samples generated from a single nucleofection experiment, therefore the overall Col1a1 expression levels will in part be dependent on the sum of the nucleofection efficiencies in each of the three nucleofections carried out in a single experiment possibly introducing a significant amount of variability. The RT-PCR assays represent another point at which variation can be introduced due to in difference in cDNA synthesis, PCR efficiencies, sample preparation, pipetting error, and inter-run variation.

Non parametric ANOVA analysis was carried out in which Col1a1 mRNA expression levels in HeLa cells transfected with shCol2, shCol4 and shGFP or shLuciferase were compared. Each ANOVA i.e. comparison of shCol2, shCol4 and shGFP or shCol2, shCol4 and shLuciferase Col1a1 expression data sets generated a p value of $p<0.0001$ indicating that 3 Col1a1 data sets compared differed significantly. This result was generated for both GAPDH and β-Actin normalised Col1a1 expression data set (Table 4.1 and Table 4.2). In order to determine which of the data sets differed, Dunn's multiple comparisons post test analysis was for both the GAPDH β-Actin normalised Col1a1 expression data sets. This analysis executes pairwise comparisons of each of the three data sets in the 4 ANOVAs carried out, e.g. shGFP to shCol2, shGFP to shCol4 and shCol2 to shCol4. It was found that comparison of shCol2 to shGFP or shLuciferase generated a p value of $p<0.001$ indicating that the Col1a1 expression data sets differed significantly. Comparison of shCol2 and shCol4 data sets generated a p values of $p>0.05$ indicating that Col1a1 expression levels in HeLa cells following transfection with shCol2 or shCol4 did not differ significantly, thus it appears that neither shCol2 nor shCol4 effects more potent suppression of Col1a1 expression in HeLa cells (Table 4.3).

The final set of analysis carried out was to determine if the shRNA (shCol2 and shCol4) counterparts of the Col1a1 targeting siRNAs (siCol2 and siCol4) more potently reduced Col1a1 expression in HeLa cells. This was achieved by carrying out a two tailed student t-test in which the total Col1a1 expression data (i.e. total GAPDH and β-Actin data sets) from HeLa cells transfected with siCol2 and shCol2 or siCol4 and shCol4 were compared. The average Col1a1 expression level in siCol2 and shCol2 transfected HeLa cells was 35.45% (±10.85%) and 23.29 (±15.54%) respectively. Comparison of siCol2 and shCol2 Col1a1 expression data generated a p value of $p=0.0017$, this indicated that the two data sets are significantly different, suggesting that shCol2 is a more potent mediator of Col1a1 suppression in HeLa cells. The average Col1a1 expression level in siCol4 and shCol4 transfected HeLa cells was 38.48% (±12.7%) and 16.97% (±8.70) respectively. Comparison of siCol4 and shCol4 Col1a1 expression data generated a p value of $p< 0.0001$, this indicated that the two data sets are significantly different, suggesting that shCol4 is a

much more potent suppressor of Col1a1 expression in HeLa cells than its siRNA counterpart, siCol4.

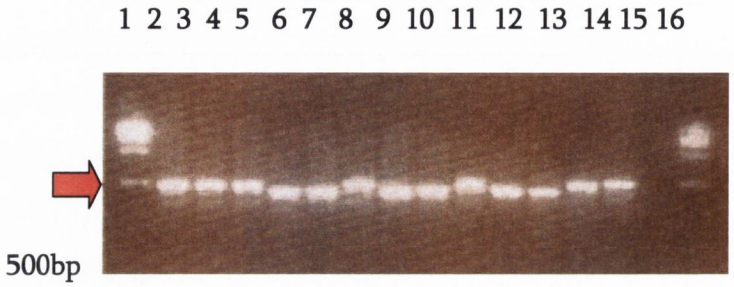


Figure 4.5 shRNA bacterial plasmids PCR. PCR analysis of a selection of bacterial colonies was carried out in order to determine which colonies contained plasmids in which the shRNA has been inserted. PCR was carried out using the primers pBluescript1201 and pBluescript1224, colonies which contain plasmids into which the shRNA has been successfully cloned will generate a PCR product of 504 bp and empty vectors will generate a PCR product of 440 bp. Lane 1:1kb ladder, Lane 2 positive PCR control, Lanes 3, 4,7,10, 13 & 14 plasmids which contain the shRNA insert (PCR product 504bp), Lanes:5, 6, 8, 9, 11, 12, empty plasmid (PCR product 440bp), Lane 15:1kb DNA Ladder.

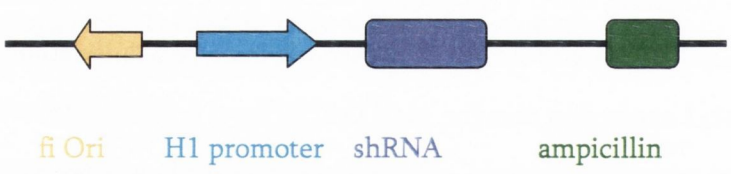


Figure 4.6 Schematic of the H1-shRNA plasmid. Expression of the shRNA is under the control of the H1 promoter, the plasmid also contains an f1 origin of replication and ampicillin resistance gene which are used in bacterial replication and selection. In each of the constructs generated either the shCol2, shCol4, shGFP and shLuciferase targeting sequences are present.

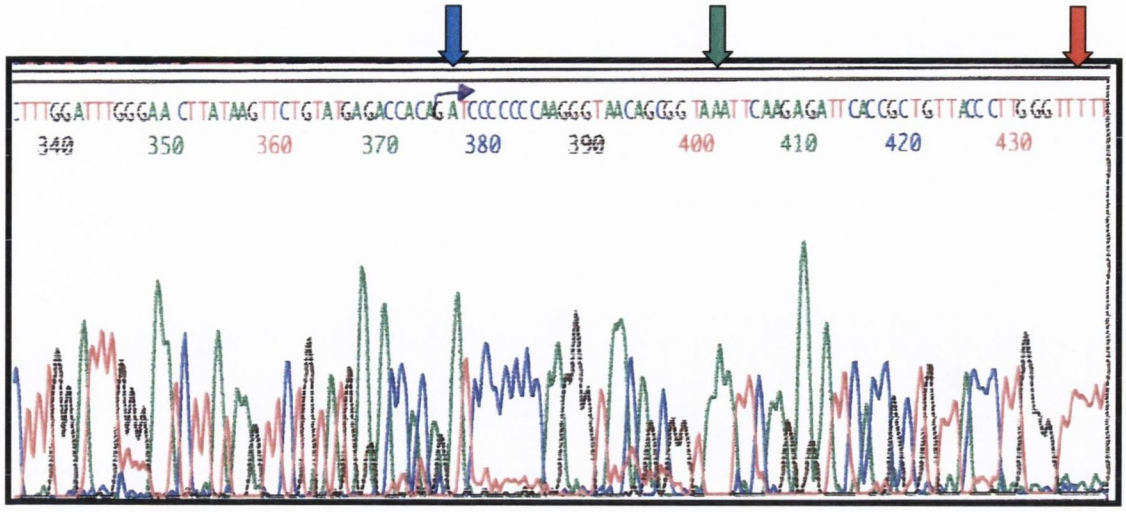


Figure 4.7 Chromatogram of Col1a1 targeting shRNA, shCol2 . Chromatogram obtained following sequencing of the shCol2 maxiprep using the forward primer pBluescript1201. The blue arrow indicates the HindIII insertion site and beginning of the shRNA sequence, the green arrow indicates the 9 nucleotide hairpin sequence and the red arrow indicates the 5 T termination signal.

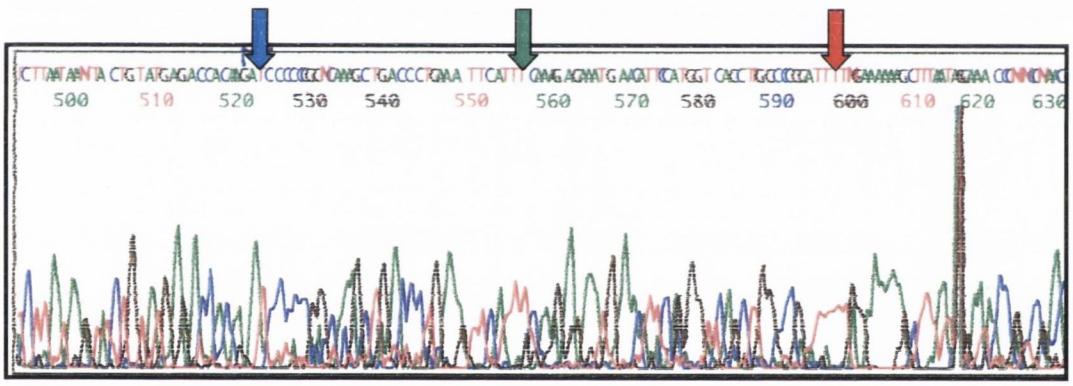


Figure 4.8 Chromatogram of the non targeting shRNA, shGFP. Chromatogram obtained following sequencing of shCol4 maxiprep using the forward primer pBluescript 1201. The blue arrow indicates the HindIII insertion site and beginning of the shRNA sequence, the green arrow indicates the 9 nucleotide hairpin sequence and the red arrow indicates the 5 T termination signal.

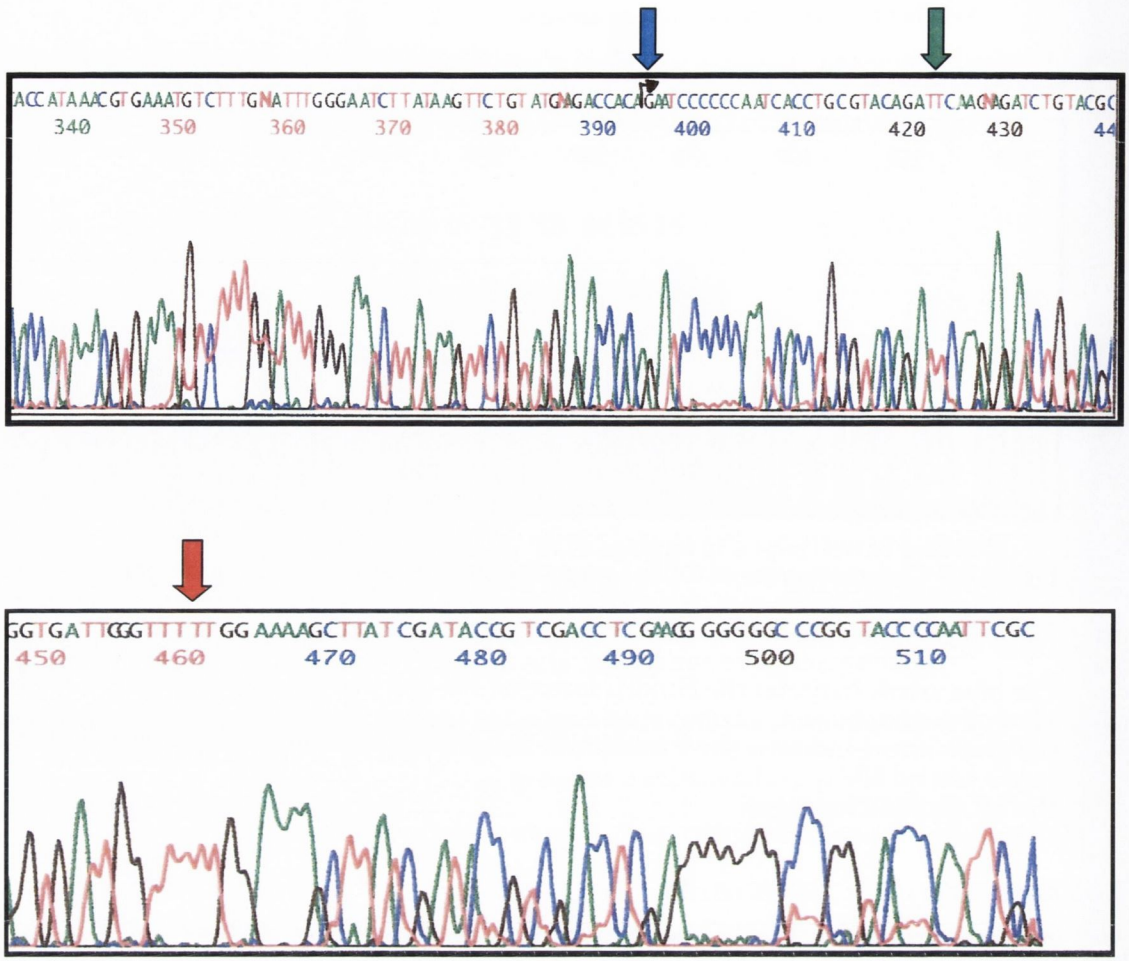


Figure 4.9 Chromatogram of the Col1a1 targeting shRNA, shCol4. Chromatogram obtained following sequencing of shCol4 maxiprep using the forward primer pBluescript1201. The blue arrow indicates the HindIII insertion site and beginning of the shRNA sequence, the green arrow indicates the 9 nucleotide hairpin sequence and the red arrow indicates the 5 T termination signal.

		shGFP	*shLucif*	*shCol2*	*shCol4*	*ANOVA*
GAPDH	*shGFP*	100%		29.11%	16.78%	p<0.0001
	StdDev			18.32	6.04	
	shLucif		100%	27.79%	15.85%	p<0.0001
	StdDev			18.32	5.43	

Table 4.1 GAPDH normalised Col1a1 expression levels in HeLa cells transfected with Col1a1 targeting shCol2, shCol4 and the non targeting controls shGFP or shLuciferase. HeLa cells were lipofectamine transfected with 4μg of each shRNA shCol2, shCol4, shGFP or shLuciferase plasmids. RNA was extracted 24 hours post transfection and Col1a1 expression levels evaluated using real time RT-PCR. Col1a1 expression values from HeLa cells transfected with shCol2 or shCol4 were normalised with GAPDH. Col1a1 expression levels in cells transfected with shGFP (green bar) and shLuiferase (blue bar) are set to 100% and the Col1a1 expression values in HeLa cells transfected with shCol2 and shCol4 determined via direct comparison. Both shCol2 and shCol4 elicited significant suppression of Col1a1 mRNA. shCol2 reduced Col1a1 expression levels to between 36.43% and 34.30% and shCol4 reduced Col1a1 expression levels to 16.78% to 15.85%. Statistical analysis of the data was carried out using GraphPad InStat3. The mean and standard deviation of each data set was calculated. Non parametric ANOVA anlaysis was carried out in which each the Col1a1 data sets generated following transfection with shCol2, shCol4 and shGFP or shLuciferase were compared. A p value of p<0.0001 was generated indicating that the 3 data sets compared differed significantly. Post test analysis was carried out using the Dunn's multiple comparison test in which pairwise analysis of each of the data sets was carried out in order to determine which of the data sets differed (see table 4.3 for details)

		shGFP	*shLucif*	*shCol2*	*shCol4*	*ANOVA*
β-Actin	***shGFP***	100%		17.28%	16.23%	p<0.0001
	StdDev			10.46	10.76	
	shLucif		100%	16.71%	18.50%	p<0.0001
	StdDev			10.39	11.89	

Table 4.2 β-Actin normalised Col1a1 expression data from HeLa cells transfected with Col1a1 targeting shCol2, shCol4 and non targeting controls shGFP and shLuciferase. HeLa cells were lipofectamine transfected with 4μg of each shRNA, shCol2, shCol4, shGFP or shLuciferase plasmids. RNA was extracted 24 hours post transfection and Col1a1 expression levels evaluated using real time RT-PCR. Col1a1 expression values from HeLa cells transfected with shCol2 or shCol4 were normalised with β-Actin. Col1a1 expression levels in cells transfected with shGFP (green bar) and shLuiferase (blue bar) were set to 100% and the Col1a1 expression levels in HeLa cell transfected with shCol2 and shCol4 determined via direct comparison. Both shCol2 and shCol4 elicited significant suppression of Col1a1 mRNA. shCol2 reduced Col1a1 expression levels to between 16.71% and 17.28 % and shCol4 reduced Col1a1 expression levels to 16.23% to 18.50%. Statistical analysis of the data was carried out using GraphPad InStat3. The mean and standard deviations were calculated of each data set was determined. Non parametric ANOVA anlaysis was carried out in which each the Col1a1 data sets generated following transfection with shCol2, shCol4 and shGFP or shLuciferase were compared. A p value of $p<0.0001$ was generated indicating that the 3 data sets compared differed significantly. Post test analysis was carried out using the Dunn's multiple comparison test in which pairwise analysis of each of the data sets was carried out in order to determine which of the data sets differed (see table 4.3 for details)

Figure 4.10 Graph of Col1a1 expression data in cells transfected with Col1a1 targeting shRNAs. HeLa cells were lipofectamine transfected with 4μg of each shRNA shCol2, shCol4, shGFP or shLuciferase plasmids. RNA was extracted 24 hours post transfection and Col1a1 expression levels evaluated using real time RT-PCR. Col1a1 expression values for all RNA samples are normalised using the housekeepers GAPDH or β-Actin. The Col1a1 expression values for HeLa cells transfected with shCol2 or shCol4 were then determined by direct comparison to the Col1a1 expression values in shGFP or shLuciferase transfected cells which were set at 100%. GAPDH normalised Col1a1 expression values compared to Col1a1 expression levels in cells transfected with shGFP is represented by the (Plain green bars) or shLuciferase is represented by the (Plain blue bars). β-Actin normalised Col1a1 expression compared to Col1a1 expression values in HeLa cells transfected with shGFP is represented by the (green hatched bars) and shLuciferase is represented by the (blue hatched bars). Col1a1 expression in HeLa cells transfected with shCol2 values ranged from 16.71% – 29.11% to and Col1a1 expression in cells transfected with shCol4 ranged from 15.85% to 18.50% compared to Col1a1 expression levels in HeLa cells transfected with the non targeting controls shGFP and shLuciferase, which were set at 100%. Statistical analysis of the data was carried out using GraphPad Instat3. Non parametric ANOVA analysis of GAPDH and β-Actin normalised data was carried out, in which the Col1a1 expression data sets from HeLa cells transfected with shCol2, shCol4 and shGFP or Luciferase were compared, each comparison generated a p value of $p<0.0001$ indicating the Col1a1 data sets differed significantly (For detail see Tables 4.1 and 4.2). Dunn's post test analysis was carried out to determine which data sets differed significantly, see table 4.3. The error bars on the graph represent the standard deviation the data set.

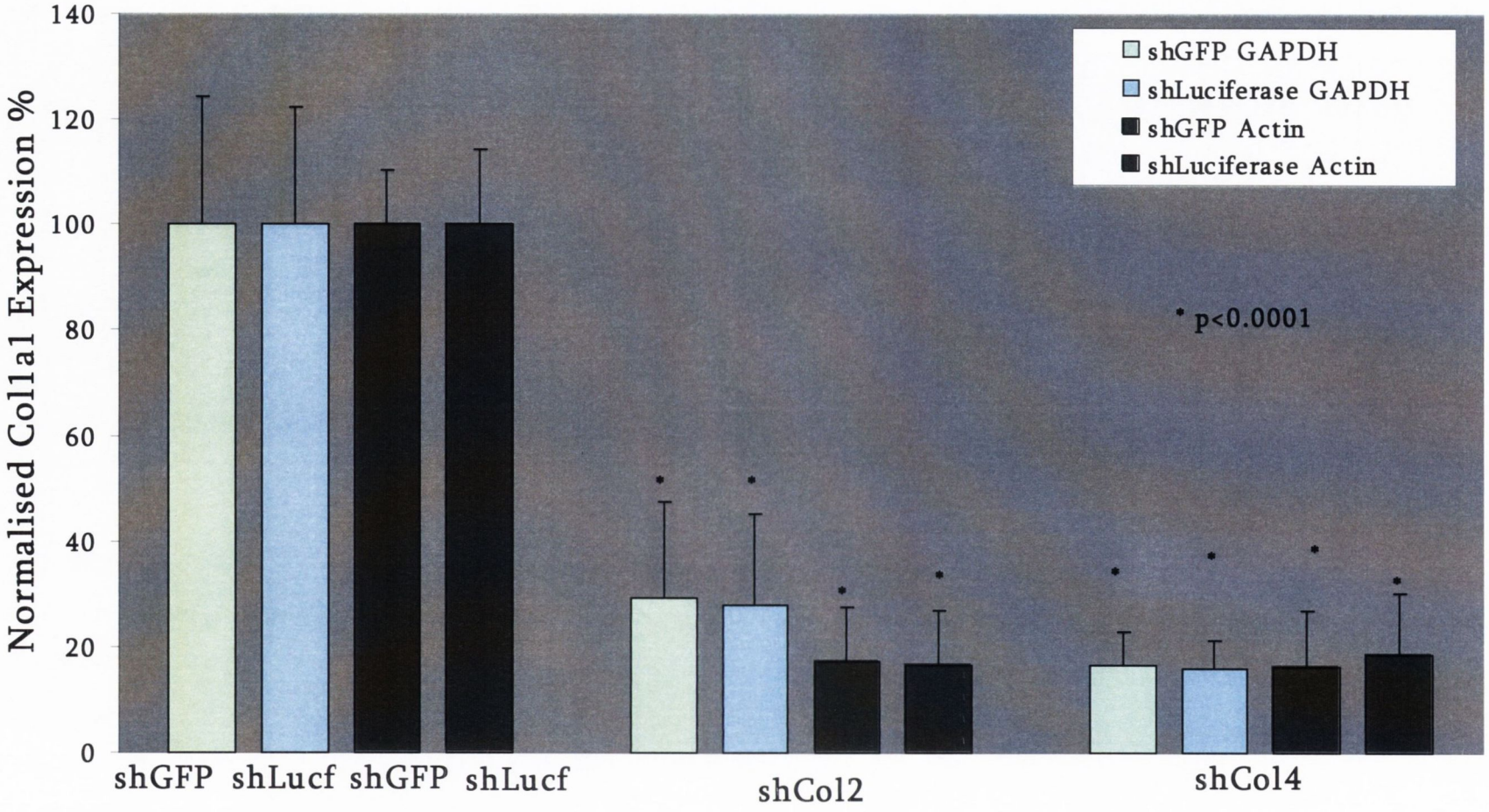


Figure 4.10. Col1a1 expression in HeLa cells transfected with shCol2, shCo4, shGFP or shLuciferase normalised to GAPDH and β-Actin.

	shCol2	*shCol4*
siGFP	p<0.001	p<0.001
siLuciferase	p<0.001	p<0.001
shCol2		p>0.05

Table 4.3 P values obtained from Dunns multiple comparison post test analysis.

Subsequent to ANOVA analysis of Col1a1 expression levels in HeLa cells transfected with shCol2, shCo4, shGFP or shLuciferase, Dunn's multiple comparisons post test analysis was carried out. This test carried out pairwise comparisons of the shGFP, shCol2, shCol4 and shLuciferase Col1a1expression data sets. Comparison of shGFP or shLuciferase to shCol2 or shCol4 generated a p value of $p<0.001$ indicating that that the levels of Col1a1 expression in HeLa cells transfected with shGFP or shLuciferase differed significantly from the levels of Col1a1 expression in HeLa cells transfected with shCol2 or shCol4. Comparison of Col1a1 expression levels in HeLa cells following transfection with shCol2 or shCol4 generated a p value of $p>0.05$, indicating the levels of Col1a1 expression do not differ, therefore neither shCol2 nor shCol4 more potently reduces Col1a1 expression in HeLa cells.

	siCol2 v shCol2	siCol4 v shCol4
Total data	p=0.0017	p<0.0001

Table 4.4 P values obtained from two tailed student t-test comparing total data set of Col1a1 expression levels in HeLa cells transfected with siCol2 or siCol4 to the total data set of Col1a1 expression levels in HeLa cells transfected with shCol2 or shCol4. The total data sets compared for each suppressor is the Col1a1 expression values calculated from both the GAPDH and β-Actin normalised data compared to the non targeting control data sets (siGFP and siNon-Tgt) or (shGFP and shLuciferase). Comparison of the total siCol2 Col1a1 expression data set to the total shCol2 Col1a1 expression data set generated a p value of p = 0.0017 indicating that the levels of Col1a1 suppression achieved with siCol2 or shCol2 differ significantly, therefore shCol2 appears to effects higher levels of Col1a1 suppression in HeLa cells. Comparison of the total siCol4 Col1a1 expression data sets to the total shCol4 Col1a1 expression data set generated a p value of p<0.0001 suggesting that the data sets differed significantly, therefore shCol4 is a more efficient suppressor of Col1a1 suppression in HeLa cells than siCol4.

4.3 Discussion

Whilst siRNA duplexes are effective mediators of RNAi and the 5 siRNA duplexes evaluated in chapter 3 of this thesis potently suppressed endogenous Col1a1 expression levels in HeLa cells as evaluated by real time RT-PCR, their utility is hampered primarily by the transience of the RNAi effect. In order to overcome this limitation plasmid based siRNA expression systems have been developed (Brummelkamp, *et al.*, 2002b; Paddison, 2002; Paul, *et al.*, 2002). As described in the introduction to this chapter, these plasmids have been shown to effectively mediate gene suppression *in vivo* and *in vitro* (Amarzguiou, *et al.*, 2005; Leung and Whittaker, 2005; Paddison, 2002; Siolas, *et al.*, 2005). Plasmid based RNAi expression systems have a variety of advantageous properties such as enabling long term suppression studies through generation of stable cell lines and the inclusion of markers such as GFP reporter genes. The inclusion of tissue specific or inducible promoters can in principle provide an additional level of control with regard to what tissues are targeted for RNAi. In addition, as the expression cassette may potentially be cloned into any viral vector due to the compact size of the shRNA and Pol III promoter sequences, viral delivery may help to overcome the limitations associated with difficult to transfect cell lines.

The H1 shRNA expression system described by Brumelkamp *et al* 2002 (Brummelkamp, *et al.*, 2002a) was employed in the generation of plasmids which expressed siCol2 and siCol4 as shRNAs. Based upon the results of the suppression study carried out in chapter 3 of this thesis no single siRNA was found statistically to be a more potent suppressor of Col1a1 expression in HeLa cells. Numerically siCol2 and siCol4 reduced Col1a1 expression levels the most and consequently were selected as the sequences which would be explored in the shRNA study. In addition to generating the Col1a1 targeting shRNA constructs, shCol2 and shCol4, a negative control shRNA was generated, shGFP. The target sequence of this shRNA is identical to the target sequence of siGFP used in chapter 3 of this thesis. The second negative control shLuciferase was kindly provided by Dr. A. Palfi, Trinity College Dublin. This negative control shRNA had also been generated according to the design reported by Brummelkamp et al 2002. shGFP and shLuciferase target non-mammalian

genes have no known homology to any human gene sequence. In principle, the shRNAs should proceed through the RNAi pathway as any shRNA, therefore identical one would expect that similar molecular and biochemical processes to occur following transfection of Col1a1 targeting shRNAs or the non targeting shRNAs, aside from sequence specific target suppression. Therefore, the use of such controls should in principle enable one to distinguish between sequence specific and non specific RNAi effects in the experiment.

Notably, both shCol2 and shCol4 were found to be potent mediators of Col1a1 suppression In HeLa cells, shCol2 reduced Col1a1 expression to 16.71% - 29.91% (±10.39-18.32) and shCol4 reduced Col1a1 expression levels to 15.85% - 18.50% (±5.43 – 11.89) (Table 4.1, Table 4.2 and Figure 4.10). Non parametric ANOVA analysis was carried out for both the GAPDH and β-Actin normalised data sets, in the Col1a1 expression data sets from shCol2, shCol4 and shGFP or shLuciferase transfected HeLa cells were compared. A p value of $p<0.0001$ was generated from each of the 4 ANOVAs (i.e. GAPDH and β-Actin normalised shCol2, shCol4 and shGFP or shCol2, shCol4 and shNon-tgt) indicating that the Col1a1 expression levels differed between the data sets. Dunns multiple comparisons post test analysis was carried out in order to specifically determine which data sets differed. This test carried out pairwise comparisons of each of the data sets in the ANOVA, e.g. shGFP to shCol2. shGFP to shCol4 and shCol2 to shCol4. Pairwise comparison of the shGFP or shLuciferase Col1a1 expression data sets to the shCol2 to shCol4 Col1a1 expression data sets generated a p value of $p<0.001$, indicating that Col1a1 expression levels differed significantly between HeLa cells transfected with the Col1a1 targeting shRNAs and the non targeting shRNAs. Comparison of shCol2 to shCol4 generated a p value of $p>0.05$. indicating that the data sets do not differ, therefore neither shCol2 nor shCol4 effects more potent suppression of Col1a1 expression in HeLa cells (Table 4.3).

The shRNAs and siRNA Col1a1 expression data was compared to determine which was more potent, siCol2 suppression was on average 29.32- 42.64% (± 8.27-12.90%) and shCol2 suppression was on average 16.71%- 29.11 (±10.39-18.32%) of the Col1a1 expression levels in HeLa cells transfected with the non targeting control siRNAs or shRNAs. Comparison of

the siCol2 expression data to the shCol2 expression data revealed that the data sets were significantly different (p=0.0017) (Table 4.3). Therefore it appears that shCol2 effects more potent Col1a1 suppression in HeLa cells than siCol2. siCol4 reduced Col1a1 expression in HeLa cells to 28.75 – 42.97% (±9.50 – 14.64%) and shCol4 reduced Col1a1 expression levels in HeLa cells to 15.35 – 18.50% (±5.43 – 11.89%). Comparison of the siCol4 expression data to the shCol4 expression data revealed that they were significantly different ($p<0.05$), (Table 4.4). Therefore shCol4 appears to be significantly more potent than its siRNA counterpart in the analysis undertaken in HeLa cells during the course of this study.

The superior potency of the shRNA construct contradicts the finding of a study in which a comparison of target gene expression levels in HeLa cells 24 hours after transfection with a 21mer siRNA (19bp stem) or its corresponding shRNA (19mer stem) revealed little difference in suppression levels (Siolas, *et al.*, 2005). Therefore how can the increased levels of suppression achieved by shCol4 and shCol2 explained? In the case of the H1 system used in this current study, it has been reported that the shRNAs generated can have variable 3' overhangs which vary in length by 1 – 4ntds, this in turn can result in the production of a diverse population of siRNAs by DICER from the shRNA substrates (Siolas, *et al.*, 2005; Vermeulen, et al., 2005). It is possible that a mixed population of siRNA duplexes (21mers – 23mers) was generated for shCol4 which differ in the length of 3' overhangs. This mixed population of Col1a1 targeting siRNAs may possibly be of superior potency than the corresponding homogeneous pool of siCol4 21mers alone. This may possibly explain the increased potency of shCol4 observed in HeLa cells in this study although this may not be the case as other factors may possibly have been involved.

The primary aim of the work detailed in this chapter was to generate a series of shRNA plasmids which effectively reduced Col1a1 expression levels. The suppression potency of these Col1a1 targeting shRNAs, shCol2 and shCol4 was effectively demonstrated in HeLa cells with Col1a1 expression levels reduced to 16.71%- 29.11 (±10.39-18.32%) and 15.35 – 18.50% (±5.43 – 11.89%) respectively (Table 4.1, Table 4.2 and Figure 4.10). This data together with siRNA Col1a1 suppression data generated in the study described in

chapter 3 of this thesis indicates the potency with which Col1a1 expression can be reduced using these effectors of RNAi. This stimulated further exploration of these Col1a1 suppressors as potentially therapeutic molecules for the treatment of Col1a1 linked Osteogenesis Imperfecta. Clearly further investigation is required, in particular validation of the efficacy of shCol2 and shCol4 at the protein level. Whilst HeLa cells are a useful model, they are not the target cell type for an Osteogenesis Imperfecta therapeutic strategy; therefore it was decided to carry out any further explorations in a more relevant cell type for Osteogenesis Imperfecta, mesenchymal stem cells.

As described in the introduction (section 1.3) one of the primary characteristics of Osteogenesis Imperfecta is the presence of brittle bones and a frequent fracture rate. OI is caused by mutations in either the Col1a1 or Col1a2 genes which encode the α1(I) and α2(I) chains of type I collagen. This is the most abundant type of collagen found in bone and is responsible for its flexibility and tensile strength. When either of the α1(I) and α2(I) chains is mutated the integrity of the type I collagen triple helix is impaired and consequently so too is flexibility and tensile strength. Osteoblasts are the "bone producing" cells of our body which synthesise and secrete type I collagen, therefore if the amount of mutant Col1a1 or Col1a2 produced by these cells could be reduced there may be a concomitant improvement in the severity of the OI phenotype. It is essential to evaluate the Col1a1 targeting suppressors described above and in chapter 3 in an appropriate cell type for OI, osteoblasts are unsuitable as they have a finite life span and would only provide limited therapeutic benefit. In contrast, mesenchymal stem cells (MSCs) which have the ability to self renew and proliferate extensively giving rise to amongst other cell lineages, osteoblasts. In theory, following delivery a population of MSCs should establish and provide a life long supply of normal cells to the recipient. Hence it was decided to explore the potential application of MSCs as both the therapeutic target and delivery vehicle for an RNAi focused approach to the treatment of Col1a1 linked Osteogenesis Imperfecta. To this end it would be necessary to assess the potency of a selection of the Col1a1 targeting siRNAs described in Chapter 3 of this thesis and the Col1a1 targeting shRNAs, shCol2 and shCol4 described in this chapter of

the thesis in MSCs both at the RNA and protein level. This study was carried out and is described in the following chapter of this thesis.

Chapter 5

Evaluation of RNAi in Mesenchymal Stem Cells

5.1 Introduction

As discussed in the introduction (section 1.2) MSCs have a variety of characteristics that makes them attractive in terms of therapeutic application, they can self renew and proliferate, they are easily isolated and readily cultured. MSCs appear to preferentially migrate to and engraft in areas of tissue damage, are thought to be of low immunogenicity and can potentially differentiate into a variety of cell types under the direction of the local biochemical milieu at the site of engraftment. In relation to OI, in theory following infusion of a normal or genetically corrected MSC population, the cells may preferentially migrate to bone, as OI bone is damaged bone and engraft in the bone tissue providing a life long supply of normal cells through self renewal and proliferation. Consequently the recipient should in principle be equipped with a life long supply of osteoblasts which produce wild type I collagen.

The potential therapeutic utility of MSCs for the treatment of Osteogenesis Imperfecta has been illustrated in a variety of reports. Early studies *in vivo* demonstrated that MSCs isolated from the OI Col1a1 minigene mouse retain their capacity to home to and engraft in bone and other tissues (Pereira, et al., 1995). Additionally, systemic administration of allogenic wild type murine MSCs resulted in a small but significant increase in bone mass in the Col1a1 minigene mouse (Pereira, et al., 1998). Reports of osteoblast specific gene expression following the differentiation of systemically administered MSCs further increased the attractiveness of MSCs as a therapeutic for Osteogenesis imperfecta (Hou, et al., 1999).

Two studies were carried out in humans involving a total of 6 children with OI. Initially a HLA matched bone marrow transplantation was carried out, donor cell engraftment and a subsequent increase in bone mineralisation was observed. A concurrent albeit short term increase in growth velocity was also observed (Horwitz, et al., 1999; Horwitz, et al., 2001). A follow up study was carried out in which each of the same 6 children received allogenic MSC transplants obtained from the same donors used in the first trial. Five of the children displayed engraftment in one or more sites such as skin, bone and marrow stroma and

the thesis in MSCs both at the RNA and protein level. This study was carried out and is described in the following chapter of this thesis.

presented with accelerated growth velocity, in addition, an increase in bone mineral content was observed in one recipient (Horwitz, et al., 2002).

Based upon the apparent successes of the Horwitz trials (Horwitz, et al., 2002; Horwitz, et al., 1999; Horwitz, et al., 2001) discussed above, Le Blanc and co-workers carried out an *in-utero* transplantation of allogenic MSCs. Ultrasound examination of the female foetus at weeks 15, 24 and 27 revealed that fractures were present and all limbs lengths were below the fifth percentile for length. A provisional diagnosis of OI was made. Umbilical cord fibroblasts were harvested at birth, DNA extracted and the Col1a1 and Col1a2 genes sequenced, the patient was found to be heterozygous for a missense mutation in exon 46 of the Col1a2 gene. The parents were given the therapeutic option of an *in-utero* transplantation of MSCs, which they accepted. MSCs were isolated from a male fetal liver (10 weeks gestation) and the procedure carried out at 32 weeks gestation. Bisphosphonate treatment was commenced at 4 moths of age. At 9 months of age 7.4% of bone biopsy cells were found to be of donor origin, there was no detectable immune response to the allogenic cells and trabeculae were found to be regularly arranged. Evaluation of skeletal mineralization revealed that it was 48% that of aged matched healthy controls at 3 months, 56% at 12 months and 76% at 22 months, it is notable that this may be due to the bisphosphonate treatment. The extensive variability in clinical phenotype and treatment with bisphosphonates makes delineation of the actual contribution of the engrafted MSCs to clinical improvement difficult. Nevertheless, the child grew appreciably in the time from transplant to biopsy at 9 months this may possibly be attributed to the presence of the transplanted MSCs. Notably only 3 factures occurred up to the age of two. Whilst more rigorously controlled studies are required, the success of the *in-utero* transplant itself, the lack of immunological response to the allogenic MSCs, the presence of detectable donor MSC engraftment and differentiation, accelerated growth and low fracture number observed in this patient are encouraging (Le Blanc, 2005).

Whilst transplantation of MSCs may provide a means of bone regeneration in some instances, it may be necessary to suppress the expression of mutant proteins, for example in the case of dominantly acting mutations in genes implicated in disorders such as OI. A

mutant human Col1a1 gene, the Cola1 minigene was successfully down regulated *in vivo,* this was achieved by crossing the human Col1a1 minigene mouse with a line of mice expressing an antisense transcript of the human Col1a1 minigene, this decreased the incidence of the lethal phenotype from 92% to 27% (Khillan, *et al.*, 1994). This clearly demonstrates the effectiveness of reducing the amount of negatively acting Col1a1 protein. An innovative therapeutic strategy for the inactivation of the mutated Col1a1 allele has been reported in which adeno-associated virus mediated homologous recombination resulted in the insertion of exogenous viral DNA sequences in the first intron of the mutated gene. Between 31% and 90% of integration events occurred at the Col1a1 locus with even distribution between the mutant and wild type allele. The phenotype of cells targeted at the mutant allele was converted to mild Type I OI. These cells produced normal type I collagen, retained their differentiation capacity and formed normal mineralised bone tissue when seeded in ceramic cubes following subcutaneous implantation in immunocompetent mice (Chamberlain, *et al.*, 2004).

Suppression of Col1a1 expression in MSCs has been achieved using a hammerhead ribozyme which targets the T allele of a C/T polymorphism located in the 3' UTR of the Col1a1 gene. As 40% of Caucasians are heterozygous for this particular polymorphism, 20% of Col1a1 associated OI patients will harbour a mutation on the cleavable T allele, in addition the homozygous T/T population will occur in an additional 40% of OI patients (Millington-Ward, et al., 1997). This hammerhead ribozyme was previously shown to preferentially cleave the T allele Col1a1 transcripts *in vitro* (Millington-Ward, *et al.*, 1999; Millington-Ward, *et al.*, 1997; Millington-Ward. 1999). The natural progression of the study was to determine if the ribozyme was equally effective in MSCs, the target cell type for the treatment of OI. Following retroviral delivery of the T targeting ribozyme to a C/T MSC cells line, Col1a1 expression was found to be down regulated by an average of 45% at the mRNA level and protein level (Millington-Ward, *et al.*, 2002). The same polymorphism was then targeted using siRNAs in MSCs, the siRNAs targeting both the C allele and T allele effectively reduced expression from their target allele, but allelic specificity was not achieved as the siRNAs could also suppress the alternative allele (Millington-Ward, 2004).

Whilst targeting of the mutant transcripts is the traditional approach toward reducing the impact of dominant negative mutations, Pochampally et al (2005) demonstrated in an *in vitro* study that over expression of a normal Col1a1 allele in MSCs isolated from a patient with Type IV OI relieved the OI phenotype with the increased production of wild type Col1a1 RNA and protein. In addition MSCs expressing the Col1a1 transgene had an increased capacity to differentiate into mature mineralising osteoblasts (Pochampally, *et al.*, 2005). Clearly the research undertaken to date lends support to the view that MSCs represent a potentially powerful cell type for the development of therapeutics for disorders such as OI.

Traditional transfection methods are not ideally suited for plasmid delivery to MSCs resulting in low transfection efficiencies and high cell mortality (Reiser, *et al.*, 2005). However, viral vectors have been successfully used for the delivery of a variety of transgenes to MSCs, including reporter genes such as GFP (Lee, *et al.*, 2001) and potentially therapeutic genes such as the BMPs (Cheng, *et al.*, 2001; Cowan, *et al.*, 2005; Dayoub, *et al.*, 2003; Dragoo, 2003; Dumont, 2002; Gysin, 2004; Tsuda, 2005; Zhang, 2004). A variety of vector types such as retroviruses (Devine, *et al.*, 2001; Devine, *et al.*, 2003; Gysin, 2004), lentiviruses (Lee, *et al.*, 2001), adenoviruses (Cheng, *et al.*, 2001; Dayoub, *et al.*, 2003; Kurozumi, *et al.*, 2005; Tsuda, 2005) and adeno-associated viruses (Chamberlain, *et al.*, 2004; Kumar, *et al.*, 2004; Kurozumi, et al., 2005; Tsuda, 2005) have been used successfully. However, some gene delivery tools including many viral vectors have associated safety concerns *in vivo*, such as insertional mutagenesis and frequent induction of strong cytotoxic and immunological responses (Glover, *et al.*, 2005).

Consequently, there has been significant impetus toward development of alternative non-viral gene delivery systems. Extensive development has occurred in the area of physical delivery technologies. Efficient plasmid delivery was achieved using electroporation technology with transgene expression detectable in 98% of cells following 12 days enrichment in selective media (Peister, 2004). A vibration based transfection system the Gene Symphoniser achieved MSC transfection efficiencies of 74% with 40% cell mortality. (Song, *et al.*, 2004b). It is notable that for each transfection method used the proliferation

and differentiation capacity of the cells was largely maintained (Peister, 2004; Song, et al., 2004b)

Another recently developed transfection technology is the Amaxa Nucleofector™. Nucleofection is an electroporation based technology which utilises specifically optimised cell type specific solutions and electrical parameters to effect direct delivery of nucleic acids to the nucleus of the cell. The efficacy of the reaction is independent of the mitotic status of the cell, this technique has been shown to effectively delivery nucleic acids into stem cells and other primary cell lines (Distler, *et al.*, 2005; Gartner, *et al.*, 2006; Hoelters, 2005; Quenneville, *et al.*, 2004; Siemen, *et al.*, 2005; Yin, *et al.*, 2006). Notably, the nucleofection process does not appear to negatively impact on the proliferative or differentiation capacity of the mesenchymal stem cells (Aluigi, *et al.*, 2006; Haleem-Smith, 2005; Hamm, *et al.*, 2002). A recent publication details the use of the Nucleofector™ technology to deliver BMP2 and BMP9 transgenes to MSCs which formed eptopic bone following *in vivo* implantation in mice (Aslan, *et al.*, 2006).

In relation to the current study as traditional lipid based transfection reagents did not result in high levels of nucleic acid delivery to MSCs (data not shown), an alternative transfection method was required. Based upon the success reported in the literature it was decided to explore the efficacy of the Amaxa Nucleofector™ for the delivery Col1a1 targeting siRNAs and shRNAs. siRNA duplexes are effective mediators of sequence specific RNA silencing and have been shown to down regulate endogenous Col1a1 expression in MSCs in our laboratory (Millington-Ward, 2004). This chapter of the thesis describes the culturing of mesenchymal stem cells and optimisation of mesenchymal stem cells nucleofection. In addition evaluation of Col1a1 expression levels is MSCs at both the mRNA and protein level following nucleofection with Col1a1 targeting siRNAs and shRNAs is described and discussed.

5.2 Materials and Methods

Evaluation of Col1a1 siRNA and shRNA potency in Mesenchymal Stem Cells

The specific methods and materials employed to carry out the research presented in this chapter are detailed below. All general methods and sequences of any primers used in the work carried out in this chapter are described in full in chapter 2.

5.2.1 Mesenchymal stem cell culturing conditions

The mesenchymal stem cells used in this study were a kind gift of Dr. G. Jordan of Musgrave Park Hospital Belfast. The cells were grown from bone explants harvested with consent from a femur fracture of a 21 year old male (personal communication Dr. G. Jordan). The MSCs were provided at passage 1. The MSCs were then expanded (section 2.8.2) and frozen (section 2.8.3) down to generate a stock of MSCs which were used throughout this project. Cells were not used beyond passage 7.

Cells were grown in α MEM medium containing 5% L-Glutamine, 5% Sodium Pyruvate, 20% Fetal Bovine Serum (South American), 100μM ascorbic acid and 5% penicillin. A frozen vial containing approximately $1x10^6$ cells was thawed and plated as described in section 2.8.1. The cells were incubated at 37°C in the presence of 6 % CO_2, when cells reach 70% – 80% confluencey they were passaged as described in section 2.8.2

5.2.2 Mesenchymal stem cell transfection using the Amaxa Nucleofector device.

The MSC specific nucleofection kit VPE-1001 was used for all MSC nucleofection reactions. The nucleofection was initially carried out as outlined in the Amaxa MSC specific nucleofection protocol, it was found that using the cell numbers recommended that cell survival rates were low. The protocol was adapted and optimized to the protocol described below. Note: The electrical parameters of the Amaxa programs and cell solution contents are considered the intellectual property of the company and no details are provided.

5.2.2.1 Optimisation of Mesenchymal stem cell nucleofection protocol

The Amaxa protocol supplied with the MSC nucleofection kit advised to use $5x10^5$ cells and to deliver 2μg of plasmid DNA. It was found following these parameters that cell mortality was high. Therefore it was necessary to optimise the nucleofection parameters. It is known that large amounts of DNA can be toxic, this is particularly relevant to Nucleofection as the nucleic acids have to be of high purity, all plasmid preps used were endotoxin free. Therefore the effect of halving the amount of plasmid delivered (1μg of plasmid DNA) was explored. The final optimization step involved increasing the number of MSCs per nucleofection, the rational was that presence of an increase cell number, and (double the number $1x10^6$) may provide a level of protection to the cells and reduce the impact of the nucleofection on the cells. Three nucleofections was carried out per optimisation step.

5.2.2.2. Optimised MSC nucleofection protocol

Each nucleofection reaction requires $1x10^6$ MSCs. Cells were passaged 6 to 7 days prior to nucleofection such that on the day of nucleofection the cells were 85% confluent. The cells were harvested by trypsinisation and counted (section 2.8.2 and 2.8.3). Individual aliquots of $1x10^6$ cells were centrifuged @ 130g for 10 minutes, the media aspirated and cells resuspended in 100μl of pre-warmed MSC specific Amaxa nucleofection solution.

For all siRNA experiments 40pm of each siRNA siCol2, siCol4, the negative controls siGFP or siNon-tgt was added to an aliquot of $1x10^6$ cells. The MSC and siRNA solution was gently mixed and transferred directly to the specialized Amaxa nucleofection cuvette which was then placed into the Nucleofector device and program U23 a human mesenchymal stem cell specific program optimised for achieving high nucleofection efficiencies was selected. For all shRNA experiments, 1μg of endotoxin free shCol2, shCol4 or the negative control shGFP plasmid was added to each aliquot of $1x10^6$ cells and the suspension transferred directly to a specialized Amaxa nucleofection cuvette, placed into the Nucleofector device and program U23 selected. A separate aliquot of $1x10^6$ MSCs was nucleofected with a GFP expression plasmid, pmaxGFP to determine transfection efficiencies; this plasmid was supplied in the MSC nucleofection kit. Once completed 500μl of pre-warmed media was added to the

cuvette, the cells were then transferred to a single well of a 6 well plate. The pmaxGFP nucleofected cells were grown on three 1% gelatin coated coverslips in a well of a 6 well plate. The cells were then incubated in a humidified 37°C, 5% CO_2 incubator. All Col1a1 targeting siRNA and shRNA nucleofections were carried out in triplicate in a single nucleofection experiment and the control siRNA and shRNA transfections carried out in duplicate. Each individual experiment was repeated 3 times.

5.2.3 RNA extraction from siRNA and shRNA nucleofected Mesenchymal Stem Cells

MSCs were nucleofected with either the siRNAs or shRNAs according to the optimized protocol described in 5.2.2. As the electroporation procedure is quite an insult to the cells, the MSCs were allowed to recover for 24 hours and an additional 24 hours allocated for the RNAi effect to occur. At 48 hours post-nucleofection RNA was extracted as described in section 2.9.3 using the RNeasy RNA extraction kit which was purchased from Qiagen UK. This kit was used to optimize the RNA yield from the nucleofected MSCs. As three separate nucleofections of Col1a1 targeting siRNA (siCol2 and siCol4) or shRNA (shCol2 and shCol4) were carried out in a single nucleofection experiment, the 3 RNA samples were pooled to generate a single sample representative of the three independent nucleofections. This generated a total of 3 pooled RNA samples for siCol2, siCol4, shCol2 and shCol4 from a total of 9 nucleofections per suppressor. In the case of the non-targeting controls the duplicate nucleofections were pooled per transfection experiment generating a total of 3 RNA samples for siGFP, siNon-Tgt, shGFP and shLuciferase which were generated from total of 6 independent nucleofections per control.

5.2.4 Real time RT-PCR evaluation of Col1a1 mRNA expression levels in Mesenchymal Stem Cells nucleofected with Col1a1 targeting siRNAs or shRNAs.

The Col1a1 expression levels in the RNA samples isolated and pooled as described in 5.2.6 was evaluated using real time RT-PCR. All real time RT-PCR reactions were set up as described in 2.9.4. The housekeeping genes used in this study as for the studies described in Chapters 3 and chapters 4 of this thesis were GAPDH and β-Actin. In all cases each RNA

sample was analyzed in triplicate (i.e. 3 wells per RNA) per primer set in a single 96 well plate RT-PCR assay. Each 96 well plate real time RT-PCR assay was run 3 times. Col1a1 expression levels were evaluated using the relative quantification method. This method first involves the normalization of Col1a1 expression levels; achieved by dividing the Co1a1 expression value in a given sample by the housekeeper expression value (in this case GAPDH and β-Actin) in the same sample. The relative expression values are then determined comparing the normalized Cola1 expression values from cells transfected with the Col1a1 targeting siRNAs (siCol2 and siCol4) or Col1a1 targeting shRNAs (shCol2 and shCol4) against the normalised Col1a1 expression values in samples which were transfected with control siRNAs (siGFP and siNon-tgt) or control shRNA (shGFP and shLuciferase). The normalised Col1a1 expression value in the control nucleofected MSCs was set at 100%.

The total data sets generated for each siRNA and shRNA study described in this chapter were analyzed using GraphPad InStat3. The mean and standard deviation was calculated, additionally a non parametric ANOVA was carried out in order to determine if the Col1a1 expression levels in MSCs nucleofected with siCol2, siCol4 and siGFP or siNon-tgt differed significantly. This analysis was repeated for the shRNA studies in order to determine if Col1a1 expression levels nucleofected with shCol2, shCol4 and shGFP or shLuciferase differs significantly. If a p value of $p<0.05$ implies that the data sets analysed differ significantly. In order to determine which of the data sets were different it was necessary to carry out post test analysis. The post test employed was Dunn's multiple comparison test; this analysis compares each of the data sets, e.g. siGFP to siCol2, siGFP to siCol4 and siCol2 to siCol4. A p value of $p <0.05 - p<0.0001$ indicates that the two data sets compared differ significantly, whilst a p value of $p>0.05$ indicates that the two data sets compared do not differ significantly. Therefore use of this post test facilitates the determination if the Col1a1 expression levels in MSCs nucleofected with siCol2 or siCol4 are significantly different from Col1a1 expression levels in MSCs nucleofected with siGFP or siNon-tgt, additionally comparison of siCol2 to siCol4 enables one to determine if either siCol2 or siCol4 more potently suppresses Col1a1 expression in MSCs, if this were the case a p value of $p<0.05$ would be generated from the comparison. Student t-test analysis was also carried out, in

which the siRNAs were compared to their shRNA counterparts to determine if the shRNAs or siRNAs were more effective at reducing Col1a1 expression levels in MSC. P values were considered to be significant if $p<0.05$. This analysis was carried out when Col1a1 expression values were normalized using GAPDH and β-Actin. All graphs were generated using Excel.

5.2.8 Evaluation of Type I collagen protein production in Mesenchymal Stem cells nucleofected with Col1a1 targeting siRNA or shRNAs using the Metra CICP ELISA System.

Levels of *de novo* synthesised type I collagen protein in MSCs following siRNA and shRNA nucleofection were evaluated using the Metra CICP ELISA kit, purchased from Quidel Corporation. The assay is based upon the fact that type I collagen as described previously (Section 1.3) is synthesised as a pro-collagen, which consists of mature collagen with N and C terminal pro-peptides. These pro-peptides are cleaved from the collagen molecule by specific proteases prior to incorporation of the collagen triple helix into a collagen fibril (Figure 5.1). The release of these peptides into the media in which the cells are growing provides a stiochiometric representation of collagen production. The Metra CICP assay measures the *C*-Terminal of type I collagen *P*rotein. 24 hours after nucleofection with either the siRNAs or shRNAs the MSC culture media was changed for serum free α MEM. The MSCs were allowed to secrete Type I collagen into the medium for 24 hours. Medium was then collected and analysed for *de novo* Type I collagen production according to the manufacturer's guidelines.

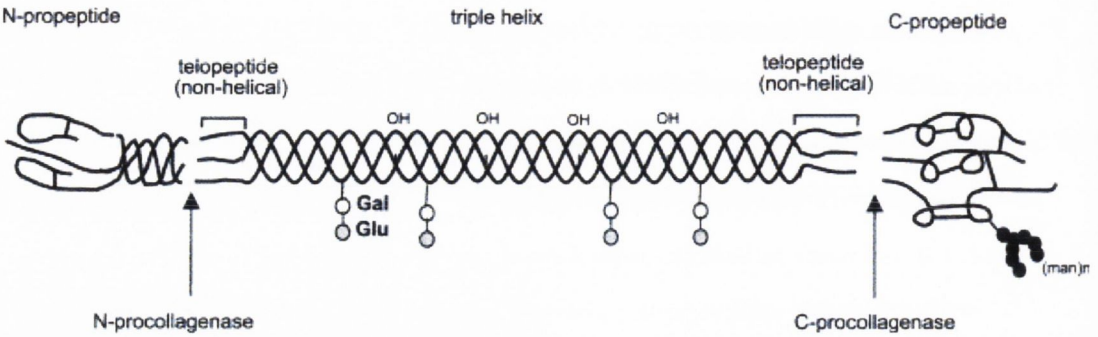


Image taken from Baum and Brodsky 1999

Figure 5.1 Type I collagen triple helix. It consists of N & C propetides which are cleaved by N & C procollagenase generating the mature collagen triple helix. The C propeptide cleavage product is measured in the Metra CICP ELISA (Baum and Brodsky, 1999)

The Metra CICP assay is a sandwich enzyme immunoassay in a micro-titre format, in this type of assay the plate is coated with a capture antibody the test sample is added and any antigen present binds to capture antibody. The detecting antibody is then added and binds to antigen. The enzyme-linked secondary antibody is added next this binds to detecting antibody. Finally the substrate is added and is converted by enzyme to detectable form.

In the case of this particular assay, the monoclonal anti-CICP capture antibody is coated on the plate. The medium collected from the nuclefected cells is applied, any antigen, i.e. C terminal propetide should be captured. The rabbit anti-CICP is then added, this should bind to the captured C terminal propeptide. The goat anti-rabbit alkaline phosphatase conjugate is added, which should bind to the rabbit anti-CICP. Finally the pNitrophenly phosphate substrate is added the optical density of the solution measeured at 405nm and a four parameter fit curve equation used to calculate the concentatation of CICP present.

All reagents used were provided in the Metra CICP kit. The test samples were diluted in assay buffer between 1 and 48 fold such that they fall within the detection limits of the

assay. Following equilibration of the anti-CICP coated strips to room temperature, 100µl of the Metra CICP concentration standards, test samples or controls were plated in duplicate and incubated at room temperature for 2 hours. The strips were then inverted and washed 3X with 300µl of 1X wash buffer per well and blotted dry on tissue paper. 100µl of the rabbit anti-CICP was then added to each well and the reaction incubated for 45-50 minutes at room temperature. The strips were then washed 3X with 300µl 1X wash buffer as described. 100µl of reconstituted enzyme conjugate was added to each well and incubated for 45-50min at room temperature. 100µl of working substrate solution was then added to each well, incubated at 18°C to 25°C for 30 -35 minutes. Samples were washed 3X with 1X wash buffer as before. 50µl of stop solution was added to each well to stop the reaction and the OD read within 15 minutes at λ405nm. A four parameter fit curve was used to analyze the data and the concentration of the C terminal pro-peptide sample determined from the curve.

Three independent media samples were collected per siRNA and pooled to generate a single representative sample per transfection experiment. Three transfection experiments were carried out generating a total of 3 media samples per siRNA, one from each transfection experiment. For the controls siGFP and siNon-tgt two samples were pooled per transfection, as only two wells of cells were transfected with siGFP and siNon-tgt in each nucleofection experiment. Three pooled samples were generated for both siGFP and siNon-tgt from the three transfection experiment but only two were assayed.

A) siCol2

Target RNA *Col1a1 NM_000088*

Nucleotides 1404-1404 5' *CCAAGGGTAACAGCGGTGAA* 3'

Sense 5' r(CAAGGGUAACAGCGGUGAA)d(TT) 3'

Antiense 5' r(UUCACCGCUGUUACCCUUG)d(GG) 3'

Annealed 5' CAAGGGUAACAGCGGUGAATT 3'

3' GGGUUCCCAUUGUCGCCACUU 5'

B) siCol4

Target RNA *Col1a1 NM_000088*

Nucleotides *227-247* 5' *ACCAATCACCTGCGTACAGA* 3'

Sense 5' r(CAAUCACCUGCGUACAGAA)d(TT)

Antiense 5' r(UUCUGUACGCAGGUGAUUG)d(GT)

Annealed 5' CAAUCACCUGCGUACAGAATT 3'

3' GUGUUAGUGGACGCAUGUCUU 5'

Figure 5.2 Col1a1 siRNA target sequences, sense and antisense oligonucleotides and annealed duplexes. A) siCol2, **B)** siCol4. **(Note:** This data was presented in chapter 3 of this thesis and is presented here again for clarity)

A) siGFP

Target DNA

Nucleotides 120 -143 *5'* CGGCAAGCTGACCCTGAAGTTCAT 3'

Sense 5' r(GCAAGCUGACCCUGAAGUUCAU) 3'

Antisense 5' r(GAACUUCAGGGUCAGCUUGCCG) 3'

Duplex 5' GCAAGCUGACCCUGAAGUUCAU 3'

3' GCCGUUCGACUGGGACUUCAAG 5'

B) siNon -Targeting

Sense 5' r(CAAUCACCUGCGUACAGAA)d(TT) 3'

Anti-Sense 3'r(UUCUGUACGCAGGUGAUUG)d(GT)5'

Duplex 5' CAAUCACCUGCGUACAGAATT 3'

3' GTGUUAGUGGACGCAUGUCUU 5'

Figure 5.3 Col1a1 siRNA target sequences, sense and antisense oligonucleotides and annealed duplexes. A) siGFP, **B)** siNon-tgt **(Note:** This data was presented in chapter 3 of this thesis and is presented here again for clarity)



Figure 5.4 shRNA target sequence and annealed sense and antisense oligonucleotides. A) sCol2, B) shCol4. Each 63mer sense and antisense oligo consists of a 19mer of sense and antisense orientation separated by a 9ntd hairpin sequence (highlighted in green). The oligos are annealed and cloned into pSuper using the BglII (highlighted in pink) and HindIII (highlighted in blue) overhangs which are included in the 64mer. The design is based upon that of Brummelkamp 2002 (Brummelkamp, et al., 2002b). **(Note:** This data was presented in chapter 4 of this thesis and is presented here again for clarity)

D) *ShGFP*

Target mRNA *CGGCAAGCTGACCCTGAAGTTCAT*

ShGFP oligos

*GATCCCC*CGGCAAGCTGACCCTGAAGTTCAT*ttcaagaga* ATGAACTTCAGGGTCAGCTTGCCG*TTTTTGGAAA*

GGG GCCGTTCGACTGGGACTTCAAGTA*aagttctct* TACTTGAAGTCCCAGTCGAACGGC*AAAAACCTT*

E) *ShLuciferase*

Target mRNA *GCGTACGGGGAAAACTTCGA*

ShLuciferase oligos

GATCCCC GCGTACGGGGAAAACTTCGA *ttcaagaga* TGGTTAGTGGACGCATGTCT*TTTTTGGAAA*

GGG CGCATGCCCCTTTTGAAGCT *aagttctct* AGCTTCATAAGGCGCATCG*AAAAACCT*

Figure 5.5 shGFP target sequence, annealed sense and antisense oligonucleotides. Each 64mer sense and antisense oligo consists of a 19mer of sense and antisense orientation separated by a 9ntd hairpin sequence (highlighted in green). The oligos are annealed and cloned into pSuper using the BglII (highlighted in pink) and HindIII (highlighted in blue) overhangs which are included in the 64mer. The design is based upon that of Brummelkamp et al 2002 (Brummelkamp, et al., 2002b) **(Note:** This data was presented in chapter 4 of this thesis and is presented here again for clarity).

5.3 Results

The work described in chapter 3 and chapter 4 of this thesis details the RNAi of Col1a1 expression in HeLa cells employing a selection of Col1a1 targeting siRNAs and shRNAs. Whilst effective suppression of Col1a1 expression was demonstrated at the mRNA level, it was not validated at the protein levels. It was decided to proceed with the study in MSCs, which, as mentioned, are ideally suited for exploration of potential OI therapeutics.

5.3.1 Optimization of Mesenchymal Stem Cell nucleofection

As discussed MSCs are difficult to transfect using traditional methods, therefore it was first necessary to establish a suitable means with which the required Col1a1 targeting siRNAs and shRNAs could be delivered, to this end use of the Amaxa Nucleofector™ was explored. A series of optimisation experiments were carried, three nucleofections were carried out per optimisation step. The Amaxa protocol advised to deliver 2μg of plasmid DNA to $5x10^5$ MSCs per nucleofection. Following these parameters cell mortality was high, in the region of 70% when using program U-23 which is recommended for efficient MSC nucleofection. It was found that decreasing the amount of plasmid DNA to 1μg decreased cell mortality rates to the region of 40%. Additionally, increasing the MSC numbers to $1x10^6$ per nucleofection using program U-23, significantly decreased cell mortality to the region of 20% - 30% and nucleofection efficiencies were high in the region of 60% – 70% (Figure 5.5). For the HeLa cell experiments described in chapters 3 and 4 the RNA was extracted 24 hours after transfection, in the work described in this chapter the RNA was extracted 48 hours after transfection as it was felt that the MSCs required an initial recovery period as nucleofection is quite an insult to the cells.

5.3.2 Nucleofection of Mesenchymal Stem Cells with Col1a1 targeting siRNAs

The nucleofector technology facilitated efficient delivery of the siRNA duplexes to the MSCs, it is claimed by the Amaxa who produce the Nucleofector device that delivery is direct to the nucleus of the cells. Unfortunately the company, Amaxa, will not provide

details as to the mechanism by which this is achieved. Each Col1a1 targeting siRNA nucleofection experiment was carried out in triplicate and the controls, siGFP and siNon-tgt, in duplicate. Three independent transfection experiments were carried out. RNA was extracted from the MSCs 48 hours post nucleofection. As three separate nucleofections of Col1a1 targeting siRNA (siCol2 and siCol4) or shRNA (shCol2 and shCol4) were carried out in a single nucleofection experiment, the 3 RNA samples were pooled to generate a single RNA sample per nucleofection experiment. This generated a total of 3 pooled RNA samples for siCol2, siCol4, shCol2 and shCol4 from a total of 9 nucleofections per suppressor. In the case of the non-targeting controls, the duplicate nucleofections were pooled per transfection experiment generating a total of 3 RNA samples for siGFP, siNon-Tgt, shGFP and shLuciferase which were generated from a total of 6 independent nucleofections per control.

5.3.3 Real time RT-PCR evaluation of the potency of Col1a1 targeting siRNAs in Mesenchymal Stem Cells.

RNA was extracted from MSCs 48 hours post nucleofection with the Col1a1 targeting siRNAs, siCol2 and siCol4. Col1a1 expression levels were evaluated using real time RT-PCR. All reactions were set up as described in section 2.9.2. Each RNA sample was analysed in triplicate per run (i.e. 3 wells of a 96 well plate) and each run was carried out 3 times. Col1a1 expression levels were evaluated using the relative quantification method. Col1a1 expression levels in all samples were normalized to GAPDH and β-Actin, this is achieved by dividing the Col1a1 expression value in a sample by housekeeping gene expression value in same sample. To determine the relative Col1a1 expression values, the normalized Col1a1 expression levels in the cells transfected with the Col1a1 targeting siRNAs (siCol2 and siCol4) were then compared to normalized Col1a1 expression levels in cells transfected with the non-targeting controls (siGFP or siNon-tgt). The relative Col1a1 expression data was analyzed using GraphPad Instat 3. The mean and standard deviation was calculated, non parametric ANOVA analysis and Dunn's multiple comparison post test analysis was carried out for both GAPDH and β –Actin normalized Col1a1 expression data.

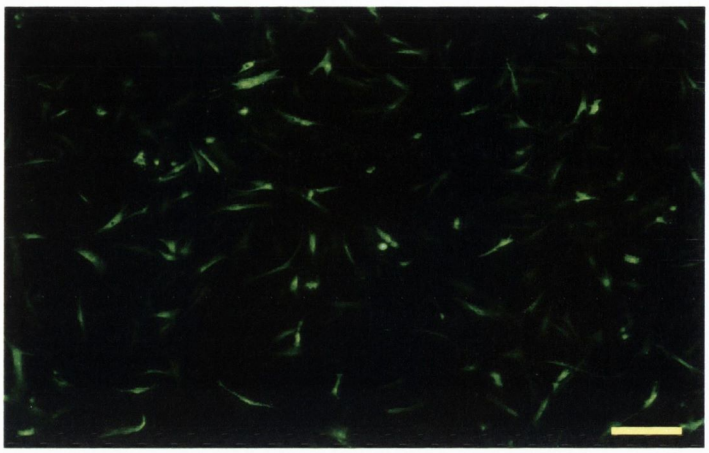
Both siCol2 and siCol4 potently reduced Col1a1 expression levels in MSCs. when normalized using GAPDH, siCol2 potently reduced Col1a1 expression levels to 2.64% (± 1.60%) and 5.08% (± 2.64%) that of Col1a1 expression levels in MSCs treated with siGFP or siNon-Tgt respectively. Similarly, when normalized using GAPDH, siCol4 reduced Col1a1 expression levels to 2.67% (± 1.25%) and 4.37% (± 2.69%) that of Col1a1 expression levels in MSCs treated with siGFP or siNon-Tgt respectively. Non parametric ANOVA analysis was carried out in which the Col1a1 expression data generated for siCol2, siCol4 and siGFP or siNon-tgt were compared, if the p value generated was $p<0.05$ - <0.0001 it indicated the data sets differed significantly. Comparison siGFP, siCol2 and siCol4 Col1a1 expression data generated a p value of $p<0.0001$, indicating that the data sets were significantly different. The Dunn's multiple comparison test revealed that the Col1a1 expression levels in MSCs nucleofected with siCol2 or siCol4 were significantly different from Col1a1 expression levels in MSCs nucleofected with siGFP, a p value of $p<0.001$ was generated from each pairwise comparison. Comparison of siCol2 to siCol4 generated a p value of $p>0.05$ indicating the neither siCol2 nor siCol4 appeared to more potently reduce Col1a1 expression in MSCs. This was reconfirmed in a second ANOVA in which the Col1a1 expression data for siCol2, siCol4 and siNon-tgt were compared, once again the ANOVA generated a p value of $p<0.0001$, the Dunn's multiple comparison test indicated that Col1a1 expression levels were significantly different in MSCs nucleofected with siCol2 or siCol4 when compared to siNon-tgt Co1a1 expression data with a p value of $p<0.001$ generated from each pairwise comparison (Table 5.1 and Figure 5.7).

The data was also normalized using β-Actin. In this instance Col1a1 expression levels in MSCs treated with siCol2 were reduced to 2.05% (± 0.68%) and 4.73% (±1.54%) of the Col1a1 expression levels in MSC treated with siGFP and siNon-Tgt respectively. Similarly, siCol4 potently reduced Col1a1 expression levels to 1.49% (± 0.86%) and 3.48% (±1.95%) that of Col1a1 expression levels in MSCs treated with siGFP or siNon-Tgt respectively ANOVA analysis revealed that the Col1a1 expression data sets siCol2, siCol4 and siGFP differed significantly, with a p value of $p<0.0001$ generated following comparison of siCol2,

siCol4 and siGFP or siCol2, siCol4 and siNon-tgt. Dunn's multiple comparison test revealed that Col2 and siCol4 Col1a1 expression data differed significantly from the siGFP Col1a1 expression data with a p values of $p<0.01$ and $p<0.001$ generated respectively from each pairwise comparison Table 5.2 and Figure 5.7.

A second ANOVA in which the Col1a1 expression data for siCol2, siCol4 and siNon-tgt were compared revealed that the 3 data sets were significantly different, the ANOVA generated a p value of $p<0.0001$, the Dunn's multiple comparison test indicated that Col1a1 expression levels were significantly different in MSCs nucleofected with siCol2 or siCol4 when compared to siNon-tgt Co1a1 expression data with a p value of $p<0.001$ generated from each pairwise comparison, comparison of Col1a1 expression levels in MSCs nucleofected with siCol2 and siCol4 generated a p value of $p>0.05$ indicating that neither siCol2 nor siCol4 appears to effect more potent suppression of Col1a1 expression in MSCs (Table 5.2 and Figure 5.7).

A)



B)

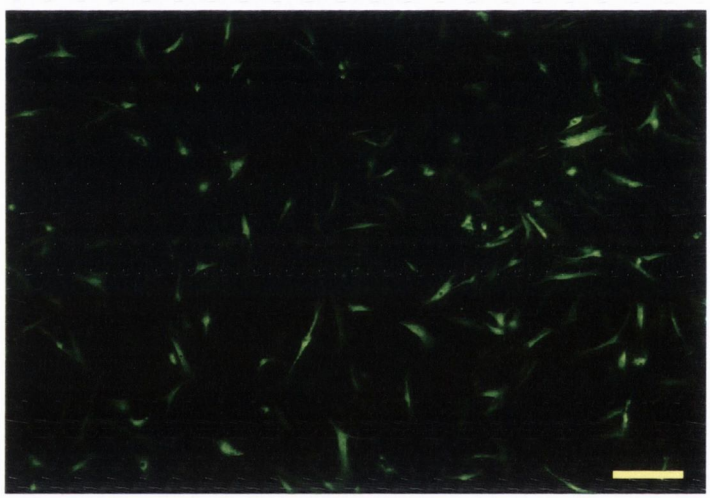


Figure 5.6. Evaluation of the efficacy of plasmid delivery to Mesenchymal Stem Cells using the Amaxa Nucleofector device. A and B) Example of optimized nucleofection $1x10^6$ MSCs were nucleofected with 1μg of an eGFP expression plasmid, plated onto coverslips and analysed for eGFP expression 24 hours post nucleofection using fluorescence microscopy. Evaluation of GFP fluorescence indicated that efficiencies were in the region of 60%-70%, scale bar 20μm

A)	*siGFP*	*siNon-tgt*	*siCol2*	*siCol4*	*ANOVA*
siGFP	100%		3.46%	2.67%	p<0.0001
StdDev	13.95		1.6	1.25	
siNon-tgt		100%	5.08%	4.39%	p<0.0001
StdDev		20.52	2.64	2.69	

B)	siCol2	siCol4
siGFP	p<0.001	p<0.001
siNon-tgt	p<0.001	p<0.001
siCol2		p>0.05
siCol4	p>0.05	

Table 5.1 GAPDH normalised Col1a1 mRNA expression levels in MSCs following nucleofection with siCol2, siCol4, siGFP or siNon-tgt. A) Col1a1 mRNA expression levels in MSCs nucleofected with siCol2 or siCol4 normalized to GAPDH expression. Relative Col1a1 mRNA expression values in MSCs transfected with siCol2 and siCol4 were determined by comparison to normalised Col1a1 expression levels in MSCs nucleofected with siGFP (green bars) or siNon-tgt (Blue bars) which were set a 100%. siCol2 reduced MSC Col1a1 expression to 3.46% (±1.6) and 5.08% (±1.6) of Col1a1 expression levels in MSCs transfected with siGFP and siNon-Tgt respectively. siCol4 reduced MSC Col1a1 expression to 2.67% (±1.25) and 4.39% (±2.69) of Col1a1 expression levels in MSCs nucleofected with siGFP and siNon-Tgt respectively. Non parametric ANOVA analysis in which the Col1a1 expression data sets for siCol2, siCol4 and siGFP or siNon-tgt were compared generated a p value of $p<0.0001$ indicating that Col1a1 expression levels in MSCs nucleofected in siCol2, siCol4 and siGFP or siNon-tgt differed significantly. **B)** p values generated from Dunn's multiple comparisons test, each pairwise comparison siGFP:siCol2, siGFP:siCol4, siNon-tgt:siCol2 and siNon-tgt:siCol4 generated a p value of $p<0.001$ indicating that Col1a1 expression levels in MSC nucleofected with siCol2 or siCol4 differed significantly from Col1a1 expression levels in MSC nucleofected with siGFP or siNon-tgt. Pairwise comparison of siCol2 to siCol4 generated a p value of $p>0.05$ indicating that Col1a1 expression levels in MSCs nucleofected with siCol2 or siCol4 do not differ significantly.

A)	*siGFP*	*siNon-tgt*	*siCol2*	*siCol4*	*ANOVA*
siGFP	100%		2.05%	1.49%	p<0.0001
StdDev	26.42		1.6	1.25	
siNon-tgt		100%	4.73%	3.48%	p<0.0001
StdDev		21.43	1.54	1.95	

B)	siCol2	siCol4
siGFP	p<0.01	p<0.001
siNon-tgt	p<0.001	p<0.001
siCol2		p>0.05
siCol4	p>0.05	

Table 5.2 β-Actin normalised Col1a1 mRNA expression levels in MSCs following nucleofection with siCol2, siCol4, siGFP or siNon-tgt. A) Col1a1 mRNA expression levels in MSCs nucleofected with siCol2 or siCol4 normalized to β-Actin expression. Relative Col1a1 mRNA expression values in MSCs transfected with siCol2 and siCol4 were determined by comparison to normalised Col1a1 expression levels in MSCs nucleofected with siGFP (green bars) or siNon-tgt (Blue bars) which were set a 100%. siCol2 reduced MSC Col1a1 expression levels to 2.05% (±1.6) and 4.73% (±1.54) of Col1a1 expression levels in MSCs transfected with siGFP and siNon-Tgt respectively. siCol4 reduced MSC Col1a1 expression to 1.49% (±0.86) and 3.48% (±1.95) of Col1a1 expression levels in MSCs nucleofected with siGFP and siNon-Tgt respectively. Non parametric ANOVAs analysis in which the Col1a1 expression data sets for siCol2, siCol4 and siGFP or siNon-tgt were compared generated a p value of p<0.0001 indicating that Col1a1 expression levels in MSCs nucleofected in siCol2, siCol4 and siGFP or siNon-tgt differed significantly. **B)** p values generated from Dunn's multiple comparisons test, each pairwise comparison, siGFP:siCol4, siNon-tgt:siCol2 and siNon-tgt:siCol4 generated a p value of p<0.001 and siGFP:siCol2 a p value of p<0.01 indicating that Col1a1 MSC nucleofected with siCol2 or siCol4 differed significantly from Col1a1 in MSC nucleofected with siGFP or siNon-tgt. Pairwise comparison of siCol2 to siCol4 generated a p value of p>0.05 indicating that Col1a1 expression levels in MSCs nucleofected with siCol2 or siCol4 do not differ significantly.

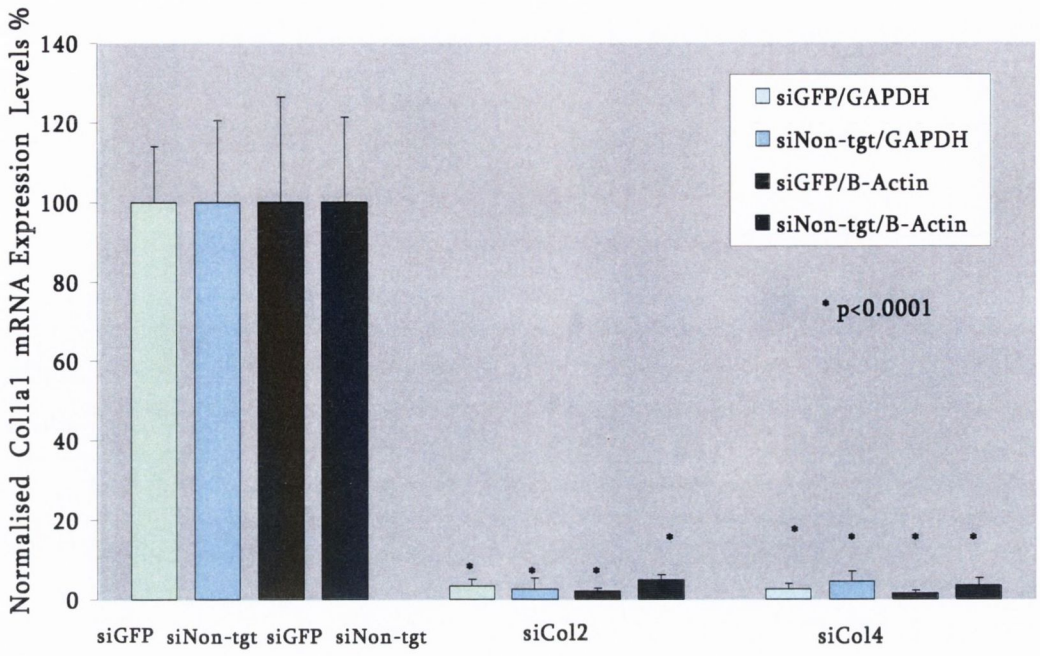


Figure 5.7 Relative Col1a1 mRNA expression levels in MSCs following siRNA nucleofection. MSCs were nucleofected with 40pm of siCol2, siCol4, siGFP and siNon-tgt, RNA was extracted 48 hours later, Col1a1 expression levels were evaluated by real time RT-PCR. Col1a1 mRNA expression levels in all samples was normalized to GAPDH and relative expression levels determine by comparison to Col1a1 expression levels in MSCs transfected with the non-targeting controls siGFP (Plain green bars) or siNon-tgt (Plain blue bars) which were set to 100%. β-Actin normalization expression was also carried out for all samples and Col1a1 mRNA expression compared to siGFP (green hatched bars) and siNon-tgt (blue hatched bars) set to 100%. Col1a1 expression in siCol2 and siCol4 nucleofected MCS ranged from 2.05% to 5.08% (± 0.68 – 2.64%) and 1.49% to 4.39% (± 0.86 – 2.69%) respectively. Non parametric ANOVA analysis which compared siCol2, siCol4 and siGFP or siNon-tgt for both GAPDH and β-Actin normalised data generated a p value of $p<0.0001$. Indicating that the Col1a1 expression data sets differed. Dunn's post test analysis involves a series of pairwise comparisons to determine exactly which data sets differed and the results are detailed in Table 5.1 and Table 5.2. Error bars represent, Standard deviations of Col1a1 expression, analyses were carried out using GraphPad InStat3.

5.3.4 Nucleofection of MSCs with Col1a1 targeting shRNAs

With the efficacy of the siRNA duplexes siCol2 and siCol4 determined at the mRNA level the next step was to assess the potency of the Col1a1 targeting shRNAs, shCol2 and shCol4 in MSCs. The optimized nucleofection protocol described in section 5.2.2.2 was used. As for the Col1a1 targeting siRNAs, MSC nucleofection with shCol2 and shCol4 was carried out in triplicate per experiment. A single control shRNA, shGFP, was used in this set of experiments which was nucleofected in duplicate; this non targeting control contains the identical targeting sequence as siGFP. It is expected that the shGFP shRNA would be processed identically as the Col1a1 targeting siRNAs, therefore comparative molecular and biochemical processes should be occurring in MSCs nucleofected with shCol2, shCol4 and shGFP to the point of mRNA target cleavage.

A total of three independent nucleofection experiments were carried out. Once again the MSCs were allowed to recover for a period of 24 hours due to the insult of the nucleofection procedure to the cells, an additional 24 hours was allowed for RNAi to occur. Therefore RNA was extracted 48 hours post-nucleofection. The 3 RNA samples generated per shRNA per nucleofection experiment were pooled. As three nucleofection experiments were carried out, this generated 3 pooled RNA samples for shCol2 and shCol4 representing a total of 9 nucleofections per shCol2 and shCol4. For the non-targeting control shGFP, a total of 3 pooled RNA samples were generated for shGFP representing 6 separate nucleofections.

Col1a1 expression levels were analysed using real time RT-PCR. Each RT-PCR reaction was set up as described in section 2.9.2. Once again, each RNA sample was analysed in triplicate per real time RT-PCR run. GAPDH and β-Actin were used as the housekeeping controls. Each run was carried out in triplicate and expression levels determined using the relative quantification method as described above in section 5.3.3 for the siRNAs nucleofection real time RT-PCR analysis. Both shRNAs effectively reduced Col1a1 expression in MSCs. When normalized to GAPDH, shCol2 reduced Col1a1 expression levels to 25.61% (± 15.55%) and shCol4 reduced

Col1a1 expression levels to 16.32% (± 14.34%) when compared to Col1a1 expression levels in MSCs nucleofected shGFP. Non parametric ANOVA analysis was carried out in which the Col1a1 expression data generated from MSCs nucleofected with shCol2, shCol4 and shGFP were compared, a p value of $p<0.0001$ was returned, indicating that there was a significant difference amongst the 3 data sets. In order to determine which data sets differed post test analysis was carried out using the Dunn's multiple comparison test. The pairwise comparisons shGFP to shCol2 and shGFP to shCol2 generated a p value of $p<0.001$ indicating that the Col1a1 expression levels in MSCs nucleofected with shCol2 or shCol4 appeared to differ significantly from Col1a1 expression levels in MSCs nucleofected with shGFP. Comparison of shCol2 and shCol4 Col1a1 expression data sets generated a p value of $p>0.05$ indicating that neither shCol2 nor shCol4 more potently suppressed Col1a1 expression in MSCs (Table 5.3).

The data was reanalyzed following normalization to β-Actin. Col1a1 expression levels in MSCs nucleofected with shCol2 were reduced to 24.21% (± 12.39%) and to 16.11% (± 10.83%) in MSC nucleofected with shCol4 compared to Col1a1 expression levels in MSCs nucleofected with shGFP. The β-Actin normalised Co1a1 expression data was also assessed using a non parametric ANOVA in which the shCol2, shCol4 and shGFP Col1a1 expression data was compared generating a p value of $p<0.0001$ indicating that the data sets differed significantly, reflecting the results described above for the GAPDH normalised data. Once again in order to specifically determine which of the Col1a1 expression data sets differed Dunn's multiple comparison post test analysis was carried out, the pairwise comparisons shGFP to shCol2 and shGFP to shCol4 generated p values of $p<0.01$ and $p<0.001$ respectively indicating that Cola11 expression levels in MSCs nucleofected with shCol2 or shCol4 were significantly different from Col1a1 expression levels in MSCs nucleofected with shGFP. Comparison of shCol2 and shCol4 Col1a1 expression data generated p value of $p>0.05$ indicating that neither shCol2 nor shCol4 more potently suppresses Col1a1 expression in MSCs, (Table 5.4).

With regard to the standard deviations error and variation can be introduced due to a variety of reasons in studies such as this involves multiple transfections, RNA extractions and RT-PCR evaluations. As each RNA sample analysed is a pool of three RNA samples generated from a single nucleofection experiment, therefore the overall Col1a1 expression levels will in part be dependent on the sum of the nucleofection efficacious in each of the three nucleofections carried out in a single experiment possibly introducing a significant amount of variability. Other factors that can potentially contribute to variation include for example, the presence of some RNA degradation. The RT-PCR assays represent another point at which variation can be introduced due to in difference in cDNA synthesis, PCR efficiencies, sample preparation, pipetting error, and inter-run variation.

The final analysis carried out was to determine if the Col1a1 targeting siRNAs or shRNAs were more potent mediators of Col1a1 suppression in MSCs. Firstly, siCol2 and shCol2 Col1a1 expression data generated following normalisation to GAPDH and comparison to their non targeting counterpart siGFP or shGFP were compared using a two tailed student t-test, the data sets, siCol2 3.46% (± 1.6%) and shCol2 25.61% (±11.05) were found to differ significantly, p=0.0005, therefore siCol2 appears to effect more potent suppression of Col1a1 expression in MSCs. Similarly, siCol4 and shCol4 Col1a1 expression data generated following normalisation to GAPDH and comparison to their non targeting counterpart siGFP or shGFP were compared using a two tailed student t-test, the data sets siCol4 2.67% (±1.25%) and shCol4 16.32% (±14.34) were found to differ significantly, p=0.0108, therefore siCol4 appears to effect more potent suppression of Col1a1 expression in MSCs (Table 5.5). The higher levels of suppression induced by the siRNAs are most likely due to the high transfection efficiencies achieved using the nucleofector.

The Col1a1 expression data for siCol2 and shCol2 or siCol4 and shCol4 was also compared following normalisation to β-Actin and comparison to the relevant non-targeting control siGFP or shGFP. Comparison of siCol2 (2.05% (±0.68%) and shCol2

(24.21% (±12.39%) Col1a1 expression data revealed that the data sets differ significantly, p=0.0017, therefore siCol2 appears to more potently effect Col1a1 suppression in MSCs than shCol2. Similarly, comparison of the siCol4 (1.49% (±0.86%) and shCol4 (16.11% (±10.83%) Col1a1 expression data revealed that the data sets differ significantly p=0.0012, therefore siCol4 appears to effect more potent Col1a1 suppression in MSCs than shCol4. Therefore, it appears that the Col1a1 targeting siRNAs are more potent mediators of Col11 suppression in MSCs but, this may in part be due to the higher transfection efficiencies achievable when delivering siRNAs rather than plasmids (Table 5.5).

A)

		shGFP	*shCol2*	*shCol4*	*ANOVA*
GAPDH	*shGFP*	100%	25.61%	16.32%	p<0.0001
	StdDev	15.87	15.55	14.34	

B)	shCol2	shCol4
siGFP	p<0.001	p<0.001
siCol2		p>0.05
siCol4	p>0.05	

Table 5.3 GAPDH normalised relative Col1a1 mRNA expression levels in Mesenchymal Stem Cells following nucleofection with shCol2, shCol4 or shGFP. **A)** GAPDH normalised Col1a1 mRNA expression values in MSCs nucleofected with shCol2 and shCol4 were 25.61 (±15.55%) and 16.11% (±14.34%) when compared to Col1a1 expression levels in MSCs nucleofected with shGFP. The mean and standard deviation of all data sets was determined. Non parametric ANOVA analysis was carried out in which the Col1a1 expression levels in MSCs following nucleofection with shCol2, shCol4 or shGFP were compared generating a p value of $p < 0.0001$, indicating that the data sets differed significantly. **B)** Post test analysis was carried out using the Dunn's multiple comparison test, both pairwise comparison of shGFP to shCol2 and shGFP to shCol2 generated a p value of $p<0.001$ indicating that Col1a1 expression levels in MSCs nucleofected with shCol2 or shCol4 differed significantly from Col1a1 expression levels in MSC nucleofected with shGFP. Comparison of shCol2 and shCol4 Col1a1 expression data generated a p value of $p>0.05$ indicating that neither shCol2 nor shCol4 effected more potent suppression of Col1a1 expression in MSCs. All data was analysed using GraphPad InStat 3.

A)		*shGFP*	*shCol2*	*shCol4*	*ANOVA*
βActin	*shGFP*	100%	24.21%	16.11%	p<0.0001
	StdDev	11.05	12.39	10.83	

B)	shCol2	shCol4
siGFP	p<0.01	p<0.001
siCol2		p>0.05
siCol4	p>0.05	

Table 5.4 β-Actin normalised relative Col1a1 mRNA expression levels in Mesenchymal Stem Cells following nucleofection with shCol2, shCol4 or shGFP.

A) GAPDH normalised Col1a1 mRNA expression values in MSCs nucleofected with shCol2 and shCol4 were 24.21 (±12.39%) and 16.11% (±10.83%) when compared to Col1a1 expression levels in MSCs nucleofected with shGFP. The mean and standard deviation of all data sets was determined. Non parametric ANOVA analysis was carried out in which the Col1a1 expression levels in MSCs following nucleofection with shCol2, shCol4 or shGFP were compared generating a p value of $p < 0.0001$, indicating that the data sets differed significantly. **B)** Post test analysis was carried out using the Dunn's multiple comparison test, both pairwise comparison of shGFP to shCol2 and shGFP to shCol2 generated a p value of $p<0.01$ and $p<0.001$ indicating that Col1a1 expression levels in MSCs nucleofected with shCol2 or shCol4 differed significantly from Col1a1 expression levels in MSC nucleofected with shGFP. Comparison of shCol2 and shCol4 Col1a1 expression data generated a p value of $p>0.05$ indicating that neither shCol2 nor shCol4 effected more potent suppression of Col1a1 expression in MSCs. All data was analysed using GraphPad InStat 3.

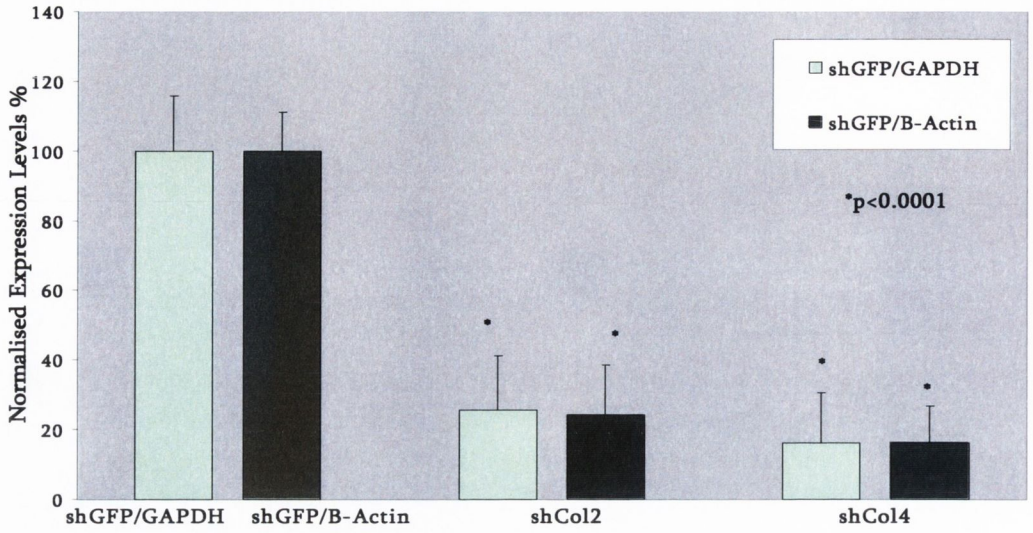


Figure 5.8 Relative Col1a1 expression levels in shRNA nucleofected Mesenchymal Stem Cells. $1x10^6$ MSCs were nucleofected with 1μg shCol2, shCol4 or shGFP. RNA was extracted 48 hours post transfection and Col1a1 expression evaluated using RT-PCR. Col1a1 expression was normalized to GAPDH (Plain green bars) or β-Actin expression (green hatched bars) and relative expression levels MSCs nucleofected with shCol2 or shCol4 determined by comparison to GAPDH or β-Actin normalised Col1a1 expression levels in cells transfected with shGFP, which was set at 100%. Expression values in MSCs nucleofected with shCol2 were 24.21% - 25.61%, (± 12.39% - 15.55%) and 16.32% - 16.11% (± 10.83% - 14.34%) in shCol4 nucleofected MSCs. Non parametric ANOVA analysis was carried out in which the Col1a1 expression levels in MSCs following nucleofection with shCol2, shCol4 and shGFP were compared, a p value of $p < 0.0001$ was generated indicating that the data sets differed significantly. Post test analysis was carried out using the Dunn's multiple comparison test, both pairwise comparison of shGFP to shCol2 and shGFP to shCol2 generated a p value of $p<0.01$ and $p<0.001$ indicating that Col1a1 expression levels in MSCs nucleofected with shCol2 or shCol4 differed significantly from Col1a1 expression levels in MSC nucleofected with shGFP. Comparison of shCol2 and shCol4 Col1a1 expression data generated a p value of $p>0.05$ indicating that neither shCol2 nor shCol4 effected more potent suppression of Col1a1 expression in MSCs. All data was analysed using GraphPad InStat 3.

A)		*siGFP/shGFP*
GAPDH	*siCol2 v shCol2*	p =0.0005
	siCol4 v shCol4	p =0.0108

B)		*siGFP/shGFP*
β-Actin	*siCol2 v shCol2*	p = 0.0017
	siCol4 v shCol4	p = 0.0012

Table 5.5. T-test analysis to determine if the Col1a1 targeting siRNAs or shRNAs more potently reduce Col1a1 expression in MSCs. A) The relative siCol2 and shCol2 and siCol4 and shCol4 Col1a1expression data generated from normalisation to GAPDH and comparison to the relevant non targeting control siGFP or shGFP were compared generating a p value of p=0.0005 and p = 0.0108. As expression levels in MSCs nucleofected with siCol2 or shCol2 are 3.46% (±1.6%) and 25.61% (±15.55%) respectively and 2.67% (±1.25%) and 16.32% (±14.34%) in MSCs nucleofected with siCol4 and shCol4 respectively, it appears that siCol2 and siCol4 are more potent suppressors of Col1a1 expression in MSCs. **B)** The relative Col1a1 expression data generated from normalisation to β-Actin and comparison to the relevant siGFP, shGFP non targeting control for siCol2 and shCol2 or siCol4 and shCol4 were compared. The p values generated for siCol2 v shCol2 was p=0.0017 and p = 0.0012 indicating that the data differs significantly, therefore, as expression levels in MSCs nucleofected with siCol2 or shCol2 are 5.08% (±2.64%) and 24.21 (±12.39%) respectively and 1.49%(±0.86%) and 16.11%(±10.83%) in MSCs nucleofected with siCol4 and shCol4 respectively, it appears that siCol2 and siCol4 are more potent suppressor of Col1a1 expression in MSCs.

5.3.5 Metra CICP ELISA evaluation of *de novo* type I collagen production in MSCs nucleofected with Col1a1 targeting siRNAs

In order to assess whether the levels of siRNA mediated Col1a1 suppression were reflected at the protein levels the Metra CICP ELISA system was used. As described (section 5.2.8) this assay measures the amount of C terminal pro-peptide released by the MSCs into the media over a 24 hour period. This is a direct indication of the amount of *de novo* type I collagen production which occurred over the 24 hour assay period. Each siRNA was transfected in triplicate per transfection experiment. For the ELISA, media samples collected from each of the 3 wells nucleofected with either siCol2 or siCol4 were pooled. In the case of siGFP and siNon-tgt, media was collected from the two wells of MSCs nucleofected per siRNA. Thus, one pooled media sample was generated from a single transfection for each of siCol2, siCol4, siGFP and siNon-tgt. As three transfection experiments were carried out a total of 3 times, a total of 3 media samples were generated per siRNA. Each sample was analyzed in duplicate per ELISA and three independent ELISA assays were carried out. As two negative controls were used, only two of the pooled media samples were included in the ELISA analysis.

The amount of *de novo* type I collagen expression in the cells transfected with siCol2 and siCol4 was determined by direct comparison with the controls siGFP and siNon-tgt, in which expression levels were set as 100%. As an illustrative example the calculation carried out for siCol2 is presented in Table 5.7. Expression levels in cells transfected with siCol2 were 12.17% (±3.09%) when compared to siGFP and 13.00% (±3.22%) when compared to siNon-tgt. siCol4 was also found to potently reduce *de novo* Col1a1 expression levels, type I collagen production levels in cells transfected with siCol4 were 10.15% (±1.89%) and 10.78% (±2.01%) when compared to siGFP and siNon-tgt respectively. Each data set was analyzed using GraphPad Instat 3, the mean and standard deviation was calculated

Non parametric ANOVA analysis was carried out in which type I collagen expression levels in MSCs nucleofected with siCol2, siCol4 and siGFP or siNon-tgt were

compared. A p value of $p<0.0001$ was generated from each ANOVA (i.e. comparison of siCol2, siCol4 and siGFP or siCol2, siCol4 and siNon-tgt) indicating that type I collagen expression data sets differed significantly. Dunn's multiple comparisons test was employed for the post test analysis, this analysis carried out pairwise Col1a1 expression data comparisons, siGFP to siCol2, generated a p value of $p<0.01$ and siGFP to siCol4, siNon-tgt to siCol2, siNon-tgt to siCol4, generated a p value of $p<0.001$ indicating that Col1a1 expression levels in MSCs nucleofected with the Col1a1 targeting siRNAs were significantly different from Col1a1 expression levels in MSCs nucleofected with the non targeting controls siGFP or siNon-tgt. Additionally the siCol2 and siCol4 type I collagen expression data was also compared generating a p value of $p>0.05$ indicating that neither siCol2 nor siCol4 appeared to more potently reduce Col1a1 expression levels in MSCs (Table 5.7 and Figure 5.9).

5.3.6 Metra CICP ELISA evaluation of d*e novo* type I collagen production in MSCs nucleofected with Col1a1 targeting shRNAs

Analysis of shRNA mediated Col1a1 suppression was also evaluated at the protein level using the Metra CICP ELISA system. As described for the siRNAs, media from each well of each triplicate shRNA nucleofection was pooled generating a single pooled representative sample per shRNA per nucleofection experiment. As three nucleofection experiments were carried out a total of 3 samples were generated for each shRNA shCol2, shCol4 and shGFP. Each sample was analyzed in duplicate in a single ELISA plate and the assay carried out three times. The amount of type I collagen protein produced by the MSCs cells transfected with shCol2 and shCol4 was determined by direct comparison with the amount of type I collagen produced by MSCs nucleofected with the control shRNA plasmid, shGFP. Type I collagen protein expression levels in MSCs transfected with shCol2 was 31.06% (±11.08%) and in MSCs transfected with shCol4, 22.8% (±8.15%) respectively. The data was analysed using GraphPad Instat 3, the mean and standard deviation of each data set was calculated. Non parametric ANOVA analysis was carried out in which type I collagen expression levels in MSCs nucleofected with shCol2, shCol4 and shGFP were

compared. A p value of $p<0.0001$ was generated indicating that the 3 data sets differed significantly. Dunn's multiple comparisons test was employed for the post test analysis. The pairwise Col1a1 expression data comparisons, shGFP to shCol2, shGFP to shCol4 generated a p value of $p<0.01$ and $p<0.001$ respectively, indicating that type I collagen expression levels appeared to be significantly different in MSCs nucleofected with the Col1a1 targeting shRNAs, shCol2 and shCol4 when compared to Col1a1 expression levels from MSCs nucleofected with the non targeting control shGFP. Type I collagen expression levels in MSCs nucleofected with shCol2 or shCol4 were also compared generating a p value of $p<0.05$ indicating that neither shCol2 nor shCol4 effects more potent suppression of type I collagen in MSCs (Table 5.8 and Figure 5.9)

The levels of *de novo* type I collagen expression in MSCs nucleofected with siCol2 12.17% – 13.00% (±3.09 – 3.22%) was compared to that in MSCs nucleofected with shCo2, 31.06% (±11.08%) using a two tailed student t-test generating a p value of $p<0.0001$. This indicated that the data sets differed significantly and that siCol2 appears to more potently suppress of type I collagen expression in MSCs than shCol2. Similarly the levels of *de novo* type I collagen in MSCs nucleofected with siCol4, 10.15%– 10.78% (±1.89 – 2.01%) were compared to that in MSCs nucleofected with shCo4, 22.8% (±8.15%) using a two tailed student t-test generating a p value of $p<0.0001$. This indicated that the data sets differed significantly and that siCol4 appeared to more potently suppress type I collagen in MSCs than shCol4 (Table 5.9). The superior potency of the Col1a1 targeting siRNAs, siCol2 and siCol4 at the protein level mirrors the results generated at the mRNA level, with the Col1a1 targeting siRNAs more potently reducing Col1a1 expression than the Col1a1 targeting shRNAs at the mRNA level.

siGFP	*siGFP/siGFP*		*siNon-Tgt*	*siN-tgt/siN-tgt*
998.4	107		668.64	76.27
1136.4	122		684.24	78.05
801.12	86.3		843.84	96.26
942.72	101		1071.12	122
805.92	86.88		1118.36	127.58
794.4	85.58		902.4	102.9
1000.56	107		889.44	101.46
946.32	101.94		958.08	109.29
928.23	*99.71%*		753.12	85.91
			876.58	*99.96%*
siCol2	*siCol2/siGFP*		*siCol2*	*siCol2/siN-tgt*
63.75	6.86		63.75	7.27
68.04	7.32		68.04	7.8
94.2	10.14		94.2	10.7
98.88	10.65		98.88	12.14
131.88	14.2		131.88	16.19
129.48	13.94		129.48	14.7
162.84	17.54		162.84	18.5
147.84	15.92		147.84	16.86
80.16	8.6		80.16	9.2
86.4	9.3		86.4	9.85
110.54	11.9		110.54	12.6
110.36	11.88		110.36	12.58
113.22	12.12		113.22	12.8
136.8	14.73		136.8	15.6
137.52	14.81		137.52	15.6
129.12	13.9		129.12	14.72
121.8	13.12		121.8	13.89
	12.17%			*13.00%*

Table 5.6 ELISA data based calculation of type I collagen expression in MSCs nucleofected with siCol2. MSCs were nucleofected with siCol2, siGFP or siNon-Tgt and de novo production of type I collagen over a 24 hour period determined. To determine the relative levels of type I collagen production the type I collagen expression values obtained for MSCs nucleofected with siCol2 (peach column) is divided by the average type I collagen expression value from MSCs nucleofected with

siGFP (green column, average 928.23). Each siCol/siGFP data point generated is multiplied by 100 so that the data generated is percent expression. Type I collagen expression in cells nucleofected with siCol2 is then presented as an average of this expression data (12.17%). The data siGFP data is also divided by the average of the siGFP column, and multiplied by 100, the value obtained will be close to 100%. The identical process is carried out when the relative type I expression values are being determined by comparison to type I collagen expression levels in cells nucleofected with siNon-Tgt. Each of the siCol2 type I collagen expression values is divided by the average of the type I collagen expression values in MSCs nucleofected with siNon-tgt (blue column average 876.58). The data points generated are expresses as a percentage and the overall type I collagen expression in MSCs nucleofected with siCol2 relative to type I collagen expression levels in MSCs nucleofected with siNont-tgt 13.00%. The siNon-tgt data is also divided by the average of the siNon-tgt column, and multiplied by 100, the value obtained will be close to 100%, and is illustrated as 100% on the graph for clarity.

	siGFP	*siN-tgt*	*siCol2*	*siCol4*	*ANOVA*
siGFP	***100%***		***12.17%***	***10.15%***	***p<0.0001***
StdDev	4.79		3.09%	1.89%	
siNon-Tgt		***100%***	***13.00%***	***10.78%***	***p<0.0001***
StdDev		18.02%	3.22%	2.01%	

B)	siCol2	siCol4
siGFP	p<0.01	p<0.001
siNon-tgt	p<0.001	p<0.001
siCol2		p>0.05
siCol4	p>0.05	

Table 5.7. Type I collagen protein expression levels following nucleofection with Col1a1 targeting siRNAs as evaluated using the Metra CICP ELISA.

A) Type I collagen expression in MSCs following nucleofection with Col1a1 targeting siRNAs siCol2, siCol4 was evaluated using the Metra CICP ELISA system. The relative protein expression levels were determined via direct comparison of the levels of type I collagen production in cell nucleofected with siCol2 or siCol4 to the levels of type I collagen production in cells nucleofected with the negative control siRNAs, siGFP (plain green bars) and siNon-tgt (blue bars). siCol2 was found to reduce Col1a1 expression to 12.17% and 13.00% of the levels of type I collagen expressed by MSCs nucleofected with siGFP and siNon-Tgt respectively. siCol4 reduced Col1a1 expression to 10.15% and 10.78% of the levels of type I collagen expressed by MSCs nucleofected with siGFP and siNon-Tgt respectively. Non parametric ANOVA analysis was carried out in which type I collagen expression levels in MSCs following nucleofection with siCol2, siCol4 and siGFP or siNon-tgt were compared, each ANOVA (i.e. comparison of siCol2, siCol4 and siGFP or siCol2, siCol4 siNon-tgt Col1a1 expression data) generated a p value of $p<0.0001$ indicating that Col1a1 expression levels in MSCs nucleofected with siCol2 or siCol4 differed from Col1a1 expression levels in MSCs nucleofected with siGFP or siNon-tgt. **B)** Dunn's multiple comparison analysis was carried out pairwise comparison of type I collagen expression data was undertaken, siGFP to siCol2 generated a p value of $p<0.01$ and siGFP to siCol4, siNon-tgt to siCol2, siGFP to siCol4, $p<0.001$ indicating that in all 4 cases the

type I collagen expression data sets differed. Comparison of siCol2 and siCol4 type I collagen expression data sets generated p value of $p>0.05$ indicating that neither siCol2 nor siCol4 effects more potent suppression of Col1a1 expression in MSCs.

	shGFP	*shCol2*	*shCol4*	*ANOVA*
shGFP	*100%*	30.97%	22.88%	p<0.0001
StdDev	4.79%	11.08%	8.15%	

B)	shCol2	shCol4
shGFP	p<0.01	p<0.001
shCol2		p>0.05
shCol4	p>0.05	

Table 5.8. Type I collagen protein expression levels following nucleofection with Col1a1 targeting shRNAs as evaluated using the Metra CICP ELISA.

A) Type I collagen protein expression in MSCs was evaluated using the Metra CICP ELISA system following nucleofection with Col1a1 targeting shRNAs shCol2 and shCol4. The relative protein expression levels were determined via direct comparison of the levels of type I collagen production in cell nucleofected with shCol2 or shCol4 to the levels of type I collagen production in cells nucleofected with the negative control siRNAs, shGFP. shCol2 and shCol4 reduced type I collagen protein expression to 30.97% (±11.08) and 22.88% (±8.15) respectively when compared to type I collagen protein expression levels in MSCs nucleofected with shGFP. The data was evaluated using GraphPad InStat3, the standard deviation is shown. Non parametric ANOVA analysis was carried out in which type I collagen expression levels in MSCs following nucleofection with shCol2, shCol4 and shGFP were compared, generating a p value of $p<0.0001$ indicating that Col1a1 expression levels in MSCs nucleofected with shCol2 or shCol4 differed from Col1a1 expression levels in MSCs nucleofected with shGFP **B)** Dunn's multiple comparison analysis was carried out IN WHICH pairwise comparison of type I collagen expression data, shCol2 to shGFP and shCol4 to shGFP generated a p value of $p<0.01$ and $p<0.001$ indicating that in both case the type I collagen expression data sets differed significantly. Comparison of shCol2 and shCol4 type I collagen expression data sets generated p value of $p>0.05$ indicating that neither shCol2 nor shCol4 effects more potent suppression of Col1a1 expression in MSCs.

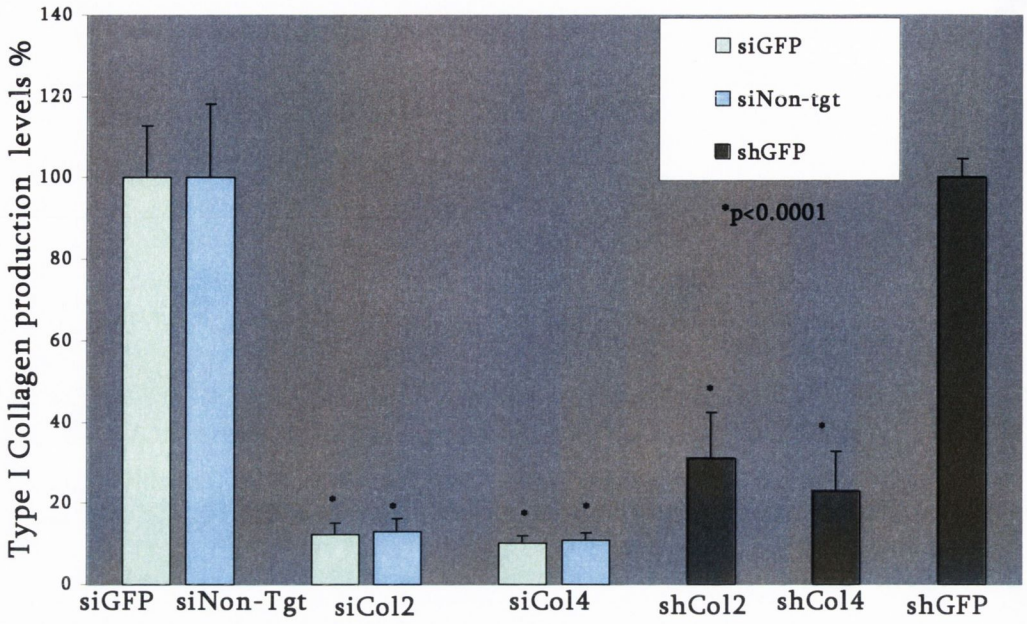


Figure 5.9. *De novo* Type I collagen protein production in MSCs following nucleofection with siRNAs or shRNAs as evaluated using the Metra CICP ELISA. Type I collagen production in MSCs was evaluated using the Metra CICP ELISA system following nucleofection with Col1a1 targeting siRNAs siCol2, siCol4 or Col1a1 targeting shRNAs, shCol2 and shCol4. Relative type I collagen expression levels were determined via direct comparison with type I collagen expression in MSCs nucleofected with the negative controls siRNAs, siGFP (plain green bars) and siNon-tgt (blue bars). MSCs type I collagen expression was reduced to 12.17% (±3.09%) and 13.00%(±3.22%) by siCol2 and to 10.15%(±1.89%) and10.78% (±2.01%) by siCol4 relative to type I collagen production levels in MSCs nucleofected with siGFP or siNon-Tgt respectively. Relative type I collagen expression in MSCs nucleofected with shCol2 or shCol4 was reduced to 30.97% (±11.08%) and 22.88% (±8.15%) relative to type I collagen production levels in MSCs nucleofected with shGFP. Non parametric ANOVA analysis of both the siRNA and shRNA data was carried out. The siRNA data sets, (siCol2, siCol4 and siGFP or siNon-tgt) were found to differ significantly $p<0.0001$, as were the shRNA data sets (shCol2, shCol4 and shGFP) $p<0.0001$. Dunn's multi-comparison post test analysis was carried out and is described above in Tables 5.7 and 5.8. Data was analysed using GraphPad InStat 3.

	siGFP/shGFP
siCol2 v shCol2	p <0.0001
siCol4 v shCol4	p <0.0001

Table 5.9. P values generated from t-test comparison of the relative levels of *de novo* type I collagen production in MSCs nucleofected with siCol2, siCol4, shCol2 or shCol4. Comparison of siRNA and shRNA type I collagen expression data. Type I collagen expression levels in MSCs nucleofected with siCol2 (12.17 – 13.00% (± 3.09 – 3.22) were compared to type I collagen expression levels in MSCs following nucleofection with shCol2 (30.97% (±11.08%). Similarly type I collagen expression levels in MSCs subsequent to nucleofection with siCol4 (30.97% ±11.08%) and shCol4 (22.88% ±8.15%) were compared. Both t-tests generated a p value of $p<0.0001$ indicating that the data sets are significantly different, therefore the siRNAs appear to more potently reduce type I collagen levels in MSCs than their shRNA counterparts. Notably this evaluation does not take into account transfection efficiencies which may be responsible for the superior levels of type I collagen suppression observed in MSCs nucleofected with the siRNAs.

5.5 Discussion

The work described in this PhD thesis is focused upon the development of a novel gene therapy for treatment of Col1a1 linked Osteogenesis Imperfecta. As described (section 1.3) OI is a dominant negative disorder, incorporation of mutant Col1a1 or Col1a2 chains into type I collagen triple helices disrupts and weakens the type I collagen fibrils responsible for the tensile strength and flexibility of bone. The initial aim of the current study was to determine if the Col1a1 targeting siRNAs and their shRNA counterparts were effective suppressors of endogenous Col1a1 expression, this was demonstrated in chapters 3 and 4 of this thesis. The next logical progression was to determine the potency of the suppressors identified in a more appropriate cell type, the target cell type for OI, mesenchymal stem cells.

Mesenchymal stem cells are an adult stem cell which gives rise to the bone forming cells termed osteoblasts. Whilst the delivery of allogenic MSCs has shown to result in *in vivo* bone formation (Jorgensen, *et al.*, 2004; Meinel, *et al.*, 2005; Pereira, *et al.*, 1995), the data regarding the immunogenic properties is conflicting (Eliopoulos, *et al.*, 2005; Horwitz, *et al.*, 2002; Inoue, 2006; Koc, 2003; Sudres, *et al.*, 2006). For this reason it is likely that autologous MSCs may be used in. In principle, the cells would be harvested from a patient, expanded *in vitro* and the therapeutic nucleic acids delivered to the expanded cells. In principle, they should migrate to and engraft into the bone tissue (Devine, 2002; Shirley, 2005; Taguchi, *et al.*, 2005) and through their self renewal capacity should provide a life long source of MSCs which express the therapeutic nucleic acids. Using the current study as an example, the MSCs would be engineered to express the Col1a1 suppression agents described and discussed in chapters 3 and 4 of this thesis.

Given this potential scenario it was necessary to evaluate the efficacy of the Col1a1 targeting siRNAs, siCol2 and siCol4 in MSCs. Due to the difficulties generally associated with transfection of primary cell lines and the poor transfection efficiencies observed in preliminary Lipofectamine based MSC transfection studies, in which transfection efficiency was in the region of 30% (data not shown), an alternative means of nucleic acid delivery to MSCs was required. A review of the most recent

literature revealed that much success had been reported using the Amaxa Nucleofector device (Aluigi, *et al.*, 2006; Aslan, *et al.*, 2006; Distler, *et al.*, 2005; Gartner, *et al.*, 2006; Quenneville, *et al.*, 2004; Siemen, *et al.*, 2005). The manufacturers of this electroporation device claim that the device can be used to deliver nucleic acids directly into the nucleus of the cell using cell type specific solutions and electroporation programs as directed by the provided protocols. Initially, the protocol provided with the human MSC specific kit was followed, $5x10^5$ cells, 2μg of plasmid DNA and program U23, which is for high transfection efficiencies. The control Amaxa GFP (pmaxGFP) plasmid provided in the kit was used in the optimization experiments; cell mortality was high, in the region of 70%. Due to the known toxic effects of the introduction of exogenous DNA into delicate primary cells such as MSCs the concentration of plasmid DNA nucleofected was reduced from 2μg to 1μg. This had a significant impact, increasing cell survival rate to 60%. It was thought that the presence of an increased number of cells in the Amaxa nucleofection cuvette may afford some level of protection to the cells, therefore, the number of MSCs per nucleofection was doubled to $1x10^6$ cells to see if post-nucleofection survival rates could be further increased. This had a significant impact, decreasing cell mortality rates increased to 20% - 30% whilst retaining transfection efficiencies in the region of 60% to 70%.

It appears that the RNAi machinery is intact and functional in mesenchymal stem cells with gene silencing achieved using both siRNA (Forte, 2006; Hoelters, 2005; Millington-Ward, 2004; Neth, *et al.*, 2006; Piersanti, *et al.*, 2006; Xu, 2006) and shRNA (Cho, *et al.*, 2006; Kahler, 2006). In the current study both siCol2 and siCol4 were found to be highly effective mediators of endogenous Col1a1 suppression in MSCs reducing expression to between 1.49% - 5.08 % (± 0.86 – 2.64%) when compared to Col1a1 expression levels in MSCs nucleofected with the non targeting controls siGFP and siNon-tgt. Non parametric analysis was carried out in which the Col1a1 expression data generated following nucleofection with siCol2, siCol4, siGFP or siNon-tgt was compared. This was carried out for both the GAPDH and β-Actin normalised data sets, each ANOVA generated a p value of $p<0.0001$ indicating that Col1a1 expression was significantly different in MSCs nucleofected with the Col1a11

targeting siRNAs when compared to Col1a1 expression in MSCs nucleofected with the non targeting siRNAs, siGFP and siNon-tgt (Table 5.1 A and Table 5.2 A). Post test analysis was carried out in order to determine exactly which of Col1a1 data sets differed significantly from one another. Dunn's multiple comparison test was employed to carry out pairwise comparisons of each of the Col1a1 expression data sets, i.e. siGFP to siCol2, siGFP to siCol4, siNon-tgt to siCol2, siNon-tgt to siCol4 etc. Pairwise comparisons of the MSC Col1a1 expression data generated following nucleofection with siGFP or siNon-tgt to siCol2 or siCol4 generated p values in the range of $p<0.01$ to $p<0.001$ indicating that the non targeting and Col1a1 targeting siRNA Col1a1 expression data sets were significantly different. Pairwise comparison of the siCol2 and siCol4 Col1a1 expression data generated a p value of $p>0.05$ indicating that the data sets do not differ and that neither siCol2 or siGFP appears to be a more potent suppressor of Col1a1 expression in MSCs (Table 5.1 B, Table 5.2 B and Figure 5.7).

The high levels of Col1a1 suppression (>95%) suggested that the nucleofector may achieve siRNA transfection efficiencies of greater than 90% in MSCs. It is most likely that the use of an eGFP plasmid to assess nucleofection efficiencies (as described in the previous paragraph) may be the reason that the levels of transfection achieved (60% to 70%) are lower than the observed levels of siRNA mediated Col1a1 mRNA suppression. Additionally, the relatively low standard deviations of the data set indicated that the nucleofector achieved consistently high siRNA transfection efficiencies of the Col1a1 targeting siRNAs. However, as discussed in chapter 4 of this thesis whilst siRNAs are effective mediators of RNAi their effect is transient as they are diluted from the cell population over successive cell divisions and are in addition susceptible to nuclease degradation. This transience will in particular limit the application of pre-synthesised siRNAs in MSCs, as these cells will give rise both to MSCs through self renewal and differentiate into osteoblasts which will ultimately produce the corrected bone matrix. Moreover, this most probably makes siRNAs non-applicable as a tool for the treatment of OI as patients with OI would require life long suppression of mutant Col1a1 or Col1a2 alleles. Therefore efficacy of the

corresponding shRNA counterparts of siCol2 and siCol4 termed shCol2 and shCol4 were evaluated MSCs.

The nucleofector device was also used to deliver the shRNA plasmids to the MSCs. Expression levels were found to be reduced to (24.21% - 25.61% (±12.39 – 15.55%) in cells transfected with shCol2 and to (16.11 % – 16.32% (±10.83 – 14.34%) in MSCs transfected with shCol4 when compared to Col1a1 expression levels in MSCs transfected with shGFP and normalised to GAPDH and β-Actin respectively (Table 5.3 A, Table 5.4 A and Figure 5.8). Comparison of the Col1a1 expression data generated following nucleofection with shCol2, shCol4 and shGFP using a non parametric ANOVA generated a p value of $p<0.0001$ indicating that the data sets differed. Dunn's multiple comparison post test analysis revealed that the reduced levels of Col1a1 expression in MSCs nucleofected with shCol2 and shCol4 differed significantly from Col1a1 expression levels in MSCs nucleofected with shGFP with $p<0.01$ – $p<0.001$ generated from each pairwise comparison. Comparison of the Col1a1 expression data generated from MSCs nucleofected with shCol2 to that generated from shCol4 nucleofected MSCs generated a p value of $p>0.05$ indicating that neither shCol2 nor shCol4 more potently reduced Col1a1 mRNA expression in MSCs. Statistical analysis of both the GAPDH and β-Actin data generated identical p values in each case (Table 5.3 B and Table 5.4 B). Student t-test analysis was carried out in which the Col1a1 expression data generated following nucleofection with siCol2 and shCol2 or siCol4 and shCol4 was compared for both the GAPDH normalised and β-Actin normalised data sets (relative to siGFP or shGFP). Each pairwise comparison generated p value of $p<0.05$ (Table 5.5) indicating that the data sets differed and as the Col1a1 expression ranges in MSCs nucleofected with the Col1a1 targeting siRNAs was (1.49 to 5.08% (±0.86 – 2.69%) and for the shRNAs was (16.11 to 25.61% (±10.83 – 15.55%) it appears that the Col1a1 siRNAs more potently suppress Col1a1 expression in MSCs. This may in part be due to the superior nucleofection efficiencies achievable with siRNAs given their small size relative to the shRNA expression plasmids.

The potency of both the Col1a1 targeting siRNA duplexes and the shRNA plasmids were evaluated at the protein level. The Metra CICP ELISA was used, as described previously type I collagen is produced in a pro-collagen form from which the N and C pro-peptides are cleaved prior to incorporation into the type I collagen fibrils. This assay quantifies the amount of C pro-peptides cleavage products, as type I collagen is a secreted protein the cleavage products are present in the medium and are easily quantified through collection of the media. Both the Col1a1 targeting siRNAs and shRNAs effectively reduced type I collagen expression levels in MSCs. siCol2 and siCol4 reduced type I collagen production to 12.17 – 13.00% (±3.09 – 3.22%) and 10.15 – 10.78% (±1.89 – 2.01%) respectively when compared to MSCs nucleofected with siGFP and siNon-Tgt. In the case of the Col1a1 targeting shRNAs, type I collagen in MSCs nucleofected with shCol2 or shCol4 was reduced to 30.97% (±11.08%) and 22.88% (±8.15%) respectively, relative to type I collagen production levels in MSCs nucleofected with shGFP.

ANOVA analysis of the type I collagen expression data generated from the ELISA analysis of *de novo* type I collagen in MSCs following nucleofection with Col1a1 targeting siRNAs, siGFP and siNon-tgt was carried out. Comparison of the siCol2, siCol4 and siGFP or siNon-tgt generated a p value of $p<0.0001$, indicating that the data sets differed significantly. Dunn's multiple comparison analysis of siCol2, siCol4 and siGFP or siNon-tgt data sets revealed that type I collagen expression levels in MSCs nucleofected with siCol2 or siCol4 differed significantly from type I collagen levels in MSCs nucleofected with siGFP or siNon-tgt, a p value of $p<0.001$ was generated from each pairwise comparison. Additionally, comparison of the siCol2 and siCol4 type I collagen expression data generated a p value of $p>0.05$ indicating that neither siCol2 nor siCol4 effects more potent suppression of type I collagen expression in MSCs, (Table 5.7).

Similarly, ANOVA analysis of the type I collagen expression data generated from the ELISA analysis of *de novo* type I collagen in MSCs following nucleofection with Col1a1 targeting shRNAs and shGFP was carried out, comparison of the shCol2,

shCol4 and shGFP type I collagen expression data sets generated a p value of $p<0.0001$ indicating that the data sets differed significantly. Dunn's multiple comparison test was used to carry out pairwise comparison of the data sets compared in the ANOVA, comparison of shCol2 to shGFP and shCol4 to shGFP generated a p value of $p<0.001$ indicating that the levels of type I collagen produced in MSCs subsequent to nucleofection with shCol2 and shCol4 differed significantly from type I collagen in MSCs nucleofected with shGFP. Comparison of the shCol2 and shCol4 data sets generated a p value of $p>0.05$ indicating that the data sets were not significantly different, therefore neither shCol2 nor shCol4 effects more potent suppression of type I collagen expression in MSCs (Table 5.8). Finally two tailed student t-tests were carried out in which levels of *de novo* type I collagen in MSCs nucleofected with siCol2 or shCol2 or siCol4 and shCol4 were compared, each test generated a p value of $p<0.0001$. This indicated that the data sets differed significantly and that siCol4 appeared to more potently suppress type I collagen in MSCs than shCol4 (Table 5.9).

The RNA and protein analysis carried out demonstrates that the siRNAs and shRNAs are effective mediators of endogenous Col1a1 expression in MSCs. Based upon the statistical analysis of the real time RT-PCR data is appears that the Col1a1 targeting siRNAs more potently suppress Col1a1 expression in MSCs that their shRNA counterparts shCol2 and shCol4 (Table 5.5). In addition, analysis revealed that neither siCol4 nor siCo2 is a more potent suppressor of Col1a1 expression in MSCs (Table 5.3 and Table 5.4). Analysis of the ELISA type I collagen expression data revealed a similar trend. The Col1a1 targeting siRNAs (siCol2 and siCol4) more potently suppress Col1a1 expression in MSCs than the corresponding shRNAs (shCol2 and shCol4) (Table 5.7) Akin to the RT-PCR results neither siCol2 nor siCol4 or in the case of the shRNAs shCol2 nor shCol4 a more potent mediator of type I collagen suppression (Table 5.8). The Col1a1 targeting siRNA effect more potent type I collagen expression than their shRNA counterparts (Table 5.9), as mentioned previously this is most likely due to the higher transfections efficiencies achievable when delivering siRNAs.

In this regard, for any study using RNAi it is important to critically evaluate the parameters used to measure suppression, for example the negative controls used in the study and the housekeeping genes used to normalise the data sets. The non targeting controls used in the siRNA and shRNA experiments, siGFP, siNon-tgt, shGFP and shLuciferase were selected for use as they are thought to be incorporated into the RISC complex as a Col1a1 targeting siRNA or shRNAs would. Therefore, one would expect similar molecular and biochemical events to be occurring in MSCs nucleofected with shCol2 or shCol4 or shGFP. The difference is that the non targeting siRNAs and shRNAs have no known mRNA target in mammalian cells and therefore in principle do not participate in an mRNA cleavage event. Thus use of such controls should enable one to distinguish between sequence specific and non specific RNAi effects in the experiment. In the current study two housekeeping genes GAPDH and β-Actin were used to normalise the RT-PCR data sets. In this regard it is reassuring that the data obtained from each of the housekeeping genes was very similar suggesting that the suppression obtained with siCol2 and siCol4, as well as the shCol2 and shCol4 may indeed represent what is occurring in the nucleofected MSCs.

The Col1a1 targeting siRNA and shRNA suppressors should in principle induce sufficient down regulation of mutant Col1a1 expression to affect a phenotypic benefit. However, this is dependent upon a variety of additional parameters being fulfilled. Firstly as integration of the shRNA plasmids into the genome would be essential for sustained heritable expression of the shRNAs in the progeny of the transfected MSCs it would be important to use an integrating vector to deliver shRNAs to MSCs whilst at the same time trying to limit the potential occurrence of insertional mutagenesis. A novel non-viral site specific gene delivery system based upon the phage ΦC31 integrase has been shown to facilitate site specific integration of exogenous DNA into the human genome (Thyagarajan, *et al.*, 2001). The second requirement as previously mentioned is the provision of a modified Col1a1 replacement gene that is not suppressed by the shRNAs due to sequence alterations at the wobble bases within the target sequence region such that the protein sequence of the Col1a1 transgene remains unchanged but the transcripts from this gene are refractory to suppression. In

this regard the work described in Chapter 6 is focused upon the potential used of the ΦC31 system to effect site specific integration of exogenous DNA sequences into the human genome and Chapter 7 is focused upon the development and testing of a wild type Col1a1 transgene.

Chapter 6

Exploration of ΦC31 Integrase system for site specific genomic integration of therapeutic nucleic acids

6.1 Introduction

The efficacy of any gene therapy approach is dependent upon the availability multiple factors, one of which is a safe and efficient means of nucleic acid delivery to the target tissue. Regardless of the efficacy of delivery if sustained expression of the nucleic acids at therapeutic levels is not achieved then the therapeutic may be of limited use. In order to reach this requirement it is likely that the therapeutic construct may need to integrate into the genome of the target cell such that it is transmitted to all subsequent cellular generations.

Whilst both viral and non-viral vectors have been explored extensively, viral vectors have frequently been the preferred delivery tool largely due to the efficiency with which these vectors can infect cells. Adenoviruses, retroviruses and adeno-associated virus have all been shown to facilitate efficient gene delivery to MSCs (Reiser, *et al.*, 2005). However, despite their efficacy as delivery tools the potential for insertional mutagenesis, cytoxicity and induction of strong immune responses has limited their use *in vivo*. In relation to the possibility of insertional mutagenesis, integrating viruses such as retroviruses and lentiviruses were thought to randomly insert into the genome. However, it has become clear that the reality is of even greater concern; both retroviruses and lentiviruses appear to preferentially integrate into transcriptionally active regions of the genome. MLV preferentially integrates within promoter regions and lentivirus within transcription units (De Palma, *et al.*, 2005).

Unfortunately these potential risks associated with viral vectors were highlighted when a patient with partial ornithine transcarbmaylase (OTC) deficiency developed a severe and fatal inflammatory response following systemic infusion of an adenoviral vector containing the OTC gene (Raper, et al., 2003). In addition, three children with X-linked severe combined immunodeficiency (SCIDX1) developed leukaemia following participation in a clinical trial due to the integration of the retroviral vector employed

for gene delivery within the LMO2 locus (a T-Cell oncogene associated with childhood leukaemia's) in the autologous haematopoietic cells the children received (Hacein-Bey-Abina, *et al.*, 2003).These developments have lead to an increased level of interest and in development of non viral gene delivery systems.

Non-viral systems are in principle safer and potentially have an unlimited transgene capacity. However, poor delivery efficiency and transient gene expression has limited their use in clinical trials to date. Once a nucleic acid expression based therapeutic has crossed the cell membrane and reached the cytoplasm the formidable task of reaching the nucleus where the therapeutic gene should be expressed remains. The work described in the previous chapter of this thesis clearly demonstrated the efficacy of the Nucleofector technology which delivers nucleic acids directly to the nucleus of the cell. For the work carried out in this doctoral thesis the delivery efficiencies achieved by the Amaxa Nucleofector in MSCs were sufficiently high to enable evaluation of RNAi-mediated suppression of Col1a1. However, in order to achieve long term expression of the therapeutic nucleic acids a means of safely integrating therapeutic DNA into the genome will most likely be required.

The most common mechanism of integrating new material into the human genome is random integration of transfected DNA (Thyagarajan, *et al.*, 2001), however this also presents the risk of insertional mutagenesis. In contrast, homologous recombination mediates highly site specific integration but at frequencies too low (<0.1%) to be of use for therapeutic purposes (Goncz, 2002). It is notable that bacteriophage can integrate into specific locations in bacterial genomes following infection using site-specific recombinase enzymes, for example, the phage ΦC31 integrase enzyme facilitates site-specific recombination between the phage and bacterial genomes (Chalberg, *et al.*, 2006). The recombination event occurs between the phage attachment site *attP* and the bacterial attachment site *AttB*. The *AttB* sequence has a TTG core sequence and it is at this site that the *AttB* sequence divides and the linearised phage genome inserts into the

bacterial genome (Groth and Calos, 2004; Thyagarajan, *et al.*, 2001). Integration generates two new sites termed *attL* and *attR* (Figure 6.1) which differ in sequence from the *AttB* site and consequently are not susceptible to integrase mediated excision reactions.

Fortuitously it was found that the human genome contains sites with partial identity to the phage *attP* sites (termed pseudo *attP* sites) which can participate in integrase mediated recombination with *AttB*, thereby facilitating the use of ΦC31 integrase enzyme to effect site specific gene integration in human cells (Thyagarajan, et al., 2001). Since the discovery of human pseudo *attP* sites, similar sites have been identified in the genomes of drosophila (Groth, *et al.*, 2004), mice (Held, *et al.*, 2005), rats (Chalberg, *et al.*, 2005), cows (Ma, e*t al.*, 2006), mosquitoes (Nimmo, *et al.*, 2006) rabbits (Keravala, 2006b) and silkworms (Nakayama G, 2006).

The range of potential applications for the system is extensive, for example, it has been used to generate transgenic organisms (Groth, *et al.*, 2004; Hollis, 2003) and in both *in vivo* (Chalberg, *et al.*, 2005; Held, *et al.*, 2005; Keravala, 2006b; Olivares, *et al.*, 2002) and *ex vivo* (Ishikawa, 2006; Keravala, 2006b) gene therapy explorations.ΦC31 integrase mediated integration of the Col7a1 or LAMB3 cDNA into the genomes of epidermal progenitor cells isolated from patients with recessive dystrophic or junctional epidermolysis bullosa respectively, restored gene expression *in vitro.* In addition in each study it was found that the modified epidermal progenitor cells produced functionally and morphologically normal skin following xenotransplantaion to nude mice. (Ortiz-Urda, *et al.*, 2003; Ortiz-Urda, *et al.*, 2002).TheΦC31 integrase system has been explored as an alternative to the retroviral vectors used in the SCIDX1 clinical trials. None of ΦC31 integrase mediated integration events in Jurkat cells (human T cell leukaemia cell line) were found to be in the vicinity of the LMO2 locus or any other known proto-oncogenes. Encouragingly high level expression of a γc transgene (the affected gene in SCIDX1) was achieved following ΦC31 integrase mediated integration in a SCID X1 T-

cell line. It appears that the integrase system may possibly provide a safer therapeutic vector alternative for use in development of SCIDX1 therapeutics (Ishikawa, 2006).

In addition to effecting site specific integration, use of the ΦC31 system increases integration frequencies above that of random integration alone with increases of between 2 and 35 fold reported (Chalberg, et al., 2005; Keravala, 2006b; Ortiz-Urda, et al., 2003; Ortiz-Urda, et al., 2002). The ΦC31 integrase system has a higher integration efficiency (10 – 100 fold higher) than recombinases such as Cre, FLP and β recombinase (Groth and Calos, 2004; Groth, et al., 2000; Keravala, 2006a; Thyagarajan, et al., 2001). The generation of the *AttL* and *AttR* recombination junctions following integration which are refractory to any ΦC31 integrase excision reaction most likely contributes to this superior integration efficiency. Additionally transgene expression levels have been found to be higher from ΦC31 integrase integrated plasmids compared to randomly integrated plasmids (Chalberg, *et al.*, 2006). A number of independent *in vivo* studies reported that in comparison to controls, which did not receive the ΦC31 integrase plasmid expression levels were significantly higher when the transgene *AttB* plasmid was co-delivered with ΦC31 integrase, for example, GFP expression was 85 fold higher in rat retina (Chalberg, *et al.*, 2005) and 10 fold higher in rabbit synovium cells (Keravala, 2006b). Expression of human factor IX gene was 12 fold higher than in mice that received the factor IX plasmid alone (Olivares, *et al.*, 2002).

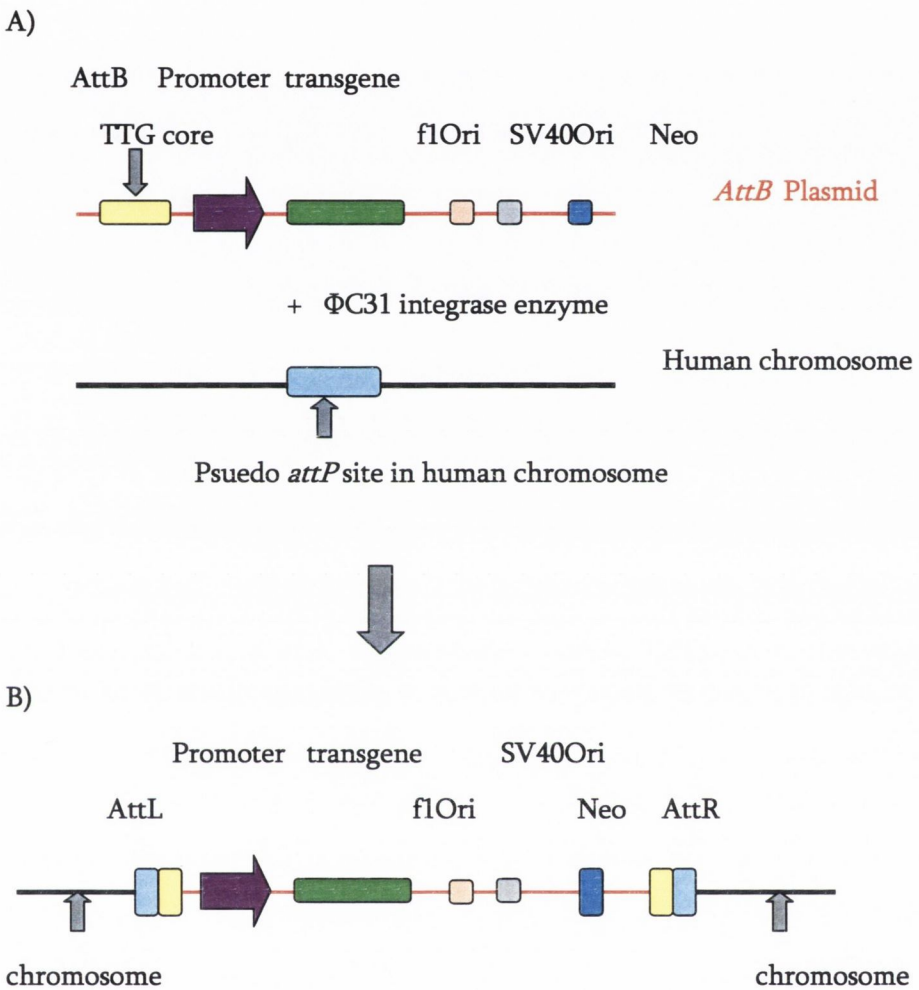


Figure 6.1 ΦC31 integrase mediated plasmid integration into the human genome. A) In the presence of the ΦC31 integrase enzyme the plasmid — with the *AttB* site ▯ is integrated into the genome ▬ at the pseudo*attP* site ▯ **B)** The integrated plasmid is flanked by two hybrid Att sites AttL and attR ▯▯ which are resistant to a ΦC31 integrase mediated back reaction. The TTG core recombination point is also highlighted.

It is clear that the ΦC31 system is highly effective (Bertoni, *et al.*, 2006; Chalberg, *et al.*, 2005; Held, *et al.*, 2005; Hollis, 2003; Keravala, 2006b; Ortiz-Urda, *et al.*, 2003; Quenneville, *et al.*, 2004), one of the advantages of this and any non-viral systems is the lack of size limitation facilitating inclusion of large transgenes, tissue specific promoters and upstream regulatory sequences. The ΦC31 integrase system in conjunction with the nucleofector technology has successfully been used to deliver a 17.2kb dystrophin expression plasmid into myogenic precursor cells (Quenneville, *et al.*, 2004). This transgene size is particularly relevant to OI as the Col1a1 and Col1a2 genes are also large at 18kb and 32kb respectively.

A major obstacle in the development of gene therapeutics is the transience of transgene expression particularly from viral vectors in stem cells, this is thought in part to be caused by epigenetic modifications such as *de novo* methylation and histone modifications (Ellis and Yao, 2005; Yao, *et al.*, 2004). The vast majority of studies using ΦC31 integrase system have reported prolonged transgene expression (Chalberg, *et al.*, 2005; Held, *et al.*, 2005; Ishikawa, 2006; Ortiz-Urda, *et al.*, 2003; Ortiz-Urda, *et al.*, 2002). Such long term expression is a prerequisite for any gene therapeutic treating a menedelian disorder which would require life long expression of the therapeutic.

It has been estimated that the human genome contains approximately 370 pseudo *AttP* sites and it appears that there may be preferential usage of sites with up to 60% of observed integration events seen multiple times in a single transfection experiment (Chalberg, *et al.*, 2006). Site usage is thought to be influenced by the level of identity with the native *attP* site, its genomic context and the transcriptional activity profile of the cell. Integration can occur in either orientation and has been frequently associated with deletions of 5 – 15bp, most likely due to the partial identity with the native bacteriophage *attP* site. The most frequently used sites are 8q22 and 19q13.31

(Chalberg, *et al.*, 2006). Selective integration into pseudo *attP* sites will be of little utility if these sites are located within essential genes such as cell cycle regulators or oncogenes. A comprehensive list of known cancer causing genes was screened for *attP* sites located within 1MB, only 3 were identified. In each case the genes are oncogenic following a translocation event, there were no detectable *attP* sites near any of the translocation partners (Chalberg, et al., 2006). Some reassurance may be taken from the lack of any mutagenic events reported to date from the *in vivo* animal studies (Chalberg, *et al.*, 2005; Held, *et al.*, 2005; Keravala, 2006b; Olivares, *et al.*, 2002) or transgenic animals generated using the ΦC31 integrase system (Groth, *et al.*, 2004; Hollis, 2003). The most serious event reported thus far from the *in vivo* animal studies is the presence of an *attP* site in the 3'UTR of the p21 gene which is involved in the cell cycle. Comparison of the human and murine 3' UTR sequences of their respective p21 genes has shown very little homology and no evidence of an *attP* site in human p21 3'UTR (Held, *et al.*, 2005). Nevertheless, it is likely that as not all *attP* sites have been identified to date that there may be an *attP* site located within important sequences in the human genome.

When assessing literature in terms of the potential risks of insertional mutagenesis one must take into account the fact that the studies have been carried out in human cell lines. Whilst excellent model systems, cell lines only reflect a minor subset of the *in vivo* properties of the cell or tissue type from which they were derived. The majority of *attP* site identified to date have been rescued from transformed cell lines such as 293T (Chalberg, *et al.*, 2006; Groth, *et al.*, 2000; Thyagarajan, *et al.*, 2001) and Jurkat cells (Ishikawa, 2006). The transcriptional profile of these cells is very perturbed in comparison to their non transformed primary cell counterpart. This is highlighted by the low number of *attP* sites detected in studies involving primary cell lines, a total of 6 sites were identified in primary keratinocytes (Ortiz-Urda, *et al.*, 2003; Ortiz-Urda, *et al.*, 2002), two in human synovial cells (Keravala, 2006b) and a single site in myoblast precursor cells (Quenneville, *et al.*, 2004). The number of sites identified in the *in vivo*

animal experiments is equally low with between two (Olivares, *et al.*, 2002) to six (Held, *et al.*, 2005) sites detected in murine liver, three in rat retina (Chalberg, et al., 2005) and three in rabbit synovial cells (Keravala, 2006b). These results suggest that due to the more restricted transcriptional repertoire of primary cell lines that the number of *AttP* sites available *in vivo* may be quite low. It is also likely that each tissue type will have a specific *AttP* site profile determined directly by the transcriptional activity of the given cell or tissue type. If indeed this is the case it may further enhance the safety profile of the ΦC31 integrase system.

It is notable that the presence of the integrase enzyme been associated with toxicity. Following *in vivo* delivery of the integrase and *AttB* plasmid with a fumarylacetoacetate hydrolase (FAH) transgene, a large number of heaptocytes displayed abnormal morphology. The number of abnormal cells was directly proportional to the amount of integrase plasmid delivered with no abnormal morphology observed in the livers of animal that received the *AttB* plasmid alone indicating an integrase mediated toxicity (Held, *et al.*, 2005). Recently it was reported that prolonged expression of the integrase enzyme induces chromosomal instability in primary cell lines with a variety of different chromosomal translocation events observed (Liu, *et al.*, 2006). 8 deletions of 1kb or greater have been reported, as yet it is unclear if these events occurred at the time of integration or are an artefact of the plasmid rescue technique employed to identify the chromosomal integration sites (Chalberg, *et al.*, 2006).

The focus of the work carried out during the course of this doctoral research project was to explore a suppression and replacement gene therapeutic approach for the treatment of Osteogenesis Imperfecta. The efficacy of the Col1a1 targeting siRNAs and shRNAs was clearly demonstrated as described in chapter 3 and chapter 4 of this thesis. Successful down regulation of Col1a1 expression using both the siRNAs and shRNAs in the target cell type, MSCs was demonstrated both at the RNA level and the protein level in chapter 5. Additionally MSC nucleofection had been optimised and established

as an effective means of delivery of therapeutic nucleic acids to the usually difficult to transfect MSCs. What was lacking from the design was an effective and safe heritable means by which sustained expression of the therapeutic could be achieved in the mesenchymal stem cell population. With a requirement for integration of the shRNA expression plasmids and based upon the success reported with the ΦC31 integrase system as described above it was decided to explore its efficacy and suitability as a delivery vector for the potential OI therapeutic under evaluation.

The following chapter describes the evaluation of the multipotency of the MSC population employed in the various studies carried out for this doctoral thesis. A selection of AttB plasmids were generated, the ΦC31 integrase system was explored in HeLa cells and in MSCs. Additionally the potential application of the ΦC31 integrase system for the delivery of shRNA plasmids was explored. To this end a series of shRNA.AttB plasmids were generated and a four week suppression study in MSCs stably expressing the plasmids carried out.

6.2 Material and Methods

This section describes the method specific to this chapter, the general methods employed and sequences of the primers used are described in chapter 2.

6.2.1 Mesenchymal Stem Cell Differentiation.

The MSCs used in this study were grown from bone explants which were isolated from the fracture site of a 21 year old male. The MSCs, a kind gift of Dr. G.Jordan, Musgrave Park Hospital Belfast, were provided at passage 1. The MSCs were expanded (section 2.8.2) and stocks frozen down at passage 2 (section 2.8.4). A single vial of cells was thawed and expanded for this differentiation study. MSCs can differentiate into a variety of different cell lineages under the influence of the appropriate biochemical stimuli. The lineages characteristic of MSC differentiation are the osteogenic, adipogenic and chondrogenic lineages (Pittenger, et al., 1999). Of these the osteoblast and adipogenic lineages were selected for study. For the differentiation experiments $2x10^5$ MSCs were plated per well of a 6 well plate, the cells were allowed to grow until confluent in normal α20 MEM medium, following which the lineage specific differentiation medium was added.

6.2.1.1 Osteogenic differentiation.

Osteoinductive medium consists of α20 MEM medium supplemented with 100μM ascorbate-2-phosphate, 100nM dexamethasone, 10mM β-glycerolphosphate. 2ml of the medium was added per well of a 6 well plate and the medium changed every second day. This regime was continued for a four week period.

6.2.1.2. Adipogenic Differentiation

To stimulate the adipogenic differentiation of the MSCs two types of adipogenic medium were used. Firstly, Adipogenic Induction Medium (AIM) was α20 MEM

medium supplemented with 0.05mg/ml of Insulin, 1μm of dexamethasome and 0.5mm of isobutylmethylxanthine (IBMX). 2ml of this was added per well of a 6 well plate, the medium was replaced 72 hours later with the Adipogenic Maintence Medium (AMM) which is the α20 MEM medium supplemented with only 0.05mg/ml of Insulin. The AMM is then replaced 24 hours later with the AIM. This regime was continued for a four week period.

6.2.2 Histological staining of differentiating MSCs

Alkaline phosphatase staining is routinely used to assess osteogenic differentiation. This enzyme accumulates within the cells which can be detected using specific kits in this case Alkaline Phosphatase (Sigma Kit No. 85) which generates amber to red stain, the colour intensity increases with increasing alkaline phosphatase concentrations. Oil Red O stains lipid vacuoles bright red and is routinely used to assess adipogenic differentiation.

6.2.2.1 Alkaline Phosphatase (Sigma Kit No. 85)

Cells were fixed by immersing in citrate buffered acetone (2:3 citrate: acetone) for 30 seconds and then gently rinsed in deionised water for 45 seconds. The slides were incubated in the Napthol AS-MX phosphate alkaline-fast violet dye solution protected from the light for 30 minutes. Slides were rinsed thoroughly in deionised water for 2 minutes, placed in Mayers Haemotoxylin solution for 10minutes and then rinsed in a mildly alkaline solution. The cells were then counterstained in toludine blue and visualised microscopically using white light. .

6.2.2.2 Oil red O Staining

Cells were fixed in 4% paraformaldehyde for an hour and stained with fresh Oil red O (Sigma) solution for 30 minutes. Oil red O was made by mixing 3 parts stock solution

with 2 parts water and filtering through a 0.2μm filter. The stain was applied to the cells for 15 minutes; cells were then washed in PBS and counter stained with toludine blue and visualized microscopically using white light.

6.2.2.3 Microscopic evaluation of MSC differentiation.

The extent of the Alkaine phosphatase or Oil red O staining of the MSCs at each of the four time points was assessed using the brightfield setting of the microscope which uses white light to view the MSC staining pattern.

6.2.3 Generation of pEGFP1.attB plasmid.

The 300bp *AttB* sequence was removed from pTOPO by EcoRI digestion (section 2.4.7), the fragment gel extracted (section 2.4.4) and ligated (section 2.4.8) into pEGFP1 which had been digested with EcoRI (section 2.4.7) and cleaned by phenol chloroform extraction (section 2.4.5). The ligated DNA was transformed into XL1 blue *E.coli* cells and plated on kanamycin selective agar plates (section 2.5.2). Bacterial PCRs were carried out to identify colonies which contained plasmids with the required *AttB* sequences (section 2.5.4) using primers pEGFPf and pEGFPr. The positive colonies generate a 432bp PCR fragment and empty vector generates 132bp fragment. A selection of positive colonies were minpreped (section 2.6.2) and sequenced (section 2.7.1), A miniprep with perfect *AttB* sequence was selected and a maxiprep carried out (section 2.7.4) generating a purified DNA stock of the plasmid pEGFP1.attB

6.2.4 Generation of pEGFP1.CMV.attB plasmid.

The 300bp *AttB* sequence was removed from pTOPO by EcoRI digestion (section 2.4.7), the fragment gel extracted (section 2.4.4) and ligated (section 2.54.8) into pEGFP1.CMV which had been digested with EcoRI (section 2.4.7) and cleaned by phenol chloroform extraction (section 2.4.5). The ligated DNA was transformed into XL1 blue *E.coli* cells and plated on kanamycin selective agar plates (section 2.5.2). Bacterial PCRs were carried out to identify colonies which contained plasmids with the required AttB

sequences (section 2.5.4) using the primers pEGFPf and pEGFPr. Positive colonies generated a 932bp PCR fragment and empty vectors generated a 632pb fragment. A selection of positive colonies were minpreped (section 2.6.2) and sequenced (section 2.7.1), A miniprep with perfect *AttB* sequence was selected and a maxiprep carried out (section 2.6.4) generating a purified DNA stock of the plasmid pEGFP1.CMV.attB.

6.2.5 Generation of pEGFP1.AttB.Col1a1 promoter plasmid.

The Col1a1 promoter was removed from the full length Col1a1 transgene by digesting with the restriction enzymes PmeI and KpnI (section 2.4.7). The 4kb promoter fragment was gel isolated (see section 2.4.4) and end filled using T4 polymerase (section 2.4.10). The Col1a1 promoter fragment was then ligated (section2.4.8) into pEGFP1.attB (generated in 6.2.1) which had been digested with the restriction enzyme SmaI. The ligation mix was transformed into XL10 Gold Ultra competent cells and plated on kanamycin selective agar plates (section 2.5.3). As the promoter could insert into the vector in either orientation it was necessary to miniprep (section 2.6.2) a selection of colonies and carry out a series of digests to determine which plasmids contained the promoter in the correct sense orientation.

Minipreps were digested with HindIII and BamH1, pEGFP-AttB has a single HindIII and BamHI site in the multiple cloning site, the Col1a1 promoter has two HindIII sites. A plasmid with the promoter in the correct sense orientation when digested with HindIII and BamHI will generate four DNA fragments of 4149bp, 2252bp, 1829bp and 277bp in size, whilst those with the promoter in the incorrect antisense orientation will generate DNA fragments of 4149bp, 2252bp, 1504bp, 602bp. A selection of minipreps which provided the appropriate DNA fragments after HindIII and BamHI digestion were sequenced (section 2.7.1) and a single minipreps then was taken forward and an endotoxin free maxi-prep carried out.

6.2.6 Generation of pcDNA.CMV$^-$.AttB

The 300bp *AttB* sequence was removed from pTOPO by EcoRI digestion (section 2.4.7), the fragment gel extracted (section 2.4.4) and ligated (section 2.4.8) into pcDNA.CMV$^-$ which had been digested with EcoRI and cleaned by phenol chloroform extraction (section 2.4.4). The ligation mix was transformed into XL1 blue *E.coli* cells and plated on ampicillin selective agar plates (section 2.5.2). Diagnostic bacterial PCRs were carried out (section 2.5.4) using primers pcDNA3.1R and pcDNA82F. Positive colonies generate a 545bp PCR fragment and empty vector a 245bp PCR fragment. A selection of positive colonies were minpreped (section 2.6.2) and sequenced (section 2.7.1). A miniprep containing perfect *AttB* sequence was selected and a maxiprep (section 2.6.4) carried out.

6.2.7 Generation of shRNA.AttB plasmids.

The H1-shRNA fragment was excised from shCol2, shCol4 and shGFP plasmids (described in chapter 4) using NotI and XhoI restriction enzymes (section 2.4.7). The resulting 300bp fragment was gel extracted (section 2.4.4) and ligated into NotI, XhoI digested pcDNA.CMV$^-$.AttB plasmid (section 2.4.8). The ligation mix was transformed into XL1 blue *E.coli* cells and plated on ampicllin selective agar plates (section 2.5.2). Positive colonies were identified using the diagnostic bacterial PCR technique using the pcDNA82F and pcDNA3.1R primers. Positive colonies generated a PCR product 300bp larger than empty pcDNA.CMV$^-$.AttB 545bp versus 845bp. Selections of shCol2, shCol4 and shGFP positive colonies were minipreped and sequenced (section 2.7.1). Individual minipreps containing perfect sequence for shCol2, shCol4 and shGFP were selected and a maxiprep of each plasmid carried out (section 2.6.2).

6.2.8 Transfection of HeLa cells with pEGFP1.CMV.attB and pCMVInt

pCMVInt is a plasmid which contains the φC31 Integrase enzyme gene under the control of a CMV promoter (Figure 2.8). $1x10^6$ HeLa cells were seeded in a 60mm plate such that the cells were 80% - 90% confluent upon transfection. 4μg of pEGFP1.CMV.attB and 4μg of pCMVInt was transfected 60mm plate of HeLa cells (as described in 2.5.5). In addition, a control plate was set up in which 3 wells were transfected with 4μg of EGFP1.CMV.AttB alone and another 3 wells with 4μg of pCMVInt.

6.2.9 Selection of HeLa cells which had integrated the pEGFP1.CMV.attB plasmid.

24 hours after transfection selective media containing 500μg/ml of G418 was added to the cells, the media was changed every second day until the control untransfected cells had completely died. The stable cell colonies that formed within each plate were pooled generating a mixed stable cell line; this was carried out for each of the 3 plates that were transfected with pEGFP1.CMV.attB and pCMVInt. G418 selection was continued for a further two weeks. This generated a total of 3 HeLa cell lines into which the pEGFP1.CMV.attB plasmid had integrated, either randomly or specifically into a pseudo *AttP* site facilitated by the ΦC31 Integrase enzyme.

6.2.10. Evaluation of plasmid integration sites in HeLa cells.

In order to determine if the stable cell lines generated were due to random integration or ΦC31 mediated integration events it was necessary to determine the genomic location of the integrated plasmids. One of the techniques employed in the literature to do this in the literature is the plasmid rescue (Figure 6.2) (Chalberg, et al., 2005; Ortiz-Urda, et al., 2002; Thyagarajan, et al., 2001). In order to rescue integrated plasmids from their chromosomal context, genomic DNA was first extracted from the stable HeLa cell lines selected for in 6.2.9. The stable HeLa cell line was split 1:3 into a 60mm tissue

culture plate genomic DNA was then extracted from these cells 24 hours later (section 2.8.6). The DNA was then digested (section 2.4.7) with EcoRV and PmeI restriction enzymes that do not cut within the plasmid and that generate digest fragments with compatible ends. Consequently the genomic DNA digest fragments may contain an intact linearised vector. Digested DNA was cleaned via phenol chloroform extraction (section 2.4.5) and self ligated (section 2.4.7). The ligation mix was then transformed and plated onto selective agar plates (section 2.5.2). Ligated fragments which contain a newly re-circularized plasmid should express the antibiotic resistance gene which permits bacterial colony growth on the selective agar plates. Minipreps of the resistant colonies were carried out (section 2.6.2) and sequenced (section 2.7.1). The sequencing primers used Att1200 and Att1201 bind within the *AttB* site. These primers flank the TTG recombination site, therefore they bind to the *AttB* sequence of the integrated vector and can be used to sequence into the genomic DNA (Figure 6.11). The aim was to assess if integrase mediated integration had occurred in HeLa cells. If indeed this was the case there should be a fragment of genomic DNA within the *AttB* site of the rescued plasmid. The use of forward and reverse primers located within the *AttB* site should enable the sequence generated from the sequencing reaction to run into the genomic sequence which would allow identification of the integration site of the plasmid (Figure 6.2).

6.2.11 Bacterial plasmid PCR of *E.Coli* colonies generated from plasmid rescue technique.

The HeLa cell transfection experiment was repeated and a second set of stable cell lines generated. An initial screening of the *E.coli* colonies generated from the plasmid rescue (section 6.2.11) was carried out in order to determine if any of the colonies contained rescued plasmids with additional genomic DNA sequences located within the *AttB* site. The bacterial PCR was carried out as described in section 2.5.4. The primers used were the forward and reverse *AttB* sequencing primers, AttB1200 and *AttB* 1201. Intact wild type *AttB* sequence will generate a PCR product of 206bp. Plasmids which have

participated in a ΦC31 Integrase mediated integration event would be expected to contain genomic DNA sequence within the *AttB* sequence as a consequence of the plasmid rescue technique. Such plasmids should generate a larger PCR product. The size of the genomic DNA insertion will depend upon the proximity of EcoRV and PmeI sites which were used to rescue plasmids from the HeLa cell genomic DNA.

6.2.12 Restriction enzyme digestion of plasmids isolated from HeLa cells by plasmid rescue.

Restriction enzyme digestion of the pEGFP1.CMV.attB plasmids rescued from the 3 HeLa cells lines generated in the second transfection experiment was carried out. The enzymes used, HindIII and BamH1 flank the *AttB* site. It would be expected that plasmids which have been involved in a ΦC31 integrase mediated integration event would have additional genomic DNA sequence present as a result of the plasmid rescue technique used to retrieve the plasmids from their genomic location. Therefore, the plasmids which have been involved in a ΦC31 integrase mediated integration event should yield larger digest DNA fragments than the control plasmids with intact wild type *AttB* sequences. The digestions were carried out as described in section 2.4.7. The wild type intact *AttB* DNA digest fragment is 852bp, due to the presence of the CMV promoter and AttB sequences in the digest fragment.

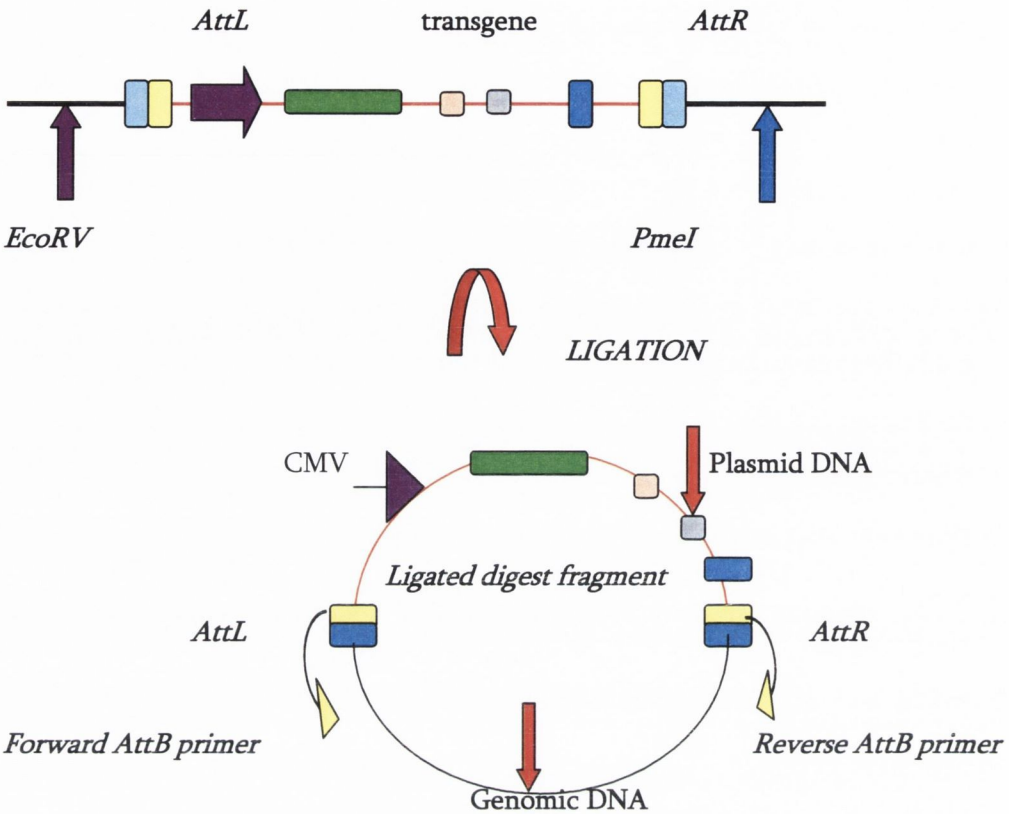


Figure 6.2 Plasmid rescue technique. This technique was used isolate plasmids which had undergone ΦC31 integrase mediated integration and to identify chromosomal integration sites. Cells were transfected with the *AttB* transgene plasmid and integrase plasmid and DNA isolated from G418 resistant population. The DNA was digested with the blunt cutters EcoRV and PmeI which did not cut within the integrated plasmid. A ligation reaction was carried out; fragments containing the intact plasmid will grow on selective agar plates following transformation. The sequencing primers Att1200 and Att1201 which bind within the *AttB* site will sequence into the genomic DNA revealing the genomic integration site identity.

6.2.13 Nucleofection of MSCs with pEGFP.AttB.Col1a1 promoter and pCMVInt.

Having determined that nucleofection is an extremely efficient means of delivering nucleic acids to MSCs, it was also employed in this instance for the delivery of the pEGFP.AttB.Col1a1 promoter and pCMVInt plasmids. $5x10^5$ MSCs were plated per reaction such that cells were 85% confluent on the day of nucleofection. Cells were harvested by trypsinization, counted and aliquoted into single reaction amounts ($1x10^6$ cells). The cells were spun @ 100g for 10 minutes, the media aspirated and the cells resuspended in 100µl of the pre-warmed nucleofection solution. 500ηg of endotoxin free pEGFP.AttB.Col1a1 promoter and 500ηg pCMVInt plasmid was added to each of the five aliquot of cells. The MSC:DNA suspension was transferred to the Amaxa cuvette, placed in the nucleofection device and program U23 selected. Once complete, 500µl of pre-warmed media was added to the cuvette and the nucleofected cells transferred to a single well of a 6 well plate. One of the pEGFP.AttB.Col1a1 and pCMVInt samples was plated onto coverslips in a single well of a 6 well plate. The cells were then incubated in a humidified 37°C, 5% CO_2 incubator. This nucleofection experiment was carried out twice.

6.2.14 Microscopic Analysis of MSC eGFP expression

48 hours after nucleofection with pEGFP.AttB.Col1a1 and pCMVInt plasmids the MSCs plated onto coverslips were rinsed with ice cold PBS, fixed with 4% paraformaldehyde for 30 minutes, washed again in PBS and analysed using the fluorescence microscope for levels of eGFP expression. The cells were mounted using a small amount of aquamount, eGFP expression was detected and evaluated using the eGFP specific filter setting of the fluorescent microscope. When assessing the stable MSC cells which had integrated the pEGFP.AttB.Col1a1 promoter plasmid the cells were treated with DAPI prior to evaluation of eGFP expression.

6.2.15 Selection of MSCs which had integrated the pEGFP.AttB.Col1a1 promoter plasmid

Cells were allowed to recover for 48 hours following which selective medium containing 250μg.ml of G418 was added. This selection media was changed every 48 hours until the stable cell colonies had formed.

6.2.16 MSC nucleofection with shRNA.AttB and pCMVInt.

To investigate the efficacy of shRNA mediated Col1a1 suppression in MSCs from a plasmid containing the *AttB* sequence and following ΦC31 mediated integration 500ηg of shCol2.AttB, shCol4.AttB or shGFP.AttB and 500ηg of the ΦC31 integrase plasmid pCMVInt was added $1x10^6$ MSCs, nucleofected using program U23 as described above (6.2.12). Each nucleofection was carried out in triplicate.

6.2.17 Selection of MSCs which had integrated the shRNA.AttB plasmids, shCol2.AttB, shCol4.AttB or shGFP.AttB

Cells were allowed to recover for 48 hours following which 250μg.ml of G418 was added. This selection media was changed every 48 hours until the control pCMVInt nucleofected cells had completely died. The stable MSC colonies that formed in each well of the shCol2.AttB, shCol4.AttB or shGFP.AttB nucleofected cells were pooled and plated into a 60mm plate. This generated a single cell line from each triplicate nucleofection, i.e. one cell line for shCol2.AttB, shCol4.AttB and shGFP.AttB

6.2.18. Four week shRNA suppression study in MSCs stably expressing shRNA.AttB plasmids, shCol2.AttB, shCol4.AttB or shGFP.AttB

The efficacy of shRNA mediated Col1a1 suppression in MSCs stably expressing the shRNA.AttB plasmids, shCol2.AttB, shCol4.AttB or shGFP.AttB was evaluated using

real time RT-PCR. As the Col1a1 RNA suppression levels were to be evaluated over a 28 day period.$1x10^5$ MSCs from each cell line was plated in triplicate per 6 well plate such that the cells were confluent per time point RNA extraction. The levels of Col1a1 expression evaluated via RT-PCR and the relative quantification method as described in chapters 3 to chapters 5 of this thesis. GAPDH was used as the housekeeping gene, therefore all data was normalized with GAPDH. A single negative control was used the stable cell line expressing shGFP.AttB. Therefore the relative Col1a1 expression levels were determined by comparison to Col1a1 expression levels in this stable MSC cell line. The data generated was evaluated using Graph Pad Instat3, the mean and standard deviation of each data set was determined and the significance of the reduced Col1a1 expression levels observed in the MSCs stably expressing shCol2.AttB and shCol4.AttB determined by comparison with Col1a1 expression levels in the MSCs stably expressing shGFP.AttB.

6.3 Results

When the studies described in this chapter were being designed a variety of parameters needed to be explored and evaluated. The overall aim of the work presented in this chapter was to assess the efficacy of the ΦC31 integrase system as a delivery vector for use in MSCs, however, in order to reach this goal a selection of issues needed to first be addressed. Firstly, in relation to the MSC population to be used in the study, was this population of cells a true mesenchymal stem cell population i.e. did it have the capacity to differentiate in a variety of more committed cell types? Secondly, was the integrase system effective in our hands? To this end it was decided to evaluate the system in HeLa cells initially. Thirdly was the system active in MSCs? Did the presence of the *AttB* site interfere with the capacity of the cells to differentiate? In the development of a gene therapy the ideal would be to have tissue specific expression of the therapeutic, in the case of Col1a1 linked OI ideally the expression of the therapeutic (e.g RNAi) and in the case of suppression and replacement both the suppressor and the replacement gene would only be expressed in tissues which express Col1a1. Therefore, the effect of the AttB sequence and ΦC31 integration upon the activity of the Col1a1 promoter was also of interest. Therefore it was decided to generate a plasmid which included the *AttB* sequence and the Col1a1 promoter sequence to drive expression of a reporter gene, in this case eGFP. In theory, this plasmid could be used to assess the transcriptional activity of the Col1a1 promoter in the MSCs as they were stimulated to differentiate into osteoblasts or adipocytes. It would be expected that eGFP expression levels would be high in the MSCs and osteoblasts but significantly reduced in the cells progressing through the adipogenic lineage. The final objective was to determine if the use of an *AttB* plasmids system was compatible with long term expression shRNAs with subsequent sustained target suppression (in this case Col1a1) in MSCs. To this end a series of shRNA.AttB plasmids were generated, delivered to MSCs, stable cell lines expressing the constructs generated and the levels of Col1a1 expression assayed over a four week period. This section of the chapter details the results of these experimental endeavours.

6.3.1 Evaluation of the multipotency of a Mesenchymal Stem Cell population.

The therapeutic approach outlined requires the delivery of therapeutic nucleic acids to MSCs but if the cells do not have the capacity to differentiate into the desired cell type (osteoblasts for OI therapeutics) the therapeutic potential of the MSCs will be significantly diminished. In addition, if the potential application of the integrase system for delivery of therapeutic nucleic acids to MSCs was going to be explored it would be necessary to determine if the presence of this non viral vector had any influence on the differentiation capacity of the cells. The lineages characteristic of MSC differentiation are the osteogenic, adipogenic and chondrogenic lineages (Pittenger, et al., 1999). It was decided to explore the capacity of the MSCs population employed in the suppression studies described in chapter 5 of this thesis to differentiate into the osteogenic and adipogenic lineages. The MSCs were successfully stimulated to differentiate into both the osteogenic and adipogenic lineages at passage 4 using standard differentiation medium and techniques described in section 6.2.1. The experiment was carried out over a four week period with histochemical analysis carried out at days 7, 14, 21 and 28.

6.3.1.1 Alkaline phosphatase valuation of Osteogenic differentiation of MSCs

Differentiation of the MSCs was evaluated using Alkaline phosphatase staining. Each well of MSCs on a 6 well plate contained 3 coverslips; therefore a total of 9 coverslips was evaluated per time point for osteogenic differentiation. In addition 6 coverslips on which control MSCs which had been grown in normal medium were also evaluated for osteogenic differentiation. Alkaline phosphatase staining of the MSCs treated with the osteoinductive medium revealed increasing levels of enzyme production over a 28 day period. At day 7 very little staining was evident, but it was clear that the cells which had been treated with the osteogenic medium were progressing through a phenotypic change as the morphology of the cells was very different from the control cells (Figure 6.3.). The control cells had a more spindle arrangement classic of MSCs (Figures 6.3 A and B), in contrast the cells treated with the osteogenic medium were beginning to take up a more rounded appearance (Figure 6.3 C and D). This pattern was evident through

days 14 – 28 also (A and B versus C and D, Figures 6.4 – 6.6). At day 14 there was a more widespread production of alkaline phosphatase albeit at low levels, by day 21 there was extensive production of alkaline phosphatase indicated by the significant amount of red staining and sustained high level of expression of alkaline phosphatase by day 28 with an further increase in the amount and intensity of the red staining resultant from the histochemcial analysis. In no case over the 28 day study duration was any alkaline phosphatase staining apparent in the control MSCs (Figures 6.3 – 6.6 A and B). Although it is notable that the control, untreated cells in figure 6.6 A and B do not appear normal with what is sometimes referred to as a fried egg like shape. It is not known why this morphological change occurs.

6.3.1.2 Evaluation of MSCs adipogenic differentiation using Oil Red O staining

Adipogenic differentiation of the MSCs was evaluated using Oil Red O staining. Each well of MSCs on a 6 well plate contained 3 coverslips; therefore a total of 9 coverslips was evaluated per time point for adipogenic differentiation. In addition, 6 coverslips on which control MSCs which had been grown in normal medium were also evaluated for adipogenic differentiation. Oil red O staining of MSCs treated with adipogenic induction and adipogenic maintenance medium revealed that the MSCs had efficiently proceeded towards the adipogenic lineage over the 28 day experiment time course. After an initial lag phase of 7 days, at which time there was only one or two cells per slide showing signs of lipid accumulation (Figure 6.7 panels C and D) there was a significant increase in the number of Oil red O positive cells by day 14 (Figure 6.8) further increasing through days 21 (Figure 6.9 panels C and D) and day 28 (Figure 6.10 panels C and D) due to both the increasing numbers of cells committed to the adipogenic lineage and the presence of fully differentiated mature adipocytes. Once again there was no evidence of adipogenic differentiation in the control MSCs at any of the time points from day 7 to day 28 evaluated (Figures 6.7 -6.10 panels A and B). It is clear from this study that this particular MSC population is multipotent and can be stimulated to differentiate into both the osteogenic and adipogenic lineages.

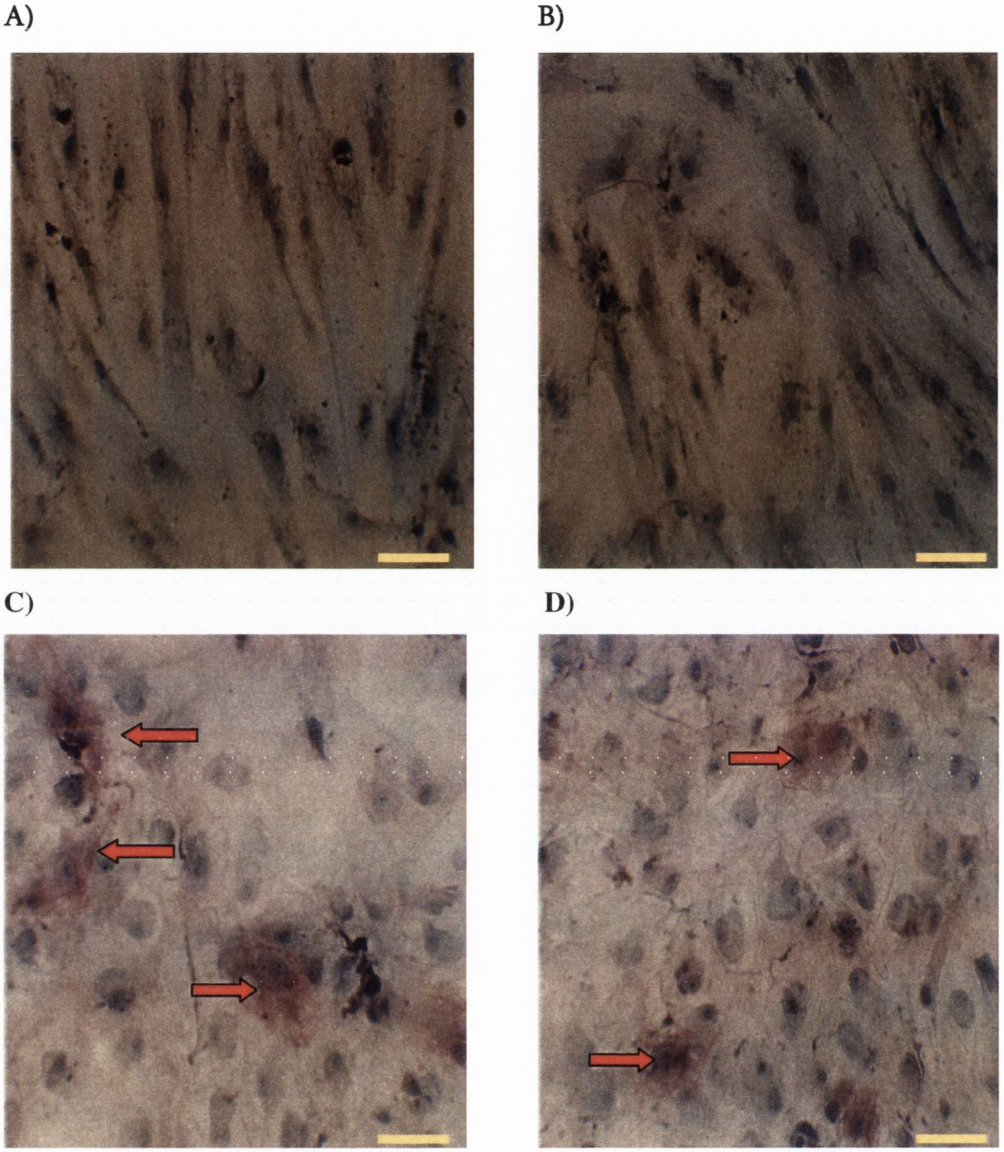


Figure 6.3 Evaluation of Osteogenic differentiation of MSCs, Day 7. 2 x10^5 MSCs were seeded into a single well of a 6 well plate. Once confluent osteogenic medium was added to the cells, cells were cultured for 7 days and stained with Alkaline Phosphatase as described in section 6.2.11. **A) & B)** Control MSCs treated with normal medium, no enzyme accumulation is evident. **C) & D)** MSCs treated with osteogenic medium for 7 days, low level alkaline phosphatase accumulation is apparent, i.e. pink staining regions, highlighted with red arrows. Scale Bar 20μm

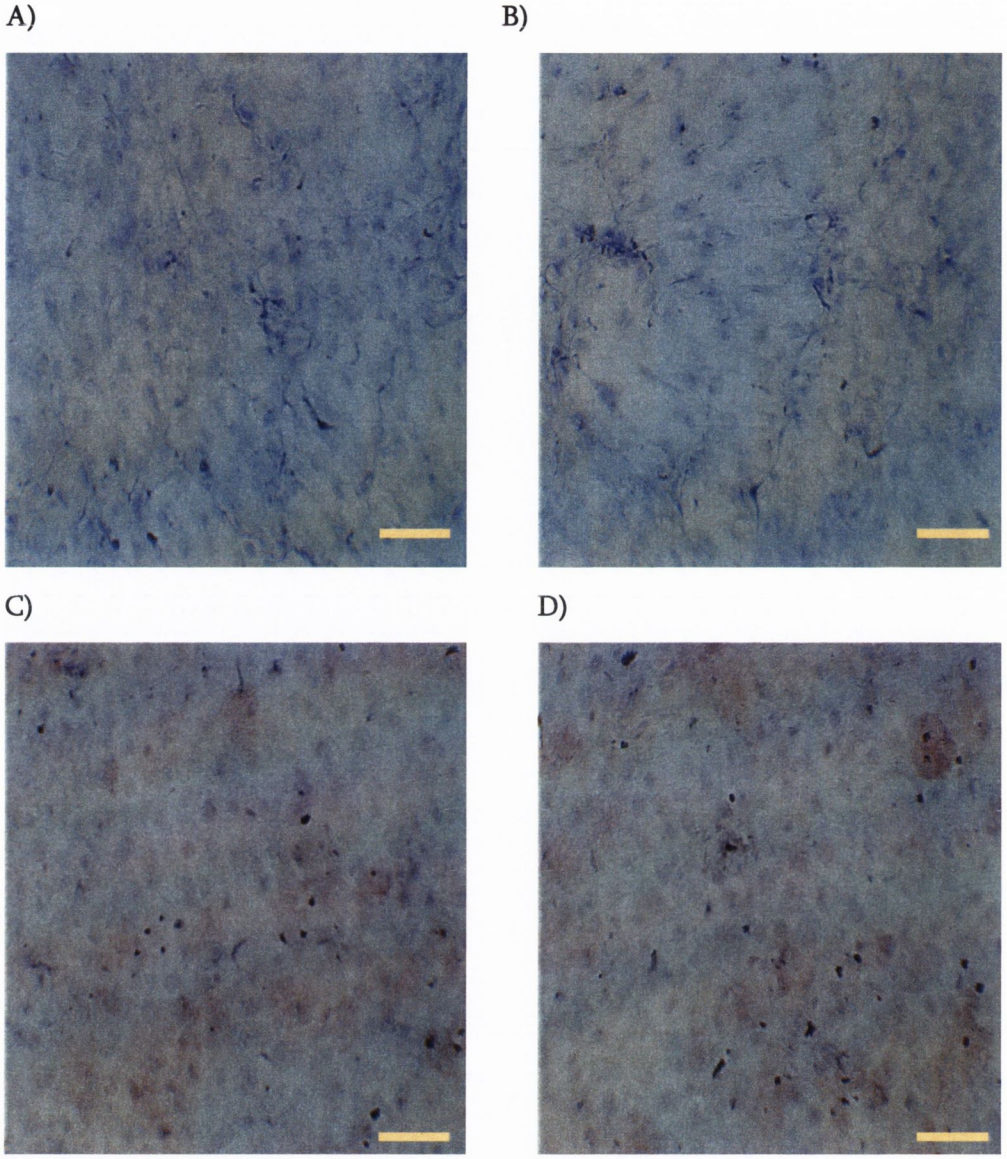


Figure 6.4 Evaluation of Osteogenic differentiation of MSCs Day 14. 2 x10^5 MSCs were seeded into a single well of a 6 well plate. Once confluent osteogenic medium was added to the cells, cells were cultured for 14 days and stained with Alkaline Phosphatase as described in section 6.2.11. **A & B** Control MSCs treated with normal medium, no enzyme accumulation is evident. **C & D** MSCs treated with osteogenic medium for 14 days, extensive alkaline phosphatase accumulation is apparent, i.e. pink staining regions. Scale Bar 40μm

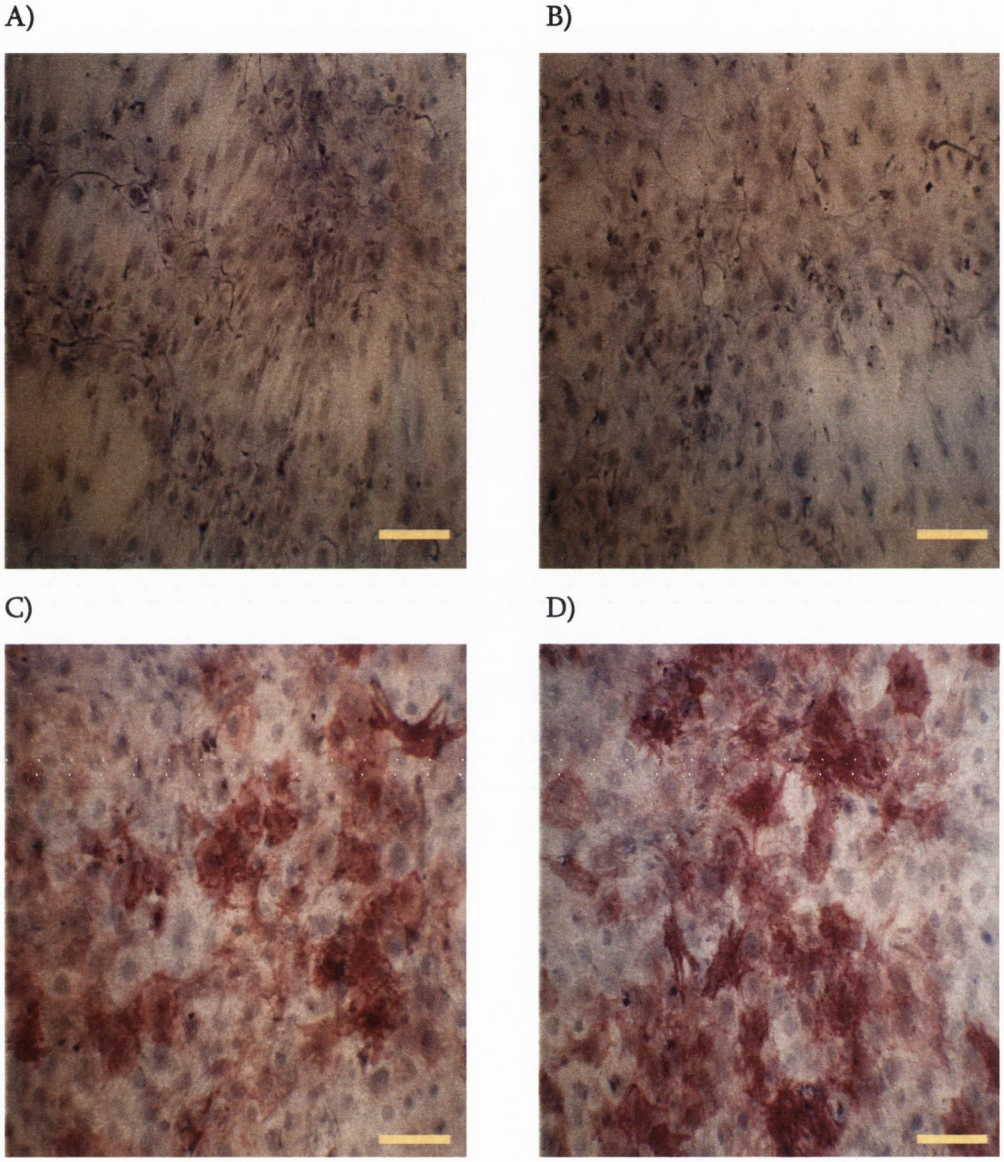


Figure 6.5 Evaluation of Osteogenic differentiation of MSCs day 21 2 x10^5 MSCs were seeded in a single well of a 6 well plate. Once confluent osteogenic medium was added to the cells, cells were cultured for 21 days and stained with Alkaline Phosphatase as described in section 6.2.11. **A & B** Control MSCs treated with normal medium, no enzyme accumulation is evident. **C & D** cells treated with osteogenic medium for 21days, high levels of alkaline phosphatase accumulation apparent, i.e. pink/red staining regions. Scale Bar 40μm

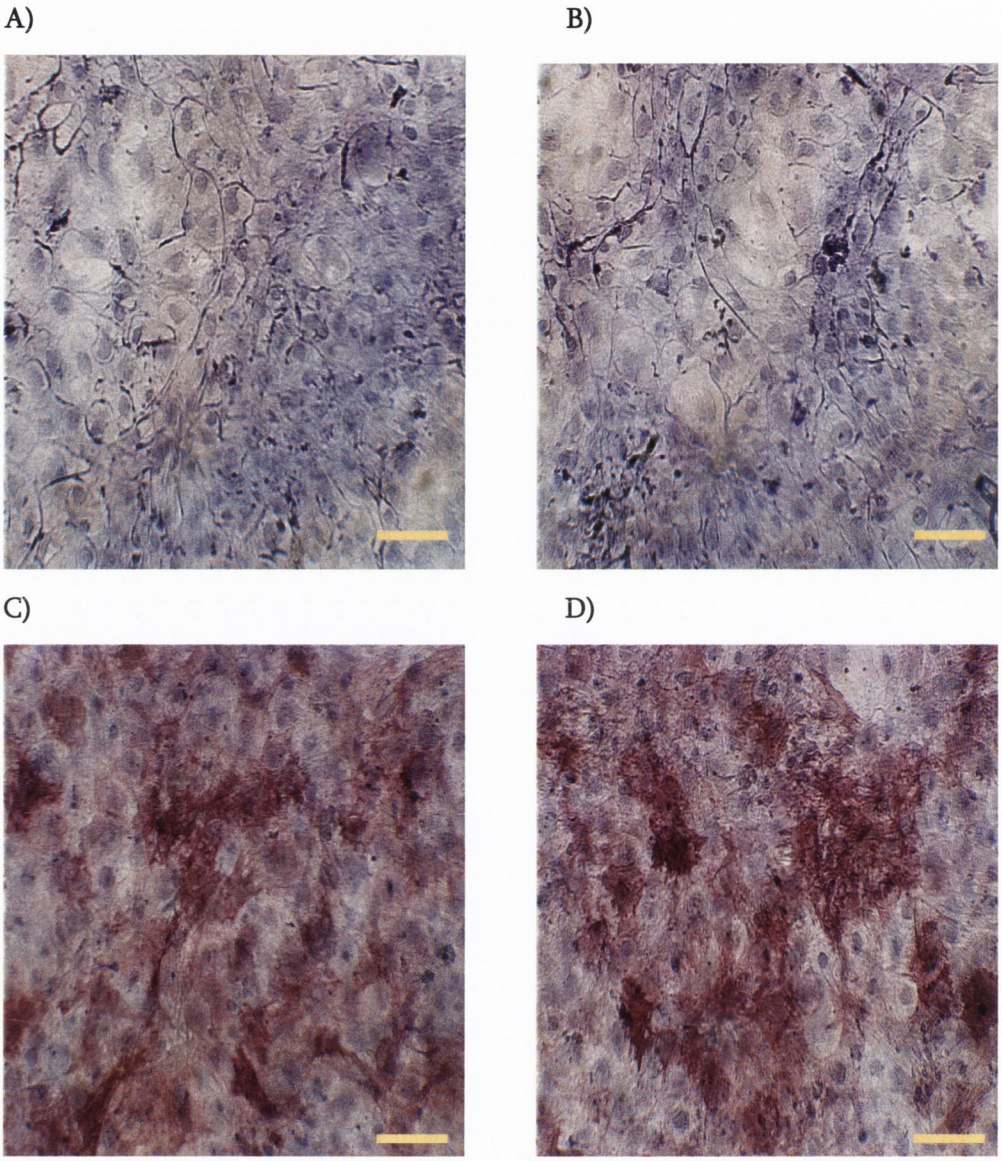


Figure 6.6 Evaluation of Osteogenic differentiation of MSCs day 28. $2x10^5$ MSCs were seeded in a single well of a 6 well plate. Once confluent osteogenic medium was added, cells were cultured for 21 days and stained with Alkaline Phosphatase as described (section 6.2.11) **A & B)** Control MSCs treated with normal medium, no enzyme accumulation is evident. **C & D** cells treated with osteogenic medium for 21days, high level alkaline phosphatase accumulation apparent, i.e. pink/red staining regions. Scale Bar 40μm

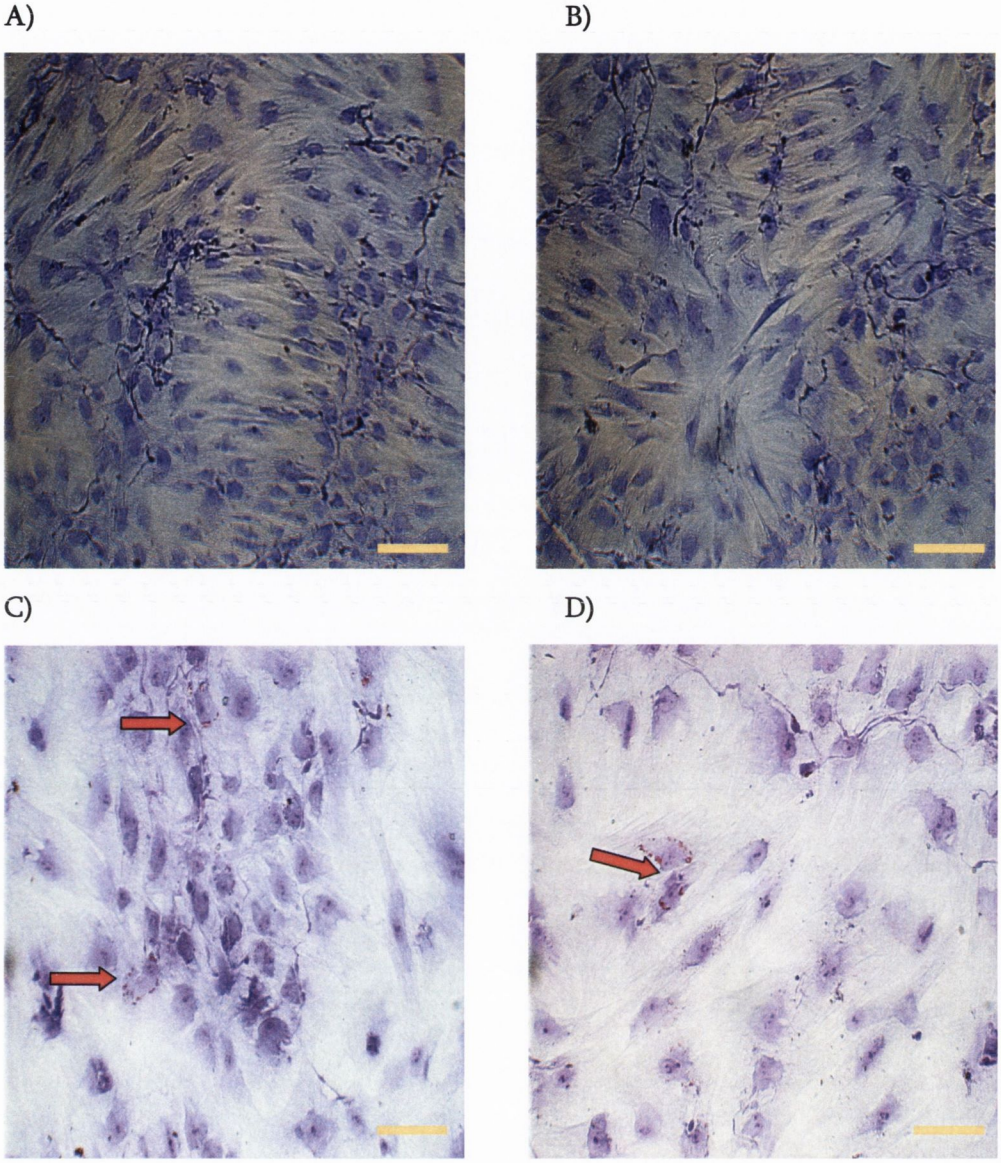


Figure 6.7 Evaluation of adipogenic differentiation of MSCs day 7. $2x10^5$ MSCs were seeded into a single well of a 6 well plate. Once confluent adipogenic medium was added to the cells, cells were cultured for 7 days and stained with Oil red O as described in section 6.2.12. **A & B** Control MSCs which were not treated with adipogenic medium, no lipid vacuoles are present. **C & D** MSCs grown in adipogenic medium for 7 days, slight development of lipid vacuoles apparent. Scale Bar 40μm

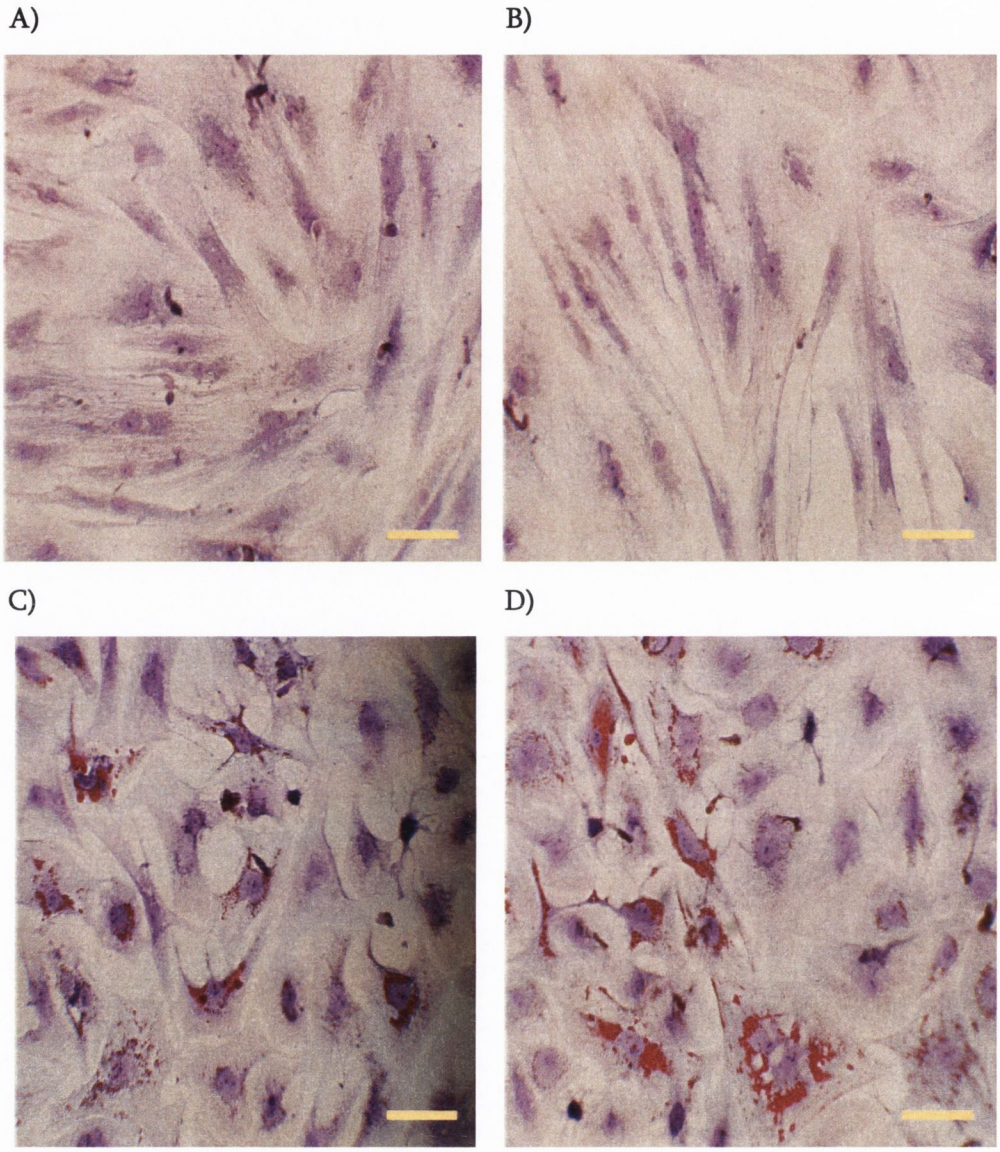


Figure 6.8 Evaluation of adipogenic differentiation of MSCs Day 14. $2x10^5$ MSCs were seeded into a single well of a 6 well plate. Once confluent adipogenic medium was added to the cells, cells were cultured for 7 days and stained with Oil red O as described in section 6.2.12. **A & B** Control MSCs which were not treated with adipogenic medium, no lipid vacuoles are present. **C & D** MSCs grown in adipogenic medium for 14 days, accumulation of lipid vacuoles stained red apparent. Scale Bar 20μm

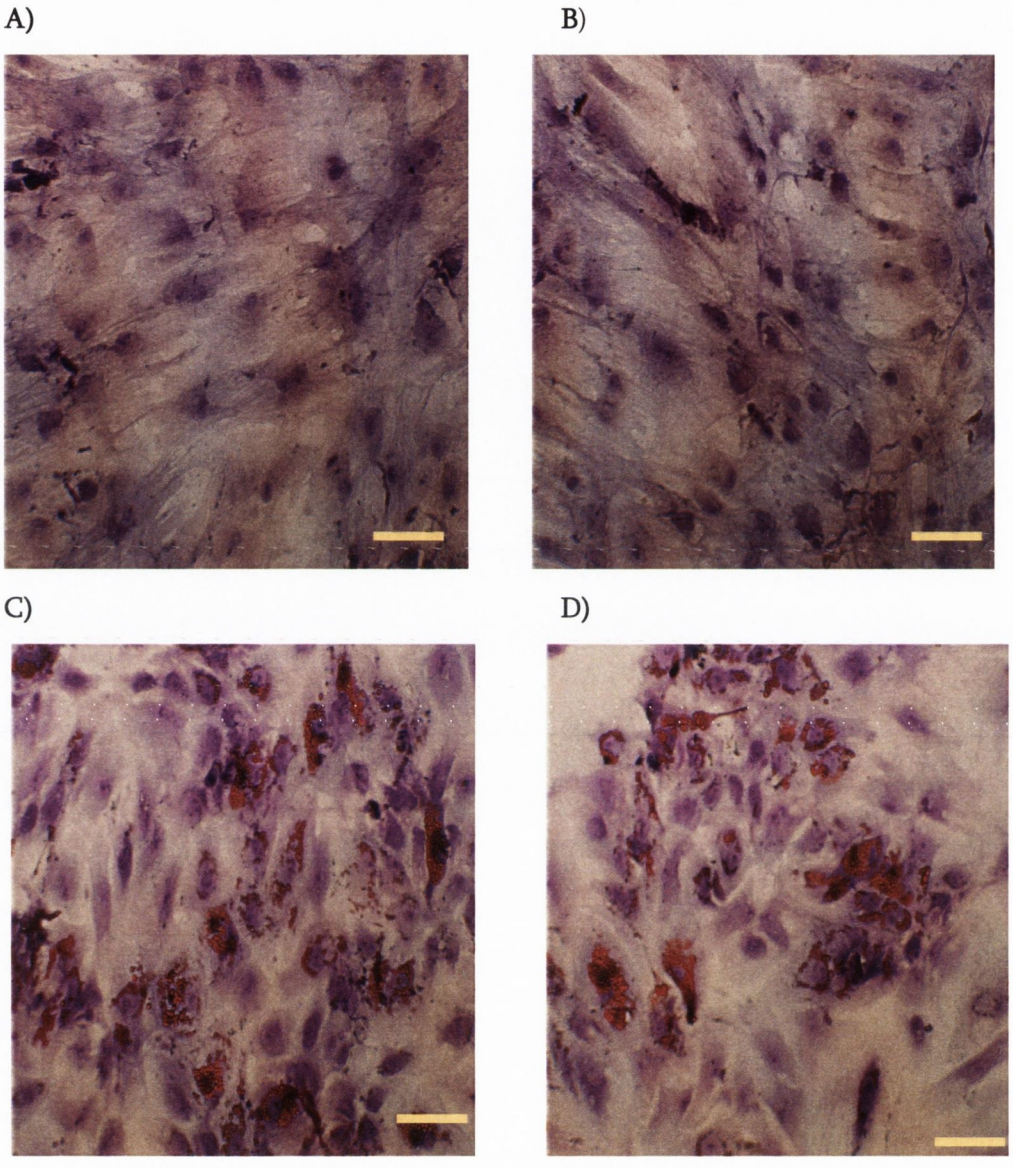


Figure 6.9 Evaluation of adipogenic differentiation of MSCs Day 14.2x10^5 MSCs were seeded in a single well of a 6 well plate. Once confluent adipogenic medium was added, cells were cultured for 7 days and stained with Oil red O (section 6.2.12). **A & B** Control MSCs which were not treated with adipogenic medium, no lipid vacuoles are present. **C & D** cells treated with adipogenic medium for 21 days, significant accumulation of lipid vacuoles (red globular stained regions) within cells is apparent. Scale Bar 20μm

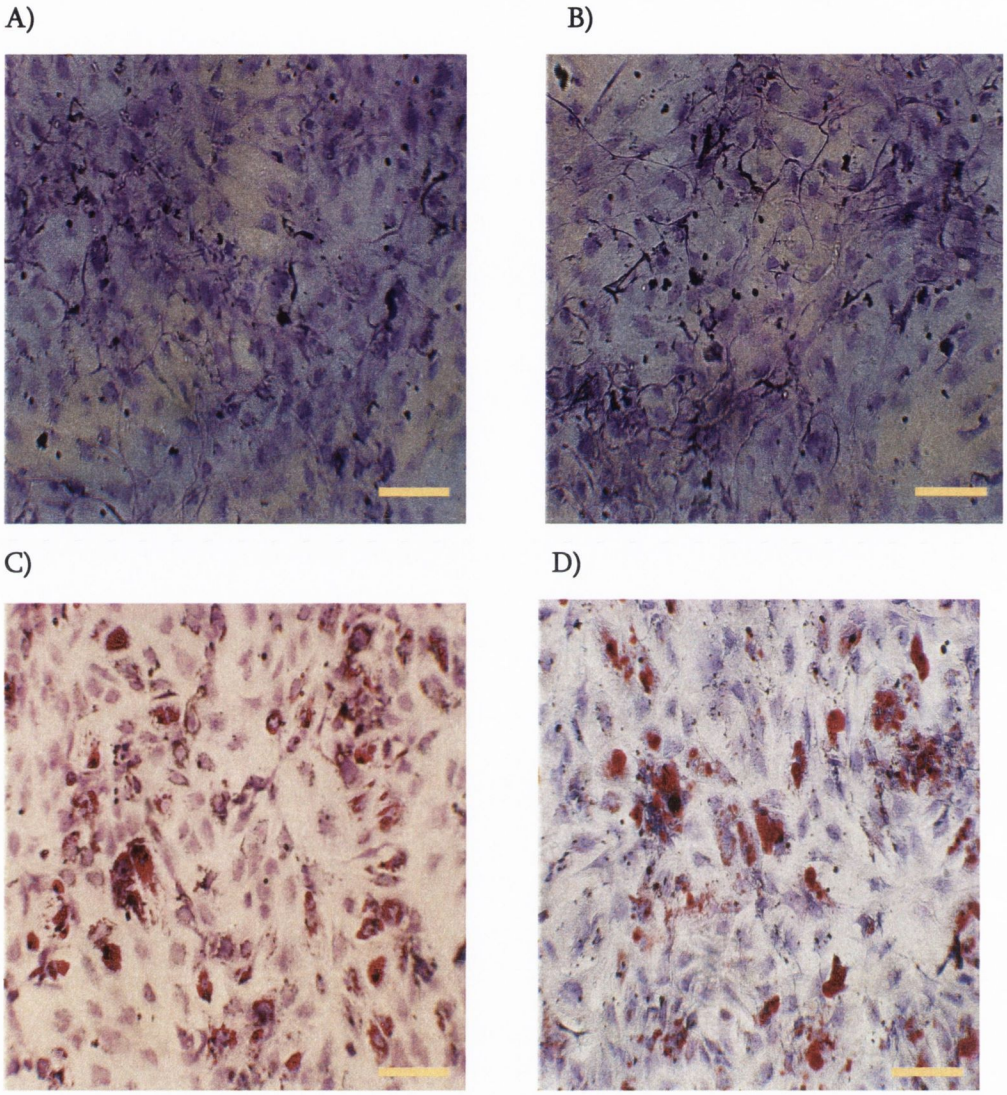


Figure 6.10 Evaluation of adipogenic differentiation of MSCs day 28. $2x10^5$ MSCs were seeded in a single well of a 6 well plate. Once confluent adipogenic medium was added, cells were cultured for 7 days and stained with Oil red O (section 6.2.12). **A & B** Control untreated MSCs, no lipid vacuoles present. **C & D** cells treated with adipogenic medium for 28 days, further increase in intracellular amount of lipid vacuoles per cell. Scale Bar 40μm

6.3.2 Generation of *AttB* containing plasmids.

Each of the required constructs containing the *AttB* sequence, pEGFP1.AttB, pEGFP1.CMV.AttB, and pcDNA.CMV^{-}.AttB were successfully generated. As integration can occur in either orientation, vectors in which the *AttB* had successfully been cloned in either orientation could be used once the sequence obtained from the *AttB* sequencing reaction was intact and correct. The sense and antisense *AttB* sequences are outlined in Figures 6.11 and 6.12. Plasmid designs and the sequence data generated for each construct generated containing the AttB is detailed in Figures 6.13 to 6.18.

6.3.3 Generation of stable HeLa cell lines

Three 60mm plates of HeLa cells were transfected with pEGFP1.CMV.AttB and pCMVInt as described in section 6.2.9. The next step was to select for HeLa cells which in which an integration event had occurred. This was achieved using G418 selection as described in section 6.2.10. 3 stable HeLa cell lines were generated in each transfection experiment, none of the HeLa stable cell lines generated were clonal and therefore most likely had a mixed population of integration events.

6.3.4 Retrieval of plasmids integrated into HeLa cell genomes using the plasmid rescue technique.

The next step was to determine whether plasmid integration in these three stable HeLa lines cells transfected with pEGFP1.CMV.AttB and pCMVInt was random or ΦC31 integrase mediated. This was determined by plasmid rescue, this technique aims to retrieve the integrated plasmids from the chromosomal locations in to which the plasmids have inserted. In order to do this the genomic DNA must first be isolated from the genome. Once the genomic DNA has been isolated the next step is to rescue the plasmids, this can in theory be achieved by digesting the genomic DNA with two restriction enzymes which have compatible ends (to facilitate ligation) but that do not

cut within the plasmid. The pool of digested genomic DNA generated should contain DNA fragments within which there is an intact linear plasmid. Ligation of the DNA should then generate intact plasmids containing genomic DNA sequences. As ΦC31 integrase mediated *AttB* recombination events occur at the TTG cross over point with the *AttB* sequence, it is possible to identify the chromosomal point of integration by sequencing from the *AttB* sequence out into the genomic DNA using primers Att1200 and Att1201 which bind within the *AttB* sequence and flank the TTG core (Figure 6.11 – Figure 6.12).

The stable HeLa cells generated using pEGFP1.CMV.AttB and pCMVInt were seeded on a 60mm dish and DNA was extracted when the cells were 70% confluent. The extracted genomic DNA was digested with the restriction enzymes PmeI and EcoRV which generate DNA digest fragments with compatible ends. The DNA digest fragments were ligated and transformed onto *E.Coli* cells and grown on kanamycin selective agar plates. The plasmid rescue technique worked efficiently with growth of kanamycin resistant bacterial colonies indicating that plasmids had been successfully rescued from the genomes of the HeLa cells. 20 of the colonies were minipreped and sequenced. The sequencing primers Att1200 and Att1201 were used (Figure 6.11). Six of the plasmids did not generate any sequence, when positive control sequencing reactions were working efficiently. Of the14 plasmids that were successfully sequenced none had any additional genomic sequences located within the *AttB* site indicating that there was no detectable ΦC31 integrase mediated events had occurred in the rescued plasmid population. It was decided to repeat the transfection experiment and generated new HeLa stable cell lines.

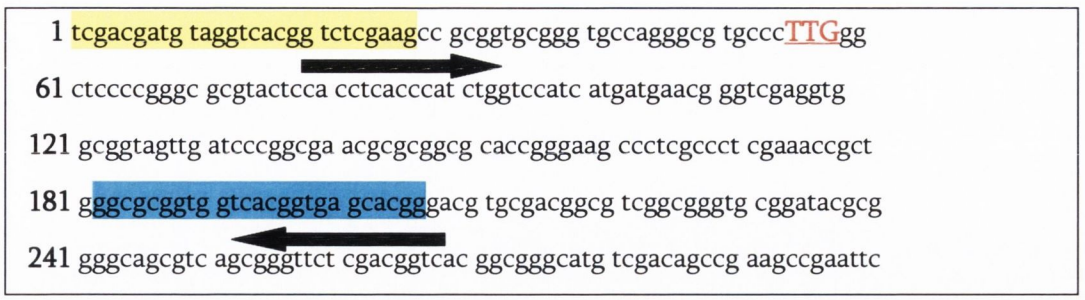


Figure 6.11 The *AttB* sequence in the sense orientation. This 300bp sequence is the *AttB* sequence. This sequence undergoes site specific homologous recombination with pseudo *AttP* sites in the human genome. The reaction is catalyzed by the ΦC31 integrase enzyme. The *AttB* sequence inserts at its TTG core sequence, highlighted in red. Highlighted in yellow is primer AttB1200 and highlighted in blue is primer AttB1201.

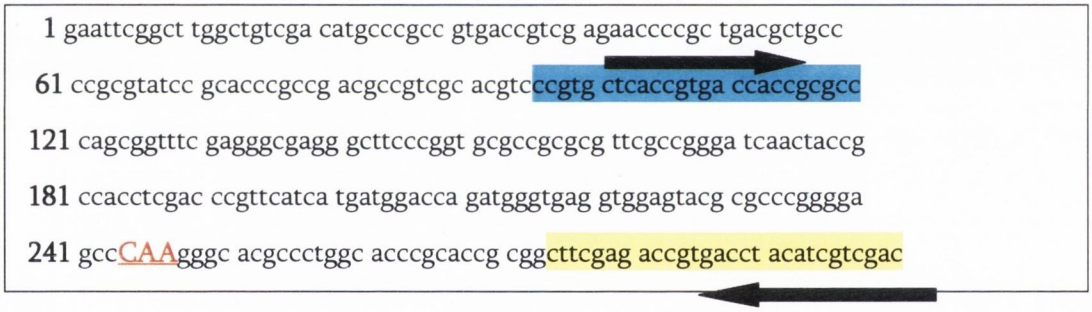


Figure 6.12. The *AttB* sequence in the antisense orientation. This 300bp sequence is the *AttB* sequence. This sequence undergoes site specific homologous recombination with pseudo *AttP* sites in the human genome. The reaction is catalyzed by the ΦC31 integrase enzyme. The *AttB* sequence inserts at its CAA core sequence, highlighted in red. Highlighted in yellow is the primer AttB1200 and highlighted in blue is the primer AttB1201.

pEGFP1.AttB plasmid

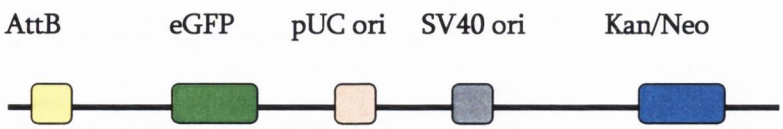


Figure 6.13 pEGFP1.AttB plasmid. This plasmid contains; the *AttB* sequence which together with ΦC31 integrase effects site specific integration into *AttP* sites in the human genome, an eGFP reporter gene, the pUC origin of replication is required for plasmid replication in bacteria and the SV40 origin of replication which is required for plasmid replication in mammalian cells. The Kanamycin /Neomycin gene facilitates selection of bacterial or mammalian cells which contain the plasmid.

AttB sequence in pEGFP1.AttB plasmid

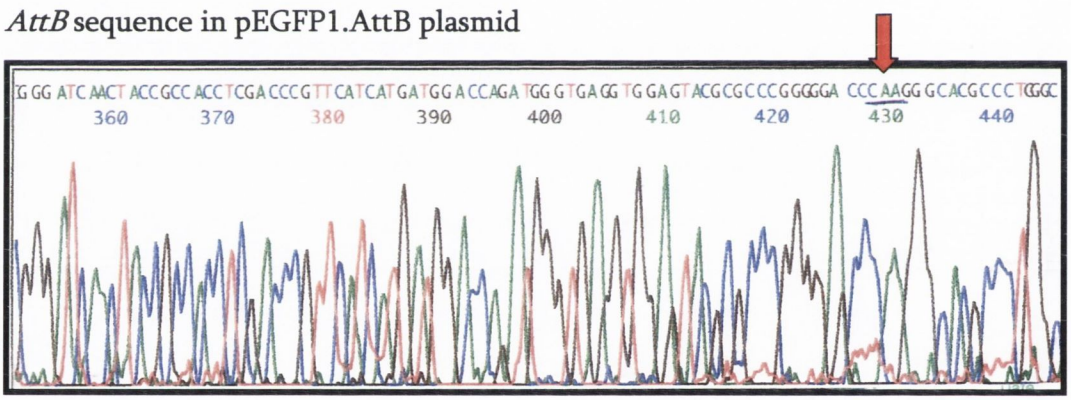


Figure 6.14 Sequence chromatogram generated from pEGFP1.AttB using pEGFPf sequencing primer. The *AttB* sequence which together with ΦC31 integrase mediates site specific integration into pseudo *AttP* sites in the human genome was cloned into the EcoRI site of pEGFP1. The AttB sequence was assessed using the pEGFPf sequencing primer. The *AttB* DNA fragment inserted in the 3' – 5' antisense orientation, the CAA recombination site is indicated by the red arrow.

pEGFP1.CMV.AttB plasmid

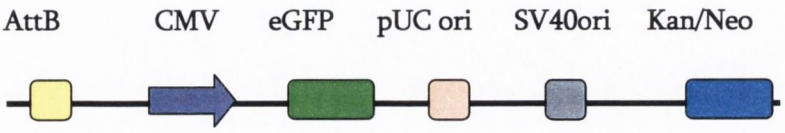


Figure 6.15 Schematic of pEGFP1-CMV.AttB plasmid. This plasmid contains; the *AttB* sequence which together with ΦC31 integrase effects site specific integration into *AttP* sites in the human genome, the CMV promoter drives eGFP reporter gene expression, the pUC origin of replication required for plasmid replication in bacteria, SV40 origin of replication required for plasmid replication in mammalian cells. The Kanamycin /Neomycin gene facilitates selection of bacterial or mammalian cells containing the plasmid.

AttB sequence in pEGFP1.CMV.AttB plasmid

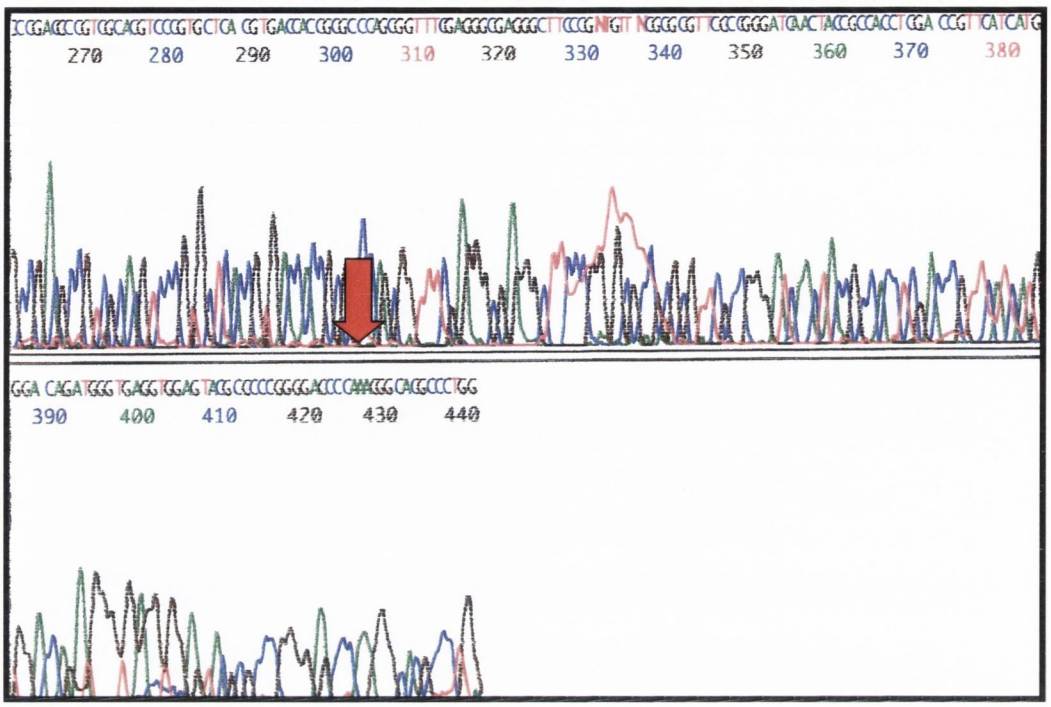


Figure 6.16 Sequence chromatogram generated from pEGFP1.CMV.AttB using pEGFPf sequencing primer. The *AttB* sequence which together with ΦC31 integrase mediates site specific integration into pseudo attP sites in the human genome was cloned into the EcoRI site of pEGFP1. The AttB sequence was assessed using the pEGFPf sequencing primer. The *AttB* DNA fragment inserted in the 3' – 5' antisense orientation, the CAA recombination site is indicated by the red arrow.

pcDNA.CMV^{-}.AttB plasmid

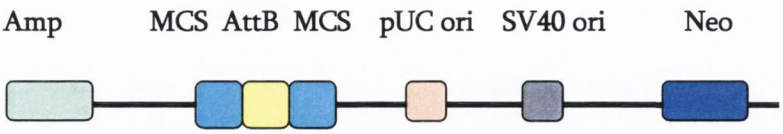


Figure 6.17 pcDNA.AttB plasmid. This plasmid contains, an ampicillin gene for selection of bacterial cells containing the plasmid, a multiple cloning site into which DNA sequences can be cloned, the *AttB* sequence which together with ΦC31 integrase effects site specific integration into pseudo *AttP* sites in the human genome, the pUC origin of replication required for plasmid replication in bacteria, the SV40 origin of replication required for plasmid replication in mammalian cells. Neomycin gene facilitates selection of mammalian cells containing the plasmid.

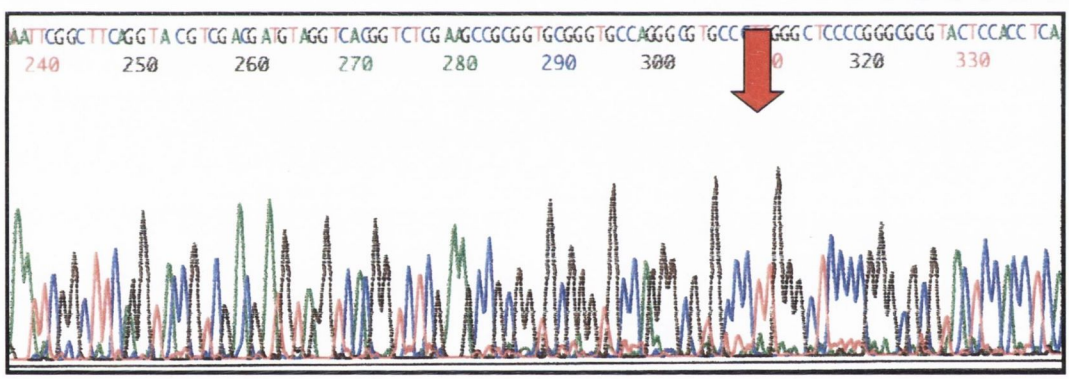


Figure 6.18 Sequence chromatogram generated from pcDNA.AttB plasmid using pcDNA82f sequencing primer. The *AttB* sequence which together with ΦC31 integrase mediates site specific integration into pseudo *attP* sites in the human genome was cloned into the EcoRI site of pcDNA3.1^{-}CMV^{-}. The *AttB* sequence was assessed using the sequencing primer pcDNA82f. The *AttB* DNA fragment inserted in the 5' – 3' sense orientation, the CAA recombination site is indicated by the red arrow.

6.3.5 Bacterial plasmid PCR screening of rescued plasmids.

The HeLa transfection experiment was repeated and three new HeLa cell lines were generated as described in section 6.2.10 and 6.2.11 of this chapter. Plasmid rescue was carried out for each of the 3 cell lines as described above (section 6.3.2). The plasmid rescue appeared to be successful with the growth of Kanamycin resistant *E.Coli* colonies following transformation with the ligated HeLa cell genomic DNA. In this study an initial PCR screen of the plasmids present within the *E.coli* colonies was carried out. The primers used were Att1201 and Att1200, these primers generated a PCR product of 206bp from the *AttB* sequence present in pEGFP1.CMV.AttB. If ΦC31 integrase mediated integration of any of the plasmids contained in the colonies screened had occurred, it would be expected that additional DNA sequences would be present within the *AttB* sequence of the rescued plasmids. If this was the case the PCR product generated from such a plasmid would be expected to be larger than the product generated from an intact, empty *AttB* sequence. A total of 45 colonies were screened using the bacterial PCR technique,15 per stable cell line no single colony generated a PCR product any larger than that generated from the control *AttB* plasmid PCR (Figure 6.19). This suggested that of the 45 plasmids screened none appeared to have any significant amount of additional DNA located within the *AttB* site. This indicated that it was unlikely that any of the rescued plasmids assessed had participated in an integrase mediated integration reaction.

6.3.6 Restriction enzyme digestion of rescued plasmids.

It was thought perhaps the ΦC31 integrase mediated integration events may be rare and additional screening was carried out. A total of 42 minipreps were carried out, 14 minipreps per HeLa cell line, the rescued plasmids were then digested using restriction enzymes Hind III and BamH1 which will excise the *AttB* and CMV promoter from the MCS generating an 852bp fragment in the absence of an integration event. If ΦC31 integrase mediated integration had occurred there should be additional sequence located within the *AttB* site which was retrieved from the genome during the plasmid

rescue, consequently the DNA digest fragment from such colonies should be larger than the fragment isolated from the control *AttB* plasmids.

Twenty one of the plasmids rescued from the genomes of the HeLa cell lines of rescued plasmids generated DNA digest fragment of about 852bp indicating that these plasmids had not been involved in an integrase mediated integration event (Figure 6.20 A and B). Nine of the rescued plasmids generated digest product of 300 – 350bp suggesting that a deletion or rearrangement may have occurred (Figure 6.21, Figure 6.22 A lane 1 and Figure 6.22 B Lane 1). Unfortunately, a digest control of the original plasmid delivered to the HeLa cells pEGFP1.CMV.AttB was not included on any of these three gels. The fact that identical digest results were generated from each of these plasmids suggests that each of the plasmids may possibly have been isolated from the identical location in the genome and could possibly represent a single integration event. This finding may be due to the fact that the original HeLa cell in which the integration event occurred was selected for using G418 and expanded, therefore the genome is present in large numbers in the cell population from which the genomic DNA was rescued

Seven of the rescued plasmids assessed appear to have linearised following digestion with HindIII and BamH1 with no detectable *AttB* digest fragment, this suggested that the *AttB* and CMV sequences some how may possibly have been lost from the plasmid, in addition the backbone of these plasmids appears to vary in size indicating that they may contain genomic DNA sequences, one could suggest that perhaps these plasmids may have been isolated from different sites in the genome, although observation may possibly be due to loading effects (Figure 6.22 A lanes 3, 4, 7 and Figure 6.22 B lanes 4, 6, 7). It is also notable that whilst the absence of any smaller digest fragment in addition to the plasmid backbone is clear in lanes 3, 4 and 7 of figure 6.22 A, the amount of digested plasmid DNA loaded in lanes 4, 6 and 7 and 8 of the gel presented in Figure 6.22 B is less than the aforementioned lanes (lanes 3,4 and 7 of Figure 6.22A), therefore it may be possible that another small DNA fragment is present but undetectable due to the low concentration of the sample analysed on the gel.

Five of the rescued plasmids have generated a DNA digest fragment of in the region of 1000 – 1100bp indicating that they may possibly contain additional sequences within the *AttB* sequence Figure 6.22 A, lanes 5, 6, 8 and Figure 6.22 B, lanes 4, 6, 7, 8.

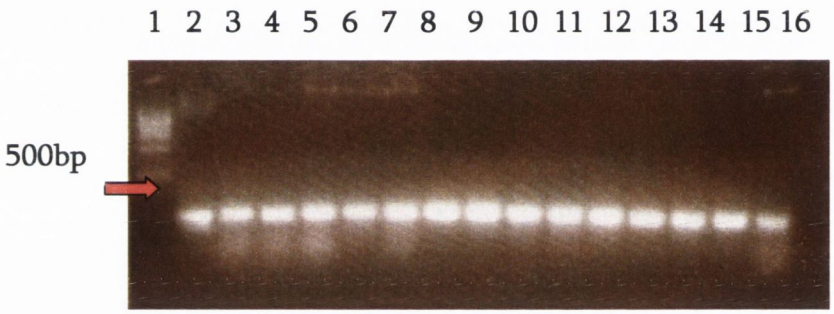


Figure 6.19 Bacterial Plasmid PCR of colonies generated from plasmid rescue from HeLa cells. Bacterial Plasmid PCR of colonies generated from plasmid rescue from HeLa cells was carried out using primers which bind within the AttB site flanking the TTG recombination core: Att1200 and Att1201. The control PCR from pEGFP1.CMV.AttB generates a PCR product of 206bp, plasmids which have been involved in an integrase mediated integration event into pseudo AttP sites in the HeLa cell genome would be expected to generated larger PCR products. **Lane 1:** 100bp ladder, **Lane** 2: control colony PCR, **Lanes 3 – 14:** rescued plasmid colony PCRs all of which generated PCR products of identical size to the control plasmid.

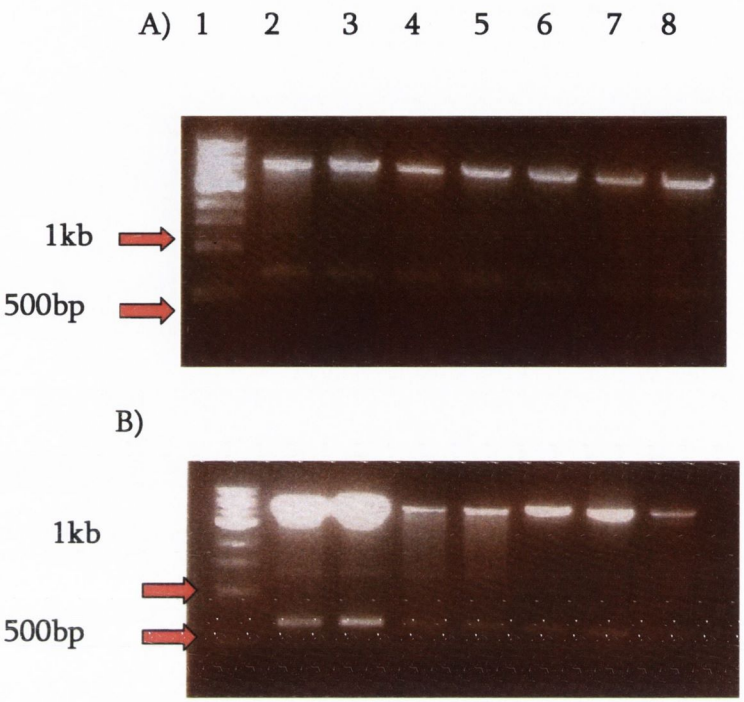


Figure 6.20 Restriction enzyme digest of plasmids rescue from genomes of HeLa into which pEGFP.CMV.AttB has been stably integrated. Restriction digest screen analysis to determine ΦC31 integrase mediated integration into the genome of the HeLa cells had occurred. This should be indicated by the presence of additional DNA sequences located within the AttB sequence of integrated pEGFP.CMV.AttB plasmids. **A)** and **B)** Sample gel of 7 of the 21 rescued plasmids, which generated a digest fragment of 850bp upon digestion with Hind III and BamHI. **Gel A Lane 1**: 1kb DNA ladder, **Lane 2:** pEGFP.CMV.AttB **Lane 3 - 8:** rescued plasmids digested with Hind III and BamHI. **Gel B)** Sample gel of 7 of the 21 rescued plasmids, which generated a digest fragment of 850bp upon digestion with Hind III and BamHI digestion with Hind III and BamHI. **Lane 1**: 1kb DNA ladder, **Lanes 2-8** Rescued plasmids digested with Hind III and BamHI.

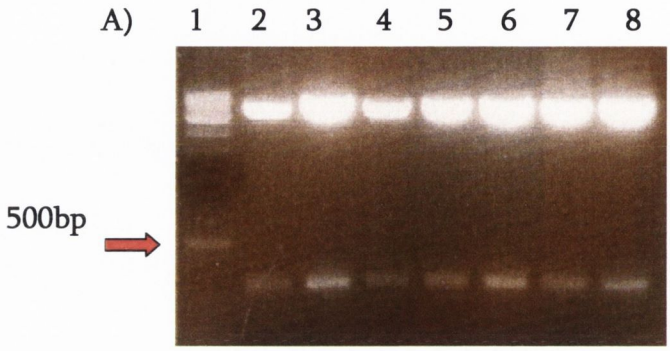


Figure 6.21 Restriction enzyme digest of plasmids rescue from genomes of HeLa into which pEGFP.CMV.AttB has been stably integrated. Restriction digest screen analysis to determine ΦC31 integrase mediated integration into the genome of the HeLa cells had occurred. This should be indicated by the presence of additional DNA sequences located within the *AttB* sequence of integrated pEGFP.CMV.AttB plasmids. **Lane 1**: 1kb DNA ladder, **Lane 2 – Lane 8**: Seven rescued plasmids which generated a digest fragments of in the region of approximately 300 -350bp upon digestion with Hind III and BamHI.

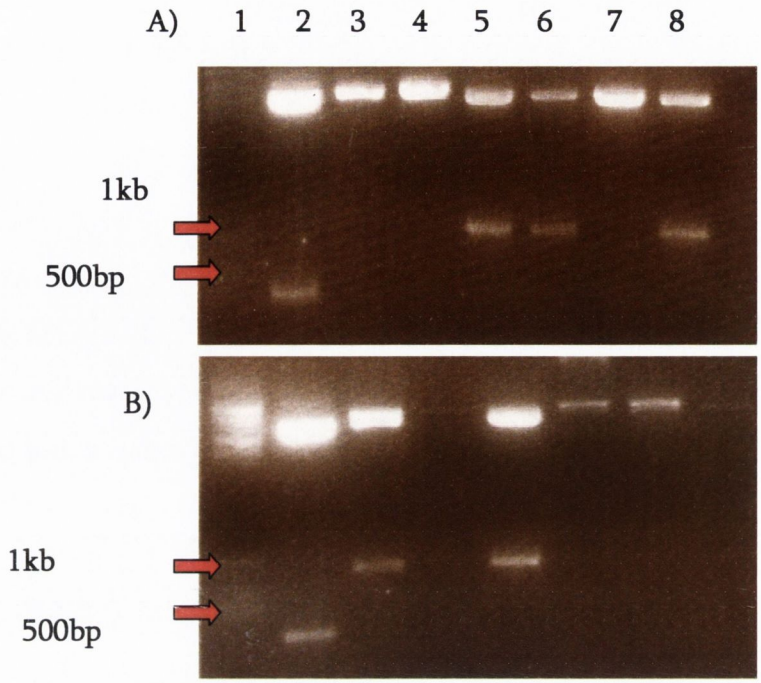


Figure 6.22 Restriction enzyme digest of plasmids rescue from genomes of HeLa into which pEGFP.CMV.AttB has been stably integrated. Rescued plasmids digested with HindIII and BamH1, **Gel A) Lane1:** 1kb DNA ladder, **Lane2:** rescued plasmid digest product of 300bp & backbone, **Lane 3 – 4:** rescued plasmid backbones may differ in size, no additional DNA digest fragment, **Lane 5 – 6:** rescued plasmids backbones of equal size, additional 1kb digest fragment, **Lane 7:** plasmid backbone, no additional digest fragment(s), **Lane 8:** rescued plasmid backbone and additional 1kb digest fragment. **Gel B) Lane1:** 1kb DNA ladder, **Lane2:** rescued plasmid digest product of 300 - 350bp & backbone, **Lane 3:** rescued plasmids and additional 1kb digest fragment, **Lane 4**rescued plasmid backbone detectable, due to low concentration of sample loaded no conclusion can be made with regard to the presence of additional smaller DNA digest fragment **Lane 5:** rescued plasmids and additional 1kb digest fragment, **Lane 6 and 7:** plasmid backbone and no apparent digest fragment, plasmid backbone of these rescued plasmids appear to be larger than those in lanes 2, 3 and 5. **Lane 8:** rescued plasmid backbone detectable, due to low concentration of sample loaded no conclusion can be made with regard to the presence of additional smaller DNA digest fragments.

6.3.7 Sequencing of plasmids selected based upon restriction digest analysis.

A selection of the rescued plasmids were sequenced, four of the plasmids did not generate any sequence data using either the forward primer Att1200 or reverse primer Att1201, as the positive control sequencing reactions worked it indicates that perhaps the *AttB* site was not present as this is the primer binding site location. Twelve of the plasmids had no evidence of an *AttB* mediated integration event as the *AttB* sequences were fully intact (Figure 6.23). Four of the plasmids only generated sequence when the primer Att1201 was used, suggesting that perhaps that half of the *AttB* site within which the primer Att1200 binds was no longer present within the rescued plasmid. There was a DNA insertion in the *AttB* site, when the sequence generated from each of the four plasmids was analysed and aligned using ClustalW it appeared that the sequence was identical (Figure 6.25) and when analysed using BLAST the sequence appeared to be that of the plasmid pEGFP1.CMV.AttB (Figure 6.25 and 6.26). The inserted sequence began in the SV40 early promoter which is located 2.2kb downstream of the *AttB* sequence. It appears that either a deletion event occurred at the time of integration into the genome or a rearrangement occurred during the plasmid rescue process. Similar uninformative recombination events have been reported by Chalberg et al when the plasmid rescue technique was employed (Chalberg, et al., 2005) and 8 deletion events of greater than 1kb have also been reported (Chalberg, et al., 2006).

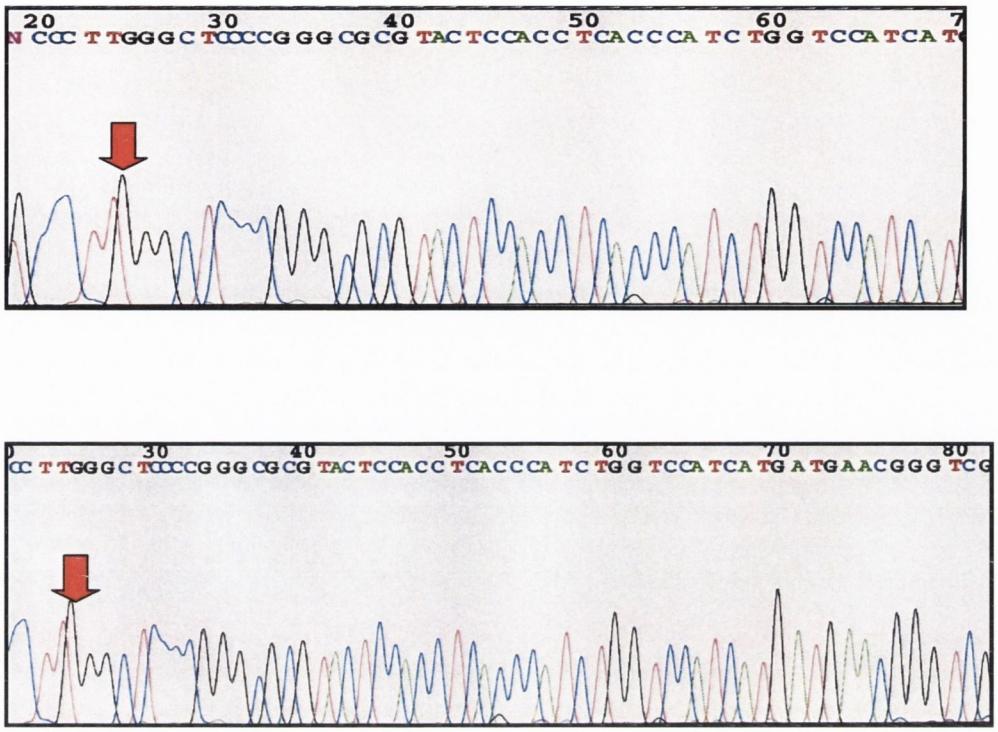


Figure 6.23. Sequence of AttB site generated from sequencing plasmids rescued from genome of HeLa cells stably expressing EGFP1.AttB. In order to determine if any of the **EGFP1.AttB** plasmids had been involved in ΦC31 integrase mediated integration into the human genome plasmids rescued from the genome of HeLa cells stably expressing EGFP1.AttB were sequenced using the primer Att1201 which binds within the *AttB* site. None of the plasmids rescued from HeLa cells stably expressing EGFP1.AttB were found to have additional sequences located within the AttB site of the plasmid. The *AttB* site is indicated by the red arrow.



Figure 6.24 ClustalW alignment of DNA sequence detected in each of the four independent plasmids rescued from the genomes of the HeLa cells

Figure 6.24 ClustalW alignment of DNA sequence detected in each of the four independent plasmids rescued from the genomes of the HeLa cells. HeLa cells were transfected with pEGFP1.CMV.AttB and integrase pCMVInt. Plasmid HeLa cells in which the pEGFP1.CMV.AttB plasmid had integrated were selected from using G418 and selection of stable cell lines generated. In order to determine if the plasmids integration was ransom or integrase mediated it was necessary to retrieve the plasmids from their genomic locations within the HeLa cells. This was achieved using the plasmid rescue technique in which HeLa cell genomic DNA is extracted and digested with two enzymes with compatible ends that do not cut within the plasmid. Hence the restriction enzymes only cut within genomic DNA. The genomic DNA fragments are ligated, transformed into *E.Coli* and plated on kanamycin selective agar plates. The genomic DNA fragments that contained an intact plasmid were able to grow on the selective agar plate as the kanamycin resistance gene is expressed from the plasmid. Four of colonies containing rescued plasmids were sequences using Att1200 each appeared to have a DNA insertion within the *AttB* site. BLAST analysis revealed that the sequences were similar from each of the four plasmids. ClustalW alignment of the four sequences revealed that all plasmids have an identical DNA sequence located within the *AttB* site.



Figure 6.25 Alignment of consensus HeLa *AttB* DNA sequence insertion with the vector EGFP1 (U55761). Alignment matched is located within the SV40 early promoter and runs into the kanamycin gene.

6.3.8 Generation of pEGFP1.AttB.Col1a1 promoter plasmid.

The ultimate aim of the study described in this chapter is to assess the suitability of the ΦC31 Integrase system as a non viral delivery tool with which nucleic acid based therapeutics targeted for the treatment of Osteogenesis Imperfecta may be delivered. As described the target cells for the potential OI therapeutics described in this thesis are Mesenchymal Stem Cells, therefore it is essential to validate the efficacy of the ΦC31 integrase system in these cells. Whilst site specific integration into pseudo *AttP* sites in the genome would be an excellent means of delivery, if the delivered therapeutic is expressed in an unregulated manner, i.e. at inappropriate levels with no level of tissue specificity the therapeutic benefit would most likely be lost to the negative effects of the perturbed expression profile of the delivered nucleic acids.

The explorative approach undertaken to evaluate the activity of the ΦC31 integrase system in MSCs was to generate an *AttB* containing construct in which expression of a reporter gene was driven by the Col1a1 promoter. There has been a variety of conflicting reports in the literature with regards to how much 5' upstream Col1a1 sequence is necessary to drive tissue specific Col1a1 expression at close to endogenous levels. The human Col1a1 minigene is a truncated form of the human Col1a1 gene in which exons 7 – 45 have been removed (Olsen, et al., 1991), this gene is known to express at close to endogenous levels in a tissue specific manner *in vivo* (Sokolov, 1993). It appears that the Col1a1 minigene contained close to the optimal 5' sequences incorporating the Col1a1 promoter, therefore it was decided to include as much of this promoter sequence in the *AttB* plasmid design. A total of 4017bp of the Col1a1 promoter was cloned into pEGFP1.AttB, generating a construct in which eGFP expression was under the control of the Col1a1 promoter. Consequently, it should be expected that the eGFP expression profile of this plasmid would reflect that of the Col1a1 gene both in terms of tissue specificity and levels of expression. Given that the target cell type human MSCs population being used in this study was demonstrated to be mutlipotent (section 6.3.1) in our hands it was of interest to establish the functionality of the ΦC31 integrase system in MSCs and to establish if the cells would

retain their multipotency following transfection with the ΦC31 integrase system plasmids. Therefore once a stable MSCs population in which the pPEGFP1.AttB.Col1a1 promoter plasmid was stably integrated was generated the next set would be differentiate the stable MSC population *in vitro* in an identical manner as the MSCs were differentiated in the validation study described in section 6.2.1 of this chapter of the thesis. The MSCs which differentiated into osteoblasts would be expected to express high levels of eGFP as they express high levels of Col1a1, in contrast eGFP expression in adipocytes would be expected to be low as Col1a1 is expressed at very low levels in adipocytes.

Insertion of the Col1a1 promoter fragment into pEGFP1.AttB involves a blunt ended ligation consequently the fragment could insert in either the sense or antisense orientation. A selection of minipreps were screened using the HindIII:BamH1 digests described in section 6.2.3 were carried out to identify a plasmid which had not undergone sequence rearrangement and contained the promoter in the sense orientation (Figure 6.27). The positive colonies were then sequenced using pEGFPf and KpnF sequencing primers to ensure that there were no sequence errors at the cloning junctions, a single miniprep with intact sequence was then selected.

pEGFP1.AttB.Col1a1 promoter plasmid

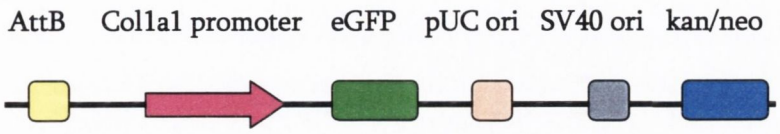


Figure 6.26 pEGFP1.AttB.Col1a1 promoter plasmid. Schematic of pEGFP1.AttB.Col1a1 promoter plasmid, this plasmid contains the *AttB* sequence which together with ΦC31 integrase effects site specific integration into *AttP* sites in the human genome, an eGFP reporter gene, the pUC origin of replication required for plasmid replication in bacteria and the SV40origin of replication required for plasmid replication in mammalian cells. The Kanamycin /Neomycin gene facilitates selection of bacterial or mammalian cells which contain and express the plasmid.

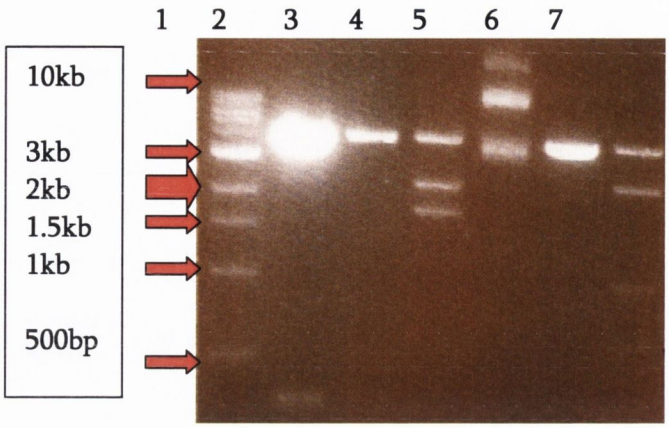


Figure 6.27 pEGFP1.AttB.Col1a1 promoter digest with HindIII:BamH1. A selection of minipreps were digested with HindIII:BamH1 to identify which plasmids contained the Col1a1 promoter sequence in the sense orientation. The DNA digest fragments were run on a 1% agarose gel. **Lane 1**:1kb ladder, **Lane 2:** pEGFP1.AttB digest control, **Lane 3**: miniprep in which AttB appears to be missing, **Lane 4:** miniprep with promoter in sense orientation (fragment sizes 1829bp, 2252bp, 4149 bp), **Lane 5:** rearranged plasmid **Lane 6**: miniprep in which AttB appears to be missing, **Lane 6**: miniprep which promoter sequence appears to have rearranged (fragment sizes in the region of 700bp, 900bp, 2500bp, 4150 bp),

6.3.9 Nucleofection of MSC with pEGFP1.AttB.Col1a1 promoter and pCMVInt

pEGFP1.AttB.Col1a1 promoter and pCMVInt plasmids were delivered to MSC via nucleofection. Plasmid delivery MSCs was efficient with in the region of 60% - 70% of cells expressing GFP 48 hours after delivery (Figure 6.28). 48 hours after nucleofection selection of the nucleofected MSCs which had integrated the pEGFP1.AttB.Col1a1 promoter plasmid was initiated with the addition of 250μg per ml G418 selection media. Despite the high nucleofection efficiencies, a low number of stable cell colonies were generated. The cells were cultured for another 7-10 days but at this stage the MSCs were in poor condition stopped dividing and were senescent.

The experiment was repeated, the nucleofection was successful, 60% to 70% of MSCs expressed eGFP under the control of the Col1a1 promoter 48 hours after nucleofection (Figure 6.28). The cells in which the plasmid had integrated were selected for using 250μg per ml G418 selection media. Similarly, in this experiment very few stable cell colonies were generated. In an attempt to expand these cells the G418 medium was replaced at this point with normal medium. The cells were dividing poorly and at this stage it was decided to access the cells microscopically. The stable MSC colonies generated in each of the 4 wells were pooled and plated onto cover slips to evaluate if the pEGFP1.AttB.Col1a1 promoter plasmid was still being expressed. Many of the MSCs did not survive, those that did were assessed for GFP expression, it was found that in comparison to the transfection efficiency control cells (Figure 6.28) that the levels of GFP fluorescence were significantly diminished as can been seen in Figures 6.29 – Figure 6.31. Notably, many of the cells do not appear to be expressing any eGFP. This suggests that expression is reduced from the pEGFP1.AttB.Col1a1 promoter plasmid this may be a protective response from the cells as high levels GFP expression is known to be toxic, or more likely these cells may be shutting down the expression machinery of the cells as they are senescent. Unfortunately, it was not possible to assess the integration sites of the plasmids as sufficient cells from which to extract DNA were not generated.

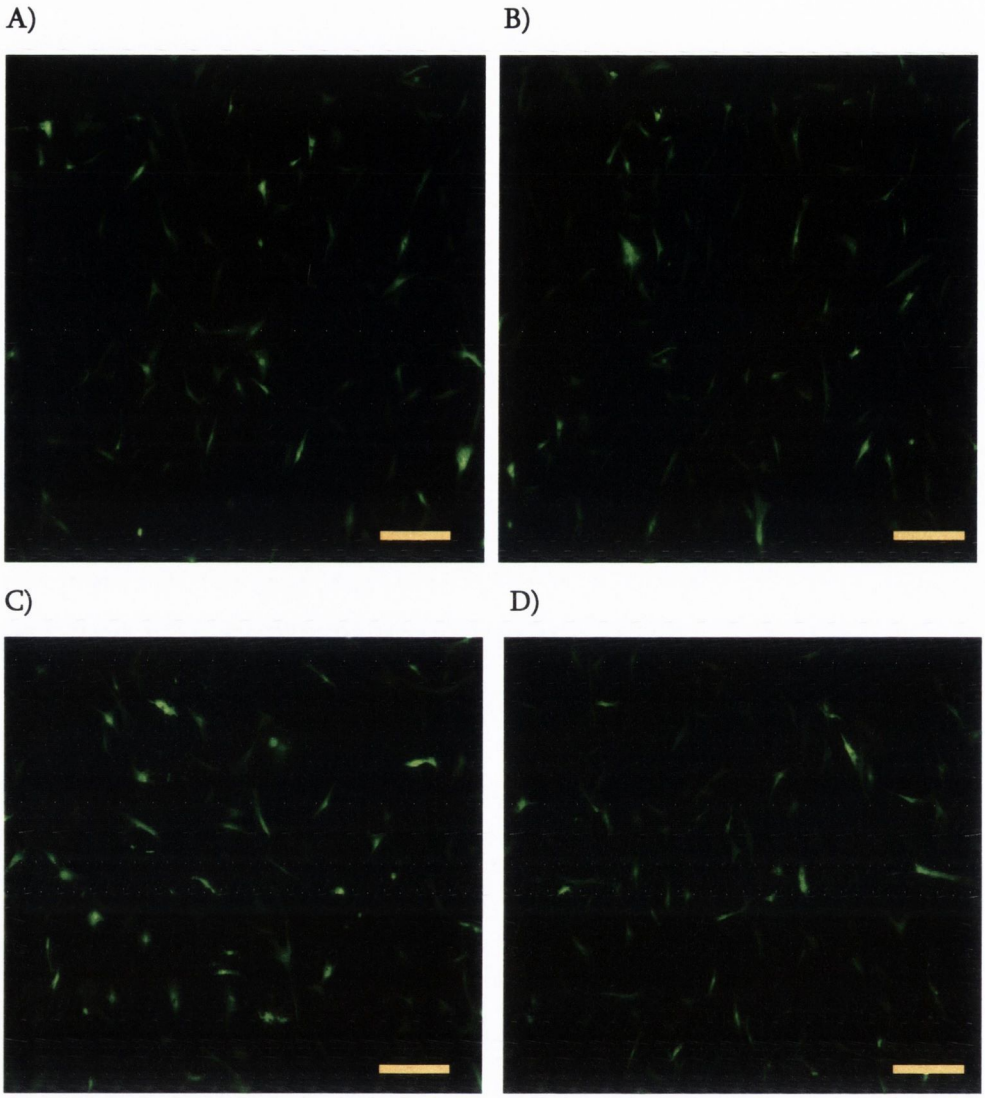


Figure 6.28 Assessment of Nucleofection efficiencies and eGFP from Col1a1 expression pEGFP1.AttB.Col1a1 promoter plasmid in MSCs. GFP expression in MSCs 2 day post nucleofection with **pEGFP1.AttB.Col1a1 promoter A & B)** MSC nucleofection experiment 1, **C & D)** MSC nucleofection experiment 2. Scale bar 20μm.

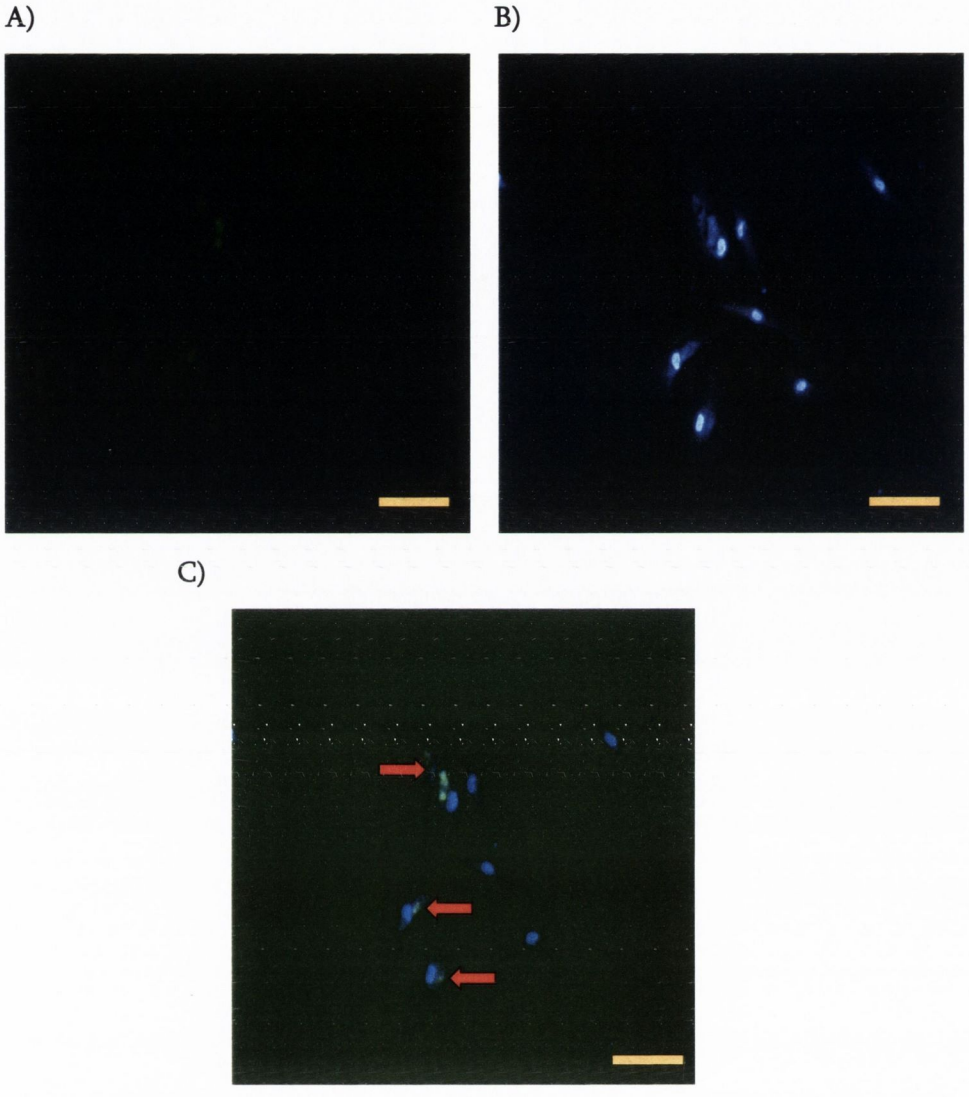


Figure 6.29. Fluorescence microscope analysis of GFP expression in MSCs MSCs were nucleofected with Col1a1-EGFP-*AttB*, pictures were taken 2 – 3 weeks subsequent to nucleofection A) GFP expressing cells B) nuclei stained with DAPI, C) overlay of both A and B, low levels of GFP expression are visible. The nuclei are stained blue with DAPI. Scale bar 40μm.

A)



B)



C)

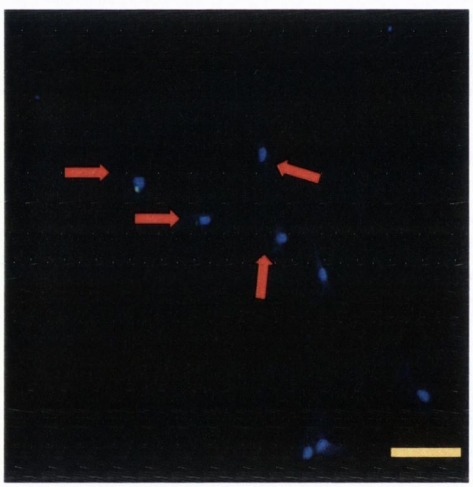
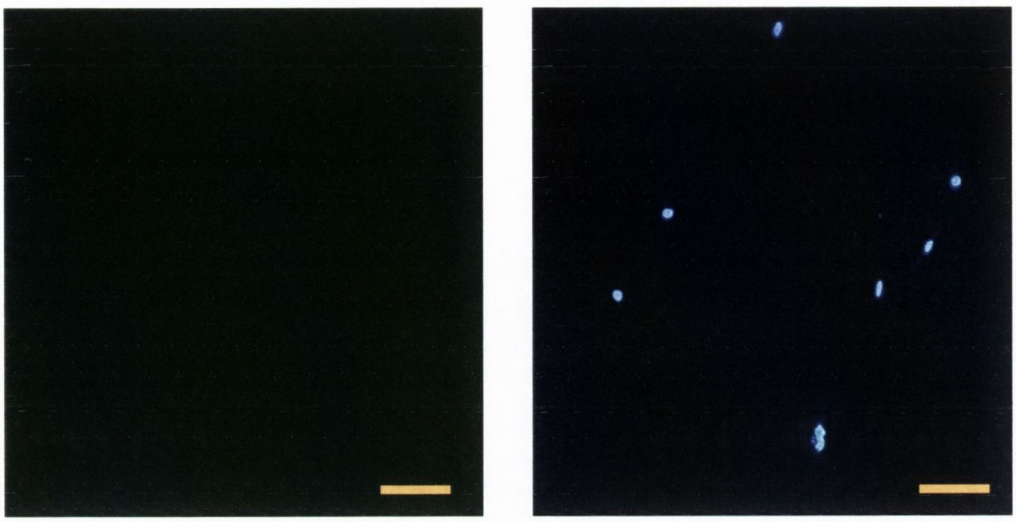


Figure 6.30. Fluorescence microscope analysis of GFP expression in MSCs MSCs were nucleofected with Col1a1-EGFP-*AttB*, pictures were taken 2 – 3 weeks subsequent to nucleofection A) GFP expressing cells B) nuclei stained with DAPI, C) overlay of both A and B,low levels of GFP expression are visible. The nuclei are stained blue with DAPI. Scale bar 40μm.



A) B)

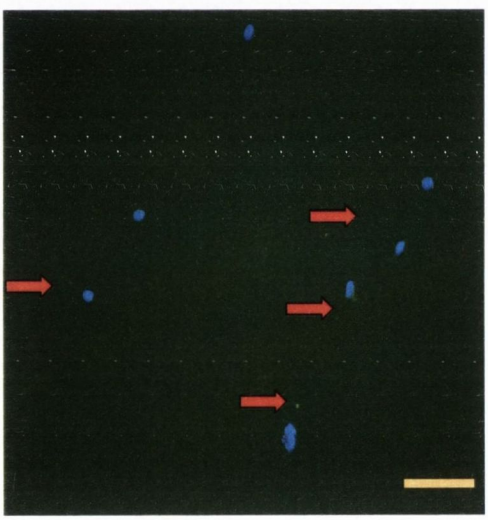


Figure 6.31. Fluorescence microscope analysis of GFP expression in MSCs MSCs were nucleofected with Col1a1-EGFP-*AttB*, pictures were taken 2 – 3 weeks subsequent to nucleofection. A) GFP expressing cells B) nuclei stained with DAPI, C) overlay of both A and B, low levels of GFP expression are visible. The nuclei are stained blue with DAPI. Scale bar 40μm.

6.3.10 Cloning of shRNA.AttB plasmids

In parallel to the MSC study described in section 6.3.9 another MSCs study was carried out in which the potential use of the ΦC31 integrase system for the delivery of Col1a1 targeting shRNAs was explored. As mentioned in the introduction to this chapter many disorders will require long term expression of therapeutic, in particular in the case of Mendelian disorders. In relation to Osteogenesis Imperfecta, if the Col1a1 targeting shRNAs described and evaluated in chapter 4 and chapter 5 of this thesis were to be incorporated into a therapeutic approach it would be important that life long sustained expression of the shRNAs in MSCs was achievable. In reality, to achieve this, it is likely that plasmids may have to be integrated into a non pathogenic location within the genome of the MSCs. In addition the potential for the occurrence of *de novo* methylation of the plasmid and histone modification of the genomic region in which integration occurred would have to be minimised to ensure that expression levels were sustained in successive cell generations.

It is claimed that the ΦC31 integrase system can maximise the possibilities of these requirements being fulfilled, as it effects site specific integration into *AttP* sites within the genome and to date none of the *AttP* sites identified have been reported to be associated with the risk of insertional mutagenesis. High levels of sustained gene expression are said to be achievable from integrase integrated plasmids. In order to explore this a series of shRNA expression constructs were generated in which the H1-shRNA expression cassette was excised from shCol2, shCol4 and shGFP, each was cloned into pcDNA.AttB generating shCol2.AttB, shCol4.AttB and shGFP.AttB (Figures 6.32 – 6.34). The presence of the G418 resistance gene on pcDNA.AttB facilitates the generation of cell lines into which the shRNA.AttB plamids shCol2.AttB, shCol4.AttB and shGFP.AttB had integrated. The ability of shCol2.AttB and shCol4.AttB to effect sustained Col1a1 suppression was investigated in MSCs.

ShRNA.AttB plasmid.

Amp AttB H1promoter shRNA pUC Ori SV40 ori Neo



Figure 6.32 Schematic of shRNA.AttB plasmid. This plasmid was generated by cloning the H1promoter-shCol sequence isolated from shCol2, shCol4 and shGFP into pcDNA.AttB plasmid. The shRNA plasmid contained, an ampicillin gene for selection of bacterial cells containing plasmid, the *AttB* sequence which together with ΦC31 integrase effects site specific integration into pseudo *AttP* sites in the human genome, the polymerase III H1 promoters which drives shRNA expression, the pUC origin of replication required for plasmid replication in bacteria and the SV40 origin of replication required for plasmid replication in mammalian cells. The Neomycin gene facilitates selection of mammalian cells containing the plasmid.

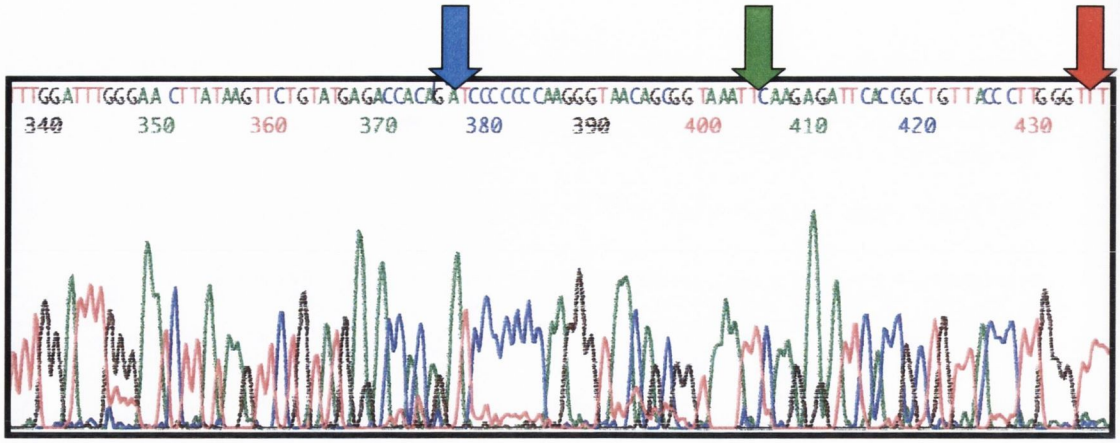


Figure 6.33 shCol2.AttB sequence chromatogram. The H1-shCol2 fragment was cloned into the NotI and XhoI restriction enzyme sites in the multiple cloning site of pcDNA.AttB. The presence of shCol2 with correct sequence was confirmed using the primer pcDNA82F. The start of the shCol2 oligonucleotide (GATC) is indicated by the blue arrow, the 9 nucleotide hairpin sequence (ttcaagaga) is indicated by the green arrow and the H1 promoter transcription termination signal (TTTTT) in red arrow. (The final TT was present in the sequence but is not illustrated in this particular chromatogram).

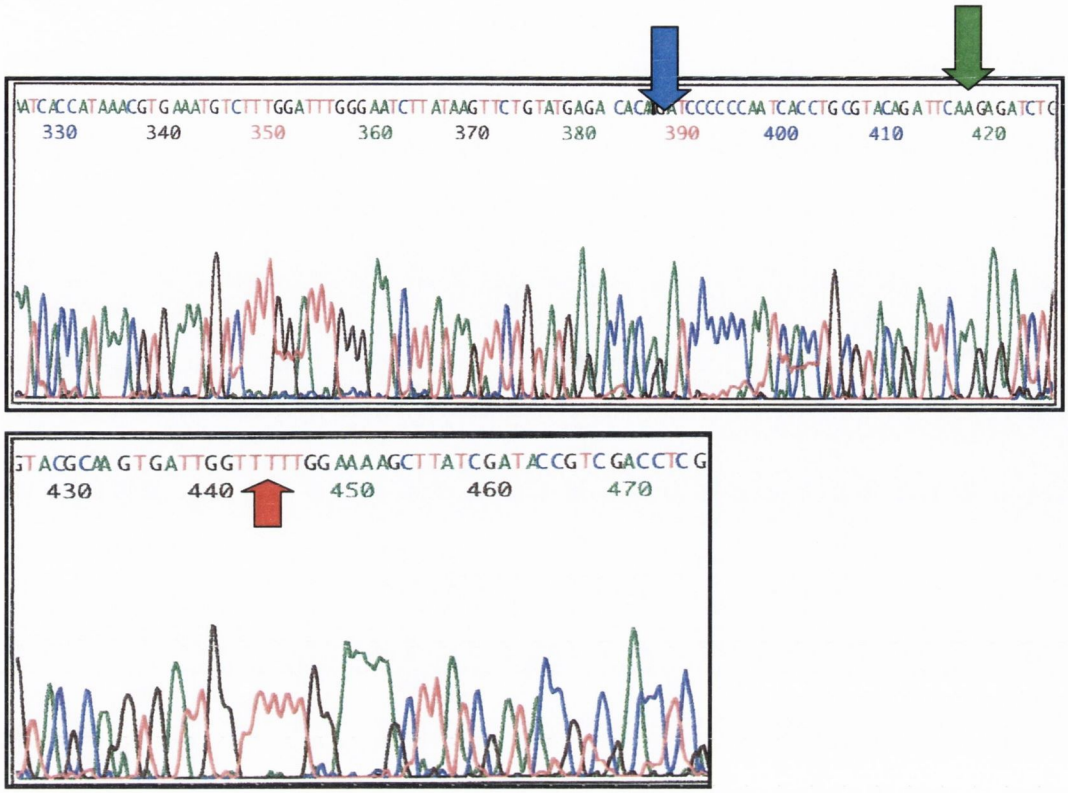


Figure 6.34 shCol4.AttB sequence chromatogram The H1-shCol4 fragment was cloned into the NotI and XhoI restriction enzyme sites in the multiple cloning site of pcDNA.AttB. The presence of shCol4 with correct sequence was confirmed using the primer pcDNA82F. The start of the shCol4 oligonucletide (GATC) is indicated by the blue arrow, the 9 nucleotide hairpin sequence (ttcaagaga) is indicated by the green arrow and the H1 promoter transcription termination signal (TTTTT) in red arrow.

6.3.11 Nucleofection of MSC with shRNA.AttB and pCMVInt plasmids.

The ultimate goal in terms of application of the ΦC31 integrase system for Osteogenesis Imperfecta is its use in the delivery of Col1a1 targeting shRNA expressing plasmids to MSCs. To initially explore the utility of the ΦC31 integrase system as a delivery option for MSCs a selection of shRNA plasmids were generated which contained the *AttB* sequences, shCol2.AttB, shCol4.AttB and shGFP.AttB.

The shRNA-AttB plasmids: shCol2.AttB, shCol4.AttB and the control plasmid shGFP.AttB were each co-nucleofected into MSCs along with the pCMVInt plasmid as described in section 6.2.16. Each nucleofection was carried out in triplicate; subsequently the MSCs underwent 250μg per ml G418 selection 48 hours after transfection. The control cells had completely died by day 8, at this stage the cells which had received shCol2.AttB, shCol4.AttB or shGFP.AttB with the integrase plasmid pCMVInt had generated a large number of colonies per well. The stable colonies were grown in the wells until 70% confluent, at this stage the triplicate wells were trypsinised, pooled and replated into a T75 flask. This generated a single cell line per shRNA, shCol2.AttB, shCol4.AttB or shGFP.AttB. The individual MSC cell lines were expanded and using these cells a 28 day suppression study was carried out. RNA was extracted then extracted from the cells at days 7, 14, 21 and 28 and the levels of Col1a1 expression in each cell line assessed using real time RT-PCR and the relative quantification method.

Col1a1 expression levels in each of the cell lines stably expressing each of the shRNA.AttB constructs shCol2.AttB, shCol4.AttB and the control plasmid shGFP.AttB were evaluated at each of the 4 time points of the study. The Col1a1 levels were assessed using real time RT-PCR and the relative quantification method using GAPDH as the housekeeping gene as described in chapters 3, 4 and 5 of this thesis. Briefly, the Col1a1 expression data in each sample was first normalised, this was achieved by dividing the Col1a1 expression value in a given sample by the GAPDH value in the same sample. The relative levels of Col1a1 expression in each sample was then

determined by comparing the normalised Col1a1 expression levels in MSCs stably expressing shCol2.AttB or shCol4.AttB to the normalised Col1a1 expression values in MSCs stably expressing shGFP.AttB. Both shCol2.AttB and shCol4.AttB plasmids were highly efficient at reducing Col1a1 expression levels which was sustained over the 4 week period analysed. Col1a1 expression levels in MSCs stably expressing shCol2 were reduced to 20.63% (± 11.69), 6.32 % (±8.51), 12.50% (±8.83) and 3.11% (±1.25) of the Col1a1 expression levels in MSCs stably expressing shGFP.AttB at days 7, 14, 21 and 28 respectively (Table 6.1). Similarly, Col1a1 expression levels in MSCs stably expressing shCol4.AttB were found to be reduced to 18.18% (±8.73), 10.31% (±7.79), 2.60% (±0.62), and 3.30% (±3.01), of the levels of Col1a1 expression in MSCs stably expressing shGFP.AttB at days 7, 14, 21 and 28 respectively (Table 6.1 and Figure 6.28). The data was evaluated using GraphPad Instat3 with which the mean and standard deviation of each data set was determined. Additionally, the statistical significance of the reduced levels of Col1a1 expression was determined using a non parametric ANOVA in which the Cola1 expression levels in MSCs stably expressing shGFP.AttB, shCol2.AttB or shCol4 were compared, it was found that the expression levels differed significantly with $p<0.0001$. Therefore, it appears that the Col1a1 targeting shRNA plasmids which had integrated into the genome of the MSCs were successfully reducing Col1a1 expression levels. .

	week 1	week 2	week 3	week 4
shGFP.AttB	100%	100%	100%	100%
StdDev	19.3	19.3	19.3	19.3
shCol2.AttB	20.63 %	6.32%	12.5%	3.11%
StdDev	11.69	8.51	8.83	1.25
ANOVA	p<0.0001	p<0.0001	p<0.0001	p<0.0001
shCol4.AttB	18.18%	10.31%	2.6%	3.30%
StdDev	8.73	7.79	0.62	3.01
ANOVA	p<0.0001	p<0.0001	p<0.0001	p<0.0001

Table 6.1 Col1a1 expression levels in MSCs nucleofected with shCol2.AttB or shCol4.AttB and pCMVInt over a period of four weeks. MSCs were co-nucleofected with either shCol2.AttB, shCol4.AttB or the control plasmid shGFP.AttB and the integrase plasmid pCMVInt. MSCs in which the plasmids had integrated were G418 selected. Col1a1 mRNA levels were assessed using real time RT-PCR, the housekeeping gene was GAPDH. Col1a1 expression levels in MSCs stably expressing shCol2.AttB or shCol4.AttB were compared to Col1a1 expression levels in MSCs stably expressing shGFP.AttB in which Col1a1 expression was set at 100%. Col1a1 expression levels MSCs stably expressing shCol2 (blue bars) was at **day 7**: 20.63%, **day 14**:6.23%, **day 21**:12.45%, **day 28**:3.11%. Expression levels in cells nucleofected with shCol4 (orange bar) were at **day 7**: 18.18%, **day 14**:10.31%, **day 21**:2.6%, **day 28**:3.3 %. All data was assessed using GraphPad InStat3 with all results extremely significant p<0.0001.

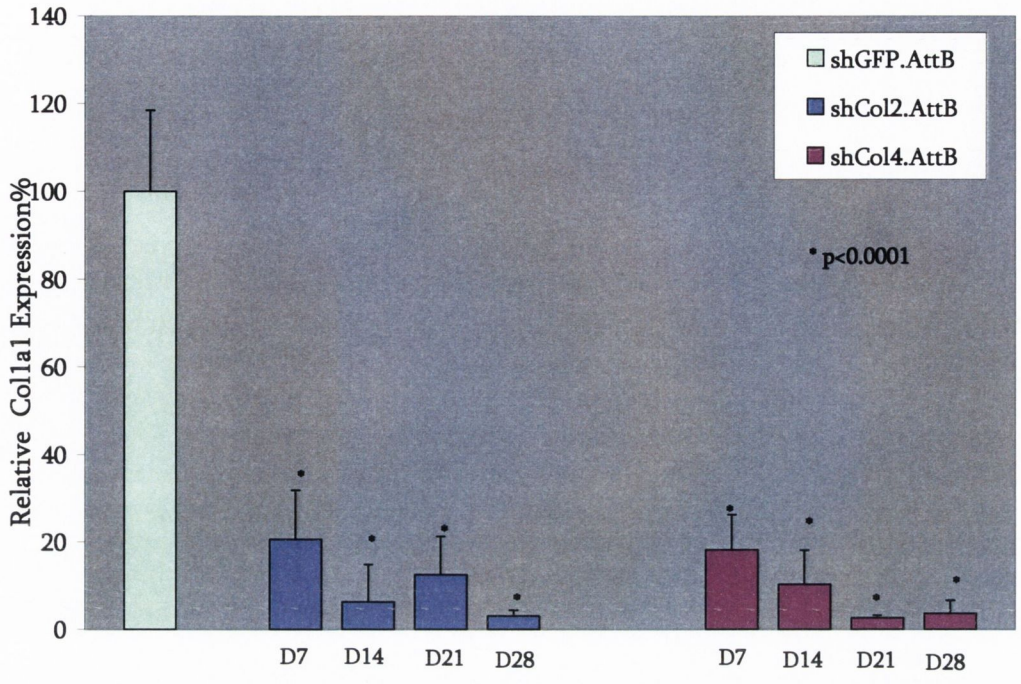


Figure 6.35 Col1a1 expression levels in MSCs nucleofected with shCol2.AttB or shCol4.AttB and pCMVInt over a period of four weeks. MSCs were co-nucleofected with either shCol2.AttB, shCol4.AttB or the control plasmid shGFP.AttB and the integrase plasmid pCMVInt. MSCs in which the plasmids had integrated were G418 selected. Col1a1 mRNA levels were assessed using real time RT-PCR, the housekeeping gene was GAPDH. Col1a1 expression levels in MSCs stably expressing shCol2.AttB or shCol4.AttB were compared to Col1a1 expression levels in MSCs stably expressing shGFP.AttB. in which Col1a1 expression was set at 100%. Col1a1 expression levels MSCs stably expressing shCol2 (blue bars) was at **day 7**: 20.63 %, **day 14**:6.32%, **day 21**:12.5 %, **day 28**:3.11%. Expression levels in cells nucleofected with shCol4 (purple bars) were at **day 7**: 18.18%, **day 14**:10.31%, **day 21**:2.6%, **day 28**:3.3%. All data was assessed using GraphPad InStat3 with all results extremely significant $p<0.0001$.

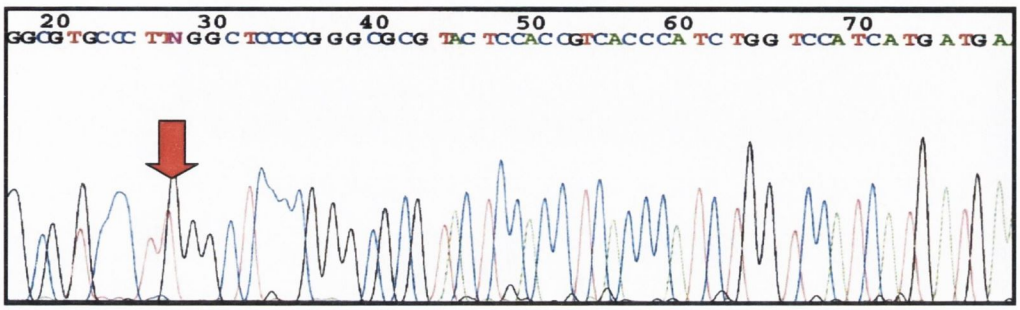
6.3.12 Evaluation of shRNA.AttB plasmid integration in MSCs.

In order to determine if the plasmids which had integrated into the MSCs had integrated randomly or a result of a ΦC31 Integrase mediated event plasmid rescue was carried out as described in section 6.2.10 of this chapter. The plasmid rescue technique worked efficiently with the generation of Ampicillin resistant *E.coli* colonies.10 colonies were selected for miniprep and subsequently sequenced using the primers AttB1200 and Att1201. Each of the plasmids generated sequence data from both primers, the *AttB* sites of each of the10 plasmids was intact with no additional sequences present indicating that none of the plasmids had been involved in an ΦC 31 integrase mediated integration event (Figure 6.35). Given the negative results from the plasmid rescue it was decided to explore if the Integrase plasmid obtained from Professor M.Calos of Stanford University was indeed functional and expressing the integrase gene.

6.3.13 Evaluation of ΦC31 integrase expression in HeLa cells.

In principle prior to embarking on the ΦC31 integrase studies the first preliminary study should have been to validate the functionality of the ΦC31 integrase plasmid to ensure it was expressing the integrase enzyme. To this end HeLa cells were transfected with the ΦC31integrase plasmid and expression determined using real time RT-PCR. The RT-PCR analysis revealed that the integrase was expressing efficiently in HeLa cells that received the plasmid in contrast to untransfected HeLa cells which did not generate a PCR product (Figure 6.37).

A)



B)

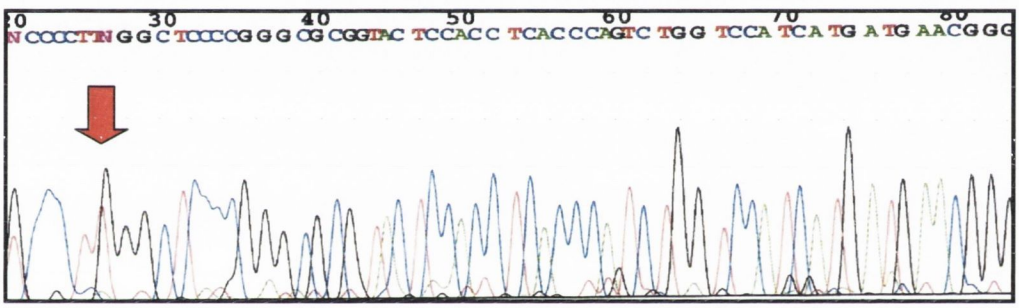


Figure 6.36. Sequence of AttB site generated from sequencing plasmids rescued from genome of MSC stably expressing shCol2 or shCol4. In order to determine if any of the shRNA.AttB plasmids had been involved in ΦC31 integrase mediated integration into the human genome plasmids rescued from the genome of MSCs stably expressing shCol2 and shCol4 were sequenced using the primer Att1200 which binds within the *AttB* site, upstream and downstream of the TTG recombination site. None of the plasmids rescued from MSCs stably expressing shCol2 or shCol4 were found to have additional sequences located within the AttB site of the plasmid. Sample chromatogram from plasmid rescued from MSCs expressing **A)** shCol2 or **B)** shCol4. The red arrow indicated the intact TTG recombination core sequence.

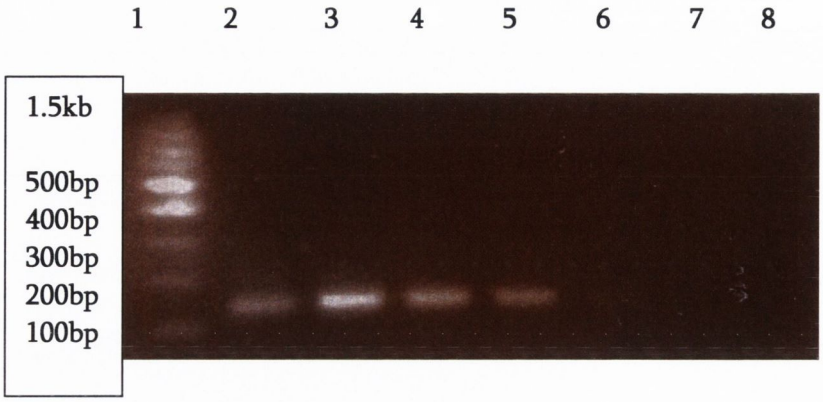


Figure 6.37 Evaluation of ΦC31 Integrase expression in HeLa cells. HeLa cells were transfected with the ΦC31 Integrase plasmid alone in order to determine if the plasmid was expressing the required integrase enzyme. RNA was extracted 24 hours after transfection and integrase expression assessed using RT-PCR. Lane 1:100bp DNA ladder, Lanes: 2 – 5 integrase RT-PCR product generated from four individual transfections of HeLa cells with the ΦC31 Integrase plasmid, Lanes 6 and 7:control untransfected HeLa cells, no integrase RT-PCR product present. Expected PCR product of 153bp.

6.4 Discussion

The research described in this chapter of the thesis was focused on a number of key objectives. Firstly, the mutlipotency of the MSC population in use was assessed. When utilizing MSCs as a therapeutic, it is essential to ensure that the population of cells is multipotent to ensure optimal therapeutic benefit is derived from these cells. For example, in the case of OI, if a MSC population was considered to be multipotent but in fact was a more committed cell line which could only progress, for example, to an adipogenic lineage, the Col1a1 targeting shRNAs described in Chapters 4, 5 and 6 of this thesis, if delivered to such cells would have little therapeutic affect as they would never form osteoblasts, the bone producing cell of the body. Whilst the ability of MSCs to differentiate into a variety of distinct lineages has been demonstrated, (De Bari, *et al.*, 2003; Long, *et al.*, 2005; Morigi, *et al.*, 2004; Pittenger, *et al.*, 1999; Rojas, *et al.*, 2005) it was important to determine if the human MSC population used in the various studies carried out for this doctoral thesis were multipotent and if they could be differentiated under the culturing conditions in our laboratory.

A four week differentiation experiment was set up in which the MSCs were promoted to differentiate down either the osteogenic or adipogenic pathways using standard differentiation protocols (section 6.2.1). The assessment of the differentiation profile of the treated cells was carried out at day 7, 14, 21 and 28. The MSC readily differentiated, there was the characteristic lag phase in which the cells were beginning to proceed through the osteogenic and adipogenic lineages, at this stage there was very little evidence histochemically of the emerging phenotypes, however the cells were beginning to look morphologically different from the control untreated MSCs. By day 14 it was clearly evident that many of the cells could no longer be classified as MSCs with significant Alkaline phosphatase and Oil red O staining in the MSCs treated with osteogenic and adipogenic inductive medium respectively (Figure6.2 – 6.10). The number of cells committed to each lineage increased steadily to day 28 at which stage at least 60% of the cells treated with the adipogenic medium became adipocytes and 65%

of those treated with osteoblastic medium had become adipocytes (Figures 6.2-6.10). In addition, at each of the four time points there was no evidence of osteogenic or adipogenic differentiation in the control untreated MSCs. It is of interest that the control cells in figure 6.6 A and B have adopted a frequently observed morphological alteration, some times referred to as a fried egg shape cell, it is unknown why this occurs. Nevertheless, it appears that the population of MSCs used in this differentiation study is at least bipotent as it can be induced to progress though at least two distinct lineages. In addition, the differentiation appears to be controlled and non-spontaneous as there was no evidence of differentiation in the control cells, when assessed for osteogenic or adipogenic differentiation (Figures 6.3 – Figure 6.10). Therefore, theoretically, it appears that this population of MSCs may be ideally suited for application in OI targeted therapeutic and should in principle readily differentiate into osteoblasts *in vivo* under the influence of the appropriate osteoinductive stimuli.

The second objective addressed in this section of the research undertaken during the course of my studies was the exploration a non viral vector system as a potential tool for delivery of nucleotide based therapeutics to mesenchymal stem cells, with the ultimate aim of potentially incorporating such a vector into a therapeutic for Col1a1 linked OI. The non viral vector system selected was the bacteriophage ΦC31 integrase system. The rationale for selecting this system was three fold, firstly non viral vectors are theoretically not restricted in terms of how much exogenous DNA sequences they can potentially accommodate. As described, when employing a mutation independent approach to target gene suppression it will most likely be necessary to provide a replacement gene which should ideally restore target gene expression levels to close to those of the endogenous gene. Therefore, the delivery vector would have to have the capacity to hold both the suppression and replacement DNA sequences. In the case of Col1a1 linked OI the replacement gene would be the Col1a1 gene, this gene is quite large at 18kb. Frequently the cDNA form of a gene is delivered but in the case of Col1a1 extensive 5' UTR and 3'UTR sequences and the first intron are required to ensure that the gene is expressed at appropriate levels in a tissue specific manner, the construct is

still large at almost 10kb. (This size optimized construct is discussed in greater detail in chapter 7 of this thesis). Therefore due to the large size of the therapeutic nucleic acids required for a potentially successful suppression and replacement strategy targeted towards the treatment of Col1a1 linked OI the unrestricted capacity of non viral vectors for exogenous DNA is a distinct advantage.

Secondly, whilst viral vectors are associated with far superior delivery efficiencies they do have a variety of safety issues as discussed in the introduction, namely, the potential to induce potentially fatal cytoxic and immunological responses (Raper, *et al.*, 2003). The third factor is also a safety issue; viral vectors carry the very real risk of insertional mutagenesis which has been reported to have occurred in two children following delivery of a retroviral vector in a clinical trial (Hacein-Bey-Abina, *et al.*, 2003). What differentiates the ΦC31 integrase system from other non viral vectors is that it can mediate site specific integration into specific location in the genome, pseudo AttP sites. The pseudo AttP sites are thought to be relatively rare, it is estimated that there are in the region of 370 in the human genome (Chalberg, *et al.*, 2006). Moreover it appears that pseudo *AttP* sites are not known to be located within any genes or near to potentially oncogenic genes (Chalberg, *et al.*, 2006). Therefore it was decided to explore the potential application of the ΦC31 integrase system as a vector to deliver Col1a1 linked OI therapeutics.

The efficacy of the ΦC31 integrase system was first explored in HeLa cells. HeLa cells were co-transfected with an eGFP reported plasmid (pEGFP.CMV.AttB) in which the *AttB* sequence had been cloned and the integrase plasmid pCMVInt. Stable cell lines in which the pEGFP.CMV.AttB plasmids had integrated were generated. DNA from these cell lines was used to determine if the pEGFP.CMV.AttB plasmids had integrated randomly or had been involved in an integrase mediated integration event. More specifically, this was explored using the plasmid rescue technique which is designed such that if ΦC31 integrase mediated integration occurred the retrieved plasmids should contain genomic DNA sequences located within the *AttB* sequence (Figure 6.2).

It is possible to identify these genomic sequences through the use of primers which bind 3' and 5' of the TTG recombination site such that they should, in principle sequence into the genomic sequence. In theory, the analysis of this sequencing data should enable identification of the chromosomal integration site as has been previously reported (Bertoni, *et al.*, 2006; Chalberg, *et al.*, 2005; Groth, *et* al., 2000; Held, *et al.*, 2005; Ortiz-Urda, *et al.*, 2003; Thyagarajan, *et al.*, 2001).

The plasmid rescue technique seemed to work efficiently, but none of the rescued plasmids had any evidence of participation in an integrase mediated integration event The majority of the rescued plasmids had intact *AttB* sites suggesting that the plasmid had not participated in integrase mediated integration. A selection of the rescued plasmids did not generate sequence results suggesting that perhaps that the *AttB* site, within which the primers bind, was no longer present in the plasmid. One possible explanation is perhaps that plasmids were involved in a reaction with the ΦC31 integrase enzyme involving other plasmids rather than the human genome which resulted in excision or rearrangement of the *AttB* site. The selection of enzymes used in the plasmid rescue technique introduces an element of bias, as plasmids which have integrated in the vicinity of the restriction enzyme sites are more likely to be successfully rescued, as most likely the plasmids used will not be capable of replicating if the rescued plasmids contain extremely large fragments of genomic DNA.

A total of 107 plasmids were assessed, either by sequencing, PCR or restriction enzyme digestion, from two transfection experiments in which a total of 6 transfections were carried out, four plasmids had an insertion in the *AttB* site. Upon analysis of the sequence contained in all four plasmids it appeared that each contained identical DNA sequence inserted within the *AttB* site, that of the SV40 promoter and neomycin resistance gene of the pEGFP1 plasmid (Figures 6.25 and 6.26). It appears that some rearrangement of the plasmid may have occurred, this could have been a rearrangement event within the HeLa cells, it may have been an aberrant integrase mediated integration event in which the plasmid rearranged or it may be an artifact of the

plasmid rescue technique. Whilst the fact that stable HeLa cell colonies were generated from each of the 6 transfections that was carried out demonstrates that integration did occur, it appears that the integration events in all cases were predominantly, if not all, random integration events.

Given that the target cell type human MSCs population being used in this study was demonstrated to be mutlipotent in our hands, it was of interest to establish the functionality ΦC31 integrase system in MSCs and if the cells would retain their multipotency following transfection with the ΦC31 integrase system plasmids. To explore this a plasmid containing the *AttB* sequence in which expression of an eGFP gene was under the control of the Col1a1 promoter was generated specifically for use in MSCs. It was intended to integrate the plasmid into the genome of the MSCs and following selection of the cells in which an integration event(s) have occurred, the cells would then be differentiated *in vitro* in an identical manner as the MSCs were in the validation study described in this chapter of the thesis. As type I collagen expression is down regulated in adipogenic cells these cells should not expresses high levels of eGFP, in contrast to osteoblasts which express high levels of type I collagen. MSCs nucleofection with the pEGFP.AttB.Col1a1 promoter and pCMVInt plasmids was successful. eGFP expression in the MSCs was high when assessed 48 hours post nucleofection (Figure 6.28). At this point it was decided to select for the MSC population in which integration events had occurred. The cells were treated with G418 selective media, despite the high nucleofection efficiencies very few stable cell colonies were generated and the colonies that did form were dividing poorly. The selected MSC population began to senesce and died. The nucleofection experiment was repeated, delivery of the plasmids was successful and high levels of eGFP expression were observed 48 hours post nucleofection (Figure 6.28). G418 selection was initiated to select for MSCs in which the plasmids had integrated. The MSCs generated very few colonies, on this occasion the G418 selection was removed in the hope that perhaps the cells may proliferate. Similar, to the previous experiment the cells were dividing poorly, at this point it appeared that the MSCs would not be viable for much longer, therefore

it was decided to evaluate the MSCs microscopically. The stable cell colonies from each of the 4 wells were trypsinised, pooled and plated onto coverslips. Microscopic analysis revealed that the MSCs were now expressing very low levels of eGFP in comparison to the levels of expression observed 2 days after nucleofection, in fact many of the cells no longer expressed GFP at any levels detectable via fluorescence microscopic analysis (Figure 6.29 – 6.31).Unfortunately, given this scenario, it was not possible to evaluate if ΦC31 integrase mediated integration had occurred as unfortunately insufficient cells were generated. One hypothesis which may explain these findings is that the strength of the Col1a1 promoter, particularly in MSCs which express high levels of Col1a1, that levels of eGFP generated under the control of the Col1a1 promoter may have been toxic to the MSCs and ultimately lead to cell death. It is notable that Pochampally *et al* (2005) used the identical promoter sequence present in pEGFP.AttB.Col1a1 promoter to drive expression of a Col1a1 transgene, it was found that Col1a1 expression levels were so high from this promoter in a wild type MSC population that similar to what was observed here the cells stopped dividing and became senescent (Pochampally, *et al.*, 2005). Whilst the genes expressed were different Col1a1 versus eGFP the comparison does highlight how high the expression levels driven by the promoter sequence used in each case are. The MSC experiment described in this chapter was carried out twice, representing a total of 8 separate MSC co-nucleofections (i.e. pEGFP1.AttB.Col1a1 promoter plasmid and pCMVInt). In parallel the MSC shRNA.AttB suppression study was set up in which the MSCs were readily selected using G418 and expanded. It was clear that the MSCs could tolerate the G418 selection process and were of sufficiently low passage to facilitate expansion of the selected cell population. To overcome this toxicity that appears to be associated with high levels of GFP expression under the control of the Col1a1 promoter it will be necessary to use a promoters such at the ubiquitin promoter which expresses at significantly lower levels in MSCs.

Finally, the efficacy of RNAi mediated suppression of Col1a1 expression was explored in MSCs which had been transfected with ΦC31 integrase based RNAi expression constructs to determine if functional shRNAs can be expressed using this system. The

ultimate goal in terms of application of the ΦC31 integrase system for Osteogenesis Imperfecta is use of the system to deliver Col1a1 targeting shRNA expression plasmids to MSCs. To initially explore the utility of the ΦC31 integrase system as a delivery option for MSCs a selection of shRNA plasmids were generated which contained the *AttB* sequences, shCol2.AttB, shCol4.AttB and shGFP.AttB. The shRNA plasmids were co-nucleofected with the integrase plasmid CMVInt. MSC stable cell lines were generated for each construct and Col1a1 expression levels assessed over a four week period. Col1a1 expression levels were found to be reduced to 3.11% - 20.63% (±1.25% - 11.69%) in cells expressing shCo2.AttB and to 3.3% - 18.18% (±0.62-8.73%) in MSCs stably expressing shCol4.AttB.

At this point it was of interest to explore if the suppression of Col1a1 expression obtained in the G418 selected MSCs population was due to the presence of randomly integrated shRNA plasmids or due to the presence shRNA plasmids which had participated in a ΦC31 integrase mediated integration events at pseudo *AttP* sites. To explore this, the plasmid rescue technique was employed. The rescued plasmids were sequenced with primers flanking the *AttB* site, Att1200 and Att1201. All 10 plasmids generated both forward and reverse sequence but there was no evidence of any integrase mediated integration events. It is possible that ΦC31 mediated integration did occur but that there were no PmeI or EcoRV enzyme sites in close proximity. The plasmid rescue technique does introduce a bias for sites of integration that have adjacent restriction enzyme sites, as the plasmids used in the studies may not be able to accommodate large DNA fragments. In future investigations a different approach to analysis of the sites of integration will be undertaken, in addition a PCR detection technique has been used in many studies (Chalberg, *et al.*, 2005; Keravala, 2006b; Quenneville, *et al.*, 2004) which may be more informative.

Despite the fact that it was not possible to demonstrate integrase mediated integration in HeLa cells or MSCs a number of the objectives of the work carried in this chapter of the thesis were realised. The multipotency of the MSC population used in the variety of

studies described in this thesis was successfully demonstrated. The 5' upstream sequences selected were shown to function effectively as a promoter in MSCs with high level GFP expression observed 48 hours post nucleofection. Three stable MSCs lines were successfully generated which expressed shCol2.AttB, shCol4.AttB or shGFP.AttB. This demonstrated that it was possible with the population of MSCs at hand to generate a selection of stable cell lines. The MSCs were co-nucleofected shCol2.AttB, shCol4.AttB or shGFP.AttB plasmids and the integrase plasmid pCMVInt. The levels of Col1a1 expression was assessed in these cells over a four week period, it was found that expression from the *AttB* shRNA plasmids was efficient and sustained over the study period. Col1a1 expression levels were successfully reduced to 3.11% - 20.63% (±0.62% - 11.69%) in cells expressing shCo2.AttB when compared to Col1a1 expression levels in MSCs stably expressing the non targeting control shGFP.AttB. Once again, it was not demonstrated that these plasmids participated in a ΦC31 integrase mediated integration event, as the plasmids rescued from the stable MSC cell lines did not have any additional sequences located within the *AttB* site.

Therefore further investigation into the efficacy of the integrase system is required in MCSs, in this regard integrase expression from the pCMVInt plasmid has been demonstrated at the RNA level, another possible route of exploration would be to confirm expression of the ΦC31 integrase enzyme at the protein level. It is notable that this same expression plasmid has been used in multiple ΦC31 studies where integrase mediated integration has been demonstrated (Ishikawa, 2006; Keravala, 2006b; Thyagarajan, et al., 2001)and it is clear that the system is functional.

It has recently been accepted that deletions of 5 -15bp at the site of integration within *AttP* sites do occur, this is not considered to be of huge concern as the *AttP* sites are regarded to be located in non essential, non coding regions of the genome. However, this is not definitive as only 101 of the 370 *AttP* sites thought to be present in the human genome have been characterized. This observation together with recent reports

of the occurrence deleterious genomic events such as deletions of >1kb (Chalberg, *et al.*, 2006) and fourteen different translocation events observed in primary human fibroblasts stably expressing the ΦC31 integrase enzyme (Liu, et al., 2006), the safety profile of the ΦC31 integrase system requires further investigation and validation. Unfortunately it seems likely, given these recent findings that an alternative vector system will be required.

In terms of exploring an alternative vector system for the suppression and replacement based Col1a1 linked OI therapeutic under exploration in the various studies described in this thesis, the system would have to be capable of containing large amount of exogenous DNA and ideally effect site specific integration. In terms of a non viral system the MSC delivery efficiencies achieved with the nucleofector device alone are sufficiently high but this may result in transient expression if integration does not occur and it in no way precludes the occurrence of insertional mutagenesis. The Herpes simplex amplicon vectors are capable in principle of containing up to 150kb of exogenous DNA sequences (Hibbitt and Wade-Martins, 2006).This vector system has been used to deliver the full length Dystrophin gene to skeletal muscle cells (Wang, 2006). The HSV system is non-integrating and consequently is associated with transient gene expression, to overcome this, a variety of groups have developed Herpes simplex/adeno-virus associated virus (HSV/AAV) hybrid amplicon vectors. These vectors have the large transgene capacity of the HSV vectors and have been shown to preferentially integrate into the AAVS1 adenoviral integration hotspot on chromosome 19 both *in vitro* and *in vivo* (Bakowska, 2003; Heister, *et al.*, 2002). Such systems are just a sample of the vector options available which could possibly be selected for use in future studies, what is apparent is that it is unlikely the ΦC31 integrase system will be explored further due to the potential for aberrant genomic events.

Chapter 7

Generation and testing of a Col1a1 replacement gene

7.1 Introduction

Osteogenesis Imperfecta is a debilitating disorder characterised by amongst other features brittle bones, frequent fracture and short stature. This pathology is due to dominant negative protein defects in either the Col1a1 or Col1a2 chains of type I collagen (Primorac, et al., 2001; Rauch and Glorieux, 2004). The phenotypic features of autosomal dominant negative forms of disorders such as Osteogenesis Imperfecta, Retinitis Pigmentosa and Epidermylosis Bullosa are due to the presence of an abnormal protein which disrupts endogenous molecular pathways; this may affect a signalling cascade, enzymatic reaction or assembly or multi-protein structures. The extreme genetic and phenotypic heterogeneity associated with OI has presented many challenges in the development of suitable therapeutics as discussed in the introduction (section 1.3). Indeed, such mutational heterogeneity is not unique to OI, over 50 mutations in the peripherin gene and over 100 mutations in the rhodopsin gene are known to cause autosomal dominant retinitis pigmentosa (Kiang, *et al.*, 2005; Palfi, 2006) and over 100 mutations in Col7a1 are known to cause epidermylosis bullosa (Pulkkinen and Uitto, 1999). The development of therapeutics targeting each specific mutation for such diseases is not a feasible option due to the huge financial, time and technical demands it presents. The work carried out for this doctoral thesis has been focused upon the exploration of the potential application of RNAi based technologies to mediate down regulation of expression of the Col1a1 gene and the generation of a replacement Col1a1 gene to restore Col1a1 expression levels to wild type levels.

As described in chapters 3, 4 & 5 of this thesis which are focused upon the suppression aspects of the development of a Col1a1 mutation independent therapeutic strategy, in order to treat a dominant negative disorder such as OI it will most likely be necessary to remove the negatively impacting protein. With a mutation independent approach, both the mutated and wild type alleles of a given gene are suppressed by the molecular tool being employed, in this case si/shRNAs, therefore a replacement gene will most likely be required to restore expression of the wild type protein. The work described in this

current chapter of the thesis is focused on the development of a Col1a1 replacement gene.

Three mutation independent suppression and replacement strategies have been developed by members of the research team in which either 3' or 5' untranslated (UTR) sequences, the coding sequence or a polymorphism is targeted by a suppression agent e.g. siRNA (Millington-Ward, *et al.*, 1997). In the first approach either the 3' or 5' UTR may be targeted by the suppression agent resulting in the suppression of transcripts derived from both the mutant and wild type alleles of the target gene. Consequently a wild type copy of the gene should be provided in parallel with the suppression agent to restore expression of the target gene within the cell; this approach has been described as suppression and replacement. The replacement gene provided will be modified at the nucleotide level at the target sequence for suppression such that transcripts from the gene are refractory to suppression. However, as the modified nucleotides are located within the 5' or 3' UTR, the replacement gene codes for the wild type Col1a1 protein.

In the second approach the suppression agent is targeted toward the coding sequence of the RNA, once again both wild type and mutant alleles are suppressed. In this case the replacement gene is engineered to contain sequence alterations at each wobble base of each codon located within the target sequence of the suppression agent. In principle transcripts from the replacement gene should encode the wild type protein yet should be refractory to suppression. The final approach is based upon designing the suppressor to target a polymorphism known to be present within the mutated allele. For this approach to be effective the suppressor needs to be highly sequence specific such that it can discriminate between the polymorphic alleles down to a single nucleotide difference. If this sensitivity can be achieved, then in principle only the mutated allele should be targeted for down regulation. In the case of Col1a1 linked OI, preventing synthesis and incorporation of the mutant Col1a1 chain into the triple helix should theoretically convert a severe OI phenotype into the mild type I OI. This is due to the fact that there should no longer be any α1 (I) mutant protein present and Col1a1

protein synthesis should continue from the remaining untargeted wild type Col1a1 allele.

The efficacy of this approach was first described in relation to ribozyme studies in which the replacement gene sequence was altered at the ribozyme cleavage site such that transcripts from the replacement gene were no longer susceptible to ribozyme targeting and cleavage (Farrar, 2002; Millington-Ward, *et al.*, 1997; O'Neill, *et al.*, 2000; Zern, 1999). In a study carried out in retinal explant cultures, alterations of the coding sequence of the periherin gene at 8 wobble bases of the codons that compromise the target site of a peripherin targeting shRNA, resulted in protection of the replacement gene from suppression whilst expression of endogenous periperhin was reduced to 17% (Palfi, 2006). Similarly Kiang et al 2005 reduced the expression of endogenous murine rhodopsin by 70% whilst expression from a murine rhodopsin replacement gene with 7 wobble base changes evaded shRNA based suppression (Kiang, *et al.*, 2005).

In relation to Col1a1 linked OI, before the development of any replacement gene incorporating the required DNA sequence changes can be undertaken it is essential that an intact wild type Col1a1 gene is generated and that it is established that this gene expresses in a tissue specific manner, at as close to endogenous levels as possible. Regardless of the capacity of any replacement gene to avoid suppression, if it does not express in the target tissue type at sufficient levels it will be of limited value. The large size (18kb) and repetitive sequences present in the gene (AF017178) have made the Co1la1 gene quite difficult to work with at the molecular level. Schneike at al 1987 reported that following *in vitro* delivery of a full length human Col1a1 gene to Mov13 cells (a transformed embryonic fibroblast cell line which contains a retroviral insertion in the first intron of the Col1a1 gene) multiple rearrangements of the transgene were observed *(Schnieke, et al., 1987).* Considerable insight into the regulatory sequences required to achieve near endogenous levels of Col1a1 expression from a replacement Col1a1 gene have been revealed from studies with the Col1a1 minigene. This human Col1a1 transgene (AY590263) consists of the first five introns and exons fused to the

terminal six introns and exons of the Col1a1 gene, together with 2.4kb of 5' and 2kb 3' of flanking sequences (Olsen, et al., 1991). The DNA construct is in frame and generates a truncated Cola1 protein which is expressed at very close to endogenous levels *in vitro* (Olsen, et al., 1991) and in a tissue specific manner *in vivo* (Sokolov, 1993). There have been many conflicting reports with regard to the exact region and amount of promoter sequence, first intron and 3' sequences that are required for the high levels of minigene expression observed (Bedalov, 1995; Breault, et al., 1997; Olsen, et al., 1991; Pavlin, et al., 1992; Sokolov, 1993; Sokolov, et al., 1995). What has become clear is that there is differential use of the regulatory sequences, particularly in the promoter in different cell types (Bedalov, 1995; Pavlin, et al., 1992; Rossert, et al., 2000) this may be part account for some of the conflicting data that has been reported (Bedalov, 1995; Breault, et al., 1997; Olsen, et al., 1991; Pavlin, et al., 1992; Sokolov, 1993). In light of the work carried out exploring the contribution of the 5'UTR, 3'UTR and the first intron toward the high levels of expression observed from the Col1a1 minigene *in vitro* and *in vivo* one could surmise that in fact the 2.4kb of 5' sequences contained in the minigene contains the necessary sequences required to drive high levels of expression, particularly in osteoblasts (Pavlin, et al., 1992) which are the bone producing cells of the body (Bedalov, 1995; Breault, et al., 1997; Olsen, et al., 1991; Pavlin, et al., 1992; Rahkonen, et al., 2004; Rossert, et al., 2000; Sokolov, 1993; Sokolov, et al., 1995). The first intron works co-operatively with the promoter and is required, ideally, in its entirety (Bedalov, 1995; Rahkonen, et al., 2004) to ensure correct mRNA processing and finally the 2kb of 3' sequence is also highly desirable from both an mRNA processing and stability perspective.

Murine models of OI are invaluable tools with which insights into the molecular mechanisms of the disease may be obtained. In addition they provide the ideal system within which the efficacy of both cellular and molecular therapies may potentially be explored and validated. The Mov13 mouse carries a retroviral insertion within intron 1 of the Col1a1 gene which prevents its transcription. Mice heterozygous for the mutation express Col1a1 from a single allele; they have a mild phenotype and are

classed as a model of the mild haploinsufficiency Type I OI. Mice homozygous for the insertion are embryonic lethal (Bonadio, et al., 1990; Jaenisch, et al., 1983; Schnieke, et al., 1983). Cell lines have been derived from the homozygous Mov13 mice, fibroblasts were isolated from embryos at day 12 of gestation and SV40 transformed in order to derive permanent cell lines (Schnieke, et al., 1987). These cells do not produce any full length Col1a1 mRNA, however Col1a2 transcription is unaffected and the α2(I) protein chains are rapidly degraded as they do not have their natural α1(I) chain partners with which to associate (Gotkin, et al., 2004). Analysis of tissue isolated from heterozygous and homozygous Mov13 mice revealed that Col1a1 is expressed in odontoblasts and in a small population of osteoblasts (Kratochwil, et al., 1989; Moser, et al., 1996). In addition it appears that low level transcription of the 3' region of the gene under the control of an internal cryptic promoter located downstream of the retroviral insertion occurs, generating shortened mRNA transcripts (Barker, 1991; Hartung, et al., 1986). This makes the cells an ideal model within which to assess the Col1a1 expression construct developed for the current study.

In relation to OI, the efficacy of gene supplementation has been shown both *in vitro* and *in vivo* (Niyibizi, 2001; Pochampally, et al., 2005; Schnieke, et al., 1987; Wu, 1990). As there is only a 4% divergence between human and mouse Col1a1, Schnieke and co-workers explored the possibility of using a human Col1a1 replacement gene (39.6kb) to restore type I collagen synthesis in homozygous Mov13 cells. The human replacement gene was expressed within Mov13 cells, production of both α1(I) and α2(I) chains was detected but were found to have abnormal electrophoretic motilities due to excessive post-translational modification (Schnieke, et al., 1987). In another study the homozygous Mov13 cells were transduced with a retrovirus containing a 4.8kb Col1a1 cDNA called pHUC. Type I collagen was detectable from the transduced homozygous Mov13 cells, however in this case biochemical and morphological abnormalities in the type I collagen were reported (Stacey, 1987), which was later contributed to a defective cell line as any type of collagen expressed in this particular Mov13 cell line was found to be over modified (Wu, 1990). In an *in vivo* study retroviral delivery of human Col1a1

replacement gene (pHUC) rescued the embryonic lethal phenotype of the homozygous Mov13 mice. However the pups did not survive for more than 36 hours which may be due to profuse bleeding despite the fact that human mouse interspecies heterotrimers were indistinguishable from wild type murine type I collagen. The reason the pups did not survive may be due to the fact that the interspecies heterotrimers were of reduced strength and flexibility compared to wild type mouse Type I collagen (Wu, 1990). Type I and III collagens account for 80%-90% of arterial wall collagen and OI patients frequently suffer from vascular disease. In addition, more recent reports have revealed that mice with a deletion of intron 1 from the Col1a1 gene suffer from arterial rupture with expression of type I collagen reduced by 32% in heterozygotes at age 12 months (Rahkonen, et al., 2004). It is notable that the first intron is not present in pHUC the full length Col1a1 cDNA which mediated rescue of the lethal Mov13 phenotype, one could hypothesise that based on the more recent analysis of the first intron just described that, perhaps delivery of a human gene which contained that first intron may have facilitated rescue beyond the 36 hours of survival observed in the study by Wu and colleagues as the intron appears to be essential for expression of sufficient type I collagen in arteries.

A spontaneous naturally occurring model of col1a2 linked OI the *oim* mouse has been described. Mice heterozygous for the mutated Col1a2 allele have a mild OI phenotype whilst mice homozygous for the Col1a2 mutation do not produce any α2(I) chains (Chipman, et al., 1993) and there is an intracellular accumulation of αI(I) homotrimers (Saban, et al., 1996). Akin to the work carried out in Mov13 mice, a gene replacement strategy has also been explored in the *Oim* mouse. Niyibizi et al 2001 demonstrated that adenoviral delivery of the murine Col1a2 gene restored production of type I collagen heterotrimers with the correct 2:1 ratio of α1 (I):α2 (I) chains in bone marrow stromal cells isolated from homozygous *Oim* mice. Additionally injection of the adenoviral Col1a2 construct into the skin of the *Oim* mice also resulted in the production of normal murine type I collagen (Niyibizi, 2001). It has been suggested that over expression of the required wild type Col1a1 or Col1a2 alleles in the cells of OI patients

could potentially shift the equilibrium in favour of normal collagen production to levels sufficient to effect therapeutic benefit (Niyibizi, et al., 2004). This indeed was found to be the case, Pochampally et al 2005 generated a Col1a1 replacement gene which was delivered to MSCs isolated from a Type III OI patient. The human Co1a1 replacement gene was highly expressed in the patients MSCs, Type I collagen production increased three fold and differentiation of the transfected cells into mineralising osteoblasts also increased significantly (Pochampally, et al., 2005). Together these reports support the theory that delivery of a replacement gene to the affected cells or tissue type can indeed restore production of wild type protein which should in turn contribute to amelioration of the disease phenotype.

The first step toward generating any replacement gene is to construct a replacement gene of intact wild type sequence which is shown to have a transcriptional and translational profile which mirrors that of the wild type Col1a1 alleles. The work described in this chapter focuses upon the generation of a replacement gene in which the DNA coding sequence is identical to the wild type Col1a1 allele. As mentioned previously the Col1a1 gene is quite difficult to work with at the molecular level due to its large size and highly repetitive sequence. In order to achieve high levels of tissue specific Co1a1 replacement gene expression it was necessary to include as much of the Cola1 regulatory sequences as possible in the construct. In light of the large size of the transgene and the extensive molecular rearrangement that had occurred in another study following delivery to Mov13 cells (Schnieke, et al., 1987) and the difficulties associated with vector based delivery of such a large construct, a more compact design was required. The design that was adopted was to create a hybrid genomic/cDNA replacement gene. The sequences that would be essential to include in the construct in addition to the protein encoding sequences would be the 5' and 3' UTRs and the first intron present in the human Col1a1 minigene construct which is associated with close to endogenous expression levels both *in vitro* and *in vivo* (Olsen, et al., 1991; Sokolov, 1993; Sokolov, et al., 1995). The Col1a1 minigene construct along with a Col1a1 cDNA construct pHUC were used to provide the required sequence.

Once the Col1a1 genomic/DNA construct was generated the next step was to assess if the construct was functional. Mov13 cells were used in the initial analysis as they do not produce type I collagen due to the presence of a retroviral insertion in both Col1a1 alleles (Schnieke, et al., 1983) and hence they are an ideal model in which to detect expression from the Col1a1 replacement gene. In this regard it is notable that low level expression of truncated Col1a1 transcripts from a cryptic promoter located downstream of the retroviral insertion has been reported (Barker, 1991; Kratochwil, et al., 1989; Kratochwil, 1993). The Col1a1 replacement gene construct was then assessed in a Type I OI cell line which is deemed a far superior model as these are human cells with an OI genotype. The expression of the Co1la1 replacement gene was assessed in these cells at the RNA level via RT-PCR and also at the protein level by ELISA, immunohistochemistry and western blotting.

7.2 Materials and Methods

Methods and materials specific to this chapter are described below. All general methods and materials and the sequences of all primers used in the work carried out in this chapter are described in chapter 2.

7.2.1 Generation of Col1a1 genomic/cDNA replacement gene.

To ensure that the Col1a1 replacement gene generated was expressed in as close to endogenous levels as possible and in a tissue specific manner, it was essential to include the optimal amount of Col1a1 regulatory sequences. Based upon the studies carried out with the human Col1a1 minigene discussed in the introduction of this chapter it was decided to include the first intron and as much of the 5' and 3' UTR sequences as possible. Two constructs the human Col1a1 minigene and pHUC together contained all of the individual Col1a1 sequence components required. The human Col1a1 minigene construct contains 2.4kb of 5' sequence, introns and exons 1-5 and 47 -52 and 2kb of 3' sequence (Olsen, et al., 1991). pHUC is a cDNA encoding the full length Col1a1 protein, it does not have any additional regulatory sequences (Tromp, 1988). Therefore a cloning approach had to be designed such that all the required elements could be contained within a single Col1a1 replacement gene which would in essence be a hybrid of genomic and cDNA Col1a1 sequences.

7.2.1.1 Sequencing of the Col1a1 Minigene 5' UTR & 3' UTR.

The Col1a1 minigene was known to contain introns and exons 1-5 and 47 -52 (Olsen, et al., 1991) but the exact amount and sequence of the 5' UTR and 3' UTR was unknown. Therefore it was necessary to sequence the 5' and 3' UTR of the Col1a1 minigene to facilitate the cloning design process. The first 5' sequencing primer used to achieve this pBluescript 1201 was located within the multiple cloning site of the Bluescript vector which contained the Col1a1 minigene (Olsen, et al., 1991) a primer walk was continued

in a 5' to 3' direction until the known Col1a1 sequence AF017178 was reached. For the 3'UTR the first primer used (101 3'F) was located within the last 100bp of the AF017178 sequence and a primer walk continued for 1005bp after the second poly A site of the Col1a1 gene. This Col1a1 minigene sequence generated in this manner was submitted and accepted by NCBI AY590263 see appendix 1. This minigene construct was used to supply the required 3' UTR and 5' UTR sequence required for generation of a genomic/DNA Col1a1 replacement gene, see sections 7.2.1.2 to 7.2.1.8.

7.2.1.2 Removal of KpnI restriction site from multiple cloning site of pcDNA3.1$^-$

As the construct cloning design required the use of the enzyme KpnI for insertion of the 5' sequences it was necessary to first remove the KpnI site from the multiple cloning site of pcDNA3.1$^-$ the vector in which the Col1a1 genomic/cDNA was to be assembled. pcDNA3.1$^-$ was digested with the restriction enzymes Acc65I and HindIII. The vector was gel isolated (section 2.4.4) and the 3' overhangs filled in using Klenow (section 2.4.9). The vector was phenol chloroform extracted (section 2.4.5) and self ligated (section 2.4.8). The ligation mix was transformed in to XL1 blue *E.Coli* cells and plated on ampicillin selective agar plates (section 2.5.2). A selection of colonies were minipreped (section 2.6.2) and a series of Acc65I and Hind III restriction enzyme digests carried out to identify plasmids in which both the KpnI and HindIII restriction sites had been destroyed. A single positive miniprep was selected and maxipreped (section 2.6.4). This plasmid was called pcDNA3.1$^-$ K$^-$ H$^-$

7.2.1.3. Removal of Col1a1 cDNA from pHUC and generation of pcdCol1a1.

The majority of the Col1a1 encoding sequences were obtained from the pHUC, which contains a full length Col1a1 cDNA. pHUC Col1a1 cDNA was digested with the restriction enzyme EcoRI (section 2.4.7) and the 4376bp Col1a1 cDNA fragment gel isolated (section 2.4.4). The pcDNA3.1$^-$ plasmid in which the KpnI and Hind III sites were removed (generated in 7.2.1.2) was digested with EcoRI (section 2.4.7), dephosphorylated (section 2.4.11) and phenol chloroform extracted (section 2.4.5). The 4376bp Col1a1 cDNA fragment and the pcDNA3.1$^-$ plasmid in which the KpnI and

Hind III sites were removed were ligated (Figure 7.1.1) transformed in to XL1 blue *E.Coli* cells and plated on ampicillin selective agar plates (section 2.5.2). A selection of positive colonies was minipreped (section 2.6.2). Restriction digests were carried out using the restriction enzymes BamHI and HindIII (section 2.4.7). Plasmids with the Col1a1 cDNA insert in the sense orientation generated fragments of 8173bp and 1477bp and those in the antisense orientation fragments of 6831bp and 2816bp. A selection of minipreps which generated the correct size DNA fragments following digestion with the enzymes BamHI and HindIII were sequenced (section 2.7.1) to confirm the presence of the insert and a single miniprep then selected and maxiprep was carried out (section 2.6.4). This plasmid was called pcdCol1a1.

7.2.1.4 Addition of Col1a1 3'Untranslated Region (3'UTR)

The Col1a1 3' UTR sequence was PCR amplified (section 2.4.2) using the forward primer AccIII and reverse primer AflIIR. The AccIII primer binding site encompasses an AccIII site present in the exon 49 of the Col1a1 sequence (AF017178) and at position 5876 – 5903 in the minigene. The AflIIR primer binds at position 8499 – 8521 in the minigene. The AflII restriction site is incorporated at the 3' end of the PCR amplified DNA sequence, this facilitated use of the AflII site in the multiple cloning site of pcdCol1a1 for cloning of the 3'UTR PCR fragment. The 2.45kb PCR product was gel isolated (section 2.4.4), digested with AccIII and AflIIR (section 2.4.7) and phenol chloroform extracted (section 2.4.5). This fragment was then ligated (section 2.4.8) into pcdCol1a1 which had been digested with AccIII and AflII (Figure 7.1.2), gel isolated (section 2.4.4), dephosphorylated (section 2.4.11) and phenol chloroform extracted (section 2.4.5). The ligation mix was transformed in to XL-10 *E.Coli* cells and plated on ampicillan selective agar plates (section 2.5.3). A selection of colonies was minipreped (section 2.6.2) and digested with AccIII and AflII to identify those which were positive for the 3'UTR sequence. Positive minipreps were sequenced (section 2.7.1) using the forward primer 3' insert detect F. A miniprep in which the 3' insert was present and of intact sequence was then maxipreped (section 2.6.2). This plasmid was called

pcdCol1a13'. This work was carried out with the assistance of Dr. S. Millington-Ward TCD.

7.2.1.5 Addition of Col1a1 5'Untranslated Region (5 'UTR)

The required 5' UTR sequence was removed from the minigene plasmid using the restriction enzymes KpnI and NotI (section 2.4.7). The KpnI site is located in exon two of the Col1a1 sequence and the NotI site in the multiple cloning site of the Bluescript minigene plasmid. The 4017bp fragment was gel isolated (section 2.x). The pcdCol1a1 plasmid was digested with KpnI and NotI (section 2.4.7), gel isolated (section 2.4.4), dephosphorylated (section 2.4.11) and phenol chloroform extracted (section 2.4.5). A ligation reaction containing the 4017bp Col1a1 fragment and the KpnI, NotI digested pcdCol1a13' plasmid was carried out (Figure 7.1.3). The ligation mix was transformed into XL 10 *E.Coli* cells and plated on ampicillan selective agar plates (section 2.5.3). A selection of positive colonies were minipreped (section 2.6.2) and a KpnI and NotI restriction enzyme digests were carried out (section 2.4.7) to identify plasmids which contained the 5'UTR insert. A selection of positive colonies were sequenced using KpnIF and pcDNAF primers (section 2.7.1), a maxiprep of a single plasmid with intact was carried out (sequence 2.6.4). This plasmid was called pcdCol1a1 Full Length.

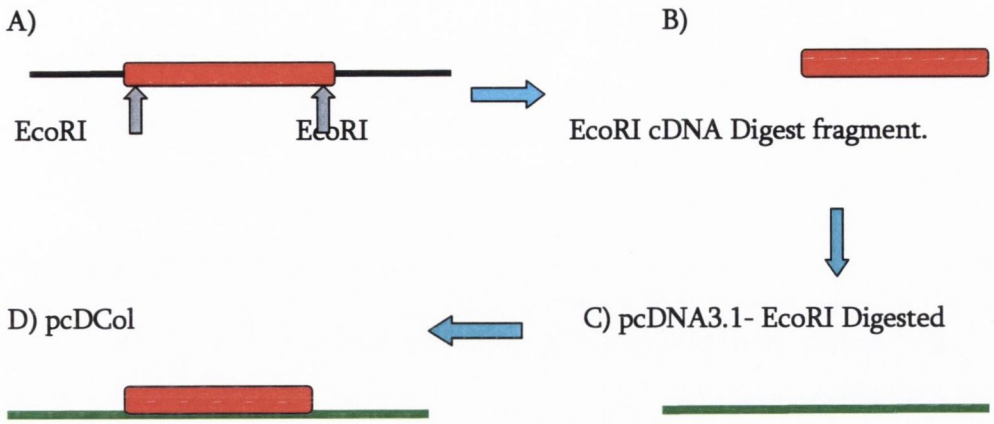


Figure 7.1.1 Col1a1 genomic/cDNA replacement gene cloning step 1: A and B) The required Col1a1 cDNA sequences were removed from pHUC, a full length Col1a1 cDNA construct, by EcoRI digestion. **C and D)** The Col1a1 cDNA fragment was cloned into pcDNA3.1- which was digested with EcoRI, generating pcdCol1a1 which is exon 1-51 of the Col1a1 cDNA.

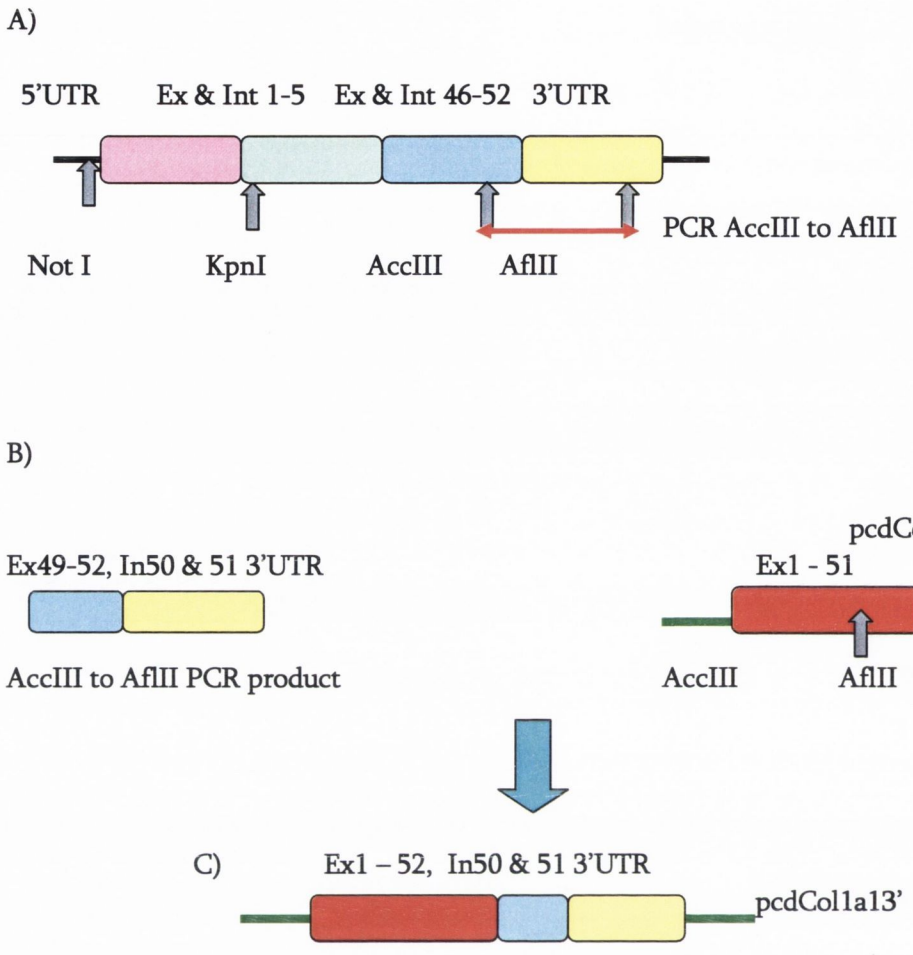


Figure 7.1.2 Col1a1 genomic/cDNA replacement gene cloning step 2: Addition of the 3'sequnces. A) The required sequences were PCR amplified from the human Col1a1 minigene. The forward and reverse PCR primers used had an AccIII or AflII site respectively incorporated to facilitate insertion of **B)** the PCR product containing exons 49 – 52, introns 50 and 51 and 1.5kb of 3'UTR into pcdCol1a1 which had been digested with AccIII and AflII. The ligated fragment generated **C)** the construct pcdCol1a13' which contains exons 1-52, introns 1, 50 and 51 and 1,5kb of 3'UTR.

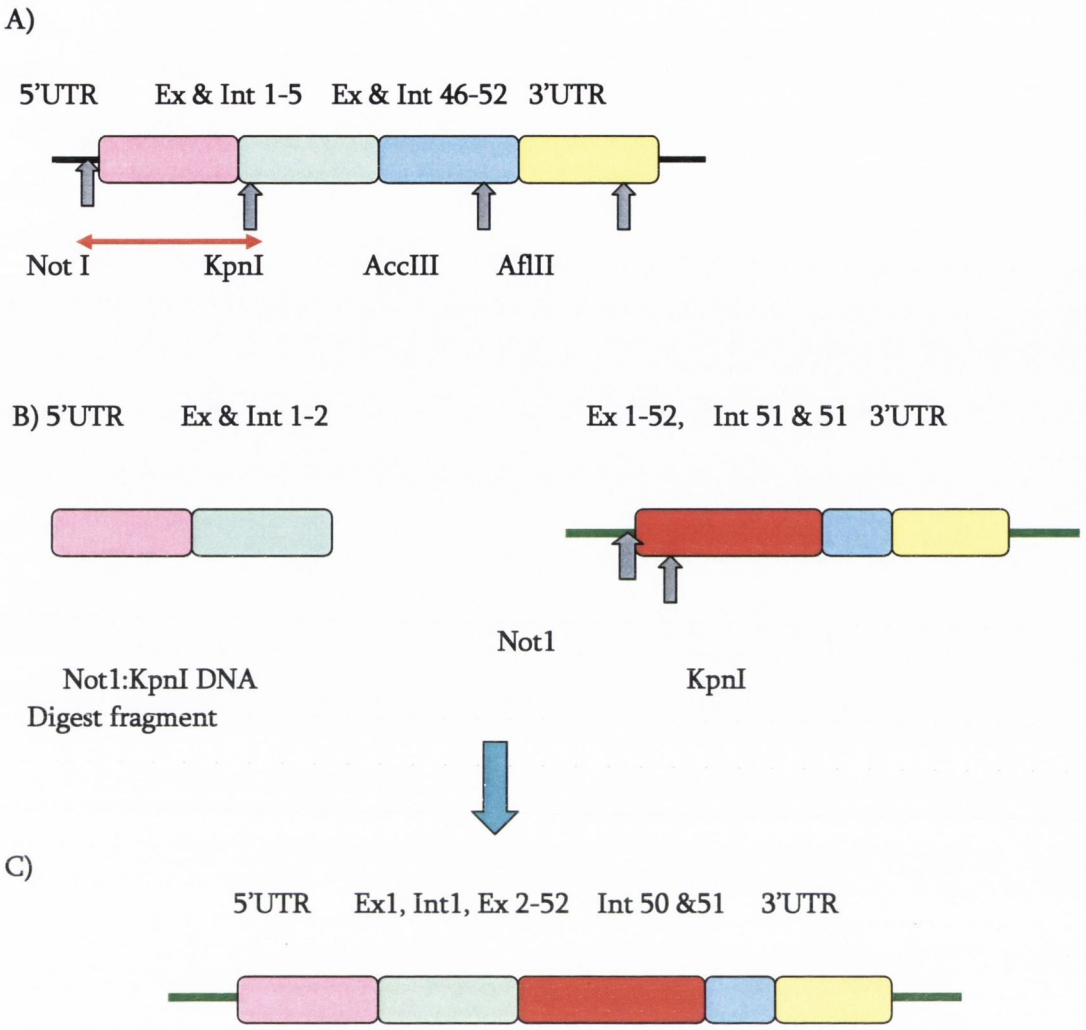


Figure 7.1.3 Col1a1 genomic/cDNA replacement gene cloning step 3, addition of 5'seqeunces. A) The required 5'UTR, exon1, introns1 and part of exon two was removed from the Col1a1 minigene using a Not1:Kpn1 digest. **B)** the NotI:KpnI DNA digest fragment was ligated into pcdCol1a13' which was digested with NotI and KpnI, generating **C)** the Col1a1 genomic/cDNA replacement gene which contains 2.5kb of 5'UTR, Exons 1-52, Intron 1, 50 and 51 and 1.5kb of 3'UTR.

7.2.1.6 Removal of CMV promoter from pcdCol1a1 Full length.

pcdCol1a1 Full Length was digested with restriction enzymes NruI (cuts pcDNA3.1- at position 209) and NotI (cuts pcDNA3.1- at position 980) (section 2.4.7). The 15kb DNA fragment was gel extracted (section 2.x), end filled using T4 DNA polymerase (section 2.4.10), phenol chloroform extracted (section 2.4.5) and self ligated (section 2.4.8). The ligation mix was transformed into XL1 blues *E.Coli* cells and plated on ampicillan selective agar plates (section 2.5.3). A selection of colonies were PCR screened using the forward primer pcDNA82F and the reverse primer Col1a1 CMV-detect R (section 2.7.1) to identify colonies which contained plasmids in which the CMV promoter has been destroyed. Plasmids which contained the CMV promoter generated a PCR product of 1821bp and those without the CMV promoter a product of 1050bp. A selection of minipreps were sequenced (section 2.7.1) using the primer pcDNA82F to identify a plasmid in which the CMV promoter had been removed. An endotoxin free maxiprep of the plasmid was then carried out (section 2.6.5). This plasmid contained all the required sequences and was called Col1a1 genomic/cDNA replacement.

7.2.1.7 Sequencing of the entire Col1a1 genomic/cDNA replacement.

The genomic/cDNA Col1a1 replacement gene of 10,575bp contains the 3' UTR encompassing the desired Col1a1 promoter sequences, intron 1, exons 1–52, introns 50 & 51 and 1927bp of 3' sequence. It was first necessary to ensure that no mutations had arisen in the generation of the construct, the plasmid was sent to MWG Biotech UK to be sequenced in full. The predicted Col1a1 genomic/cDNA was supplied to MWG Biotech with the replacement. The Col1a1 genomic/cDNA sequence was returned aligned to the supplied consensus sequence (appendix 2).

7.2.1.8. Identification of mutations in Col1a1 genomic/cDNA replacement. Analysis of the aligned sequence revealed the presence of four mutations (see section 7.3.2). The mutations were a nonsense mutation at codon 6 of exon 21, (CAA – TAA), a single nucleotide deletion at codon 17 of exon 27, a single nucleotide deletion at codon 14 of exon 28 and the final mutation was a missense mutation in which a Glycine was

converted to an Asparginine (GGT to AAT) at codon 31 of exon 40. The presence of these mutations was confirmed by sequencing the construct using specifically designed mutation detection primers termed StopF, StopR Gly-AsnF, Gly-AsnR, DelF and DelR. Sequencing was carried out by MWG biotech UK. The additional sequencing data confirmed that the missense and nonsense mutations were present but neither of the deletions was found to be present in the in Col1a1 genomic/cDNA replacement.

7.2.1.8 Correction of the mutations identified in the Col1a1 genomic/cDNA replacement gene construct.

It was essential to correct the identified mutations. The first step was to establish if they had occurred spontaneously or if they were present in the pHUC construct from which the original Col1a1 cDNA was excised (section 7.2.1.3). The pHUC was sequenced (MWG Biotech) using the mutation detection primers StopF, StopR Gly-AsnF, Gly-AsnR, DelF and DelR. It was found that the nonsense(exon 21) and missense (exon 40) mutations did indeed occur spontaneously during the cloning process, as expected the deletions were not present in the pHUC. Therefore it was possible to use the pHUC cDNA to correct the mutations; this was carried out in a single cloning step.

Briefly, the pHUC Col1a1 cDNA was digested (section 2.4.7) with restriction enzymes KpnI and ClaI which cut at nucleotide 267 and 4445 respectively. The 4178bp fragment was gel isolated (section 2.4.4). The mutated Col1a1 genomic/cDNA replacement gene was also digested with KpnI and ClaI, the required 11 kb fragment was gel isolated (section 2.4.4), dephosphorylated (section 2.4.11) and phenol chloroform extracted (section 2.4.5). The Col1a1 genomic/cDNA plasmid and pHUC Col1a1 cDNA insert were ligated (section 2.4.8) the ligation mix was transformed into XL10 *E.Coli* cells and plated on ampicillin selective agar plates. A selection of colonies were minipreped, and two were sent to be sequenced (GATC Biotech Germany) using the mutation detection primers StopF, StopR Gly-AsnF, Gly-AsnR to determine if either contained the correct

wild type sequence. A single plasmid with the correct sequence was selected and a maxiprep carried out (section 2.6.5).

7.2.2 Culturing of Mov13 cells.

The Mov13 cells were a kind gift of Dr. R. Janeisch, Dept. of Biology, Massachusetts Institute of Technology, USA. The fibroblasts are homozygous for the retroviral insertion in intron one of the endogenous murine Col1a1 gene, were isolated from 12 day gestation embryos and SV40 transformed. This particular cell line is designated M13-5, and are called Mov13 for the work described in this thesis. Mov13 cells were grown at 37°C and 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 5% L-Glutamine, 5% Sodium Pyruvate and 10% fetal calf serum and 100μM ascorbic acid. The cells were cultured until about 85% confluent at which point they were sub-cultured as described in section 2.8.2. Briefly, the medium was pipetted off and the cells rinsed with 10ml PBS to remove traces of medium. 2mls of Trypsin-EDTA was added and the plate incubated at 37°C for 5 minutes. 10ml of warm DMEM (FCS in the medium inactivates the Trypsin-EDTA) was added and the remaining cells detached by pipetting the medium against the monolayer. The cell suspension was centrifuged at 1300g for 5 minutes.

If the cells were to be plated into a 6 well plate the supernatant was discarded and the cells were re-suspended in 2mls of medium, counted and $5x10^5$ cells plated per well. Alternatively the cells were resuspended in 6mls, split 1:3 and replated into 60mm tissue culture dishes.

7.2.3 Nucleofection of Mov13 cells.

The Amaxa nucleofection kit VPI-1002 was used for all Mov13 nucleofection reactions. This kit is designed for nucleic acid delivery to a multitude of primary fibroblasts

isolated from different tissues (e.g. lung or dermis) and from different mammalian species (e.g. mouse, human, rat).

7.2.3.1 Optimisation of Mov13 nucleofection protocol

The Amaxa VPI-1002 protocol recommends a selection of 5 programs that can potentially be selected for use for the nucleofection reaction. 2 programs, A-24 and A-23 were selected for use based upon the information and kit specifications detailed on the Amaxa website (www.amaxa.com). These programs are optimised for nucleofection of transformed mouse embryonic fibroblasts. Duplicate nucleofections with the supplied pmaxGFP plasmid were carried out per program i.e. 2 for A-23 and 2 for A-24. A single optimization experiment was carried out to determine which program achieved the highest nucleofection efficiencies and cell survival rates.

Briefly, Mov13 cells were passaged 2 to 3 days prior to nucleofection such that cells were 85% confluent upon nucleofection. The cells were harvested by trypsinisation and counted (section 2.8.4). The recommended optimal cell number range is $2x10^5$ to $2x10^6$ per nucleofection. In this cells were divided into individual aliquots of $1x10^6$ cells, centrifuged @ 100g for 10 minutes, the media aspirated and the Mov13 cells re-suspended in 100µl of pre-warmed fibroblast specific Amaxa nucleofection solution. 2µg of pmaxGFP plasmid was added to each aliquot of $1x10^6$ cells and the suspension transferred directly to a specialised Amaxa nucleofection cuvette which was placed into the Nucleofector and either program A-23 or A-24 selected. Once completed 500µl of pre-warmed media was added to the cuvette, the cells were then transferred to a single well of a pre-warmed 6 well plate containing 1% gelatin coated coverslips and placed in a humidified 37°C, 6% CO_2 incubator. 24 hours later the medium was changed, removing any of the cells which had not survived the nucleofection process. 24 hours later pmaxGFP expression was assessed via fluorescence microscopy to determine transfection efficiencies.

7.2.3.2 Fluorescence microscopy of Mov13 nucleofected cells.

In order to determine the nucleofection efficiencies of the Mov13 cells with the pmaxGFP plasmid it was necessary to evaluate the number of cells which were expressing eGFP. The cells were rinsed thoroughly with PB and fixed for an hour in 4% paraformaldehyde following which the cells were rinsed thoroughly with PBS. The coverslips were mounted using aqua polymount. The cells were first visualized using the brightfield stetting of the microscope, an image of the cell population in that view was taken, the eGFP specific filter setting was then applied so that the cells in the population which were expressing eGFP would be detected and a picture of this was taken. In order to determine the nucleofection efficiencies the brightfield and eGFP images were overlaid using the microscope software. The total cell population in a given image was counted, the eGFP expressing cells counted and the percentage nucleofection determined.

7.2.3.3 Optimised Mov13 nucleofection protocol

Cells were passaged as described above, $1x10^6$ were transfected with 2μg of endotoxin free Col1a1 replacement gene using program A-23 as described above (7.2.3.1). In addition a transfection efficiency control reaction was carried out in which $1x10^6$ were transfected with 2μg of pmaxGFP plasmid (provided in the Amaxa kit VPI-1002). The pmaxGFP nucleofected cells were plated onto coverslips in the well of a six well plate. The Col1a1 nucleofections were carried out in triplicate in addition to a single control pmaxGFP. In addition a single aliquot of cells were nucleofected in the absence of DNA. Media was changed 24 hours after nucleofection and allowed to recover for an additional 24 hours before any additional analysis or manipulations were carried out. This nucleofection experiment carried out a total of three times, GFP nucleofection controls were not carried out for the final two nucleofection experiments. In total 9 Mov13 nucleofections with the Col1a1 genomic/cDNA replacement gene were carried out.

7.2.4 Generation of Mov13.Col1a1 stable cell lines

48 hours after nucleofection of the Col1a1 genomic/cDNA replacement gene the medium was replaced with G418 selective medium (500μg/ml). The medium was changed every second day until the control cells had died (cells nucleofected in the absence of DNA). At this point the selective media (G418 @ 500μg/ml) was changed every second day until the nucleofected cells formed distinct colonies in each well. These were then trypsinised, pooled and replated into a 60mm plate and grown in G418 (500μg/ml) selective medium for another two weeks, generating a Mov13 cell line which should, in principle, stably express the human Col1a1 genomic/cDNA replacement. This selection procedure was carried out for all three Mov13 Col1a1 nucleofection experiments generating a total of 3 independent Mov13, MovCol1, MovCol2 and MovCol3 each of which should in principle express human Col1a1 genomic/cDNA replacement.

7.2.5 Real time RT-PCR analysis of Col1a1 expression in Mov13 and MovCol1a1 cell lines.

The 3 stable Mov13 cell lines generated MovCol1, MovCol2, MovCol3 were assessed for Col1a1 expression. For each cell line MovCol1, MovCol2, MovCol3 and Mov13 three wells of a 6 well plate were seeded with $5x10^5$ cells. 48 hours later RNA was Trizol extracted (section 2.8.2). Col1a1 expression levels were assessed using real-time RT-PCR. The Col1a1 primers used were designed such that they amplified both mouse and human Col1a1, this was possible as the human and mouse sequences are only 4% divergent. The housekeeping gene used was GAPDH and reactions were set up as described in section 2.x. Each individual RNA sample was plated in triplicate per 96 well plate. Each RT-PCR reaction was carried out a total of three times. The relative quantification method was used to determine Col1a1 expression levels. As mentioned previously although the homozygous Mov13 mice are embryonic lethal and there is no detectable Col1a1 expression, the Mov13 mice are known to express Col1a1 in odontonblasts and osteoblasts (Kratochwil, et al., 1989; Moser, et al., 1996). In addition

very low levels of transcription of downstream regions of the Mov13 Cola1 alleles from a cryptic promoter has been reported (Barker, 1991; Hartung, et al., 1986). This low level of Col1a1 expression in the Mov13 cells is readily detectible using the Col1a1Duo primers which bind in the 3' region of the Col1a1 RNA. Each RNA sample was pipetted in triplicate on a single 96 well plate per primer set when the data was analysed the average of the 3 values obtained for the triplicate was used. Statistical analysis was carried out using Graph Pad InStat3, the average, standard deviation were calculated. In addition the statistical significance of the data was determine, a two tailed t-test with two degrees of freedom was carried out in which the Col1a1 expression values determined for the stable cell lines MovCol1 – MovCol3 were compared to the Col1a1 expression values in the Mov13 cells.

7.2.6 Culturing of Type I OI cell lines

Two OI cell lines are dermal fibroblasts, these cells were a kind gift of Dr. M. Valli, Centro per lo Studio aelle Malattie del Tessuto Connettivo, Pavia, Italy. Cell line OI479 is from 5 year old child, the child's father was also affected. The major clinical signs of Osteogenesis imperfecta in the child include multiple fractures, mild scoliosis and blue sclerae. Notably the childs' stature was almost normal with no long bone deformities and at the time of biopsy the hearing was normal whereas the father had hearing impairment. The ratio type III/type I in the dermal fibroblasts isolated from the child is higher than in an age matched control suggesting a reduced level of type I collagen synthesis. The mutation is Col1a1 IVS17 +1G>A a base substitution at a highly conserved splice site which results in a null Col1a1 allele (Personal communication Dr. M. Valli).The second cell line OI579 was isolated from an 6 year old sporadic case of OI. The patient presented with a frequent fracture rate, absence of scoliosis and blue sclera. The childs' stature was almost normal, hearing was normal and there were no long bone deformities at the time of biopsy. The ratio type III/type I in the dermal fibroblasts is higher than in an age matched control suggesting that the levels of type I collagen is reduced. The mutation is Col1a1 IVS 29C5-A which is a base substitution at a highly

conserved splice site results in a null Col1a1 allele (personal communication Dr. M. Valli). Both dermal fibroblast cell lines were obtained at passage 6. Dr. Valli indicated that the cells can potentially be passaged to P20 - P30, but advised not to use the cells beyond P10-12 as their growth characteristics change and most likely the gene expression profile of the cells. A single T75 flask of each cell line at P6 was obtained, the cells were split 1:3 to expand the cell lines. From the 3 flasks generated 5 vials were frozen down and one was used for the first nucleofection experiment.

Type I OI cells were grown at 37°C and 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 5% L-Glutamine, 5% Sodium Pyruvate, and 10% fetal calf serum and 100μM ascorbic acid. The cells were cultured until about 85% confluent at which point they were sub-cultured as described in section 2.8.2. Briefly, the medium was pipetted off and the cells rinsed with 10ml PBS to remove traces of medium. 2mls of Trypsin-EDTA was added and the plate incubated at 37°C for 5 minutes. 10ml of warm DMEM (FCS in the medium inactivates the Trypsin-EDTA) was added and the remaining cells detached by pipetting the medium against the monolayer.The cell suspension was centrifuged at 1300g for 5 minutes. If the cells were to be plated into a 6 well plate the supernatant was discarded and the cells were re-suspended in 2ml of medium, counted and $5x10^5$ cells plated per well. Alternatively the cells were resuspended in 6mls, split 1:3 and replated into T175 tissue culture dishes.

7.2.7 Nucleofection of Type I OI cell lines

The Amaxa kit VPI-1002 was also used for the delivery of plasmid DNA to the Type I OI cells. The Mov13 optimised protocol was used for nucleofection of the Type I OI cells. The cells were cultured and nucleofected as described in section 7.2.3.2. The only difference was that for this series of nucleofection experiments the Human Dermal Fibroblast specific program U-23 was used. The nucleofection experiment was carried out a total of 3 times and the cells G418 selected as described below in 7.2.9. pmaxGFP

transfection controls were not carried out. A fresh vial of passage 7 cells for both 47 OI and 57 OI cells was expanded and used for each nucleofection experiment

7.2.8 Generation of stable Type I OI cell lines

48 hours after nucleofection the medium was replaced with G418 selective medium (250μg/ml). The medium was changed every second day until the control cells had died (cells nucleofected in the absence of DNA). At this point the cells in each well of the 6 well plate were grown in G418 selective medium (250μg/ml) until distinct cell colonies had formed in each well. The cells were trypsinised, pooled and replated into a single 60mm plate. Cells were grown for another two weeks in G418 (250μg/ml) selective medium.

7.2.9 Real time RT-PCR analysis of Col1a1 expression in Type I OI cell lines.

It was not possible to generate any OI47 cell lines stably expressing the Col1a1 genomic/cDNA replacement. In contrast 3 stable cell lines were generated from the OI579 Type I OI dermal fibroblasts termed, 57OI-C1, 57OI-C2, 57OI-C3 which should express the introduced Col1a1 genomic/cDNA replacement gene. For each cell line 57OI-C1, 57OI-C2, 57OI-C3 and 57OI, 2 wells of a 6 well plate were seeded with $5x10^5$ cells. 48 hours later RNA was Trizol extracted (section 2.x). This was repeated 2 times, generating a total of 4 RNA samples per cell lines Col1a1 mRNA expression levels were assessed. It was not possible to generate large sample numbers due to the limited passage capacity of this primary cell line, generating extensive sample numbers would have reduced the quality and reproducibility of the data. The fragility of these cells was highlighted by the inability to generate a stable cell line from the 47OI cell line and consequently extreme caution was taken in the handling and passaging of the 57OI cells). The housekeeping gene used was GAPDH and reactions were set up as described in section 2.9.4. Each individual RNA sample was plated in triplicate per 96 well plate. Each RT-PCR reaction was carried out three times. The relative quantification method was used to determine Col1a1 expression levels. In this method the Col1a1 expression value of a sample is divided by the expression value of a housekeeper in the same

sample, in this case GAPDH, this is normalisation. The normalised Col1a1 values in the 57OI-C1, 57OI-C2 and 57OI-C3 cells are then compared to the normalised Col1a1 expression values in the 57OI cell to determine the relative expression levels. Statistical analysis of the results was carried out using GraphPad InStat3, the mean and standard deviations were calculated. A two-tailed student t-test was carried out in which the relative Col1a1 expression in each of the stable cell lines was compared to the relative Col1a1 expression values in the wild type 57OI Type I OI cells, in order to determine if the differences in Col1a1 expression between cell lines were significant.

7.2.10 Metra CICP ELISA analysis of Col1a1 expression in Type I OI cell lines

Levels of *de novo* synthesised Col1a1 protein in Type I OI cells following nucleofection of the Col1a1 genomic/cDNA replacement gene were assessed using the Metra CICP ELISA system. The assay is based upon the fact that type I collagen is synthesised as a pro-collagen, which consists of mature collagen with N & C terminal pro-peptides. These pro-peptides are cleaved from the collagen molecule by specific proteases prior to incorporation of the collagen triple helix into a growing collagen fibril. The release of these peptides into the media in which the cells are growing provides a stoichiometric representation of collagen production. The Metra CICP assay measures the *C*-Terminal of type I collagen *P*rotein.

$5x10^5$ cells from each stable Type I OI cell line 57 OI-C1, 57 OI-C2, 57 OI-C3 and 57OI media were plated in duplicate and allowed to grow for 48 hours. The media was changed for serum free α DMEM. The cells were allowed to secrete type I collagen into the medium for 24 hours. Medium was then collected, the duplicate samples from each cell line were pooled generating a single representative sample per cell line. The pooled samples were analysed for *de novo* type I collagen production according to the manufacturer's guidelines. As described in section 5.2.8, the Metra CICP assay is a sandwich enzyme immunoassay in a micro-titer format. In the case of this particular

assay, the monoclonal anti-CICP capture antibody is coated on the plate. The medium collected from the nucleofected cells is added to the wells of the microtitre plate, in principle any antigen, i.e. C terminal propetide should be captured. The rabbit anti-CICP is then added, this should bind to the captured C terminal propeptide. The goat anti-rabbit alkaline phosphatase conjugate is added, which should bind to the rabbit anti-CICP. Finally the pNitrophenyl phosphate substrate is added, the optical density of the solution measured at 405nm and a four parameter fit curve equation used to calculate the concentatation of CICP present. All reagents used were provided in the Metra CICP kit.

The test samples were diluted in assay buffer between 1 and 48 fold such that the samples fall within the detection limits of the assay. Following equilibration of the anti-CICP coated strips to room temperature 100µl of the Metra CICP concentration standards, test samples or controls were plated in duplicate and incubated at room temperature for 2 hours. The strips were then inverted and washed 3 times with 300µl of 1X wash buffer per well and blotted dry on tissue paper. 100µl of the rabbit anti-CICP was then added to each well and the reaction incubated for 45-50 minutes at room temperature. The strips were then washed 3 times with 300µl 1X wash buffer as described. 100µl of reconstituted enzyme conjugate was added to each well and incubated for 45-50mins at room temperature. 100µl of working substrate solution was then added to each well, incubated at 18^{0}C to 25^{0}C for 30 -35 minutes. Samples were washed 3 times with 1X wash buffer as previous. 50µl of stop solution was added to each well to stop the reaction and the OD read within 15 minutes at λ405nm. A four parameter fit curve was used to analyse the data and the concentration of the C terminal pro-peptide sample determined from the curve.

7.2.11 Western Blot analysis of Col1a1 expression.

In order to determine if the Col1a1 replacement gene was generating wild Col1a1 protein it was essential to carry out western blot analysis. This would reveal if the Col1a1 protein chains being produced from the introduced Col1a1 genomic/DNA replacement gene were of the correct size and if they have normal electrophoretic mobility. This is of particular importance with regard to OI linked Col1a1 or Col1a2 mutations as abnormal electrophoretic mobility of the Cola1a1 and Col1a2 chains has frequently been observed. This is due to the fact that the presence of a mutation in either chain may reduce the triple helix assembly time. This in turn results in increased posttranslational modification of the chains; it is the presence of additional moieties that causes the alteration in electrophoretic mobility.

7.2.11.1 Protein Extraction from Type I OI fibroblasts

$5x10^5$ cells from each stable Type I OI cell line 57OI-C1, 57OI-C2, 57OI-C3 and 57OI media were plated in duplicate and allowed to grow for 48 hours. Cells were washed twice in ice cold PBS. 500µl of ice-cold modified RIPA buffer containing a protease inhibitor cocktail was added per well of a 6 well plate and the cells removed using a chilled plastic cell scraper. The cells were further lysed by passing the RIPA:cell suspension through an 18 gauge needle. The cell suspension from both wells per cell line was pooled and transferred to a sterile tube and shaken gently at 4 ^{0}C for 15 minutes to allow the cells to lyse. The lysate was centrifuged at 17,000g in a pre-cooled centrifuge for 15 minutes. The supernatant containing the protein was immediately transferred to a fresh centrifuge tube. The protein concentration is determined using the Bradford Assay and divided into 200µl aliquots that can be stored a –20^{0}C for up to a month.

7.2.11.2 Bradford Assay

In order to determine the protein concentrations of the samples generated in 7.2.12.1 the Bradford protein quantification assay was carried out. This assay is carried out in a

96 well plate format. The Bradford reagent was prepared by diluting it 1 in 5 with dH_20. The standard curve made with IgG ranges from 0.1mg – 1.0mg/ml. The plate was prepared as follows; the first row of the microtitre plate is Bradford reagent only, 200µl in each well. The standards were then plated, 10µl of each standard was added to each well and each standard was plated in triplicate. 10µl of each test sample at a dilution 1/10, 1/20, 1/40 was added per well, each sample was plated in duplicate. Finally 190µl of the Bradford reagent is added to each well. The colorimetric assay is fully developed within 10 minutes and should not be analysed more than 30 minutes after addition of the Bradford reagent. Analysis is carried out at λ595nm.

7.2.11.3 SDS PAGE electrophoresis.

A two-phase polyacrylamide gel was prepared, with a 5% resolving gel and 7.5% separation gel. 15µg of denatured protein heated to 90^0C in the presence of denaturing agent 5% β- mercaptoethanol was loaded into each well (as determined by the Bradford). Gels were initially run at 50V until the samples reached the separating gel, the voltage was then increased to 80V for 3 - 4 hours. Once separated the next step was to transfer the proteins to a nitrocellulose membrane.

A wet transfer system was used to transfer the separated proteins from the gel to the nitrocellulose (Protran BA 83) membrane. The electrophoresis glass cassette was carefully opened, the stacking gel was cut away and the remainder of the gel immersed in transfer buffer. The appropriate size membrane was soaked in methanol for 15 seconds, soaked in sterile nuclease free water for 2 minutes and subsequently equilibrated in transfer buffer for 10 minutes.

The transfer stack was set up as follows, the cassette was opened, and a foam pad was placed down, followed by two sheets of filter paper. Next the membrane was placed on the filter paper, followed by the gel, two more filter papers and the final foam pad. All items has been previously been equilibrated in transfer buffer and any bubbles removed by rolling a Pasteur pipette over the surface of each layer in the stack. The cassette

holder was placed in the transfer tank so that the gel side of the cassette holder was facing the cathode (–) and the membrane side was facing the anode (+). Adequate transfer buffer was added to the tank to cover the cassette holder. The transfer can be run at 160mA when packed in ice for two hours.

7.2.11.4 Antibody probing

The primary antibody, a goat anti-Col1a1 antibody (Collagen Type I (C-18) SC8784) was purchased from Santa Cruz. This polyclonal antibody was raised against a peptide mapping to the internal region of the human Col1a1 protein. It is recommended for the detection of pro-collagen and mature collagen from both humans and mice. The recommended dilution range is 1:100 to 1:1000. The secondary antibody donkey anti-goat IgG-HRP (sc2020) was also purchased from Santa Cruz. The recommended working dilution range is 1:500 to 1:10,000. The blot was rinsed in TBS and blocked in a suitable blocking solution (Blotto A purchased from Santa Cruz) at room temp for an hour. The goat anti-collagen type I antibody was diluted 1:200 in blocking buffer; 10mls was added to the membrane and incubated for 48 hours at 4°C. The blot was then washed 4 times for 10 minutes in TBST. The donkey anti-goat IgG-HRP secondary antibody was diluted 1:5000 in blocking buffer. 10mls was added to the membrane and incubated at room temperature for 1 hour. The blot was then washed 4 times for 10 minutes in TBST followed by two 5 minute washes in TBS.

7.2.11.5 Chemiluminiscent detection

Santa Cruz western blotting luminol reagent was used. 2ml of solutions A (Luminol) and B (hydrogen peroxide) were mixed and added to the blot. The luminescence was allowed to develop for two minutes. The membrane was placed in the center of the Kodak imager stage. A series of exposures were taken (30 seconds – 5 minutes) until the optimal image was taken and stored.

7.2.12 Type I collagen Immunofluorescence analysis in Type I OI cells

$5x10^5$ cells per cell line were seeded into two wells containing three 1% gelatin coated coverslips and allowed to grow for 3 days. The cells were carefully washed 2 times in PBS and fixed with a 1:1 acetone: methanol solution for 30 minutes. Cells were carefully washed 3 times with PBS and incubated in a 1% BSA blocking solution for an hour at room temperature. 2 washes in PBS were carried out. 500µl of the primary antibody a monoclonal antibody mouse anti-human Col1a1 (C-2456 purchased from Sigma) diluted 1:2000 was applied to the cells which were incubated at RT for an hour. Following 3 washes in PBS 500µl of a 1:5000 dilution of the secondary antibody a goat anti mouse Cy™3 conjugated antibody (111-165-143, Jackson Immunoresearch) was applied to the cells and incubated at room temperature for an hour. The cells were then washed 3 times in PBS and incubated in a 1/5000 DAPI solution for 3 minutes. A final 3 PBS washes were carried out. The coverslips were mounted using aqua polymount and analysed using fluorescence microscopy.

7.2.13 Microscopic analysis of Type I collagen Immunofluorescence in Type I OI cells

The cells were first visulaised using the DAPI specific filter stetting of the microscope, an image of the cell population in that view was taken, the Cy™3 specific filter setting was then applied so that the type I collagen produced by the cells could be visualised. The DAPI and Cy™3 images were overlaid using the microscope software so that an estimation of the amount of type I collagen produced by the cell population could be made. The total cell population in a give image was counted so that the fluorescence intensity could be assessed relative to the number of cells present that should be expressing type I collagen.

7.3. Results

7.3.1 Generation of Col1a1 genomic/cDNA replacement gene.

The RNAi suppression strategy undertaken during the course of the work carried out in this thesis was a mutation independent approach in which the siRNAs and shRNAs were designed to target the coding sequence of both mutant and wild type Col1a1 alleles. As expression from both alleles is reduced using such a strategy it will most likely be necessary to provide a Col1a1 replacement gene refractory to suppression due to the inclusion of subtle sequence alterations at the wobble base of each codon comprising the siRNA/shRNA target site. In addition to evading suppression, it is essential that the Col1a1 replacement gene expresses close to endogenous levels in a tissue specific manner.

When designing the Col1a1 replacement gene it was decided to include as much of the regulatory sequences present in the human Col1a1 minigene (2.4kb of the 5'UTR, the first intron and 2kb of the 3'UTR sequences) which has been shown to have an almost identical expression profile as endogenous Co1a1 both *in vivo* and *in vitro* (Breault, et al., 1997; Olsen, et al., 1991; Sokolov, 1993; Sokolov, et al., 1995). As the exact sequence of the 5' UTR and 3'UTRs of the Col1a1 minigene was not known these regions of the minigene construct were sequenced first. This was carried out successfully and the full sequence of the minigene was compiled using the newly obtained sequence data generated from sequencing the 5' UTR and 3'UTRs of the Col1a1 minigene and the sequences of the first and last 6 introns and exons present in the minigene obtained from the Col1a1 gene reference sequence AF017178. The compiled sequence was submitted and accepted by the NCBI data base, it is called "Synthetic construct mutant collagen alpha 1(I) chain (Col1a1) gene" accession number AY590263 (Appendix 1). As exons 6 – 45 are not preset in the minigene an alternative source of this sequence was required. The Col1a1 cDNA construct pHUC contains the entire human Col1a1 cDNA but no regulatory sequences this construct was selected to supply the required Col1a1 coding sequences. Using the assembled Col1a1 minigene

sequence AY590263 and the Col1a1 mRNA sequence NM_000088, restriction maps were generated and a cloning strategy of the Col1a1 genomic/cDNA replacement gene construct designed (Figure 7.2).

As it was necessary to have a unique KpnI site to facilitate insertion of the 5'UTR sequences the KpnI site was removed from the multiple cloning site of pcDNA3.1- plasmid, the Hind III site was also destroyed in the process. Subsequently each of the required Col1a1 regulatory or coding sequence DNA fragments was successfully isolated and inserted into pcDNA3.1^{-} in which the KpnI and HindIII sites had been removed. The DNA fragments were inserted in following order **1)** the 4376bp EcoRI Col1a1 cDNA fragment excised from the pHUC, **2)** Introns 50, 51 and exons 51 and 52, plus an additional 1.5kb of 3' sequences amplified from the minigene via PCR 4 and **3)** 4017bp of genomic DNA incorporating the Col11a1 promoter, exons 1, part of exon 2 and intron 1 excised from the minigene were successfully ligated together generating the Col1a1 genomic/cDNA replacement gene (Figure 7.1.1 to Figure 7.1.4).

7.3.2 Sequencing of the Col1a1 genomic/cDNA replacement gene.

It was essential to ensure that no mutational events or errors had occurred during the cloning of the construct, therefore the entire 10,575bp of the Col1a1 genomic/cDNA replacement gene was sequenced. This was carried out by a commercial company MWG UK. The predicted sequence of the Col1a1 genomic/cDNA replacement gene was assembled and provided to MWG as requested. The sequencing data generated from the construct by MWG Biotech was aligned with this reference sequence and returned (see appendix 2).This sequence alignment was then analysed for the presence of any mutations or cloning errors, a total of 50 mismatches were present. The majority of these were either polymorphisms in exons (10) or sequence variations in either intronic (11), 5'UTR (2) or 3'UTR (22) sequences. The five remaining mismatches represented two codon changes, a C to T point mutation in codon 17 of exon 21 changing the codon from a CAA (Proline) to a TAA resulting in the introduction of a stop codon (non-sense

mutation). There was a deletion of a C residue in codon 17of exon 27 and codon 14 of exon 28. Finally, a missense mutation was present in codon 31 of exon 40, two consecutive G residues had been replaced by two A residues changing the codon GGT (Glycine) to AAT (Asparagine) (Figure 7.2 and appendix 2). It was necessary to confirm that the mutations detailed above were indeed present in the construct. A series of mutation detection primers were designed termed StopF, StopR, Gly-AsnF, Gly-AsnR, DelF and DelR (sequences detailed in section 2.3). The Col1a1 genomic/cDNA replacement gene was sequenced (MWG Biotech UK) using these specific mutation detection primers. It was found that the nonsense (CAA – TAA) and missense Glycine to Asparganine (GGT-AAT) mutations were present (Figure 7.3 and 7.6), but neither of the deletions were present (Figure 7.9) indicting that there must have been sequencing errors in the original sequence returned from MWG Biotech. It was essential to correct the mutations and restore the wild type coding sequence of the Co1a1 genomic/cDNA replacement gene. The remaining nucleotide changes were not confirmed by sequencing as they were polymorphisms or variations in the 5'UTR and 3'UTR and not expected to have any significant impact upon the expression of the Col1a1 genomic/cDNA replacement gene. Although it is notable that it is not possible to be certain of this and hence evaluation of the functionality of this construct was essential.

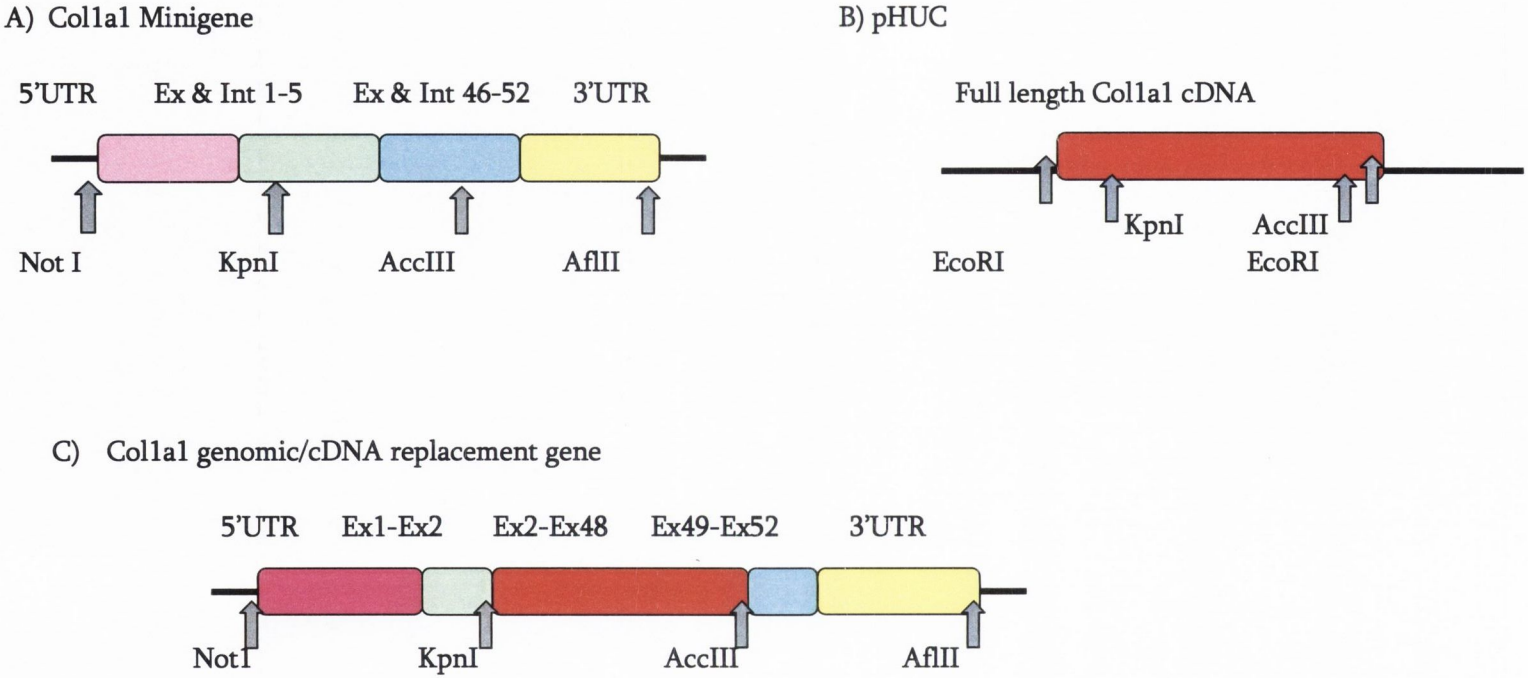


Figure 7.1.4 Cloning design for generation of the Col1a1 genomic/cDNA replacement gene. Sequences from the **(A)** Col1a1 minigene and **(B)** pHUC were used to generated the **(C)** Col1a1 genomic/cDNA replacement gene see Figure legend for details.

Figure 7.1.4 Cloning design for generation of the Col1a1 genomic/cDNA replacement gene. In order to generate a Col1a1 replacement gene it was essential to include the optimal amount of regulatory sequences as well as the protein encoding sequences. The Col1a1 minigene **(A)** contains 2.4kb of 5UTR, the first 5 and last 6 exons and 2kb of 3'UTR of the Col1a1 gene. As it expresses at close to endogenous levels *in vitro* and *in vivo* it was decided to include these regulatory sequences in the Col1a1 replacement. The exons 49 – 53, introns 50 – 51 and the 3'UTR were PCR amplified using primers which bound at the indicated AccIII and AflII sites, and inserted into the AccIII and AflII sites of pcDNA3.1- which contained the Col1a1 cDNA which had been isolated from **(B)** pHUC using the indicated EcoRI sites. The 5' UTR, Exon1, intron 1 and part of exon 2 were removed from the Col1a1 minigene using the enzymes NotI and KpnI. This 4kb fragment was cloned into pcDNA3.1- containing the 3'UTR and Col1a1 cDNA which had been digested with Not I and KpnI to generate **(C)** the Col1a1 genomic/cDNA replacement gene. This construct contained 2.5kb of 5'UTR, exons 1-52, introns 1, 50-51 and 1.5kb of 3'UTR.

7.3.3 Evaluation of the timing of the Col1a1 genomic/DNA replacement gene mutatagenesis.

At the time the mutations were identified and verified via the additional sequencing reactions described in section 7.3.2, it was unknown if the mutations had arisen spontaneously during the cloning process, or if they were in fact present in the original 4376bp EcoRI digest cDNA from the pHUC construct. This was explored by using the mutation detection primers StopF, StopR, Gly-AsnF, Gly-AsnR, DelF and DelR to sequence the pHUC Col1a1 cDNA construct, this sequencing was carried out by MWG Biotech UK. The mutations were not present in the pHUC Col1a1 DNA (Figures 7.4, Figure 7.7 and Figure 7.10). Therefore it was likely that the mutations had arisen spontaneously during the cloning process. As mentioned in the introduction, the Col1a1 gene is technically difficult to work with due to its large size and its highly repetitive sequences. Each cloning step had to be repeated many times due to the occurrence of sequence rearrangements during the cloning process.

7.3.4 Correction of the mutations which had arisen during generation of the Col1a1 genomic/cDNA replacement gene.

As the mutations were located within the region of the Col1a1 sequence that is present in the original pHUC cDNA, it was possible to correct the sequence errors by replacing the mutated Col1a1 sequence with the correct wild type sequence present in the pHUC using a single cloning step (Figure 7.10). A KpnI and ClaI digest fragment spanning exons to 2 to 52 of the pHUC cDNA sequences was excised from pHUC and ligated into a KpnI and ClaI digested Col1a1 genomic/cDNA replacement plasmid. Use of KpnI which cuts in exon 2 and ClaI which cuts in exon 52 facilitated the retention of the essential non coding sequences (i.e. the first intron, the Col1a1 promoter and the 3'UTR) in the Col1a1 genomic/cDNA replacement gene. Introns 50 and 51 were lost from the original construct as they are not present in the KpnI – ClaI fragment digested from the pHUC which was used to replace the mutation containing sequences. Notably there appears to be little evidence in the literature detailing the requirement of Introns

50 and 51 of the Col1a1 gene for high level tissue specific expression and consequently it was thought that the loss of these sequences may not have a detrimental impact of the functionality on the Col1a1 genomic/cDNA replacement gene. However, the loss of the introns highlighted the need to assess the functionality of the Col1a1 replacement gene in terms of Col1a1 expression. The presence of the correct wild type sequence was assessed using the mutation detection primers StopF, StopR, Gly-AsnF, and Gly-AsnR. The miniprep with the wild type sequence present (Figures 7.5 and 7.8) was selected and taken forward in the study. This work was carried out with the assistance of Dr. S. Millington-Ward, TCD. The full length Col1a1 replacement genomic cDNA construct is currently being sequenced in its entirety.

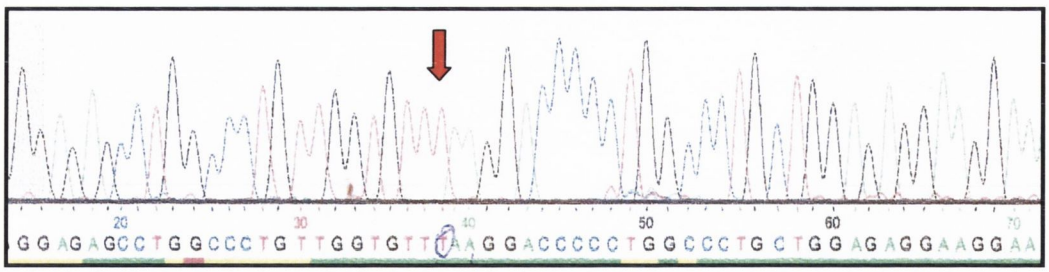


Figure 7.2 Chromatogram of sequence derived from the Col1a1 genomic/cDNA replacement gene using the mutation detection primer StopF. It was necessary to confirm the presence of the nonsense mutation in which a C – T point mutation changed codon 6, CAA – TAA in exon 21 of the Col1a1 genomic/cDNA construct. This was revealed to be the case, the T base change is indicated by the red arrow.

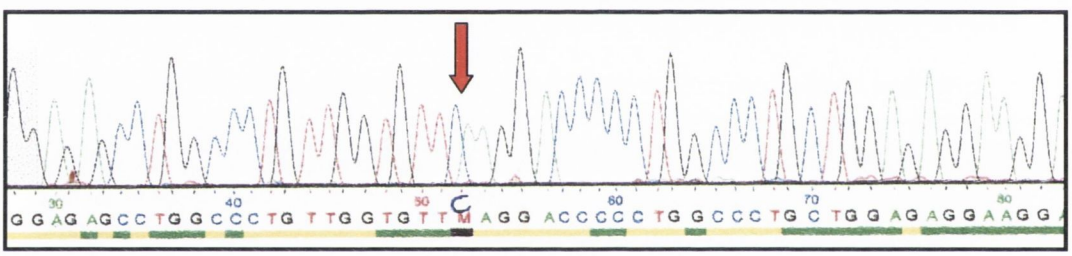


Figure 7.3 Chromatogram of sequence derived from pHUC Col1a1 cDNA construct using the mutation detection primer StopF. It was necessary to determine if the mutation detected in the Col1a1 genomic/cDNA replacement gene (Figure 7.2) had occurred spontaneously or if it was present in the original pHUC from which the sequence was originally obtained. The pHUC Col1a1 cDNA construct contains the wild type CAA sequence according to the trace. The wild type C sequence is indicated by the red arrow.

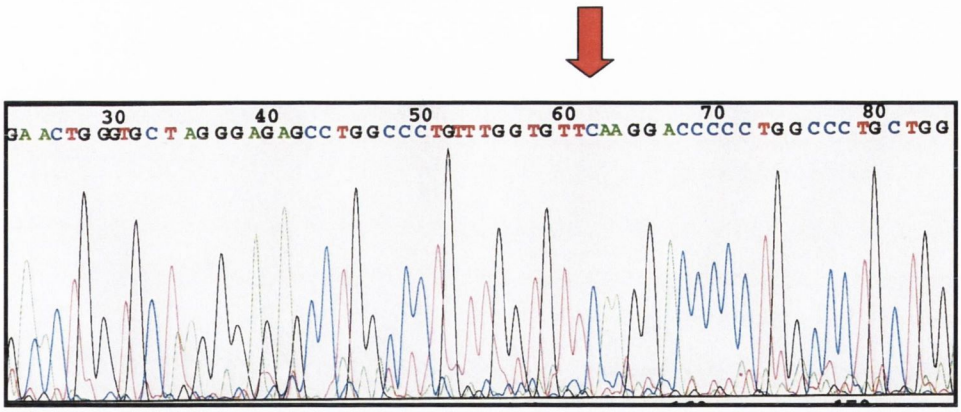


Figure 7.4 Chromatogram of sequence derived from the corrected Col1a1 genomic/cDNA replacement gene using the mutation detection primer StopF. It was necessary to confirm that the mutated sequence (TAA) had been successfully corrected (CAA) in codon 6 of exon 21 of the Col1a1 genomic/cDNA replacement gene. The red arrow indicates the required wild type CAA sequence confirming that the mutated sequence had been corrected.

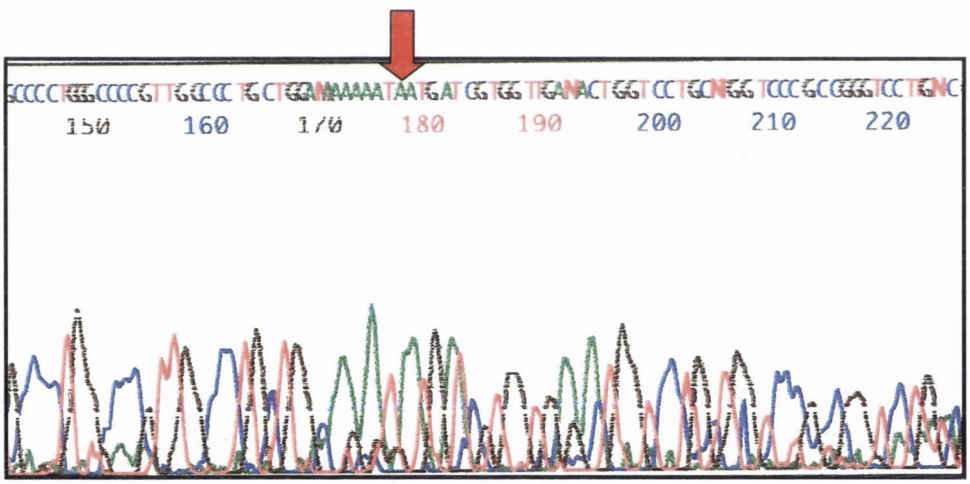


Figure 7.5 Chromatogram of sequence derived from the Col1a1 genomic/cDNA replacement gene using the mutation detection primer Gly-AsnF. It was necessary to confirm the presence of the missense mutation in which two consecutive GG nucleotides were point mutated to two AA nucleotides, changing a codon 31 Glycine to Asparagine (GGT-AAT) in exon 40 of the Col1a1 gene. The red arrow indicates the mutated AAT codon.

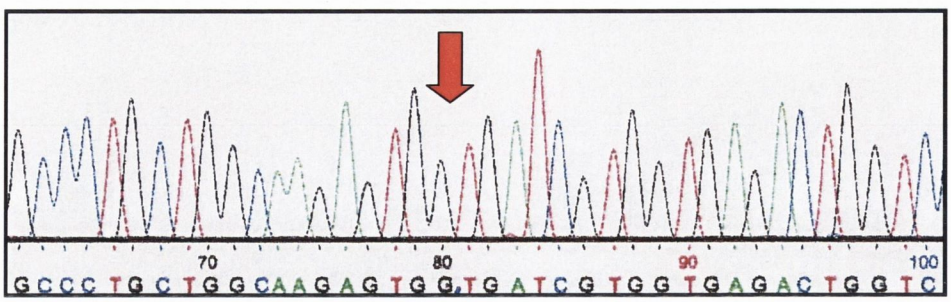


Figure 7.6 Chromatogram of sequence derived from pHUC Col1a1 cDNA construct using the mutation detection primer Gly-AsnF It was necessary to determine if the mutation detected in the Col1a1 genomic/cDNA replacement gene (Figure 7.5) had occurred spontaneously of if it was present in the original pHUC Col1a1 cDNA construct from which the sequence was originally obtained. The pHUC Col1a1 cDNA construct contains the wild type GGT sequence as indicated by the red arrow above.

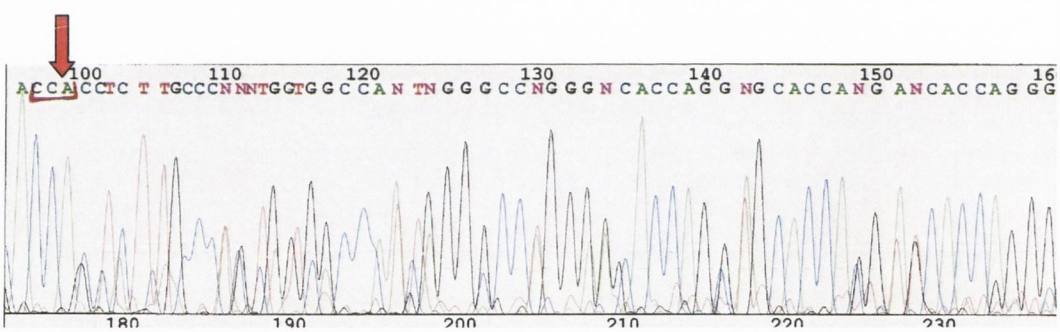


Figure 7.7 Chromatogram of sequence derived from Col1a1 genomic/cDNA replacement gene using the mutation detection primer Gly-AsnR It was necessary to determine if the mutation detected in the Col1a1 genomic/cDNA replacement gene (Figure 7.5) was now replaced by wild type sequence isolated from pHUC Col1a1 cDNA construct. The sequence contains the wild type CCA sequence as indicated by the red arrow above. This sequence was generated using a reverse primer, the ACC is the corrected GGT glycine codon 31 of exon 40

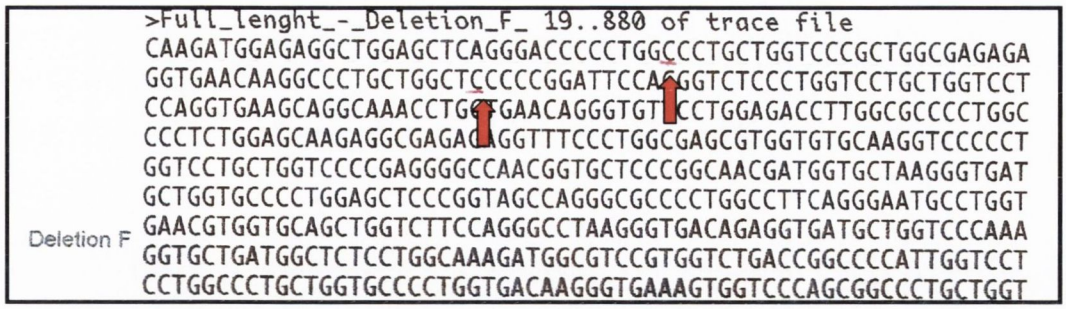


Figure 7.8 Sequence derived from the Col1a1 genomic/cDNA replacement gene using the mutation detection primer DelF. It was necessary to confirm the presence of the two separate C deletions; the first in codon 17 of exon 27 and the second in codon 14 of exon 28 sequencing revealed that these mutations were not present. The C nucleotides are indicated with the red arrows.

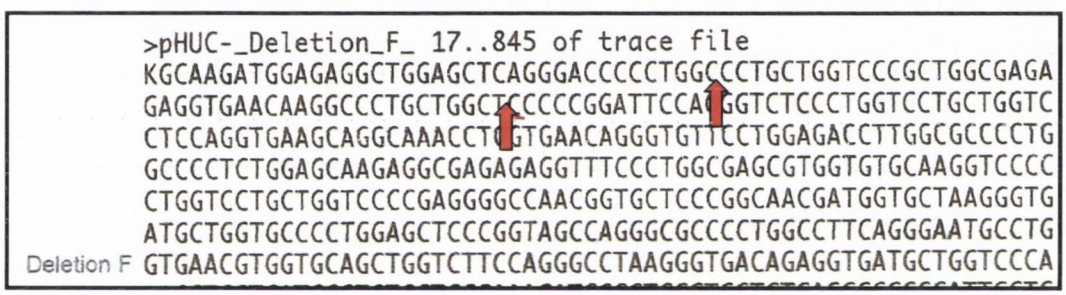


Figure 7.9 Sequence derived from pHUC Col1a1 cDNA construct using the mutation detection primer DelF. The pHUC was sequenced to determine if the C nucleotides apparently deleted from the Col1a1 genomic/cDNA replacement are actually present in the pHUC Col1a1 cDNA construct. Both C residues as indicated with red arrows are present in the pHUC Col1a1 cDNA construct.

A) Nonsense mutation Exon 21.

TTGGTGTTCAAGGA
TTGGTGTTTAAGGA

CAA – TAA
Proline to Stop codon

Codon 6 of exon 21

B) Missense mutation Exon 40

AAGAGTGGTGATCGT
AAGAGTAATGATCGT

GGT - AAT
Glycine to Aspargaine codon

Codon 31 of exon 40

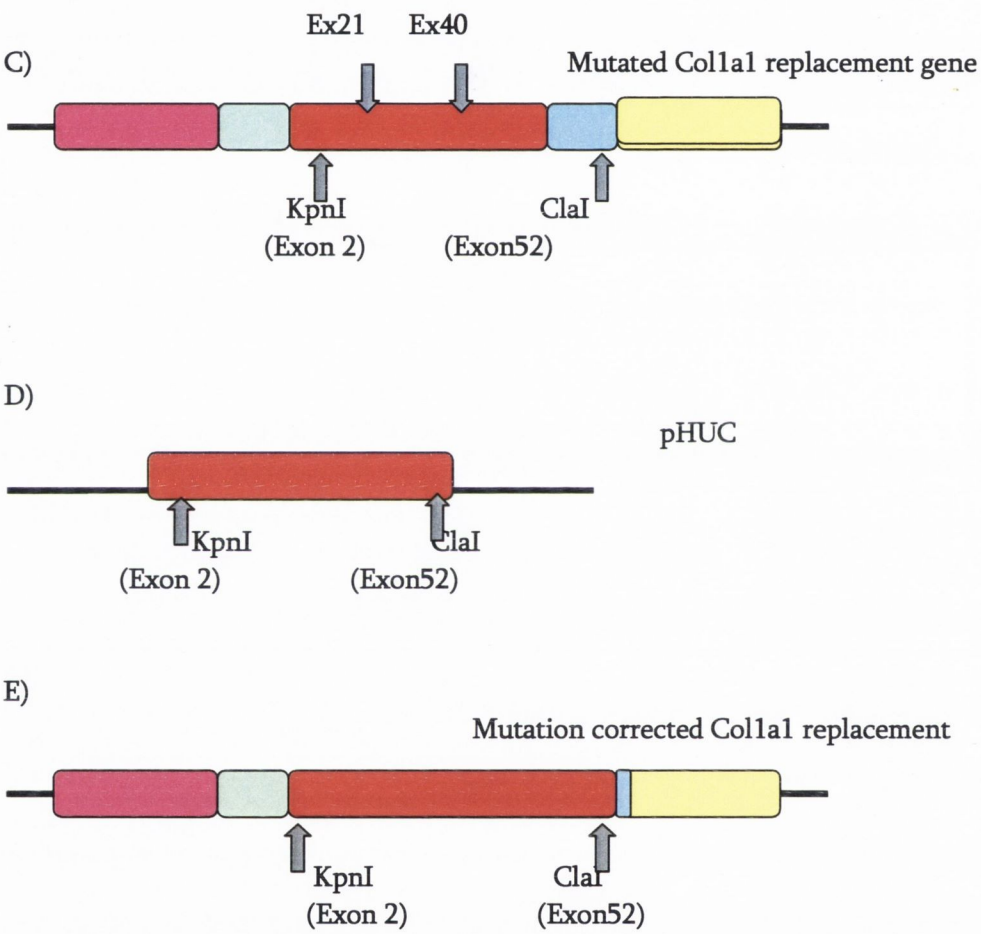


Figure 7.10 Mutated Col1a1 genomic/cDNA replacement gene and repair strategy. Sequencing of the Col1a1 genomic/cDNA replacement gene revealed the presence of two mutations **(A)** a nonsense mutation in exon 21 and **(B)** a missense mutation in exon 40. It was confirmed that the mutations arose spontaneously during the generation of the Col1a1 genomic/cDNA replacement gene and that the sequence of pHUC Col1a1

cDNA construct was the intact wild type Col1a1 sequence. Therefore it was possible to repair the sequence errors by removing the mutated sequence from the Col1a1 genomic/cDNA construct **(C)** using the indicated enzymes KpnI and ClaI which cut in exon 2 and 52 respectively. This sequence was replaced by a KpnI and ClaI digest fragment excised from pHUC **(D)**. The repaired Col1a1 genomic/cDNA construct no longer has introns 50 and 51 as they were removed from the Col1a1 genomic/cDNA replacement gene and are not present in the Col1a1 cDNA isolated from the pHUC construct which was used to replace the mutated sequences in the Col1a1 genomic/cDNA replacement gene. However, the essential 5' UTR, intron1 and 3'UTR sequences were retained.

7.3.5 Optimisation of Mov13 nucleofection.

Homozygous Mov13 mice are embryonic lethal due to the absence of type I collagen production. This is due to the presence of a retroviral insertion in intron 1 of both Col1a1 alleles which prevents transcription of the Col1a1 gene. The Mov13 cell line is a transformed foetal fibroblast cell line. The cells were isolated from homozygous Mov13 mice at day 12 of gestation and SV40 transformed generating an immortalised cell line (Schnieke, et al., 1987). This cell type is ideal for evaluation of expression of the Col1a1 genomic/cDNA replacement gene as any detectable Col1a1 mRNA expression should in principle be from the introduced replacement gene

It was first necessary to optimise plasmid delivery to the Mov13 cells using the nucleofector device. This electroporator uses cell type specific solutions and electrical parameters to deliver nucleic acids to the nucleus of the cells, the composition of the solutions, electrical parameters and delivery mechanism are not provided by the manufacturer Amaxa. Optimisation involved evaluation of the two transformed mouse fibroblast specific nucleofection programs termed A-23 and A-24. Two nucleofection reactions with pmaxGFP plasmid were carried out for each program and the cell survival and nucleofection efficiencies assessed 48 hours later. The cell survival rates for cells nucleofected using program A-23 were estimated to be the region of 65% and in the region of 50% for program A-24. Based upon the evaluation of eGFP fluorescence from the nucleofected pmaxGFP plasmid transfection efficiencies in the region of 50% were achieved using A-23 (Figures 7.13 - 7.14) and transfection efficiencies in the region of 40% were achieved using A-24 (Figure 7.11 - 7.12). This experiment was carried out a single time as cell survival rates (65%) and nucleofection efficiencies (50%) using A-23 was sufficiently high to proceed to the evaluation of the Col1a1 genomic/cDNA replacement gene. Optimal nucleofection efficiencies are 48% according to Amaxa (www.amaxa.com).

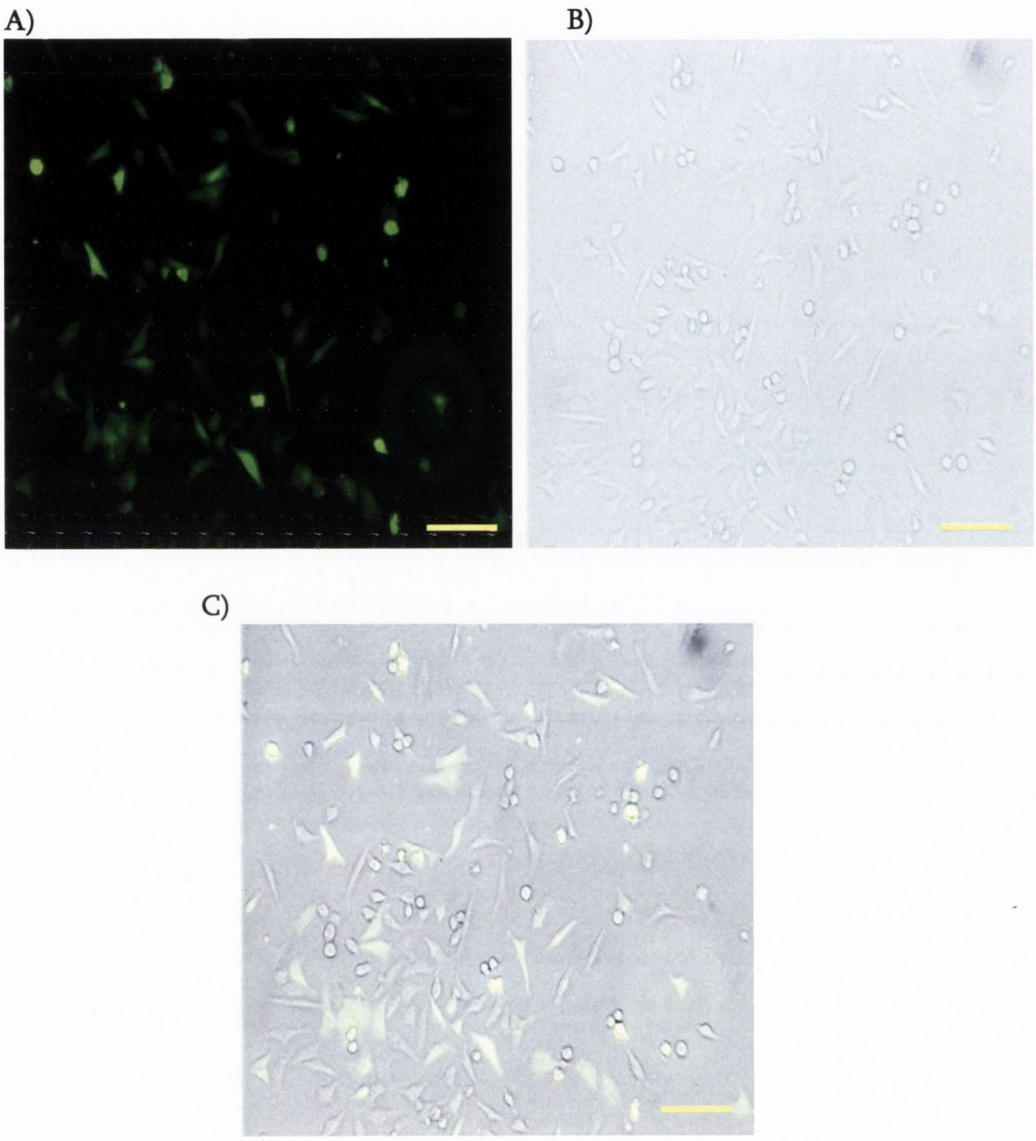


Figure 7.11 Evaluation of Mov13 nucleofection efficiencies. 1x10^6 Mov13 cells were nucleofected with 2µg of pmaxGFP plasmid using program A-24. The cells were plated onto 1% gelatin coated coverslips and GFP expression assessed 48 hours later. **A)** GFP expressing Mov13 cells, **B)** Brightfield of total cell population and **C)** Overlay of A & B, Nucleofection efficiencies in the region of 40% were achieved with program A-24. Scale bar 40µm

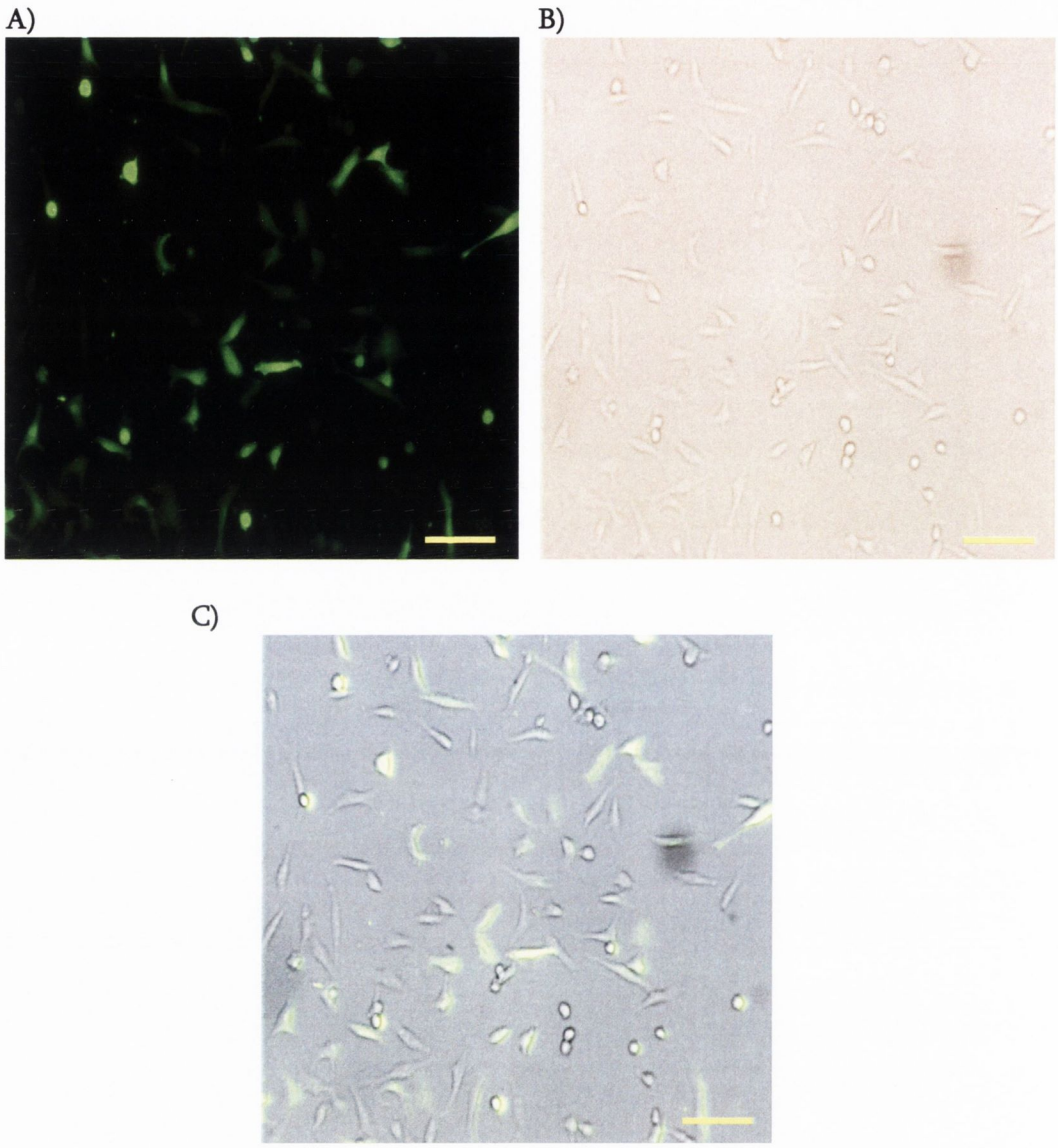


Figure 7.12 Evaluation of Mov13 nucleofection efficiencies. $1x10^6$ Mov13 cells were nucleofected with 2μg of pmaxGFP plasmid using program A-24. The cells were plated onto 1% gelatin coated coverslips and GFP expression assessed 48 hours later. A) GFP expressing Mov13 cells, B) Brightfield of total cell population and C) Overlay of A & B. Nucleofection efficiencies in the region of 40% were achieved with program A-24. Scale bar 20μm.

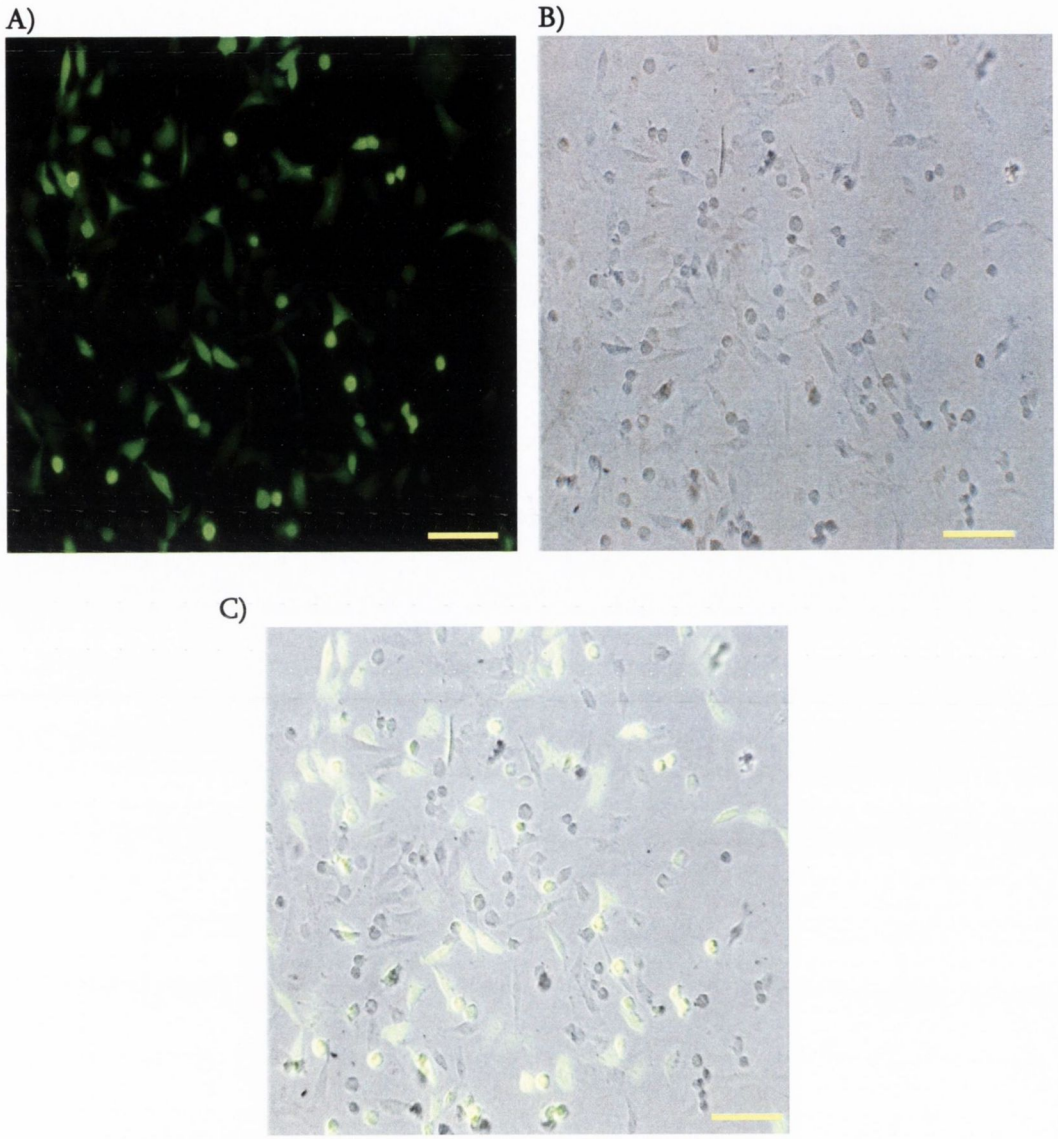


Figure 7.13 Evaluation of Mov13 nucleofection efficiencies. 1x10^6 Mov13 cells were nucleofected with 2μg of pmaxGFP plasmid using program A-23. The cells were plated onto 1% gelatin coated coverslips and GFP expression assessed 48 hours later. A) GFP expressing Mov13 cells, B) Brightfield of total cell population and C) Overlay of A & B, Nucleofection efficiencies in the region of 50% were achieved with program A-23. Scale bar 40μm.

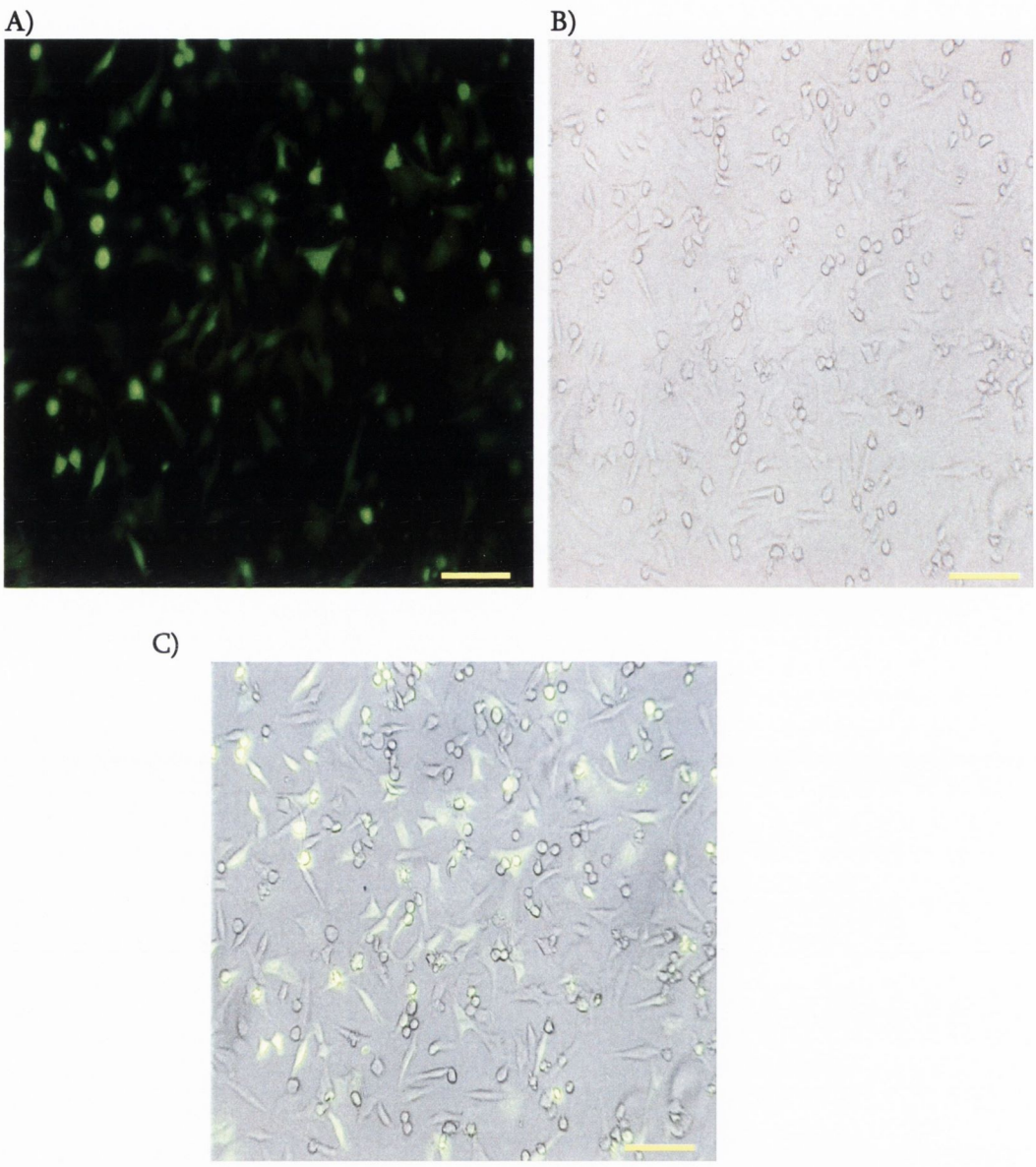


Figure 7.14 Evaluation of Mov13 nucleofection efficiencies. $1x10^6$ Mov13 cells were nucleofected with 2μg of pmaxGFP plasmid using program A-23. The cells were plated onto 1% gelatin coated coverslips and GFP expression assessed 48 hours later. A) GFP expressing Mov13 cells, B) Brightfield of total cell population and C) Overlay of A & B, Nucleofection efficiencies in the region of 50% were achieved with program A-23. Scale bar 40μm.

7.3.6 Generation of Mov13 cells stably expressing the Full Length Col1a1 genomic/cDNA replacement gene.

The optimised Mov13 protocol detailed in section 7.2.3 was used to transfect Mov13 cells with the Col1a1 genomic/cDNA replacement gene. In each experiment the Col1a1 genomic/cDNA replacement gene was nucleofected in triplicate. G418 selection was commenced 48 hours post nucleofection. Media was changed every day for 8 days until the control cells nucleofected in the absence of DNA had died. Subsequent to this G418 selective medium was changed every second day until distinct colonies had formed. Nucleofection of the Mov13 cells was efficient with a large number of colonies forming in each well. Mov13 cells are a very robust cell type with a fast doubling time, it was necessary to continue the G418 selection (500μg.ml) for two weeks after pooling of the colonies prior to analysis to ensure that the cells present were expressing the Col1a1 genomic/cDNA gene. A total of 3 cell lines were generated as described. Therefore each cell line represented 3 independent transfections, with all 3 cell lines representing a total of 9 independent transfections; consequently the cell lines were not clonal cell lines and will most likely have integrated the Col1a1 genomic/cDNA replacement gene in a variety of different locations in different cells.

7.3.2 Real Time RT-PCR evaluation of Col1a1 genomic/cDNA replacement gene expression levels in Mov13 fibroblasts

For each stable cell line, MovCol1, MovCol2 and MovCol3 3 wells of a 6 well plate were seeded with $5x10^5$ cells and 2 wells were seeded with $5x10^5$ Mov13 and the RNA extracted 48 hours later (section 2.9.1). This was carried out 3 times generating 9 RNA samples per cell line. As the cell lines were in principle stably expressing the Col1a1 genomic/cDNA replacement gene, the variability in transfection efficiencies was eliminated from this experiment. The 3 RNAs samples generated per RNA transfection experiment were pooled to generate a single RNA representing each triplicate RNA extraction. This generated a total of 3 pooled RNA samples per stable cell line MovCol1 – MovCol3. Col1a1 expression levels were assessed using real-time RT-PCR all

reactions were set up as previously described (section 2.9.4) and levels of Col1a1 expression calculated using the relative quantification method. Mouse specific GAPDH primers were not available the time this RT-PCR analysis was carried out, primers for an alternative housekeeping gene elongation factor 2 (EEF2) were present in the lab and were used instead. Col1a1 expression levels in all samples were normalised to EEF2 (i.e. Col1a1 expression value in a sample divided by the EEF2 expression value in same sample). The relative expression values were determined by comparing the Col1a1 expression values in the Mov13 cells stably expressing the human Col1a1 genomic/cDNA replacement gene (MovCol1, MovCol2 and MovCol3) to Col1a1 expression in untransfected Mov13 cells. Col1a1 expression values in the Mov13 cells were set as 1 and the relative fold change determined by direct comparison. All data generated was analysed using GraphPad InStat3, the mean and standard deviation were calculated. Col1a1 expression levels in MovCol1, MovCol2 and MovCol3 were found to be 11.9 fold (±3.32). 10.2 fold (±2.7) and 10.46 (±2.67) fold higher than Col1a1 expression in untransfected Mov13 cells respectively (Table 7.1). It is evident from this data and also clearly illustrated by the agarose gel analysis of the real time PCR products from one of the runs (Figure 7.16 and Figure 7.17) that the Col1a1 genomic/cDNA construct being expressed in the stable cell lines MovCol1 – MovCol3. In addition, a two tailed student t-test with two degrees of freedom was carried out, the levels of Col1a1 expression in each of the Stable cell lines MovCol1-MovCol3 was found to be significantly higher than the Col1a1 expression values in untransfected Mov13 control cells with p values ranging from $p=<0.0001$ – $p= 0.0003$ (Table 7.1 and Figure 7.11)

		Mov13	*MovCol1*	*MovCol2*	*MovCol3*
EEF2	*Average*	1	11.9	10.2	10.46
	Std Dev	0.227	3.32	2.7	2.67
	Ttest		P<0.0001	p=0.0003	p=0.0001

Table 7.1 Summary of Real Time RT-PCR expression data from Mov13 and MovCol cells. Three independent Mov13 cell lines MovCol1 – MovCol3 which stably express the genomic/DNA Col1a1 replacement gene were generated. $5x10^5$ cells from each cell line Mov13 and MovCol1 – MovCol3, were plated per well of a 6 well plate. 48 hours later RNA was Trizol extracted and Col1a1 expression levels assessed using Real Time RT-PCR. Each RNA was plated in triplicate on each 96 well plate, each run was set up three times. The housekeeping gene employed was elongation factor 2 (EEF2), normalised Col1a1 expression values for each cell line MovCol1 – MovCol3 were compared to normalised Col1a1 expression values from the control Mov13 cells and the fold difference in expression determined. Col1a1 expression in Mov13 cells was set to 1. Col1a1 expression was 11.9 fold higher in MovCol1 cells, 10.2 fold higher in MovCol2 cells and 10.46 fold higher in MovCol3 cells. In each case the difference between Col1a1 expression in the stable cell lines MovCol1- MovCol3 and the Mov13 control untransfected cells was found to be statistically significant when evaluated using a two tailed student t-test using GraphPad InStat, p values and standard deviations are shown for each cell line.

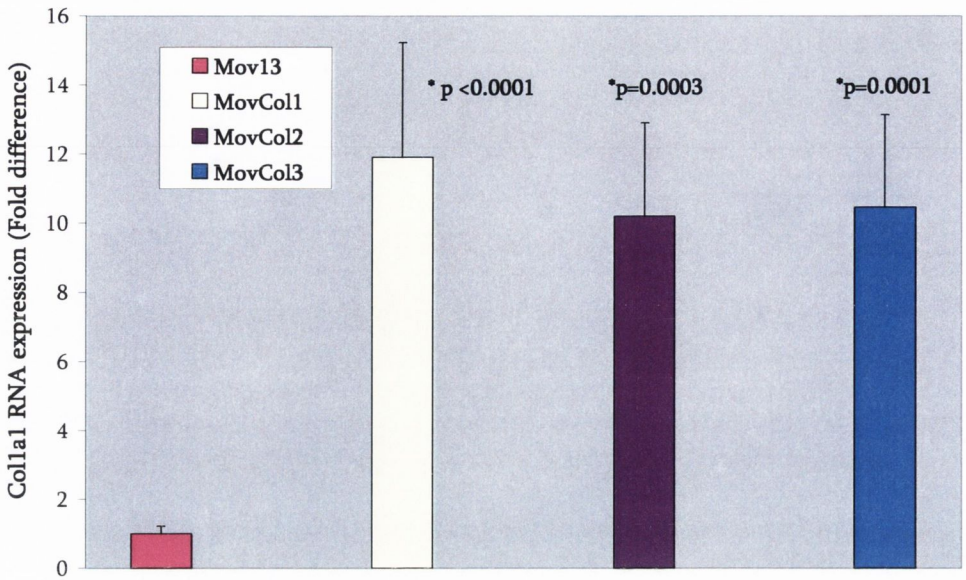


Figure 7.15 Graph of Col1a1 RNA expression levels in Mov13 cells and MovCol1a1 cells. 3 independent Mov13 cell lines MovCol1 – MovCol3 stably expressing the genomic/cDNA Col1a1 replacement gene were generated. $5x10^5$ cells from each cell line Mov13 and MovCol1 – MovCol3 was plated in triplicate in a 6 well plate. 48 hours later RNA was extracted and Col1a1 expression levels assessed using real time RT-PCR. Each of the RNA samples were plated in triplicate on each 96 well plate and each run carried out 3 times. Col1a1 expression values were normalised to EEF2 expression values for each cell line MovCol1 – MovCol3. Normalised MovCol Col1a1 expression values were compared to normalised control Mov13 Col1a1 expression values and the fold difference in expression determined. Col1a1 expression in Mov13 cells was set to 1. Col1a1 expression was 11.9 fold higher in MovCol1, 10.92 fold higher in MovCol2 and 10.46 fold higher in MovCol3. In each case the difference between Col1a1 expression in the stable cell lines MovCol1- MovCol3 and the Mov13 control untransfected cells was found to be statistically significant when evaluated using a two tailed student t-test using GraphPad InStat, p values and standard deviations are shown for each cell line.

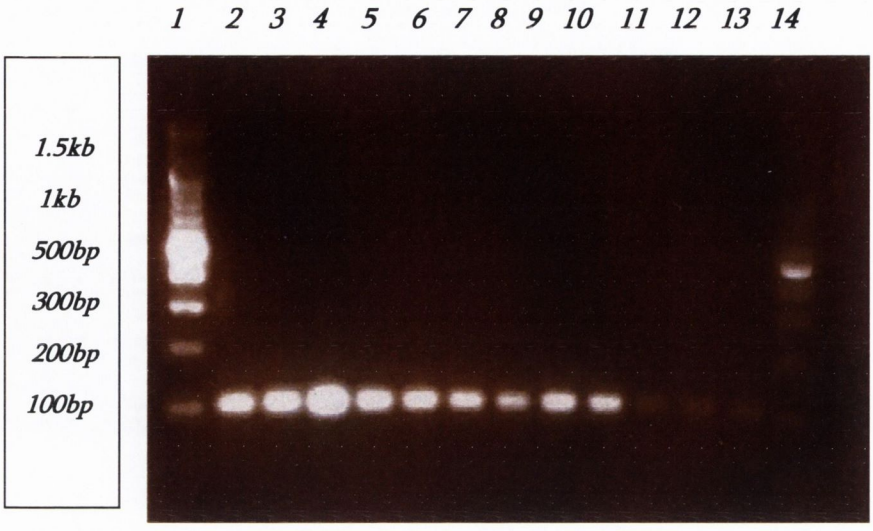


Figure 7.16 Gel electrophoresis of Col1a1 Real Time RT-PCR cDNA products generated from RNA extracted from MovCol1 – MovCol3 and Mov13 cells. RNA was extracted from Mov13 cells and Mov13 cells which stably expressed a full length Col1a1 replacement gene (MovCol1- MovCol3 cells). Col1a1 expression was assessed using Real Time RT-PCR, each sample was analysed in triplicate in the RT-PCR run and Col1a1 cDNA products generated visualised on a 2% agarose gel. Lane 1, 100bp DNA ladder, Lanes 2-4 MovCol1, Lanes 5-7 MovCol2, Lanes 8-10 MovCol3, Lanes 11-13 Mov13, Lane 14 100bp ladder. Expected RT-PCR product size 110bp

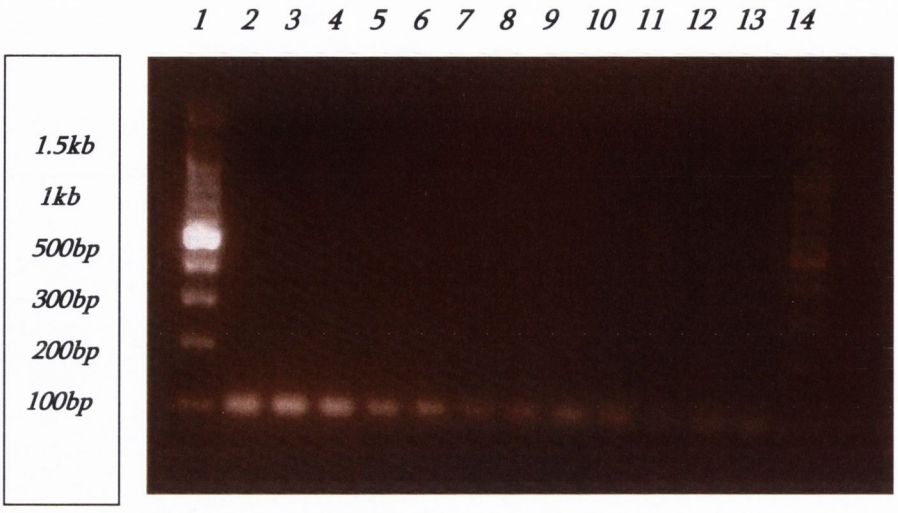


Figure 7.17 Gel electrophoresis of EEF2 Real Time RT-PCR cDNA products generated from RNA extracted from MovCol1 – MovCol3 and Mov13 cells. RNA was extracted from Mov13 cells and Mov13 cells which stably expressed a full length Col1a1 replacement gene (MovCol1- MovCol3 cells). EEF2 expression was assessed using Real Time RT-PCR, each sample was analysed in triplicate in the RT-PCR run and EEF2 cDNA products generated visualised on a 2% agarose gel. Lane 1, 100bp DNA ladder, Lanes 2-4 MovCol1, Lanes 5-7 MovCol2, Lanes 8-10 MovCol8, Lanes 11-13 Mov13, Lane 14 100bp ladder. Expected RT-PCR product size 98bp.

7.3.3 RT-PCR analysis of Col1a1 expression in Type I OI fibroblasts.

In addition to evaluation of expression of the Col1a1 genomic/cDNA replacement gene in Mov13 cells, a second cell line which originated from a patient with Type I OI was used to evaluate Col1a1 expression from the Col1a1 genomic/cDNA replacement gene. A total of 3 independent cell lines were generated (one from each Col1a1 nucleofection experiment). For each cell line 57OI-C1, 57OI-C2, 57OI-C3 and 57OI, 2 wells of a 6 well plate were seeded with $5x10^5$ cells. 48 hours later RNA was Trizol extracted (section 2.9.2). This was repeated 2 times, generating a total of 4 RNA samples per cell lines. It was only possible to carry out this low number of replicates as the Type I OI cells are primary cells which have a finite passaging capacity. The cells were received at passage 6, it was advised not to use the cells beyond passage 10 - 12 as the growth properties and expression profile of the cells begin to change (personal communication Dr.M.Valli). In addition one could hypothesise that the fragile nature of the cells may in part be responsible for the inability to generate a stable cell line using the OI47 Type I OI cell line. Therefore the passage number of the cells was kept at passage 10 or less to ensure the reproducibility of the results. The housekeeping gene used was GAPDH and reactions were set up as described in section 2.9.4. Each individual RNA sample was plated in triplicate per 96 well plate. Each RT-PCR reaction was carried out three times. The relative quantification method described in 7.2.9 was used to determine Col1a1 expression levels. The levels of Col1a1 expression in the untransfected control cells was set as 1 and the relative levels of Col1a1 expression in each of the stable cell lines 57OI-C1, 57OI-C2, 57OI-C3 was determined by direct comparison. It was found that Col1a1 expression had increased 1.64 (±0.33) fold in 57OI-C1 cells, 1.88 (± 0.39) fold in 57-OIC2 cells and 1.68 (± 0.46) fold in 57OI-C3 cells. Statistical analysis of the results was carried out using GraphPad InStat3, the mean and standard deviations were calculated. A two-tailed student t-test was carried out in which the normalised Col1a1 expression in the stable cell lines 57OI-C1, 57OI-C2, 57OI-C3 was compared to the relative Col1a1 expression values in the original 57OI Type I OI cells the increase in Col1a1 expression was found to be significant in each of the stable cell lines 57OI-C1, 57OI-C2, 57OI-C3

with p values ranging from p=0.033 to p=0.00074. This data is presented in Table 7.2 and Figure 7.14.

Type I OI is the mildest form of the disease due to a Col1a1 haploinsufficeny. The vast majority of Type I OI linked mutations in the Col1a1 allele are non sense mutations or mutations which result in premature termination of Col1a1 transcription and the Col1a1 transcripts are rapidly targeted for degradation (Willing, et al., 1996). Consequently the amount of α1 (I) chains available for incorporation into the type I heterotrimer is reduced by 50%. The dermal fibroblasts used in this study were isolated from two children with Type I OI. The children from whom the dermal fibroblasts were isolated were 2 of 20 patients with Osteogenesis Imperfecta for whom the clinical, biochemical and molecular findings of their disorder was detailed in a recent publication (Venturi, et al., 2006). This report details how of the four patients with Type I OI one was due to a nonsense mutation and the remaining three cases (which includes the two patients from which the cell lines 49OI and 57OI were obtained) were due to base substitutions in highly conserved splice sites. These mutations did not result in exon skipping indicated by the detection of only normal size Col1a1 RT-PCR product from the appropriate coding regions. The authors suggested it was most likely that a premature stop codon was introduced as a consequence of the aberrant splicing with subsequent non sense mediated decay. The cell line 49OI was isolated from patient 9 in the study and in this case the introduction of a premature stop codon has been verified, it is most likely that the null allele in the 57OI cells (isolated from patient 12 in the study) may be due to similar molecular events as the mutation is also a single base change within a highly conserved splice site. In both cases a Col1a1 null allele assay revealed heterozygosity for a 3' UTR MnlI RFLP at the genomic DNA levels but only a single allele was evident when the analysis was repeated at the cDNA level indicating that indeed each of these cell lines is only expressing functional Col1a1 allele (Venturi, et al., 2006).

As the levels of Col1a1 expression in Type I OI cells are considered to be reduced by in the region of 50%, the increase of between 1.64 (±0.33) and 1.68 (±0.46) fold in the levels of expression observed in the presence of the Col1a1 genomic/cDNA replacement gene is encouraging. It is clear that the construct is functional in human cells, in particular in cells isolated from a patient with OI. Whilst the promoter is known to drive high levels of expression it was hoped that total expression levels (i.e. expression from the functional endogenous gene and the delivered replacement gene) did not exceed the level of endogenous Col1a1 expression in wild type cells. In the case of each stable cell lines generated and analysed the expression did not exceed 2 fold. If this were the case it would be possible that the levels of Col1a1 expression may be too high from the replacement gene, which could potentially be toxic to the cells. Pochampally *et al* reported that expression of a Cola1a1 transgene in wild type MSCs resulted in significant toxicity and senescence of the cells (Pochampally, et al., 2005). Therefore it is clear that over-expression of the Col1a1 would be highly undesirable, additionally systemic sclerosis is associated with over expression of Col1a1 (Cicchillitti, 2004). Fortunately, in this *in vitro* cell model explored in the current study, Col1a1 expression levels appear to fall within what would be considered endogenous levels of expression.

		57 OI	*570I C1*	*570I C2*	*570I C3*
GAPDH	*Average*	1	1.64	1.88	1.68
	t-test		p = 0.026	P=0.0074	p=0.033
	StdDev	0.113	0.33	0.39	0.46

Table 7.2 Summary of Col1a1 RNA expression data from Type I OI and Type I OI Col1a1 stable cell lines assessed using real time RT-PCR. Three independent Type I OI cell lines, 57OI-C1, 57OI-C2 and 57OI-C3 stably expressing the genomic/cDNA Col1a1 replacement gene were generated. 5.10^5 cells from each cell line 57OI, 57OI C1, 57OI C2 and 57OI C3 was plated in duplicate in a 6 well plate. 48 hours later RNA was extracted and Col1a1 expression levels assessed using real time RT-PCR. Each RNA sample was plated in triplicate on each 96 well plate and each run carried out 3 times. Col1a1 expression values were normalised to GAPDH expression values for each cell line 57OI, 57OI C1, 57OI C2 and 57OI C3. Normalised 57OI C1 – 57OI C3 Col1a1 expression values were compared to normalised control 57OI Col1a1 expression values and the relative fold Col1a1 expression levels determined. Col1a1 expression in 57 OI cells was set at 1. Col1a1 expression was increased 1.64 fold in 57OI-C1, 1.88 fold in 57OI-C2 and 1.68 fold in 57IO-C3. In each case the difference between Col1a1 expression in the stable cell lines 57OI C1, 57OI C2 and 57OI C3 and the untransfected 57OI control cells was found to be statistically significant when evaluated using a two tailed student t-test using GraphPad InStat, p values and standard deviations are shown for each cell line.

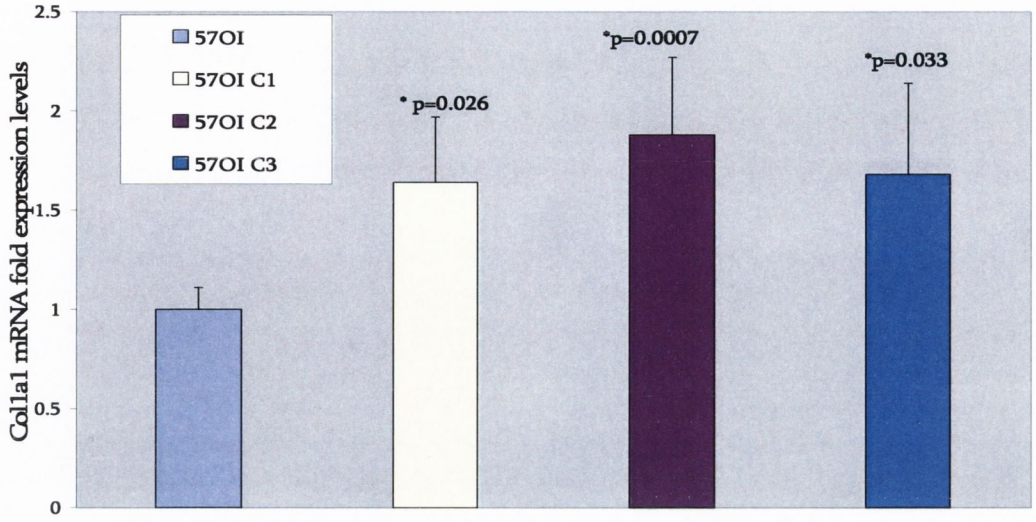


Figure 7.18 Real time RT-PCR evaluation of Col1a1 mRNA expression data from type I OI and Type I OI Col1a1 stable cell lines. Three independent Type I OI cell lines, 57OI-C1, 57OI-C2 and 57OI-C3 stably expressing the genomic/cDNA Col1a1 replacement gene were generated. $5x10^5$ cells from each cell line 57OI, 57OI C1, 57OI C2 and 57OI C3 was plated in duplicate in a 6 well plate. 48 hours later RNA was extracted and Col1a1 expression levels assessed using real time RT-PCR. Each RNA sample was plated in triplicate on each 96 well plate and each run carried out 3 times. Col1a1 expression values were normalised to GAPDH expression values for each cell line 57OI, 57OI C1, 57OI C2 and 57OI C3. Normalised 57OI C1 – 57OI C3 Col1a1 expression values were compared to normalised control 57OI Col1a1 expression values and the relative fold Col1a1 expression levels determined. Col1a1 expression in 57 OI cells was set at 1. Col1a1 expression was increased 1.64 fold in 57OI-C1, 1.88 fold in 57OI-C2 and 1.68 fold in 57IO-C3. In each case the difference between Col1a1 expression in the stable cell lines 57OI C1, 57OI C2 and 57OI C3 and the 57OI control untransfected cells was found to be statistically significant when evaluated using a two tailed student t-test using GraphPad InStat, p values and standard deviations are shown for each cell line.

7.3.4 Metra CICP ELISA of *de novo* Col1a1 expression in Type I OI fibroblasts

Levels of *de novo* synthesised Col1a1 protein in Type I OI cells following nucleofection of the Col1a1 genomic/cDNA replacement gene were assessed using the Metra CICP ELISA system which quantifies the amount of C terminal pro-peptide released by the cells over 24 hours. The release of the C pro-peptide into the media in which the cells are growing provides a stoichiometric representation of collagen production. The Metra CICP assay measures the *C*-Terminal of type I collagen *P*rotein.

$5x10^5$ cells from each stable Type I OI cell line 57 OI-C1, 57OI-C2, 57OI-C3 and 57OI media were plated in duplicate and allowed to grow for 48 hours. The media was changed for serum free α DMEM. The cells were allowed to secrete type I collagen into the medium for 24 hours. Medium was then collected, the duplicate samples from each cell line were pooled generating a single representative sample per cell line. The pooled samples were analysed for *de novo* type I collagen production according to the manufacturer's guidelines. The assay was carried out 3 times. In order to determine if there was an increase in the level of Col1a1 protein expression in the Type I OI cells stably expressing the genomic/DNA construct, the levels of C pro-peptide released by the stable cell lines 57 OI-C1, 57OI-C2 and 57OI-C3 over 24 hours was compared to the levels of C propeptide released by the control wild type 57OI Type I OI fibroblasts. It was found that the levels of Col1a1 expression had increased 1.72 (±0.26) fold in 57OI-C1 cells, 1.73 (±0.29) fold in 57OI-C2 cells and 1.63 (±0.31) fold in 57OI-C3 cells. All data was assessed using GraphPad InStat3, the mean and standard deviation was determined. In addition a two tailed student t-test with two degrees of freedom was carried out in which the levels of Col1a1 protein expression from each of the stable cell lines was compared to the levels of protein expression in the control wild type 57OI Type I OI fibroblasts. In all cases the increased levels of Col1a1 protein expression observed was considered to be very significant with p values ranging from p=0.0004 – p=0.0009. The data is presented in Table 7.3 and Figure 7.15. The assay indicates that there is between 1.63 (±0.31) and 1.73 (±0.29) fold more type I collagen production in

the cell lines stably expressing the Col1a1 genomic/cDNA replacement gene. As the Metra CICP ELISA system detects and quantifies C terminal pro-peptides which are by products of the cleavage of the pro-collagen triple helix, it appears that the Cola1a1 genomic/cDNA construct is being efficiently expressed as a functional α1(I) chains which are incorporated into type I collagen triple helices. The total level of type I collagen production does not exceed twice the level of Col1a1 production observed in the original 57OI Type I OI cell lines, yet the increases observed are significant. This result is encouraging, as mentioned previously it is essential that the total level of type I collagen production does not exceed endogenous wild type expression levels It is notable that it may have been of value to also include a normal wild type fibroblast cell line was not included in the study, however as Type I OI patients only have one functional Col1a1 allele it is likely that Col1a1 expression levels in these patients' cells will be in the region of 50% of wild type Col1a1 expression levels. Hence the Type I OI cell line more accurately represents the Osteogenesis Imperfecta patient.

	57 OI	*570I C1*	*570I C2*	*570I C3*
Average	1	1.72	1.73	1.63
t-test		p= 0.0004	p = 0.0007	p= 0.0009
StdDev	0.121	0.26	0.29	0.31

Table 7.3 Metra CICP ELSIA analysis of *de novo* Type I collagen protein production by a Type I OI fibroblast cell line and Type I OI fibroblasts stably expressing the Col1a1 genomic/cDNA replacement gene over a 24 hour period. Three independent Type I OI cell lines, 57OI C1, 57OI C2 and 57OI C3 stably expressing the Col1a1 genomic/cDNA replacement gene were generated. $5x10^5$ cells from each cell line 57OI, 57OI C1, 57OI C2 and 57OI C3 were plated in duplicate in a 6 well plate. 24 hours later medium was replaced with serum free media, the C terminal propeptide accumulated in the media for 24 hours. The media was collected and duplicate samples pooled and levels of *de novo* type I collagen production evaluated using the Metra CICP ELISA. Levels of *de novo* type I collagen production were measured as a fold change compared to control 57OI cells, in which Col1a1 production levels were set at 1. There was an increase of 1.72 fold in 57OI C1 cells, 1.73 fold in 57OI-C2 and 1.63 in 57OI-C3. In each case the difference between Col1a1 expression in the stable cell lines 57OI C1, 57OI C2 and 57OI C3 and the untransfected 57OI control cells was found to be statistically significant when evaluated using a two tailed student t-test using GraphPad InStat, p values and standard deviations are shown for each cell line.

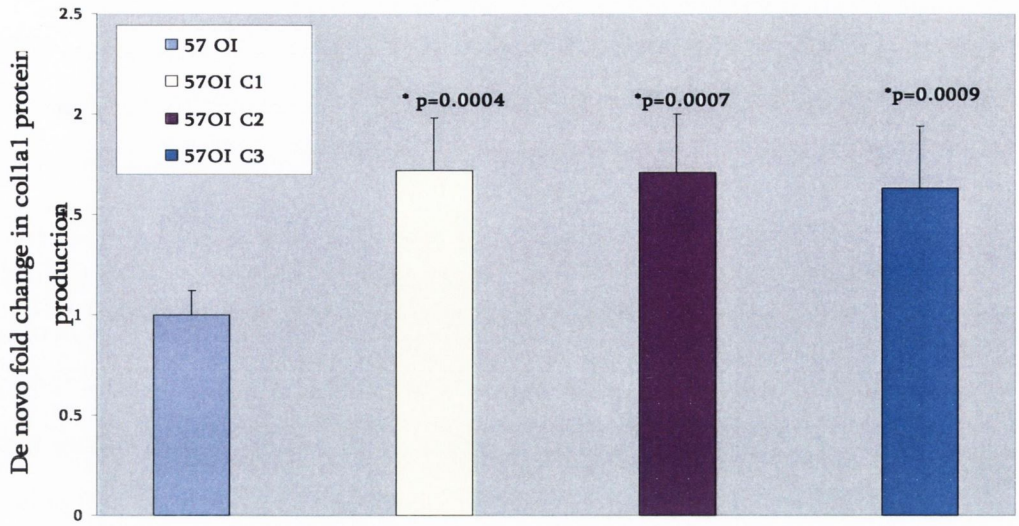


Figure 7.19 Metra CICP ELSIA analysis of *de novo* Type I collagen protein production in a Type I OI fibroblast cell line and Type I OI fibroblasts stably expressing the Col1a1 genomic/cDNA replacement gene over a 24 hour period. Three independent Type I OI cell lines, 57OI C1, 57OI C2 and 57OI C3 stably expressing the Col1a1 genomic/cDNA replacement gene were generated. $5x10^5$ cells from each cell line 57OI, 57OI C1, 57OI C2 and 57OI C3 were plated in duplicate in a 6 well plate. 24 hours later the media was replaced with serum free media, the C terminal propeptide then accumulated in the media for 24 hours. The media was collected, duplicate samples pooled and levels of *de novo* type I collagen production evaluated using the Metra CICP ELISA. Levels of *de novo* type I collagen production were measured as fold change compared to the control 57OI cells, in which type I collagen production levels were set at 1. There was an increase of 1.72 fold in 57OI C1 cells, 1.73 fold in 57OI-C2 and 1.63 in 57OI-C3. In each case the difference between Col1a1 expression in the stable cell lines 57OI C1, 57OI C2 and 57OI C3 and the 57OI control untransfected cells was found to be statistically significant when evaluated using a two tailed student t-test using GraphPad InStat, p values, standard deviations are shown for each cell line. $p<0.05$. p* values are detailed on the graph.

7.3.5. Western Blot analysis of Col1a1 expression in Type I OI fibroblasts.

It was essential to determine if the type I collagen being expressed from the Col1a1 genomic/cDNA replacement gene was a) of the correct size and b) being processed and assembled into the Type I heterotrimers in an identical manner as the Col1a1 chains being expressed from the endogenous Col1a1. This was explored using the Type I OI stable cell lines. A western blot was carried out to determine if the α1 protein chains expressed from the Col1a1 genomic/cDNA replacement gene were of the correct size. Analysis of protein extracted from each of the cell lines revealed that the detected pro-collagen, pCα1(I) collagen, pNα1(I) collagen and the mature Type I collagen (α1 (I) protein) were all of the correct size (Figure 7.20). The four forms of type I collagen, 1) full length pro-collagen, 2) pCα1(I) collagen, the pro-collagen processing intermediate which retains the carboxyterminal propeptide, 3) pNα1(I) collagen the procollagen processing intermediate which retains the amino terminal propeptide and 4) the mature type I collagen from which both the carboxy and amino terminal peptides have been removed, were detectable because the antibody epitope lies within the sequences retained in the mature fully processed protein. No additional bands were detected in the protein samples isolated from any of the 3 stable cell lines when compared to the protein isolated from the original 57OI Type I OI cells when probed with the anti – Col1a1 antibody (SC8784 Santa Cruz USA). Frequently the presence of a mutation in either the α1(I) or α2 (I) chains has been found to delay the triple helix assembly time, during this extended assembly time additional residues are modified, consequently the presence of increased post-translational modifications frequently occurs in the Col1a1 and Col1a2 chains in patients with OI (Gajko-Galicka, 2002; Primorac, et al., 2001). The presence of additional moieties on the chains can decrease the electrophoretic mobility of the chains; this can often be readily observed following separation of the protein on SDS-PAGE gels. In the currently study the Col1a1 chains detected in the western blot did not display any detectable decreased or altered mobility in comparison to the α1 (I) chains isolated from the wild type control 57OI Type I OI cells. Therefore it appears that due to the identical electrophoretic mobility and protein size of the type I collagen detected in protein samples isolated from both the original cells and stable cell lines,

that the Col1a1 replacement gene appears to be producing Col1a1 protein which is identical or very similar in size and post translational processing when compared to the endogenous protein.

7.3.6 Immunofluorescent analysis of Col1a1 protein production in Type I OI fibroblasts.

The third approach taken in the protein analysis was to assess and compare the levels of type I collagen production and secretion by the stable cell lines 57OI-C1, 57OI-C2, 57OI-C3 and the original 57OI cells to determine if there were any detectable differences between them. Based upon the RT-PCR and ELISA data it appears that the three Type I OI stable cell lines 57OI-C1, 57OI-C2, 57OI-C3 were efficiently expressing the Col1a1 genomic/cDNA replacement gene. Col1a1 expression levels in these cell lines increased by an average of 1.64 – 1.88 (± 0.33 – 0.46) fold at the mRNA level (as assessed by real time RT-PCR), and 1.63 – 1.72 (± 0.26 – 0.31) fold at the protein level (as determined using the Metra CICP ELISA system). In addition the α1 (I) chains expressed from the Col1a1 genomic cDNA replacement gene appear to be of the correct size and processed in an identical manner to the endogenous α1 (I) chains (as determined by western blot). Given the findings a difference in immunofluorescence may be detectable.

For this analysis $2x10^5$ cells per cell line were seeded into two wells containing three 1% gelatin coated coverslips and allowed to grow for 3 days. Production of Type I collagen was assessed using a monoclonal antibody mouse anti-human Col1a1 (C-2456 Sigma). The secondary antibody was a goat anti mouse Cy™3 conjugated antibody (111-165-143, Jackson Immunoresearch). The coverslips were mounted using aqua polymount and analysed using fluorescence microscopy. This was carried out 3 times. Type I collagen production and secretion was detected following probing of the slides upon which 57OI cells were growing (Figures 7.21 – 7.23). Type I collagen production and secretion was also detected from the stable Type I OI cell lines (Figures 7.24 -7.25). Based upon the RT-PCR and ELISA data the expected difference in type I collagen

production between the wild type 57OI cells and the Type I OI cell lines stably expressing the Col1a1 genomic/cDNA replacement gene is in the region of 1.7 fold but such difference are much more difficult to accurately evaluate using immunofluorescence, this analysis is not quantitative rather it simply serves to support the data already obtained from the RT-PCR and ELISA based evaluation of Col1a1 expression.

A total of 12 coverslips of cells for each of the Type I 57OI cell lines stably expressing the Col1a1 genomic/cDNA replacement gene, 57OI-C1, 57OI-C2, 57OI-C3 and the original Type I OI cell lines 57OI were assessed. Comparing the Type I collagen immunofluorescence results from the original cells 57OI (Figures 7.21 panel 1A-1C) to the immunofluorescence results from the stable cell lines 57OI-C1, 57OI-C2, 57OI-C3 it appears that there is more Type I collagen produced by the stable cell lines with a significant mesh like arrangement of the Type I collagen protein evident (Figure 7.21 panel 2A-2C). Type I collagen production observed in Figure 7.21 panel C shows cells probed with the Cy™3 conjugated secondary antibody alone which demonstrate the specificity of the 20 antibody. Therefore, one may conclude that the results from this additional assay suggest that Type I OI fibroblasts in 7.21 panel 2 appear to be expressing the Col1a1 genomic/cDNA replacement gene. It is difficult to definitively state that the stable cell lines are indeed producing more type I collagen than the original Type I OI cells the assay could be considered subjective and it is in no way quantitative. Nevertheless from the total of 36 slides evaluated for the three stable cell lines 57OI-C1, 57OI-C2, 57OI-C3, it appears that when compared to the original Type I OI 57OI immunofluorescence that these three stable cell lines are likely producing more type I collagen. Alone this result would be of little significance but with the support of the RT-PCR and ELISA results it seems reasonable to suggest that this indeed may be the case.

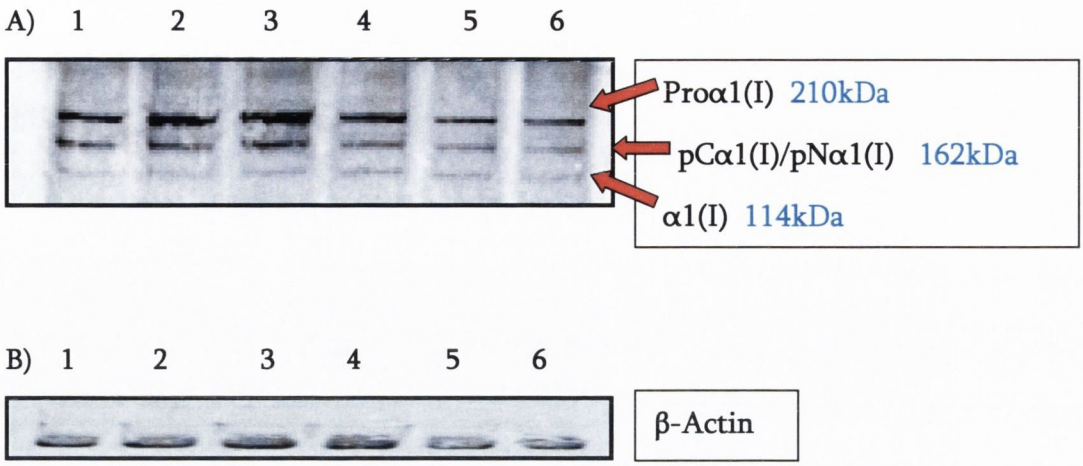


Figure 7.20 Western Blot of protein extracted from Type I OI Col1a1 stable cell lines and the original 57OI cells. Western blot analysis of type I collagen production in the Type I OI fibroblasts was carried out in order to determine if the α1(I) chains from the Col1a1 replacement gene were of the correct size and were being incorporated into the Type I collagen triple helix. $5x10^5$ cell were plated from each cell line 57OI C1, 57OI C2, 57O3 C1 and 57OI. Protein was extracted 48 hours later. 30μg of denatured protein was added per well of the SDS-PAGE, following protein separation and transfer the blot was probed using an anti-Col1a1 antibody (sc 8784 Santa Cruz), the blot was then stripped and reprobed with a β-Actin antibody. **A)** Western Blot probed with anti-Col1a1 antibody, Lanes 1-3. 57OI C1, 57OI C2, 57O3 C1, Lanes 4-6 57OI. Each lane generated an identical banding pattern when probed with the anti-Col1a1 antibody, the top band is the unprocessed proα1(I) protein, the middle band consists of both the N and C terminal cleaved α1(I) proteins which co migrate and the lowest third band is the fully mature α1(I) protein. **B)** Western blot reprobed with β-Actin antibody, Lane1: 57OI C1, Lane2: 57OI C2, Lane3:57O3 C1, Lanes 4-6 57OI. β-Actin protein was detected in each of the lanes 1 -6.

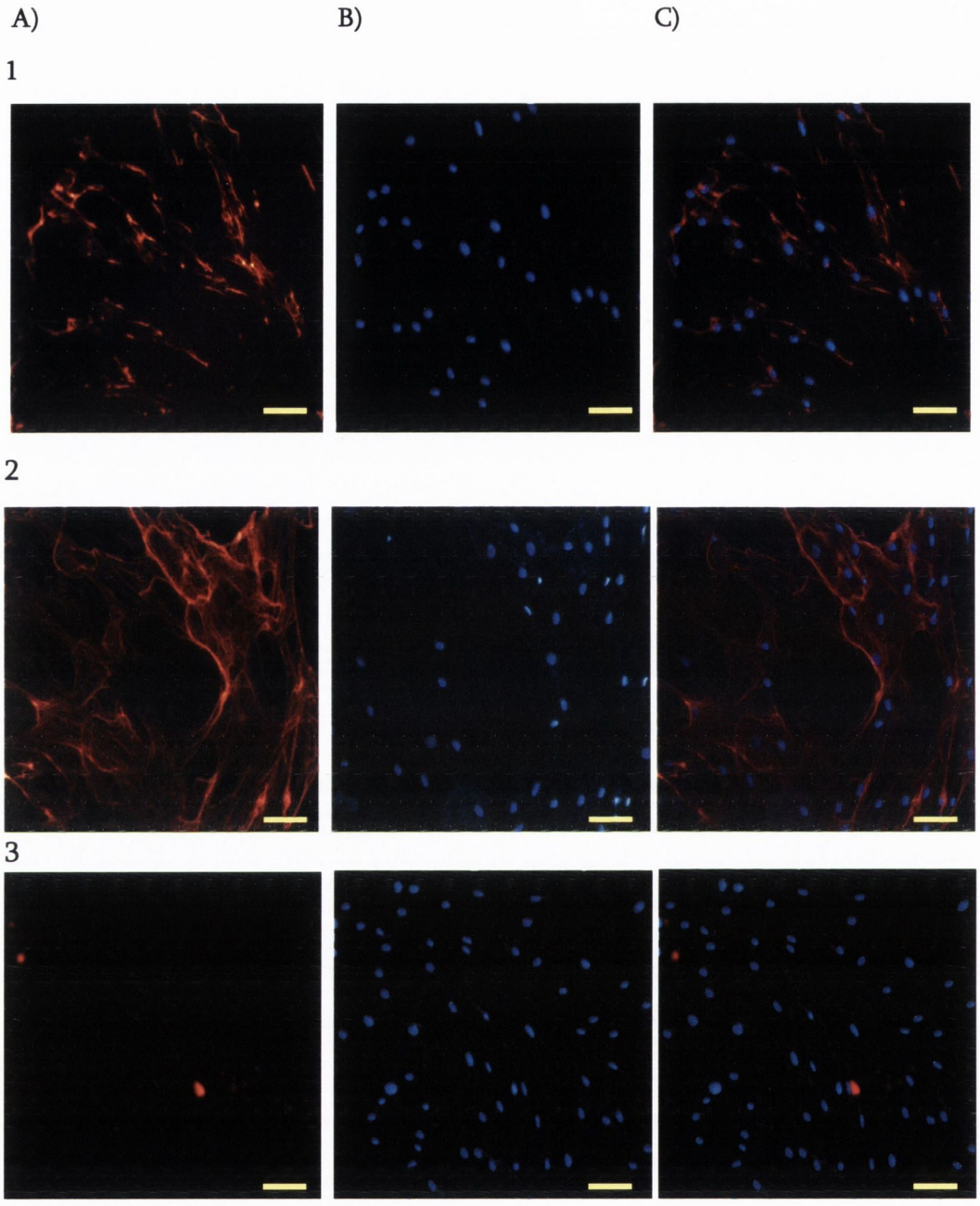


Figure 7.21 Evaluation of Col1a1 production by type I OI fibroblasts. $2x10^5$ 57OI cells were plated onto 1% gelatin coated coverslips, 72 hours later Col1a1 protein production was assesses by immunofluorescence using an antiCol1a1 antibody and Cy™3 conjugated secondary antibody. **Panel 1:** Type I OI-Col1a1 fibroblasts **A1)** Type I collagen, **B1)** DAPI stained nuclei (31) of the cells **C1)** Overlay of A & B Indicating the cell population producing and secreting the type I collagen. **Panel 2:** Type I OI-Col1a1

fibroblasts **A2)** Type I collagen **B2)** DAPI stained nuclei (44) of the cells **C2)** Overlay of A & B Indicating the cell population producing and secreting the type I collagen. **Panel 3:** Type I OI Fibroblasts **A3)** Type I collagen **B3)** DAPI (31) stained nuclei (68) of the cells **C3)** Overlay of A & B Indicating the cell population producing and secreting the type I collagen. Scale bar 20μm.

7.4 Discussion

The work described in this thesis has been focused upon the development of a mutation independent approach for the treatment of Osteogenesis Imperfecta. A selection of siRNA and shRNAs were shown to reduce Col1a1 expression efficiently approximately to 2.67 % -5.08 % (±1.25 – 2.69%) and 16.11% -25.61 % (±10.83- 15.55%) respectively in the target cell type MSCs (see chapter 5 for details). Clearly reduction of Col1a1 expression levels to 2% would be seriously detrimental and of little therapeutic benefit. Therefore in principle a Col1a1 replacement gene will be required that has the capacity to not only restore Col1a1 expression levels to close to endogenous levels but also evade suppression by the Col1a1 targeting si/shRNA.

The first step taken in the development of a Col1a1 replacement gene should be to generate a wild type replacement gene which expresses at close to endogenous expression levels in the appropriate cell or tissue type. The human Col1a1 minigene is highly expressed in a tissue specific manner both *in vitro* and *in vivo*. Based upon this it was decided to incorporate as much of the non coding Col1a1 regulatory sequences present in the Col1a1 minigene in the design of the Cola1 replacement gene. The non coding Col1a1 regulatory sequences present in the Col1a1 minigene construct are 2.3kb of 5' UTR sequences, the first intron and 2 kb of the 3' UTR. As the actual DNA sequence of the 5' and 3' UTRs was unknown these were sequenced, using this new sequence data the entire Col1a1 minigene sequence was assembled, i.e the 2.4kb of 5' UTR sequences, the first intron, the first 5 exons and introns, the last 6 exons and introns and 2 kb of the 3' UTR. This sequence was submitted the NCBI data base and accepted, it is denoted as Synthetic construct mutant collagen alpha 1(I) chain (Col1a1) gene accession number AY590263 (Appendix 1). Using this complete minigene sequence and the Col1a1 mRNA reference sequence a cloning strategy was designed. The Col1a1 replacement gene was designed to contain a total of 2.3kb of Col1a1 5' regulatory sequences, intron 1, exons 1 – 52, introns 50 and 51 and 1.5 kb of 3' UTR sequences (Figure 7.2). The 5' sequences, intron1 and exons 1 and 3 were obtained from

the Col1a1 minigene, exons 3 – 49 from pHUC a full length Col1a1 cDNA construct and finally exons 50 – 52, introns 50 and 51 and the 1.5kb of 3' UTR were also obtained from the Col1a1 minigene. Therefore the Col1a1 replacement gene generated was in fact a hybrid of Col1a1 genomic and Col1a1 cDNA sequences.

Cloning of the Col1a1 genomic/cDNA replacement gene was an arduous process, there were frequent rearrangements observed at each cloning step, each step was repeated multiple times and very large numbers of minipreps screened until plasmids without rearrangements were identified at each step. When the Col1a1 genomic/cDNA replacement gene was eventually cloned in its entirety it was then sequenced in full unfortunately revealing the presence of 2 mutations. A nonsense mutation in which a Proline (CAA) codon was converted to a stop (TAA) codon was present in codon 6 of exon 21, if this spontaneous mutation had occurred during embryogenesis it would have resulted in a premature stop of translation and the truncate protein would most likely be targeted for destruction. In principle this mutation would reduce the quantity of available full length α1(I) chains to 50%, the resulting phenotype would most likely be Type I OI. The second mutation occurred in codon 31 of exon 40 in which a Glycine (GGT) codon was mutated to an Aspargaine (AAT) codon; this would be considered a classical OI mutation. The full length α1(I) chain would be translated and possibly incorporated into the type I collagen heterotrimers. It is difficult to predict what OI phenotype (Type II – Type IV) would ensue due to the phenotypic heterogeneity associated with OI. It was not possible to proceed with this mutated construct and it was necessary to repair the construct such that the wild type Col1a1 coding sequence was restored. The mutation repair strategy undertaken (Figure 7.10) was successful and both the nonsense and missense mutations removed from the construct (Figures 7.4 and 7.7). With the wild type sequence of the construct restored, a variety of cell lines and assays were employed to evaluate the functionality of the Col1a1 genomic/cDNA replacement gene at the mRNA and protein.

The Mov13 mouse has a retroviral insertion in intron 1 of the endogenous Col1a1 allele, the heterozygotes are a model of Type I OI and the homozygotes are embryonic lethal. An embryonic fibroblast cell line was isolated from the homozygous Mov13 embryos a transformed generating a stable cell line which does not produce type I collagen. These cells have been extensively used to assess the regulation of Col1a1 expression, assembly of type I collagen and Co1a1 expression constructs (Fitzgerald, 1999; Lee, et al., 1992; Stefanovic and Brenner, 2003). Mov13 cells were used for initial analysis of the Col1a1 genomic/cDNA replacement construct. The cells are quite robust as they are a transformed cell line with a fast doubling time and are not limited in terms of passage number. The Amaxa nucleofector device was used to achieve efficient delivery of the Col1a1 genomic/cDNA replacement gene to the Mov13 cells. Three stable Mov13 cell lines (MovCol1, MovCol2 and MovCol3) were generated which expressed the Col1a1 genomic/cDNA replacement gene. Col1a1 expression in each of these three stable cell lines and in the original untransfected Mov13 cells was evaluated at the mRNA levels using Real time RT-PCR. It was found that the original Mov13 cells appeared to be producing a very low level of Col1a1 transcripts. On review of the literature it appears that this has been previously reported and is due to the presence of a cryptic promoter located downstream of the retroviral insertion in the Col1a1 allele which drives transcription of a downstream region of the Col1a1 gene (Barker, 1991; Kratochwil, 1993).The RT-PCR primers designed for the quantification of Col1a1 expression in the Mov13 and Mov13 stable cell lines, (MovCol1 – MovCol3) were designed to amplify both mouse and human Col1a1 genes, as the primers bind in the 3' region of the Col1a1 mRNA a hence a low level of Col1a1 RT-PCR product was observed from the Mov13 cells (Figure 7.16). Col1a1 mRNA expression levels in the Mov13 cell stably expressing the Col1a1 genomic/cDNA replacement, MovCol1-MovCol3 were found be 10.2 to 11.9 (±2.67-3.32) fold higher than Col1a1 expression in the untransfected control Mov13 cells indicating that the Col1a1 genomic/cDNA replacement gene is being expressed (Table 7.1 and Figure 7.15). Whilst Mov13 cells are a suitable model for the initial evaluation they do not produce endogenous type I collagen with which to compare the type I collagen generated in the presence of the Col1a1 genomic/cDNA replacement

gene, additionally Mov13 cells are mouse cells and it would be preferable to test the Co11 replacement construct in a human cell line.

Originally the next step in the investigation was to evaluate expression of the Col1a1 replacement gene in Type I OI MSCs. As discussed in Chapter 5 of this thesis, MSCs are the ideal cell type for delivery of OI therapeutics. As Type I OI is associated with a null Col1a1 allele it is expected that Col1a1 levels in these cells will be reduced by approximately 50%, but no abnormal Col1a1 will be present. Therefore one could compare the protein generated by the Type I OI cells to that produced by the Type I OI cells expressing the Col1a1 genomic/cDNA replacement gene. For example, it is well documented that the presence of a mutation on an α1(I) chain delays the assembly of the type I collagen heterotrimers, consequently the α1(I) chains are frequently subject to excessive post-translational modification (Gajko-Galicka, 2002; Primorac, et al., 2001). The presence of additional moieties on the α1(I) chains frequently may result in an abnormal electrophoretic migration pattern and this may be readily detected via SDS-PAGE and western blotting analysis. If the Col1a1 replacement gene is expressed at the correct level following introduction into a cell line such as a Type I OI MSC cell line, Col1a1 expression levels would ideally increase to a maximum of twice the level of Col1a1 expression in the original Type I OI MSCs. Unfortunately it was not possible to source such the Type I OI MSCs, however two Type I OI dermal fibroblast cell lines were successfully obtained, termed 47OI and 57OI. These cells represented an excellent alternative to the Type I OI MSC cell line described above that was originally hoped for use as they were of the same genotype. Unfortunately both of the dermal fibroblast cell lines were provided at passage 6 and it was advised not to use the cells beyond passage 10 – 12 to ensure that normal cell characteristics and reproducibility of the data was maintained (personal communication Dr.M.Valli).

The Amaxa Nucleofector device was used to deliver the Col1a1 genomic/cDNA replacement gene to the Type I OI dermal fibroblast cell lines using the primary dermal fibroblast specific program U23. Three independent nucleofection experiments were

carried out for the Type I OI dermal fibroblast cell lines 47 OI and 57 OI used in the study. Following G418 cell colonies were formed from both 47OI and 57OI cells, however, only the 57OI cells were expanded to sufficient numbers to enable further analysis. Three 57OI cell lines were generated, 57OI C1, 57OI C2, 57OI C3, which in principle should express the Col1a1 genomic/cDNA replacement gene. Given the fragility of the cells, extreme care was taken with the OI 57 stable cells to ensure sufficient cells were generated for RT-PCR, ELISA, western blot and Immunofluorescent analyses.

Real Time RT-PCR analyses of Col1a1 expression in the three stable cell lines 57OI C1, 57OI C2, 57OI C3 and the original 57OI dermal fibroblasts revealed that the stable cell lines were expressing 1.63 to 1.73 (±0.26 -0.31) fold more Col1a1 than the original 57OI cells. This is an encouraging result, it demonstrates that the construct may indeed efficiently be expressing Col1a1 and moreover the expression levels fall within the range of endogenous Col1a1 expression levels. As the Type I OI dermal fibroblasts used for this study produce in the region of 50% of the normal levels of Col1a1, which was set at 1 for this analysis, one would expect the expression to be approximately doubled in these cells following delivery of an efficiently expressing Col1a1 replacement gene. In addition to evaluation of Col1a1 expression at the RNA level, Col1a1 expressoion was evaluated at the protein level using the Metra CICP ELISA system to determine what the levels of *de novo* type I collagen production were in the stable cell lines 57OI C1, 57OI C2, 57OI C3 compared to the original 57OI cells. The ELISA results mirrored the RT-PCR results with expression levels increased in the region of 1.63 to 1.73 (±0.26 – 0.31) for each of the Type I OI cell lines stably expressing the Col1a1 genomic/cDNA replacement gene, 57OI C1, 57OI C2, 57OI C3. (Table 7.2 and Figure 7.19).

It was clear from the initial results that the Col1a1 genomic/cDNA replacement gene was being efficiently translated and there was an increase in the amount of *de novo* type I collagen production as determined by measurement of the amount of released terminal C propeptide. However, what was still unknown was if the α1(I) chains were

being processed in the same manner as wild type α1(I) chains This issue was assessed by western blotting, if the α1(I) chains were being over modified due to, for example, deceased triple helix assembly times this often is readily detectable as a decrease in mobility of the chains in comparison to the wild type α1(I) chains. If this was the case, the lanes in which protein isolated from the stable cells lines 57OI C1, 57OI C2, 57OI C3 were run may contain additional bands of different mobility when compared to the wild type protein. Western blot analysis revealed that the size and mobility of the α1(I) chains isolated from the stable cell lines was very similar if not identical to that of the α1(I) chains isolated from the original 57OI dermal fibroblasts. Therefore it appears that the α1(I) chains produced from the Col1a1 genomic/cDNA replacement gene may be processed in a similar manner as the endogenous α1(I) chains.

The final aspect of the evaluation of the expression of the Col1a1 genomic/cDNA replacement gene was to compare the levels of Type I collagen produced by the wild type 57OI cells and the Type I OI cells stably expressing the Col1a1 genomic/cDNA replacement gene using immunofluorescence. The cells were grown on slides for 3 days and type I collagen detected using an anti-Col1a1 antibody and a Cy™3 labelled secondary antibody. Based upon the RT-PCR and ELISA data described above it was expected that there would be an increase of in the order of 1.7 fold in the levels of type I collagen immunofluorescence observed for the Type I OI cell lines stably expressing the Col1a1 genomic/cDNA replacement. Whilst there appears to be an increase, it is difficult to quantify exactly how much of an increase it is as this analysis is not quantitative but merely qualitative. What is clear is that a small number of the stable cells appear to have the capacity to produce and secrete a significant amount of Type I collagen which adopts a mesh like structure typical of that observed with wild type fibroblasts. The results obtained from this additional assay serve to further support the results obtained using real time RT-PCR and the ELISA.

In conclusion it appears that the design of the Col1a1 genomic/cDNA replacement gene is amenable to drive appropriate levels of Col1a1 expression in human cells as revealed

by the RT-PCR and ELISA analyses of Col1a1 expression levels in the Type I OI stable cell lines. It is notable that in no instance did Col1a1 expression levels exceed 2 fold of what was detectable in the Type I OI cells which are believed to produce approximately 50% of the normal quantity of type I collagen. Retrospectively, a positive control of a normal human dermal primary cell line should have been included in the analyses. Nevertheless the results obtained from the Mov13 cells and the Type I OI cells suggest that the Col1a1 genomic/cDNA replacement gene serves to express the gene of interest most likely at somewhat similar levels to the endogenous Col1a1 gene.

In addition, western blot analysis revealed that given the identical banding pattern of the type I collagen extracted from the original 57OI cells with that extracted from the stable cell lines 57OI C1, 57OI C2 and 57OI C3, that the α1(I) protein appears to be, of the correct size, processed as endogenous α1(I) protein with no evidence of additional unwanted post-translational modifications. Finally, immunofluorescent analysis of the levels of type I collagen from the original 57OI Type I OI dermal fibroblasts and the stable cell lines revealed that there appears to be a detectable increase in type I collagen production and secretion from the stable cell lines in comparison to the wild type 57OI.

The work described in this chapter of the thesis is focused upon the replacement aspect of a therapeutic strategy for Col1a1-linked autosomal dominantly inherited Osteogenesis Imperfecta. In contrast the work described in chapters 3 – chapter 5 of this thesis has been focused upon the development of an RNAi based Col1a1 gene silencing strategy. By combining both suppression and replacement, in principle, a mutation independent therapeutic approach can be developed for Col1a1 linked OI. With regard to the Col1a1 replacement aspect of the potential therapeutic, a significant body of work was required to simply generate a wild type Col1a1 replacement gene. Briefly, as it was necessary to include non coding regulatory sequences a Col1a1 replacement gene was a hybrid of genomic and cDNA sequences was generated, the non coding regulatory DNA was obtained from the Col1a1 minigene construct and the protein encoding cDNA from the pHUC full length Cola1 cDNA construct. The cloning

process was quite a difficult process, a non sense and missense mutation was spontaneously introduced into the Col1a1 cDNA. This had to be corrected before any further investigations could be carried out. Once the wild type sequence was restored in the Col1a1 genomic/cDNA replacement gene, a detailed evaluation of the Col1a1 expression from the construct was assessed at both the RNA level in Mov13 cells and a Type I OI dermal fibroblast cell line. Type I collagen protein expression in Type I OI dermal fibroblasts stably expressing the Col1a1 genomic/cDNA replacement gene was also carried out using ELISA western blotting and immunofluorescence techniques. The results obtained from the experiment are extremely encouraging, the Col1a1 genomic/cDNA replacement gene was found to express Col1a1 both at the RNA and protein levels. Moreover the preliminary evidence obtained indicated that the Col1a1 expression levels are in the order of what may be appropriate for an Osteogenesis Imperfecta therapeutic.

Given the results obtained in the study the way is now clear to proceed to the next logical step in the development of a mutation independent suppression and replacement therapeutic strategy for Osteogenesis Imperfecta. To recap, the overall strategy is RNAi-mediated Col1a1 suppression in combination with provision of a replacement Col1a1 gene, sequence modified such that transcripts from the replacement gene are protected from RNAi-mediated suppression. It is proposed to utilise the degeneracy of the genetic code to achieve the latter. The Col1a1 genomic/cDNA replacement gene will be sequence modified at degenerate sites over the RNAi effector target sequence. Transcripts from such a sequence modified replacement gene will encode wild type Col1a1 mRNA that should be refractory to RNAi suppression. More specifically the Col1a1 genomic/cDNA replacement gene will be sequence modified at the degenerate positions within the target sequences for the Cola1 targeting shRNAs shCol2 and shCol4 described in chapter 4 and chapter 5 of this thesis. In principle the sequence modified version of the replacement will be incorporated into a mutation independent suppression and replacement therapeutic for OI. Thus it is clear that the work described in this chapter involving the wild type Col1a1 genomic/cDNA replacement was a

necessary prerequisite prior to proceeding to generating a sequence modified version of the Col1a1 genomic/cDNA replacement.

Discussion and concluding remarks

The studies carried as part of this doctoral thesis have been focused upon the development of a potential RNAi based suppression and replacement therapeutic approach for the treatment of Col1a1 linked OI. Since the discovery of RNAi there has been a flurry reports detailing the potential therapeutic application of RNAi in which the expression of mutant genes can be down regulated, for review see (Leung and Whittaker, 2005). In fact the significance of the discovery of RNAi has recently been highlighted with the awarding of the Noble prize for medicine to Fire and Mello for their elegant demonstration of RNAi in *c. elegans*. Like many dominant negative disorders OI is genetically heterogeneous, to overcome such heterogeneity a mutation independent therapeutic approach termed suppression and replacement has been developed by members of the group here at the Dept of Genetics TCD (Millington-Ward, et al., 1997; Millington-Ward. S, 1999)

The mutation independent approach adopted in the RNAi studies described in chapters 3 – 6 of this thesis target suppression effectors to the coding sequence of the Col1a1 gene. The first series of studies were focused upon the design and assessment of Col1a1 targeting siRNAs. Five Col1a1 targeting siRNAs (siCol1 – siCol5) which targeted the coding sequence of the Col1a1 gene were initially assessed in HeLa cells which endogenously express Col1a1 at sufficiently high levels to facilitate meaningful evaluation of the 5 siRNAs. The potency of the suppressors was assessed at the RNA level using real time RT-PCR analysis. Each of the 5 Col1a1 targeting siRNAs (siCol1 – siCol5) effectively reduced Col1a1 expression levels to between 28.75% and 47.34% (±5.90 – 24.63%) $p<0.001$.

Whilst effective mediators of suppression, siRNA duplexes can only induce short term suppression which lasts in the region of 3 – 5 days (Amarzguiou, et al., 2005). In order to treat a dominant negative mendelian disorder such as OI, life long suppression of the

mutant protein will most likely be required. Therefore, the next logical progression in the study was to incorporate a selection of the effective Col1a1 siRNAs into an siRNA expression system. The Brummelkamp H1 short hairpin RNA (shRNA) expression system was adopted and its application is described in chapter 4 of this thesis (Brummelkamp, et al., 2002; Paddison, 2002). Two Col1a1 targeting shRNAs were generated, shCol2 and shCol4, the targeting sequence of which are identical to siCol2 and siCol4. HeLa cells were used as the *in vitro* model and the potency of shCol2 and shCol4-based suppression of Col1a1 expression was assessed at the RNA level using real time RT-PCR. Both shCol2 and shCol4 effectively reduced Col1a1 expression in HeLa cells to between15.85% to 29.91% (±5.43% - 18.32%), $p<0.001$. At this stage in the experimental proceedings is was clear that the siRNAs and shRNAs that had been designed, generated and evaluated were effectively and potently reducing the expression of Col1a1 at the RNA level, but that further analysis was required, particularly at the protein level. Whilst an excellent *in vitro* model for preliminary validation studies, HeLa cells are not a relevant cell type in which to explore potential OI therapeutics.

In contrast, Mesenchymal stem cells (MSCs) which have the ability to self renew and proliferate extensively giving rise to a variety of precursor cell types including osteoblast precursor cells, the target cell type for OI. In theory following delivery of a population of MSCs (which had been engineered to express the desired therapeutic), the MSCs should establish and provide a life long supply of corrected cells to the recipient. Hence it was decided to explore the potential of MSCs as both the therapeutic target and vehicle for the RNAi-based suppression and replacement strategy for the treatment of Col1a1 linked OI under development.

Primary cell lines in particular are associated with low transfection efficiencies when traditional transfection methods are used to deliver nucleic acids. A relatively new electroporation technology the Amaxa Nucleofector device was explored as a delivery tool for MSCs and it was found, following optimisation, that siRNA transfection

efficiencies of almost 90 - 100% could be achieved as revealed by the reduction of Col1a1 mRNA expression to 2.57% to 5.08% (±1.6 – 2.64%) following nucleofection with siCol2 or siCol4. Plasmid nucleofection efficiencies were in excess of 70% with Col1a1 mRNA expression reduced to between 16.32% - 25.61 % (±14.32 – 16.32). Therefore, it appears that nucleofection is a highly efficient means of nucleic acids delivery to MSCs and both the siRNAs (siCol2 and siCol4) and shRNAs (shCol2 and shCol4) effectively reduced Col1a1 expression in MSCs. The effectiveness of these suppressors was also explored at the protein level, the Metra CICP ELISA system which was used to measure the amount of *de novo* type I collagen production by the nucleofected MSCs over a 24 hour period revealed that the siRNAs (siCol2 and siCol4) reduced type I collagen expression to 10.15% - 13.00% (±1.89% -3.27%) and the shRNAs reduced type I collagen production to 22.88% - 30.97% (±8.15% - 11.08%).

At this stage in the project a suitable means of nucleic acid delivery to MSCs had been optimised and highly potent mediators of Col1a1 RNAi had been evaluated. Therefore it was possible to deliver highly potent mediators of Col1a1 suppression to MSCs and most likely through random integration of the H1-shRNA plasmids achieve sustained expression of the shRNA. This however is not the ideal situation, random integration carries a similar risk as many viral vectors for insertional mutagenesis. Hence what was required was a method by which site specific integration into non pathogenic locations in the human genome could be achieved. For this reason the ΦC31 integrase system which facilitates site specific integration into *AttP* sites in the human genome was evaluated as a potential delivery vector for the Col1a1 linked OI suppression and replacement strategy.

An initial concern was that ΦC31 integrase mediated integration into MSCs may possibly have a negative impact upon the differentiation capacity of the MCS. Given that it was vital that the MSCs used for the therapeutic approach under development remain multipotent, the multipotency of the human MSC population employed in the various studies carried out for this doctoral thesis was evaluated. Notably, the MSCs

were successfully stimulated to proceed through the adipogenic and osteogenic lineages as described in chapter 6. With the multipotency of the cells established the next phase of the study was to assess the efficacy of the ΦC31 integrase system in our hands. This was first explored in HeLa cells. A total of 107 plasmids were assessed for evidence of integration into *AttP* sites within the genome, based upon PCR, restriction enzyme digest analysis and sequencing analysis, none of the plasmids evaluated appeared to have undergone integrase mediated integration into the genome of the HeLa cells. The efficacy of the system was also explored in MSCs, the construct delivered to the MSCs was a reporter plasmid in which eGFP expression was under the control of sequences considered to contain the Col1a1 promoter, termed the pEGFP1-AttB.Col1a1 promoter plasmid. It was found that eGFP expression from the construct was high at 48 hours post nucleofection, however, it was not possible to generate an MSC cell line as the eGFP expression levels from the Col1a1 promoter appeared to be toxic to the cells, the MSCs became senescent and died. Hence unfortunately it was not possible to determine if ΦC31 mediated integration had occurred in MSCs.

In parallel to the Col1a1 promoter study in MSCs, an *attB* plasmid based RNAi study in MCSs was established. Three shRNA plasmids, shCol2.AttB, shCol4.AttB and shGFP.AttB were generated and subsequently co-nucleofected into MSCs together with the ΦC31 integrase plasmid. Generation of MSC stable cell lines expressing each construct was readily achieved illustrating the possible toxicity of the pEGFP1-AttB.Col1a1 promoter plasmid in the previous experiment, in which it had not been possible to generate MSC cell lines stably expressing the pEGFP1-AttB.Col1a1 promoter plasmid. Col1a1 expression in each of the cell lines expressing shCol2.AttB, shCol4.AttB or shGFP.AttB was evaluated over a four week period to determine if the plasmids could effect sustained suppression of Col1a1 expression in MSCs. It was found that shCol2 and shCol4 reduced Col1a1 expression to between 2.35% - 15.00% (±0.45 – 11.69%), therefore sustained suppression of Col1a1 expression appeared to be achievable in MSCs. It is notable that this experiment would not be amenable to a longer time course as MSCs have a limited passaging capacity, and in addition it is likely

that such a reduction in Col1a1 expression may impact negatively on the cells. For this reason these MSC cells were also not used to establish if they retain their differentiation capacity subsequent to genetic modification. However, in this regard it is indeed planned that such an experiment will be undertaken in MSCs in the future subsequent to delivery of a suppression and replacement therapeutic. In this case the replacement gene should, in principle, provide levels of Col1a1 expression similar to that of the endogenously expressed gene.

Thus far the studies described have focused upon the delivery and suppression aspects of the suppression and replacement gene therapeutic approach for Col1a1 linked OI. As mentioned the suppression strategy adopted is a mutation independent approach, where the Col1a1 coding sequence of both alleles is targeted. Therefore, it is likely that a Col1a1 replacement gene will be required. A first step towards the generation of such a replacement gene involved the generation of a wild type Col1a1 gene which ideally is expressed in a tissue specific manner at close to endogenous levels Col1a1 expression (as described in Chapter 7).

Despite the many difficulties encountered in the cloning process, including extensive rearrangement and the introduction of mutations in the cloning process, the wild type Col1a1 genomic/cDNA construct was generated and the functionality of this construct evaluated at the RNA and protein levels. Expression of the Col1a1 replacement was first evaluated in Mov13 cells which contain a retroviral insertion in both Col1a1 alleles. A series of Mov13 cell lines stably expressing the Cola1 gene were generated, the Col1a1 replacement gene appeared to be expressed in the Mov13 cell line with expression levels ranging from 10.9 to11.9 (±2.67 – 3.32) fold that of untransfected Mov13 cells. Although it is generally stated that Mov13 cells do not produce any Col1a1 RNA, the RT-PCR analysis carried out detected low level Col1a1 expression. On review of the literature it appears that there is a cryptic promoter located down stream of the retroviral insertion in intron 1 and low level expression of truncated Col1a1 transcripts is driven from this promoter. Whilst a low level of Col1a1 expression was detected in

untransfected Mov13 cells, it is clear that much higher levels of expression are present in Mov13 cells transfected with the wild type Col1a1 construct.

At this point it was felt that a human cell line that more accurately represented the OI genotype was required. Dermal fibroblasts isolated from a patient with Type I OI in which Col1a1 expression levels are reduced to 50% of endogenous levels were used. Of the two Type I OI cell lines at hand, 49OI and 57OI, only the latter was used to successfully generate cell lines stably expressing the Col1a1 genomic/cDNA replacement gene. Three 57OI derived cell lines were generated and Col1a1 expression evaluated at the mRNA level. It was found that in comparison to the original untransfected 57OI cells, that Col1a1 mRNA expression had increased 1.64 – 1.88 (± 0.33 – 0.46) fold. ELISA analysis, of type I collagen expression in each of the 57OI Type I OI cell lines stably expressing the Col1a1 genomic/cDNA replacement gene revealed that the presence of the Col1a1 gene increase Col1a1 expression levels 1.63 – 1.73 (±0.26 – 0.31) fold compared to the unmodified Type I OI human dermal fibroblasts. Western blot analysis suggested that the Col1a1 genomic/cDNA replacement gene appears to be expressing α1(I) chains of the correct size which appear to be post-translationally modified in a similar manner as the endogenously produced α1(I) chains. This is indicated by the identical banding patterns generated from protein extracted from unmodified 57OI cells and the 57OI cell lines stably expressing the Col1a1 genomic/cDNA replacement gene. The final evaluation of expression of the Col1a1 genomic/cDNA replacement gene at the protein level that was undertaken was immunofluorescence - this assay is qualitative rather than quantitative. Nevertheless, it appears when comparing the levels of type I collagen immunofluorescence generated by the unmodified 57OI cells to that generated by the 57OI cells expressing the Col1a1 genomic/cDNA replacement gene, the stable 57OI cells appear to be producing more type I collagen. Taking the results generated from RT-PCR, ELISA, western blotting and immunoflouresence analyses together, it appears that the Col1a1 genomic/cDNA replacement gene is being expressed efficiently. It is notable that expression levels at the RNA or protein level in the 57OI cells stably expressing the Col1a1 genomic/cDNA

replacement gene did not exceed twice the level of Col1a1 or type I Col1a1 expression in the unmodified 57 OI cells. As the 57OI cells only produce functional RNA from one allele and 50% or the normal amount of α1 (I) chains, it appears that the Col1a1 genomic/cDNA replacement gene is being expressed at the appropriate levels in cells isolated from a patient with OI. This finding is very encouraging, if the construct was expressing at too low a level sufficient type I collagen would not be produced but more notably over expression of the Col1a1 genomic/cDNA replacement gene could have quite a negative impact as over expression of Col1a1 has been associated with systemic scleroderma (Cicchillitti, 2004). Moreover, in addition to Col1a1 expression being within the range of endogenous expression levels, the α1 (I) chains appear to be processed and incorporated into the type I collagen heterotrimer in a similar manner as the endogenously expressed α1 (I) chains.

Therefore, the studies carried out as part of this doctoral thesis have resulted in the identification and evaluation of a selection of highly potent mediators of Col1a1 RNAi in the target cell type for OI, mesenchymal stem cells, as evaluated at the RNA and protein level. A highly effective means with which siRNAs, shRNA expression plasmids and large constructs such as the Col1a1 genomic/cDNA replacement gene has been efficiently applied and optimised for use in difficult to transfect cell lines such as mesenchymal stem cells and human dermal fibroblasts. A Col1a1 genomic/cDNA replacement gene was generated which appears to contain the ideal combination of non coding sequences to effect appropriate levels of Col1a1 expression, the α1 (I) chains appear to be processed as endogenous chains are and Col1a1 and type I collagen expression levels were efficiently restored to within range of wild type levels in cells isolated from a Type I OI patient. The aspects of the project involving the use of the ΦC31 integrase system were more problematic and indeed site specific integration was not demonstrated either in HeLa cells or MSCs using this approach. As the integrase construct does seem to be expressing the ΦC31 integrase enzyme (as evaluated only by RT-PCR), it may be possible that the assays employed to detect integration were not ideally suited to do so. At the time this work was being undertaken a number of reports

of genomic deletions and rearrangements associated with the ΦC31 integrase system emerged in the literature and hence prompted a re-evaluation of the approach. It was decided that work using the ΦC31 integrase system would be suspended for the moment due to the uncertainty of the impact of this method of delivery on the human genome and that alternative vectors for delivery of the OI therapeutic being explored would be considered.

As discussed previously, before any vector system can be considered for use in the Col1a1 suppression and replacement therapeutic strategy described, it is essential that it can accommodate in excess of 11kb, this requirement limits the potential pool of vectors available, The herpes simplex adeno-associated vector systems are one system that may be considered, such hybrid vectors can accommodate up to 100Kb of exogenous DNA sequences. Additionally these vectors preferentially integrate into AAVS1 site in the human genome. Other systems with large transgene capacities are human and bacterial artificial chromosomes, whilst they can accommodate large transgenes and remain episomal, which is important from a safety perspective, delivery efficiencies are low. As the more traditional viral vector systems can not accommodate the OI replacement gene and are associated with insertional mutagenesis and strong cytotoxic and immune responses investigation of new and emerging systems is required

Despite difficulties associated with the vector aspects of the project, it is worth highlighting again that both the suppression aspect of the studies carried out and the Col1a1 replacement studies generated encouraging results. At this stage a logical progression for the work would involve the generation of a replacement Col1a1 gene in which the required sequence changes at degenerate sites have been engineered into the gene at the target sites for siCol2 and siCol4-mediated suppression. In principle transcripts from this replacement gene should be refractory to siCol2 and siCol4-mediated suppression and the gene should provide a source of wild type protein. In addition, clearly the efficacy of this replacement construct in combination with the Col1a1 suppressors in a suppression and replacement strategy targeting Col1a1 linked

OI should be determined. This analysis would ideally be undertaken initially in an MSC cell population isolated from patients Type III or Type IV OI. Subsequent to evaluation *in vitro*, to progress the work further, clearly *in vivo* evaluation of the approach will need to be considered. In this regard unfortunately a number of the key biological tools that would enable such research to be undertaken are not currently available. More specifically, animal models presenting with autosomal dominantly inherited forms of OI due to the presence of a mutant human Col1a1 transgene would ideally be required to test the efficacy of RNAi-mediated suppression (given the stringent sequence requirement for potent suppression using siRNAs). Such an animal model would enable the *in vivo* evaluation of the efficacy of RNAi molecules relevant to human patients. At present, the mouse models available such as the Brtl III and BrtlIV express mutant mouse collagen transcripts, with the exception of the Col1a1 minigene mouse which expresses a truncated human Col1a1 gene, unfortunately these models are not ideally suited for the evaluation of suppressors targeting the human genes.

In summary, during the course of the study a number of essential components required for the development and evaluation of a suppression and replacement therapeutic for autosomal dominantly inherited Col1a1-linked OI have been developed and indeed the results from these aspects of the work are really quite encouraging as has been discussed in detail above. However, additional components are clearly still required, not least an appropriate vector for delivery and an animal model expressing a mutant human Col1a1 transgene as also highlighted above. In this regard the technologies that enable the development of such components are clearly now available. For example, a wide range of viral and non-viral vectors have been developed and rapid generation of transgenic animals has become a fairly readily achievable. It is therefore planned that these biological tools and techniques will be used in the future development of vectors and animal models which will enable detailed *in vivo* exploration of the therapeutic strategies described in this thesis. It is hope that the energy invested by this and other groups will generate a successful therapeutic for this debilitating disorder.

Appendix 1

Synthetic construct mutant collagen alpha1(I) chain

Pubmed Accession number (AY590263)

NCBI Entrez Nucleotide My NCBI [Sign In] [Register]
PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books
Search Nucleotide for Go Clear
Limits Preview/Index History Clipboard Details
Display GenBank Show 5 Send to Hide: sequence all but gene, CDS and mRNA features
Range: from begin to end Reverse complemented strand Features: + Refresh

1: AY590263. Reports Synthetic constru...[gi:48994843] Links

Features Sequence

```
LOCUS       AY590263               10461 bp    DNA     linear   SYN 21-JUN-2004
DEFINITION  Synthetic construct mutant collagen alpha 1(I) chain (Col1a1) gene,
            complete cds.
ACCESSION   AY590263
VERSION     AY590263.1  GI:48994843
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 10461)
  AUTHORS   Olsen,A.S., Geddis,A.E. and Prockop,D.J.
  TITLE     High levels of expression of a minigene version of the human pro
            alpha 1 (I) collagen gene in stably transfected mouse fibroblasts.
            Effects of deleting putative regulatory sequences in the first
            intron
  JOURNAL   J. Biol. Chem. 266 (2), 1117-1121 (1991)
   PUBMED   1702430
REFERENCE   2  (bases 1 to 10461)
  AUTHORS   McMahon,H.P., Millington-Ward,S. and Farrar,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2004) Genetics, Trinity College Dublin, Dublin 2,
            Ireland
REFERENCE   3  (bases 1 to 10461)
  AUTHORS   Olsen,A.S., Geddis,A.E. and Prockop,D.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2004) Department of Biochemistry and Molecular
            Biology, Jefferson Institute of Molecular Medicine, Jefferson
            Medical College, Thomas Jefferson University, Philadelphia, PA
            19107, USA
FEATURES             Location/Qualifiers
     source          1..10461
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
                     /note="derived from human Col1A"
     gene            434..>10461
                     /gene="Col1a1"
                     /note="minigene"
     mRNA            join(434..2515,3978..4172,4311..4345,4457..4483,
                     4574..4663,5068..5121,5482..5589,5681..5963,6096..6286,
                     6575..6817,6947..>10461)
                     /gene="Col1a1"
                     /product="mutant collagen alpha 1(I) chain"
     5'UTR           434..2412
                     /gene="Col1a1"
     TATA_signal     2266..2272
                     /gene="Col1a1"
     CDS             join(2413..2515,3978..4172,4311..4345,4457..4483,
                     4574..4663,5068..5121,5482..5589,5681..5963,6096..6286,
                     6575..6817,6947..7093)
                     /gene="Col1a1"
                     /note="mutant Col1a1 protein; consists of 11 of the 52
                     exons; 5' sequence, first 5 and last 6 introns and exons
                     and 3' sequence"
                     /codon_start=1
                     /transl_table=11
                     /product="mutant collagen alpha 1(I) chain"
```

```
                     /protein_id="AAT48109.1"
                     /db_xref="GI:48994844"
                     /translation="MFSFVDLRLLLLLAATALLTHGQEEGQVEGQDEDIPPITCVQNG
                     LRYHDRDVWKPEPCRICVCDNGKVLCDDVICDETKNCPGAEVPEGECCPVCPDGSESP
                     TDQETTGVEGDTGPRGPRGPAGPPGRDGIPGQPGLPGPPGPPGPPGPPGPPGSPGEQGPSG
                     ASGPAGPRGPPGSAGAPGKDGLNGLPGPIGPPGPRGRTGDAGPVGPPGPPGPPGPPGP
                     PSAGFDFSFLPQPPQEKAHDGGRYYRADDANVVRDRDLEVDTTLKSLSQQIENIRSPE
                     GSRKNPARTCRDLKMCHSDWKSGEYWIDPNQGCNLDAIKVFCNMETGETCVYPTQPSV
                     AQKNWYISKNPKDKRHVWFGESMTDGFQFEYGGQGSDPADVAIQLTFLRLMSTEASQN
                     ITYHCKNSVAYMDQQTGNLKKALLLKGSNEIEIRAEGNSRFTYSVTVDGCTSHTGAWG
                     KTVIEYKTTKTSRLPIIDVAPLDVGAPDQEFGFDVGPVCFL"
     3'UTR           7094..10461
                     /gene="Col1a1"
     polyA_signal    7341..7355
                     /gene="Col1a1"
     polyA_signal    8471..9476
                     /gene="Col1a1"
ORIGIN
        1 ggatccccaa atatccttca gaaccccagg cccatgatgt agcaaccccc aattcacacc
       61 ttggaggttt caactcttct ttaagatggg cgtgggaaag cctggatggg aaacatatgg
      121 ggaggggcgg ggagctgcag gcaggagccc ttcttactac gaaaacccaa gaagcaagga
      181 agtggacagg tcactaaccc tcatactacc aagccctgcg gcaccctgcc ctagaccacc
      241 actctaaatg tctgttccct ccaaaaacag gacccctgtc gcctattagg gaggggtct
      301 cttggaactg acccacagta gggggcagga ctttggtggg ttcaagaact gccatctcag
      361 cacctcagcc ccctagtcct gccctgcagt cgctggcact aggcgggggc agaccctggg
      421 ccacaagttg ctgccacatg gtcgggataa ttgatgaagg tccatccctc cattgctgtc
      481 tccagccctg cctctctgga aactctatat tttcccttta attatagccc ctgcagtctc
      541 cctctgctgc cccacccgca ccgctcatcc tggctgccca cggccagccg gccagccgac
      601 gtggctccct ccccttctgt tccttttttt tccccctttgc cttcgttgca caaaaccagc
      661 tgggggaggg cgtggagagg ggcgggggga ggcaatggaa tcttggatgg tttgggggag
      721 gcgggactcc ccgcttccac gtttgcagct ctggagcacc cggggtgggg agctgcacag
      781 gagggagaga aatgaacagg gcactgcaag gagaccccca ggccttctct cagccctaca
      841 gagtttctca ggacgaggta gattggggtt gaggcagagc cttgttgggg gaatgggaca
      901 tggaggaaga aaggacgtgg agttctagag ccatcttcct tagatatagc ctgctgtcct
      961 tcgggtcccc agacccttte agagtgtaca gatgattctc tctggttcct aaggcataga
     1021 gcaatgaccg ggattttcaa gaaagagatg aggcagtggg aagtagcccc taaaacaaag
     1081 tcaatcatcc tctgcagccc atcccacacc cccaaaggaa agtttcaccc agacacccaa
     1141 aatatcccat acatccccaa cactgagtcc aggtacaact ggagaagggg ctttatgcag
     1201 ctcccagaaa gacacccctt tagctaagtg ccctcctcc acccaggttc tctctggttt
     1261 gactgtgctg ggaaggaggg tctctaagca gcccctggcc acagccatgg caaacaaaac
     1321 tcttctctaa gtcaccaatg atcacaggcc tcccactaaa aatacttccc aactctgggg
     1381 tggaagagtt tgggggatga atttttaggg gattgcaagc cccaatcccc acctctgtgt
     1441 ccctagaatc ccccaccccc accttggctg ctccatcacc caaccaccaa agctttcttc
     1501 tgcagaggcc acctagtcat gtttctcacc ctgcacctca gcctccccac tccatctctc
     1561 aatcatgcct agggtttgga ggaaggcatt tgattctgtt ctggagcaca gcagaagaat
     1621 tgacatcctc aaaattaaaa ctcccttgcc tgcacccctc cctcagatat ctgattctta
     1681 atgtctagaa aggaatctgt aaattgttcc ccaaatattc ctaagctcca tcccctagcc
     1741 acaccagaag acacccccaa acaggcacat ctttttaatt cccagcttcc tctgttttgg
     1801 agaggtcctc agcatgcctc tttatgcccc tcccttagct cttgccagga tatcagaggg
     1861 tgactggggc acagccagga ggacccccte cccaacaccc ccaacccttc caccttgga
     1921 agtctcccca cccagctccc cagttcccca gttccacttc ttctagattg gaggtcccag
     1981 gaagagagca gaggggcacc cctacccact ggttagccca cgccattctg aggacccagc
     2041 tgcaccccta ccacagcacc tctggcccag gctgggctgg ggggctgggg aggcagagct
     2101 gcgaagaggg gagatgtggg gtggactccc ttcctcctc ctccccctct ccattccaac
     2161 tcccaaattg ggggccgggc caggcagctc tgattggctg gggcacgggc ggccggctcc
     2221 ccctctccga ggggcagggt tcctccctgc tctccatcag gacagtataa aaggggcccg
     2281 ggccagtcgt cggagcagac gggagtttct cctcggggtc ggagcaggag gcacgcggag
     2341 tgtgaggcca cgcataagcg gacgctaacc ccctccccag ccacaaagag tctacatgtc
     2401 tagggtctag acatgttcag ctttgtggac ctccggctcc tgctcctctt agcggccacc
     2461 gccctcctga cgcacggcca agaggaaggc caagtcgagg gccaagacga agacagtaag
     2521 tcccaaactt ttgggagtgc aaggatactc tatatcgcgc cttgcgcttg gtcccggggg
     2581 ccgcggctta aaacgagacg tggatgatcc ggagactcgg gaatggaagg gagatgatga
     2641 gggctcttcc tcggcgccct gagacaggag ggagctcacc ctggggcgag gttggggttg
     2701 aacgcgcccc gggagcggga ggtgagggtg gagcgcggcg tgagttggtg caagagagaa
     2761 tcccgagcgc gcaaccgggg aagtggggat ctgggtgcag agtgaggaaa gcacgtcgaa
     2821 gatgggatgg gggcgccgag cggggcattt gaagcccaag atgtagaagc aatcaggaag
     2881 gccgtgggat gattcataag gaaagattgc cctctctgcg ggctagagtg ttgctggggc
     2941 cgtgggggtg ctgggcagcc gcggaggggg tgcggagcgt gggcgggtgg aggatgagaa
     3001 actttggcgc ggactcggcg gggcggggtc cttgcgcccc ctgctgaccg atgctgagca
     3061 ctgcgtctcc cggtccaacg cttactgggg caggagccgg agcgggaaga cccgggttat
     3121 tgctgggtgc ggacccccac ctctagatct ggaaagtaaa gccagggatg gggcagccca
     3181 agcctcttaa agaggtagtc gggccggtga ggtcggcccc gccccggccc cattgcttag
     3241 cgttgcccga cacctagtgg ccgtctgggg agccgctagc gcggtgggag tggttagcta
```

```
3301 acttctggac tatttgcgga ctttttggtt ctttggctaa aagtgacctg gaggcattgg
3361 ctggctttgg gggactgggg atggccccga gagcgggctt ttaagatgtc taggtgctgg
3421 aggttagggt gtctcctaat tttgaggtac atttcaagtc ttgggggggc gtcccttcca
3481 atcagccgct cccattctct tagccccgcc cccgccaccc cacatgccca gggaatgggg
3541 gcgggatgag ggatggacct cccttctctc ctccctcgcc ctcctcctgt ctctaccacg
3601 caagccactc cccacgagcc tgccctcccg atggggcccc tcctattctc cccccgccct
3661 ccccctctca ccctgtggtt ttatttcact tggcttcagc gccaatgggc tgaggttgga
3721 gttggaagcc accgcggact aaagctttgt ttaaattcct gagaactgga aagagttaca
3781 gcctccctgg ccaggcgcct cggcgctgtc acccgcgctg atgaggagca ggcgagcttt
3841 taaggatttg aggaaagaag aacgggggga ggggcgggaa gtgaaaaatc caagtgtgcc
3901 tcttagaccc gggggaaagg tggttaagct gggggttgca gtcactactg acaacgcccc
3961 tcttccgcct gtcccagtcc caccaatcac ctgcgtacag aacggcctca ggtaccatga
4021 ccgagacgtg tggaaacccg agccctgccg gatctgcgtc tgcgacaacg gcaaggtgtt
4081 gtgcgatgac gtgatctgtg acgagaccaa gaactgcccc ggcgccgaag tccccgaggg
4141 cgagtgctgt cccgtctgcc ccgacggctc aggtgcgctg cgctcggcct ggggcctggg
4201 gctggggctg ggggtggtcg gcgctcgctg gccctccgtg ctggaggcct ctgccgacgg
4261 gagcagcatt agcaaacctt ggctctaacg ggcgtctctt cgtcccctag agtcacccac
4321 cgaccaagaa accaccggcg tcgaagtaat ctcctgccct cgaattttgc ccctgcgcgg
4381 cccgtgactc ctcacagtcc tcccttctct aacctggcct cttgtttctt ctcccccaat
4441 cccacaggga cccaagggag acactggccc ccgaggccca agggtaagcg ttgcactctg
4501 ggctgtgggg ggctgcaggt gggcatggct ctcggcccca cgctcacccc ggccccgccc
4561 tctccccctg cagggacccg caggcccccc tggccgagat ggcatccctg gacagcctgg
4621 acttcccgga ccccccggac cccccggacc tcccggaccc ccggcctcgg aggagtaagt
4681 ggagaggcct tgtgtgtcca ctctcccctg ttttgttttt gtttttggc agatgacata
4741 ttttatactt tgaggtcgac cctggtgctg tgggatagga gggggagctt cgcctcagtt
4801 tcccccctctg gatagtcatt ctttcccctc cctagtgggg actggggtct gaagatttgt
4861 gggcatgtcc aagtagcttc tgagagggtg aggggtacac agagagggat tatgggagag
4921 gtctctgcct atggacaccc tcgggctaga tttccagaat aatgaagggg catgggttgc
4981 ccacactgcc cttgtctctc cagccaggcc ctcaggctac atttgacgct cactgggcct
5041 gaactgcctt ttttatctgt ccttcagggc tctcctggtg aacaaggtcc ctctggagcc
5101 tctggtcctg ctggtccccg agtaagtcat gccttctctc tcctcttcct gagccccaag
5161 cccaggctca cctcggggac ccttgccagg acccaggcac cctttgcctc tctggagaag
5221 ggttcaggga cagggagtgg gcaaagaaag gaagaatcct gaacaaacaa tctgatctag
5281 ctttggcctc tctgctcccc aatccgtcct cccctggctc agcggctggg aggagctatg
5341 gcatgtccta tggaaagagg ctgaggctgg ctctatgagc cgtggggcca gagccagcag
5401 ggagggtggt gggcctctcc tccagagctg gggttgttcg ggcttctggc agcctttctc
5461 aaaccatttc ccccactcca gggtcccccc ggctctgctg gtgctcctgg caaagatgga
5521 ctcaacggtc tccctggccc cattgggccc cctggtcctc gcggtcgcac tggtgatgct
5581 ggtcctgttg tatgtagccc ctcatccccc ctgctcatgg ccctccagcc cccatagcac
5641 ttggatgccg gaatccccac tctcttccct ctctgtgcag ggtccccccg gccctcctgg
5701 acctcctggt ccccctggtc ctcccagcgc tggtttcgac ttcagcttcc tgccccagcc
5761 acctcaagag aaggctcacg atggtggccg ctactaccgg gctgatgatg ccaatgtggt
5821 tcgtgaccgt gacctcgagg tggacaccac cctcaagagc ctgagccagc agatcgagaa
5881 catccggagc ccagagggaa gccgcaagaa ccccgcccgc acctgccgtg acctcaagat
5941 gtgccactct gactggaaga gtggtgtggg cctgccctag cctctccctc cctcctactc
6001 ctgccatgcc agggtcccca tgcccatatg tgcccctacc atatggtgct ggctgctccc
6061 tttccctgac tccatcttgc cctgccctac cacaggagag tactggattg accccaacca
6121 aggctgcaac ctggatgcca tcaaagtctt ctgcaacatg gagactggtg agacctgcgt
6181 gtaccccact cagcccagtg tggcccagaa gaactggtac atcagcaaga accccaagga
6241 caagaggcat gtctggttcg gcgagagcat gaccgatgga ttccaggtgc gtgagctgga
6301 cctcagagcc agtgttagga gatgggctag cccagtgctc agaagggaca tgaagtcctg
6361 gagtaggtct ctgctaaggg tgatggacag agctgggctg ggaggcaggg gtctcaggtc
6421 cctgctagtg gttcagacac aggctgccga tgggcaggtg gtgcccctct gatataacgg
6481 tgcattgggc agctctctga ggaccctgga caggaggcca gcaggactag aggttcccgc
6541 atagctcact cttccctctc tctcctccct gcagttcgag tatggcggcc agggctccga
6601 ccctgccgat gtggccatcc agctgacctt cctgcgcctg atgtccaccg aggcctccca
6661 gaacatcacc taccactgca agaacagcgt ggcctacatg gaccagcaga ctggcaacct
6721 caagaaggcc ctgctcctca agggctccaa cgagatcgag atccgcgccg agggcaacag
6781 ccgcttcacc tacagcgtca ctgtcgatgg ctgcacggtg agtgcccaga atccccaggc
6841 agggccccac ctctccggcc ttgggcattt tggccaggcc atagtgccct ctctccatca
6901 ctcccacgtg gtaatgcccc ctcccgttgt ctccgcccca ccccagagtc acaccggagc
6961 ctggggcaag acagtgattg aatacaaaac caccaagacc tcccgcctgc ccatcatcga
7021 tgtggccccc ttggacgttg gtgccccaga ccaggaattc ggcttcgacg ttggccctgt
7081 ctgcttcctg taaactccct ccatcccaac ctggctccct cccacccaac caactttccc
7141 cccaacccgg aaacagacaa gcaacccaaa ctgaaccccc ccaaaagcca aaaaatggga
7201 gacaatttca catggacttt ggaaaatatt tttttccttt gcattcatct ctcaaactta
7261 gtttttatct ttgaccaacc gaacatgacc aaaaaccaaa agtgcattca accttaccaa
7321 aaaaaaaaaa aaaaaaaaaa gaataaataa ataacttttt aaaaaaggaa gcttggtcca
7381 cttgcttgaa gacccatgcg ggggtaagtc cctttctgcc cgttgggtta tgaaacccca
7441 atgctgccct ttctgctcct ttctccacac ccccccttggc ctcccctcca ctccttccca
7501 aatctgtctc cccagaagac acaggaaaca atgtattgtc tgcccagcaa tcaaaggcaa
7561 tgctcaaaca cccaagtggc ccccacccctc agcccgctcc tgcccgccca gcaccccccag
7621 gccctgggga cctggggttc tcagactgcc aaagaagcct tgccatctgg cgctcccatg
```

```
     7681 gctcttgcaa catctcccct tcgtttttga gggggtcatg ccgggggagc caccagcccc
     7741 tcactgggtt cggaggagag tcaggaaggg ccacgacaaa gcagaaacat cggatttggg
     7801 gaacgcgtgt catcccttgt gccgcaggct gggcgggaga gactgttctg ttctgttcct
     7861 tgtgtaactg tgttgctgaa agactacctc gttcttgtct tgatgtgtca ccggggcaac
     7921 tgcctggggg cggggatggg ggcagggtgg aagcggctcc ccatttttat accaaaggtg
     7981 ctacatctat gtgatgggtg gggtggggag ggaatcactg gtgctataga aattgagatg
     8041 ccccccccagg ccagcaaatg ttccttttttg ttcaaagtct atttttattc cttgatattt
     8101 tttctttctt ttttttttt tttgtggatg gggacttgtg aatttttcta aaggtgctat
     8161 ttaacatggg aggagagcgt gtgcgctcca gcccagcccg ctgctcactt tccaccctct
     8221 ctccacctgc ctctggcttc tcaggcctct gctctccgac ctctctcctc tgaaacccctc
     8281 ctccacagct gcagcccatc ctcccggctc cctcctagtc tgtcctgcgt cctctgtccc
     8341 cgggtttcag agacaacttc ccaaagcaca aagcagtttt tccctagggg tgggaggaag
     8401 caaaagactc tgtacctatt ttgtatgtgt ataataattt gagatgtttt taattatttt
     8461 gattgctgga ataaagcatg tggaaatgac ccaaacataa tccgcagtgg cctcctaatt
     8521 tccttctttg gagttggggg aggggtagac atggggaagg ggccttgggg tgatgggctt
     8581 gccttccatt cctgcccttt ccctccccac tattctcttc tagatccctc cataacccca
     8641 ctccccttttc tctcaccctt cttataccgc aaacctttct acttcctctt tcatttttcta
     8701 ttcttgcaat ttccttgcac cttttccaaa tcctcttctc ccctgcaata ccatacaggc
     8761 aatccacgtg cacaacacac acacacactc ttcacatctg gggttgtcca aacctcatac
     8821 ccactccccct tcaagcccat ccactctcca ccccctggat gccctgcact tggtggcggt
     8881 gggatgctca tggatactgg gagggtgagg ggagtggaac ccgtgaggag gacctggggg
     8941 cctctccttg aactgacatg aagggtcatc tggcctctgc tcccttctca cccacgctga
     9001 cctcctgccg aaggagcaac gcaacaggag aggggtctgc tgagcctggc gagggtctgg
     9061 gagggaccag gaggaaggcg tgctccctgc tcgctgtcct ggccctgggg gagtgaggga
     9121 gacagacacc tgggagagct gtggggaagg cactcgcacc gtgctcttgg gaaggaagga
     9181 gacctggccc tgctcaccac ggactgggtg cctcgacctc ctgaatcccc agaacacaac
     9241 ccccctgggc tggggtggtc tggggaacca tcgtgccccc gcctcccgcc tactcctttt
     9301 taagcttttt ctttctttct ttctttttttt tcttaaattc ttgcttcctc ttccccccctc
     9361 ttctccgtcc cttctctttc ccttcacctt ctgcctgctc cttcccagag gccctgagaa
     9421 ggcgaggccc ggagggaagg tgactgttcg gctcccccat ctcaggtaag ggggacagcg
     9481 agtgtggtca gcgggggaca gagaatccag ggcgtcagaa gtccagacag aacaggcacc
     9541 cggagccagg gagacagcga gacgacagct ggtcacctct ctcacctcat ctcaggtcta
     9601 aacccaaaaa gaacaaaaag cccgcactcc tccctaaggt ctccagagat ggacaagtct
     9661 gtagatacca tccatcagca atactaatat cgtgatgtca ttaagtattc atggccgttc
     9721 attatggcct tcaataattc atcattagca cttctaccac acttaggaca gtgtagaaag
     9781 ggaatcagtg tagagaggat aagaatgatg gcatataaat gaaggcttat tcatatgcaa
     9841 attacccaaa gttgtgtgat gcctatccat ctctctttct ctctttttttt tgagacagag
     9901 tctcactctg tcacccaggc tggagtgcgg tggcgcagtc tcagctcact gcaacctccg
     9961 cctcccaggt tcaagcaatt ctcatgcctc agcctcccaa gtaactgggg attcagccat
    10021 gcactaccac gcccagctaa tttttatatt tttagtagag tcagggtttt accatgttag
    10081 ccagcctggt ctcgaactcc tggcctcaag tgatccacct gcctcagcct ctcaaagtgc
    10141 tggatgtgcg gacatgagcc acagcaccag gctccatctc ttgatcaatg tgttgattgc
    10201 ttggtgatgc ctgccatgct gcccaggaaa cctgtccact tggacctctg ggacttttagg
    10261 accctggctg tccaggccac atccccatct gccctcgtgg gggacccagg gacatcctga
    10321 gcctggcatc caaggctcta cctggccact acccacctgc ccaggctcac tgtcccctcc
    10381 cctgcccaca tctgtcacac caaccaaaaa gggtggccaa cagttctcac aaatgcactg
    10441 catttcccca cttccaggcc t
//
```

Disclaimer | Write to the Help Desk
NCBI | NLM | NIH

Feb 20 2007 16:53:14

```
3301 acttctggac tatttgcgga ctttttggtt ctttggctaa aagtgacctg gaggcattgg
3361 ctggctttgg gggactgggg atggccccga gagcgggctt ttaagatgtc taggtgctgg
3421 aggttagggt gtctcctaat tttgaggtac atttcaagtc ttgggggggc gtcccttcca
3481 atcagccgct cccattctct tagccccgcc cccgccaccc cacatgccca gggaatgggg
3541 gcgggatgag ggatggacct cccttctctc ctccctcgcc ctcctcctgt ctctaccacg
3601 caagccactc cccacgagcc tgccctcccg atggggcccc tcctattctc cccccgccct
3661 ccccctctca ccctgtggtt ttatttcact tggcttcagc gccaatgggc tgaggttgga
3721 gttggaagcc accgcggact aaagctttgt ttaaattcct gagaactgga aagagttaca
3781 gcctccctgg ccaggcgcct cggcgctgtc acccgcgctg atgaggagca ggcgagcttt
3841 taaggatttg aggaaagaag aacgggggga ggggcgggaa gtgaaaaatc caagtgtgcc
3901 tcttagaccc gggggaaagg tggttaagct gggggttgca gtcactactg acaacgcccc
3961 tcttccgcct gtcccagtcc caccaatcac ctgcgtacag aacggcctca ggtaccatga
4021 ccgagacgtg tggaaacccg agccctgccg gatctgcgtc tgcgacaacg gcaaggtgtt
4081 gtgcgatgac gtgatctgtg acgagaccaa gaactgcccc ggcgccgaag tccccgaggg
4141 cgagtgctgt cccgtctgcc ccgacggctc aggtgcgctg cgctcggcct ggggcctggg
4201 gctggggctg ggggtggtcg gcgctcgctg gccctccgtg ctggaggcct ctgccgacgg
4261 gagcagcatt agcaaacctt ggctctaacg ggcgtctctt cgtcccctag agtcacccac
4321 cgaccaagaa accaccggcg tcgaagtaat ctcctgccct cgaattttgc ccctgcgcgg
4381 cccgtgactc ctcacagtcc tcccttctct aacctggcct cttgtttctt ctcccccaat
4441 cccacaggga cccaagggag acactggccc ccgaggccca agggtaagcg ttgcactctg
4501 ggctgtgggg ggctgcaggt gggcatggct ctcggcccca cgctcacccc ggccccgccc
4561 tctccccctg cagggacccg caggcccccc tggccgagat ggcatccctg gacagcctgg
4621 acttcccgga ccccccggac ccccccggac tcccggaccc ccggcctcgg aggagtaagt
4681 ggagaggcct tgtgtgtcca ctctcccctg ttttgttttt gttttttggc agatgacata
4741 ttttatactt tgaggtcgac cctggtgctg tgggatagga gggggagctt cgcctcagtt
4801 tccccctctg gatagtcatt ctttccccctc cctagtgggg actggggtct gaagatttgt
4861 gggcatgtcc aagtagcttc tgagagggtg aggggtacac agagagggat tatgggagag
4921 gtctctgcct atggacaccc tcgggctaga tttccagaat aatgaagggg catgggttgc
4981 ccacactgcc cttgtctctc cagccaggcc ctcaggctac atttgacgct cactgggcct
5041 gaactgcctt ttttatctgt ccttcagggc tctcctggtg aacaaggtcc ctctggagcc
5101 tctggtcctg ctggtccccg agtaagtcat gccttctctc tcctcttcct gagccccaag
5161 cccaggctca cctcggggac ccttgccagg acccaggcac cctttgcctc tctggagaag
5221 ggttcaggga cagggagtgg gcaaagaaag gaagaatcct gaacaaacaa tctgatctag
5281 ctttggcctc tctgctcccc aatccgtcct cccctggctc agcggctggg aggagctatg
5341 gcatgtccta tggaaagagg ctgaggctgg ctctatgagc cgtggggcca gagccagcag
5401 ggagggtggt gggcctctcc tccagagctg gggttgttcg ggcttctggc agcctttctc
5461 aaaccatttc ccccactcca gggtcccccc ggctctgctg gtgctcctgg caaagatgga
5521 ctcaacggtc tccctggccc cattgggccc cctggtcctc gcggtcgcac tggtgatgct
5581 ggtcctgttg tatgtagccc ctcatccccc ctgctcatgg ccctccagcc cccatagcac
5641 ttggatgccg gaatccccac tctcttccct ctctgtgcag ggtccccccg gccctcctgg
5701 acctcctggt ccccctggtc ctcccagcgc tggtttcgac ttcagcttcc tgccccagcc
5761 acctcaagag aaggctcacg atggtggccg ctactaccgg gctgatgatg ccaatgtggt
5821 tcgtgaccgt gacctcgagg tggacaccac cctcaagagc ctgagccagc agatcgagaa
5881 catccggagc ccagagggaa gccgcaagaa ccccgccccgc acctgccgtg acctcaagat
5941 gtgccactct gactggaaga gtggtgtggg cctgccctag cctctccctc cctcctactc
6001 ctgccatgcc agggtcccca tgcccatatg tgcccctacc atatggtgct ggctgctccc
6061 tttccctgac tccatcttgc cctgccctac cacaggagag tactggattg accccaacca
6121 aggctgcaac ctggatgcca tcaaagtctt ctgcaacatg gagactggtg agacctgcgt
6181 gtaccccact cagcccagtg tggcccagaa gaactggtac atcagcaaga accccaagga
6241 caagaggcat gtctggttcg gcgagagcat gaccgatgga ttccaggtgc gtgagctgga
6301 cctcagagcc agtgttagga gatgggctag cccagtgctc agaagggaca tgaagtcctg
6361 gagtaggtct ctgctaaggg tgatggacag agctgggctg ggaggcaggg gtctcaggtc
6421 cctgctagtg gttcagacac aggctgccga tgggcaggtg gtgcccctct gatataacgg
6481 tgcattgggc agctctctga ggaccctgga caggaggcca gcaggactag aggttcccgc
6541 atagctcact cttccctctc tctcctccct gcagttcgag tatggcggcc agggctccga
6601 ccctgccgat gtggccatcc agctgacctt cctgcgcctg atgtccaccg aggcctccca
6661 gaacatcacc taccactgca agaacagcgt ggcctacatg gaccagcaga ctggcaacct
6721 caagaaggcc ctgctcctca agggctccaa cgagatcgag atccgcgccg agggcaacag
6781 ccgcttcacc tacagcgtca ctgtcgatgg ctgcacggtg agtgcccaga atccccaggc
6841 agggccccac ctctccggcc ttgggcattt tggccaggcc atagtgccct ctctccatca
6901 ctcccacgtg gtaatgcccc ctcccgttgt ctccgcccca ccccagagtc acaccggagc
6961 ctggggcaag acagtgattg aatacaaaac caccaagacc tcccgcctgc ccatcatcga
7021 tgtggccccc ttggacgttg gtgccccaga ccaggaattc ggcttcgacg ttggccctgt
7081 ctgcttcctg taaactccct ccatcccaac ctggctccct cccacccaac caactttccc
7141 cccaacccgg aaacagacaa gcaacccaaa ctgaaccccc ccaaaagcca aaaaatggga
7201 gacaatttca catggacttt ggaaaatatt tttttccttt gcattcatct ctcaaactta
7261 gtttttatct ttgaccaacc gaacatgacc aaaaaccaaa agtgcattca accttaccaa
7321 aaaaaaaaaa aaaaaaaaaa gaataaataa ataacttttt aaaaaaggaa gcttggtcca
7381 cttgcttgaa gacccatgcg ggggtaagtc cctttctgcc cgttgggtta tgaaacccca
7441 atgctgccct ttctgctcct ttctccacac ccccccttggc ctcccctcca ctccttccca
7501 aatctgtctc cccagaagac acaggaaaca atgtattgtc tgcccagcaa tcaaaggcaa
7561 tgctcaaaca cccaagtggc ccccaccctc agcccgctcc tgcccgccca gcaccccccag
7621 gccctgggga cctggggttc tcagactgcc aaagaagcct tgccatctgg cgctcccatg
```

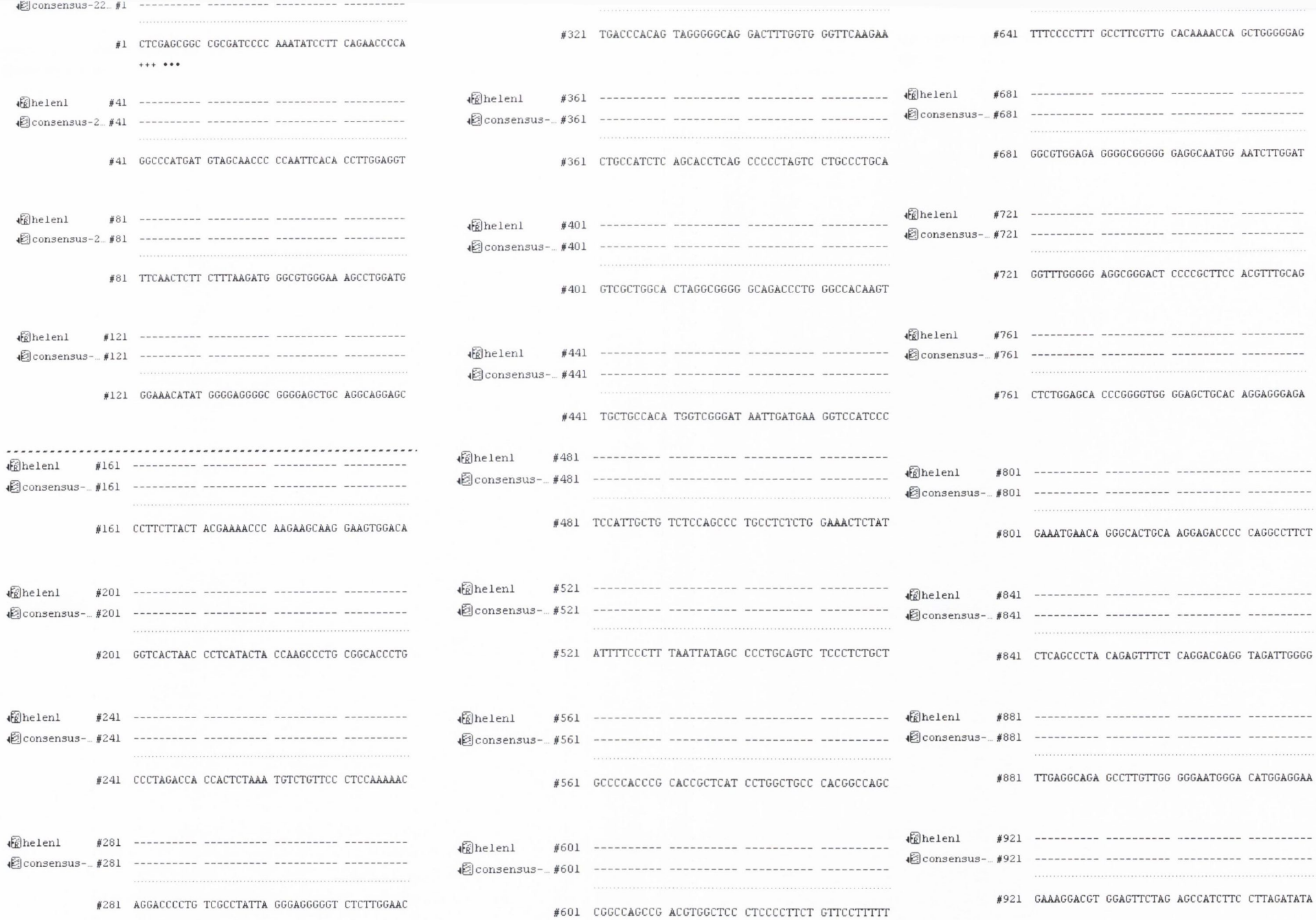
```
     7681 gctcttgcaa catctcccct tcgtttttga gggggtcatg ccgggggagc caccagcccc
     7741 tcactgggtt cggaggagag tcaggaaggg ccacgacaaa gcagaaacat cggatttggg
     7801 gaacgcgtgt catcccttgt gccgcaggct gggcgggaga gactgttctg ttctgttcct
     7861 tgtgtaactg tgttgctgaa agactacctc gttcttgtct tgatgtgtca ccgggggcaac
     7921 tgcctggggg cggggatggg ggcagggtgg aagcggctcc ccatttttat accaaaggtg
     7981 ctacatctat gtgatgggtg gggtggggag ggaatcactg gtgctataga aattgagatg
     8041 ccccccccagg ccagcaaatg ttccttttttg ttcaaagtct atttttattc cttgatattt
     8101 tttctttctt ttttttttttt tttgtggatg gggacttgtg aatttttcta aaggtgctat
     8161 ttaacatggg aggagagcgt gtgcgctcca gcccagcccg ctgctcactt tccaccctct
     8221 ctccacctgc ctctggcttc tcaggcctct gctctccgac ctctctcctc tgaaaccctc
     8281 ctccacagct gcagcccatc ctcccggctc cctcctagtc tgtcctgcgt cctctgtccc
     8341 cgggtttcag agacaacttc ccaaagcaca aagcagtttt tccctagggg tgggaggaag
     8401 caaaagactc tgtacctatt ttgtatgtgt ataataattt gagatgtttt taattatttt
     8461 gattgctgga ataaagcatg tggaaatgac ccaaacataa tccgcagtgg cctcctaatt
     8521 tccttctttg gagttggggg aggggtagac atggggaagg ggccttgggg tgatgggctt
     8581 gccttccatt cctgcccttt ccctccccac tattctcttc tagatccctc cataacccca
     8641 ctcccctttc tctcaccctt cttataccgc aaacctttct acttcctctt tcattttcta
     8701 ttcttgcaat ttccttgcac cttttccaaa tcctcttctc ccctgcaata ccatacaggc
     8761 aatccacgtg cacaacacac acacacactc ttcacatctg gggttgtcca aacctcatac
     8821 ccactccccct tcaagcccat ccactctcca ccccctggat gccctgcact tggtggcggt
     8881 gggatgctca tggatactgg gagggtgagg ggagtggaac ccgtgaggag gacctggggg
     8941 cctctccttg aactgacatg aagggtcatc tggcctctgc tcccttctca cccacgctga
     9001 cctcctgccg aaggagcaac gcaacaggag aggggtctgc tgagcctggc gagggtctgg
     9061 gagggaccag gaggaaggcg tgctccctgc tcgctgtcct ggccctgggg gagtgaggga
     9121 gacagacacc tgggagagct gtggggaagg cactcgcacc gtgctcttgg gaaggaagga
     9181 gacctggccc tgctcaccac ggactgggtg cctcgacctc ctgaatcccc agaacacaac
     9241 ccccctgggc tggggtggtc tggggaacca tcgtgccccc gcctcccgcc tactcctttt
     9301 taagcttttt ctttctttct ttctttttttt tcttaaattc ttgcttcctc ttcccccctc
     9361 ttctccgtcc cttctctttc ccttcacctt ctgcctgctc cttcccagag gccctgagaa
     9421 ggcgaggccc ggagggaagg tgactgttcg gctcccccat ctcaggtaag ggggacagcg
     9481 agtgtggtca gcgggggaca gagaatccag ggcgtcagaa gtccagacag aacaggcacc
     9541 cggagccagg gagacagcga gacgacagct ggtcacctct ctcacctcat ctcaggtcta
     9601 aacccaaaaa gaacaaaaag cccgcactcc tccctaaggt ctccagagat ggacaagtct
     9661 gtagatacca tccatcagca atactaatat cgtgatgtca ttaagtattc atggccgttc
     9721 attatggcct tcaataattc atcattagca cttctaccac acttaggaca gtgtagaaag
     9781 ggaatcagtg tagagaggat aagaatgatg gcatataaat gaaggcttat tcatatgcaa
     9841 attacccaaa gttgtgtgat gcctatccat ctctctttct ctctttttttt tgagacagag
     9901 tctcactctg tcacccaggc tggagtgcgg tggcgcagtc tcagctcact gcaacctccg
     9961 cctcccaggt tcaagcaatt ctcatgcctc agcctcccaa gtaactgggg attcagccat
    10021 gcactaccac gcccagctaa tttttatatt tttagtagag tcagggtttt accatgttag
    10081 ccagcctggt ctcgaactcc tggcctcaag tgatccacct gcctcagcct ctcaaagtgc
    10141 tggatgtgcg gacatgagcc acagcaccag gctccatctc ttgatcaatg tgttgattgc
    10201 ttggtgatgc ctgccatgct gcccaggaaa cctgtccact tggacctctg ggactttagg
    10261 accctggctg tccaggccac atccccatct gccctcgtgg gggacccagg gacatcctga
    10321 gcctggcatc caaggctcta cctggccact acccacctgc ccaggctcac tgtcccctcc
    10381 cctgcccaca tctgtcacac caaccaaaaa gggtggccaa cagttctcac aaatgcactg
    10441 catttcccca cttccaggcc t
//
```

Disclaimer | Write to the Help Desk
NCBI | NLM | NIH

Feb 20 2007 16:53:14

Appendix 2

Alignment of Col1a1 Replacement gene sequence with Col1a1 replacement reference sequence.


consensus-22... #1 ---------- ---------- ---------- ----------
#1 CTCGAGCGGC CGCGATCCCC AAATATCCTT CAGAACCCCA
+++ •••
helenl #41 ---------- ---------- ---------- ----------
consensus-2... #41 ---------- ---------- ---------- ----------
#41 GGCCCATGAT GTAGCAACCC CCAATTCACA CCTTGGAGGT
helenl #81 ---------- ---------- ---------- ----------
consensus-2... #81 ---------- ---------- ---------- ----------
#81 TTCAACTCTT CTTTAAGATG GGCGTGGGAA AGCCTGGATG
helenl #121 ---------- ---------- ---------- ----------
consensus-... #121 ---------- ---------- ---------- ----------
#121 GGAAACATAT GGGGAGGGGC GGGGAGCTGC AGGCAGGAGC
helenl #161 ---------- ---------- ---------- ----------
consensus-... #161 ---------- ---------- ---------- ----------
#161 CCTTCTTACT ACGAAAACCC AAGAAGCAAG GAAGTGGACA
helenl #201 ---------- ---------- ---------- ----------
consensus-... #201 ---------- ---------- ---------- ----------
#201 GGTCACTAAC CCTCATACTA CCAAGCCCTG CGGCACCCTG
helenl #241 ---------- ---------- ---------- ----------
consensus-... #241 ---------- ---------- ---------- ----------
#241 CCCTAGACCA CCACTCTAAA TGTCTGTTCC CTCCAAAAAC
helenl #281 ---------- ---------- ---------- ----------
consensus-... #281 ---------- ---------- ---------- ----------
#281 AGGACCCCTG TCGCCTATTA GGGAGGGGGT CTCTTGGAAC
#321 TGACCCACAG TAGGGGGCAG GACTTTGGTG GGTTCAAGAA
helenl #361 ---------- ---------- ---------- ----------
consensus-... #361 ---------- ---------- ---------- ----------
#361 CTGCCATCTC AGCACCTCAG CCCCCTAGTC CTGCCCTGCA
helenl #401 ---------- ---------- ---------- ----------
consensus-... #401 ---------- ---------- ---------- ----------
#401 GTCGCTGGCA CTAGGCGGGG GCAGACCCTG GGCCACAAGT
helenl #441 ---------- ---------- ---------- ----------
consensus-... #441 ---------- ---------- ---------- ----------
#441 TGCTGCCACA TGGTCGGGAT AATTGATGAA GGTCCATCCC
helenl #481 ---------- ---------- ---------- ----------
consensus-... #481 ---------- ---------- ---------- ----------
#481 TCCATTGCTG TCTCCAGCCC TGCCTCTCTG GAAACTCTAT
helenl #521 ---------- ---------- ---------- ----------
consensus-... #521 ---------- ---------- ---------- ----------
#521 ATTTTCCCTT TAATTATAGC CCCTGCAGTC TCCCTCTGCT
helenl #561 ---------- ---------- ---------- ----------
consensus-... #561 ---------- ---------- ---------- ----------
#561 GCCCCACCCG CACCGCTCAT CCTGGCTGCC CACGGCCAGC
helenl #601 ---------- ---------- ---------- ----------
consensus-... #601 ---------- ---------- ---------- ----------
#601 CGGCCAGCCG ACGTGGCTCC CTCCCCTTCT GTTCCTTTTT
#641 TTTCCCCTTT GCCTTCGTTG CACAAAACCA GCTGGGGGAG
helenl #681 ---------- ---------- ---------- ----------
consensus-... #681 ---------- ---------- ---------- ----------
#681 GGCGTGGAGA GGGGCGGGGG GAGGCAATGG AATCTTGGAT
helenl #721 ---------- ---------- ---------- ----------
consensus-... #721 ---------- ---------- ---------- ----------
#721 GGTTTGGGGG AGGCGGGACT CCCCGCTTCC ACGTTTGCAG
helenl #761 ---------- ---------- ---------- ----------
consensus-... #761 ---------- ---------- ---------- ----------
#761 CTCTGGAGCA CCCGGGGTGG GGAGCTGCAC AGGAGGGAGA
helenl #801 ---------- ---------- ---------- ----------
consensus-... #801 ---------- ---------- ---------- ----------
#801 GAAATGAACA GGGCACTGCA AGGAGACCCC CAGGCCTTCT
helenl #841 ---------- ---------- ---------- ----------
consensus-... #841 ---------- ---------- ---------- ----------
#841 CTCAGCCCTA CAGAGTTTCT CAGGACGAGG TAGATTGGGG
helenl #881 ---------- ---------- ---------- ----------
consensus-... #881 ---------- ---------- ---------- ----------
#881 TTGAGGCAGA GCCTTGTTGG GGGAATGGGA CATGGAGGAA
helenl #921 ---------- ---------- ---------- ----------
consensus-... #921 ---------- ---------- ---------- ----------
#921 GAAAGGACGT GGAGTTCTAG AGCCATCTTC CTTAGATATA


helen1 #961 ---------- ---------- ---------- ----------
consensus-... #961 ---------- ---------- ---------- ----------
#961 GCCTGCTGTC CTTCGGGTCC CCAGACCCTT TCAGAGTGTA
helen1 #1001 ---------- ---------- ---------- ----------
consensus... #1001 ---------- ---------- ---------- ----------
#1001 CAGATGATTC TCTCTGGTTC CTAAGGCATA GAGCAATGAC
helen1 #1041 ---------- ---------- ---------- ----------
consensus... #1041 ---------- ---------- ---------- ----------
#1041 CGGGATTTTC AAGAAAGAGA TGAGGCAGTG GGAAGTAGCC
helen1 #1081 ---------- ---------- ---------- ----------
consensus... #1081 ---------- ---------- ---------- ----------
#1081 CCTAAAACAA AGTCAATCAT CCTCTGCAGC CCATCCCACA
helen1 #1121 ---------- ---------- ---------- ----------
consensus... #1121 ---------- ---------- ---------- ----------
#1121 CCCCCAAAGG AAAGTTTCAC CCAGACACCC AAAATATCCC
helen1 #1161 ---------- ---------- ---------- ----------
consensus... #1161 ---------- ---------- ---------- ----------
#1161 ATACATCCCC AACACTGAGT CCAGGTACAA CTGGAGAAGG
helen1 #1201 ---------- ---------- ---------- ----------
consensus... #1201 ---------- ---------- ---------- ----------
#1201 GGCTTTATGC AGCTCCCAGA AAGACACCCC TTTAGCTAAG
helen1 #1241 ---------- ---------- ---------- ----------
consensus... #1241 ---------- ---------- ---------- ----------
#1241 TGCCCTCCCT CCACCCAGGT TCTCTCTGGT TTGACTGTGC
helen1 #1281 ---------- ---------- ---------- ----------
consensus... #1281 ---------- ---------- ---------- ----------
#1281 TGGGAAGGAG GGTCTCTAAG CAGCCCCTGG CCACAGCCAT
helen1 #1321 ---------- ---------- ---------- ----------
consensus... #1321 ---------- ---------- ---------- ----------
#1321 GGCAAACAAA ACTCTTCTCT AAGTCACCAA TGATCACAGG
helen1 #1361 ---------- ---------- ---------- ----------
consensus... #1361 ---------- ---------- ---------- ----------
#1361 CCTCCCACTA AAAATACTTC CCAACTCTGG GGTGGAAGAG
helen1 #1401 ---------- ---------- ---------- ----------
consensus... #1401 ---------- ---------- ---------- ----------
#1401 TTTGGGGGAT GAATTTTTAG GGGATTGCAA GCCCCAATCC
helen1 #1441 ---------- ---------- ---C------ ----------
consensus... #1441 ---------- ---------- ---------- ----------
#1441 CCACCTCTGT GTCCCTAGAA TCCTCCACCC CTACCTTGGC
helen1 #1481 ---------- ---------- ---------- ----------
consensus... #1481 ---------- ---------- ---------- ----------
#1481 TGCTCCATCA CCCAACCACC AAAGCTTTCT TCTGCAGAGG
helen1 #1521 ---------- ---------- ---------- ----------
consensus... #1521 ---------- ---------- ---------- ----------
#1521 CCACCTAGTC ATGTTTCTCA CCCTGCACCT CAGCCTCCCC
helen1 #1561 ---------- ---------- ---------- ----------
consensus... #1561 ---------- ---------- ---------- ----------
#1561 ACTCCATCTC TCAATCATGC CTAGGGTTTG GAGGAAGGCA
helen1 #1601 ---------- ---------- ---------- ----------
consensus... #1601 ---------- ---------- ---------- ----------
#1601 TTTGATTCTG TTCTGGAGCA CAGCAGAAGA ATTGACATCC
helen1 #1641 ---------- ---------- ---------- ----------
consensus... #1641 ---------- ---------- ---------- ----------
#1641 TCAAAATTAA AACTCCCTTG CCTGCACCCC TCCCTCAGAT
helen1 #1681 ---------- ---------- ---------- ----------
consensus... #1681 ---------- ---------- ---------- ----------
#1681 ATCTGATTCT TAATGTCTAG AAAGGAATCT GTAAATTGTT
helen1 #1721 ---------- ---------- ---------- ----------
consensus... #1721 ---------- ---------- ---------- ----------
#1721 CCCCAAATAT TCCTAAGCTC CATCCCCTAG CCACACCAGA
helen1 #1761 ---------- ---------- ---------- ----------
consensus... #1761 ---------- ---------- ---------- ----------
#1761 AGACACCCCC AAACAGGCAC ATCTTTTTAA TTCCCAGCTT
helen1 #1801 ---------- ---------- ---------- ----------
consensus... #1801 ---------- ---------- ---------- ----------
#1801 CCTCTGTTTT GGAGAGGTCC TCAGCATGCC TCTTTATGCC
helen1 #1841 ---------- ---------- ---------- ----------
consensus... #1841 ---------- ---------- ---------- ----------
#1841 CCTCCCTTAG CTCTTGCCAG GATATCAGAG GGTGACTGGG
helen1 #1881 ---------- ---------- ---------- ----------
consensus... #1881 ---------- ---------- ---------- ----------


consensus... #1921 ---------- ---------- ---------- ----------
#1921 TCCACCTTTG GAAGTCTCCC CACCCAGCTC CCCAGTTCCC
helenl #1961 ---------- ---------- ---------- ----------
consensus... #1961 ---------- ---------- ---------- ----------
#1961 CAGTTCCACT TCTTCTAGAT TGGAGGTCCC AGGAAGAGAG
helenl #2001 ---------- ---------- ---------- ----------
consensus... #2001 ---------- ---------- ---------- ----------
#2001 CAGAGGGGCA CCCCTACCCA CTGGTTAGCC CACGCCATTC
helenl #2041 ---------- ---------- ---------- ----------
consensus... #2041 ---------- ---------- ---------- ----------
#2041 TGAGGACCCA GCTGCACCCC TACCACAGCA CCTCTGGCCC
helenl #2081 ---------- ---------- ---------- ----------
consensus... #2081 ---------- ---------- ---------- ----------
#2081 AGGCTGGGCT GGGGGGCTGG GGAGGCAGAG CTGCGAAGAG
helenl #2121 ---------- ---------- ---------- ----------
consensus... #2121 ---------- ---------- ---------- ----------
#2121 GGGAGATGTG GGGTGGACTC CCTTCCCTCC TCCTCCCCCT
helenl #2161 ---------- ---------- ---------- ----------
consensus... #2161 ---------- ---------- ---------- ----------
#2161 CTCCATTCCA ACTCCCAAAT TGGGGGCCGG GCCAGGCAGC
helenl #2201 ---------- ---------- ---------- ----------
consensus... #2201 ---------- ---------- ---------- ----------
#2201 TCTGATTGGC TGGGGCACGG GCGGCCGGCT CCCCCTCTCC
consensus... #2241 ---------- ---------- ---------- ----------
#2241 GAGGGGCAGG GTTCCTCCCT GCTCTCCATC AGGACAGTAT
helenl #2281 ---------- ---------- ---------- ----------
consensus... #2281 ---------- ---------- ---------- ----------
#2281 AAAAGGGGCC CGGGCCAGTC GTCGGAGCAG ACGGGAGTTT
helenl #2321 ---------- ---------- ---------- ----------
consensus... #2321 ---------- ---------- ---------- ----------
#2321 CTCCTCGGGG TCGGAGCAGG AGGCACGCGG AGTGTGAGGC
helenl #2361 -------A-- ---------- ---------- ----------
consensus... #2361 ---------- ---------- ---------- ----------
#2361 CACGCATGAG CGGACGCTAA CCCCCTCCCC AGCCACAAAG
helenl #2401 ---------- ---------- ---------- ----------
consensus... #2401 ---------- ---------- ---------- ----------
#2401 AGTCTACATG TCTAGGGTCT AGACATGTTC AGCTTTGTGG
helenl #2441 ---------- ---------- ---------- ----------
consensus... #2441 ---------- ---------- ---------- ----------
#2441 ACCTCCGGCT CCTGCTCCTC TTAGCGGCCA CCGCCCTCCT
helenl #2481 ---------- ---------- ---------- ----------
consensus... #2481 ---------- ---------- ---------- ----------
#2481 GACGCACGGC CAAGAGGAAG GCCAAGTCGA GGGCCAAGAC
helenl #2521 ---------- ---------- ---------- ----------
consensus... #2521 ---------- ---------- ---------- ----------
#2521 GAAGACAGTA AGTCCCAAAC TTTTGGGAGT GCAAGGATAC
#2561 TCTATATCGC GCCTTGCGCT TGGTCCCGGG GGCCGCGGCT
helenl #2601 ---------- ---------- ---------- ----------
consensus... #2601 ---------- ---------- ---------- ----------
#2601 TAAAACGAGA CGTGGATGAT CCGGAGACTC GGGAATGGAA
helenl #2641 ---------- ---------- ---------- ----------
consensus... #2641 ---------- ---------- ---------- ----------
#2641 GGGAGATGAT GAGGGCTCTT CCTCGGCGCC CTGAGACAGG
helenl #2681 ---------- ---------- ---------- ----------
consensus... #2681 ---------- ---------- ---------- ----------
#2681 AGGGAGCTCA CCCTGGGGCG AGGTTGGGGT TGAACGCGCC
helenl #2721 ---------- ---------- ---------- ----------
consensus... #2721 ---------- ---------- ---------- ----------
#2721 CCGGGAGCGG GAGGTGAGGG TGGAGCGCGG CGTGAGTTGG
helenl #2761 ---------- ---------- ---------- ----------
consensus... #2761 ---------- ---------- ---------- ----------
#2761 TGCAAGAGAG AATCCCGAGC GCGCAACCGG GGAAGTGGGG
helenl #2801 ---------- ---------- ---------- ----------
consensus... #2801 ---------- ---------- ---------- ----------
#2801 ATCTGGGTGC AGAGTGAGGA AAGCACGTCG AAGATGGGAT
helenl #2841 ---------- ---------- ---------- ----------
consensus... #2841 ---------- ---------- ---------- ----------
#2841 GGGGGCGCCG AGCGGGGCAT TTGAAGCCCA AGATGTAGAA


helen1 #2881 ---------- ---------- ---------- ----------
consensus... #2881 ---------- ---------- ---------- ----------
#2881 GCAATCAGGA AGGCCGTGGG ATGATTCATA AGGAAAGATT
helen1 #2921 ---------- ---------- ---------- ----------
consensus... #2921 ---------- ---------- ---------- ----------
#2921 GCCCTCTCTG CGGGCTAGAG TGTTGCTGGG GCCGTGGGGG
helen1 #2961 ---------- ---------- ---------- ----------
consensus... #2961 ---------- ---------- ---------- ----------
#2961 TGCTGGGCAG CCGCGGAGGG GGTGCGGAGC GTGGGCGGGT
helen1 #3001 ---------- ---------- ---------- ----------
consensus... #3001 ---------- ---------- ---------- ----------
#3001 GGAGGATGAG AAACTTTGGC GCGGACTCGG CGGGGCGGGG
helen1 #3041 ---------- ---------- ---------- ----------
consensus... #3041 ---------- ---------- ---------- ----------
#3041 TCCTTGCGCC CCCTGCTGAC CGATGCTGAG CACTGCGTCT
helen1 #3081 ---------- ---------- ---------- ----------
consensus... #3081 ---------- ---------- ---------- ----------
#3081 CCCGGTCCAA CGCTTACTGG GGCAGGAGCC GGAGCGGGAA
helen1 #3121 ---------- ---------- ---------- ----------
consensus... #3121 ---------- ---------- ---------- ----------
#3121 GACCCGGGTT ATTGCTGGGT GCGGACCCCC ACCTCTAGAT
helen1 #3161 ---------- ---------- ---------- ----------
consensus... #3161 ---------- ---------- ---------- ----------
helen1 #3201 ---------- ---------- ---------- ----------
consensus... #3201 ---------- ---------- ---------- ----------
#3201 AAAGAGGTAG TCGGGCCGGT GAGGTCGGCC CCGCCCCGGC
helen1 #3241 ---------- ---------- ---------- ----------
consensus... #3241 ---------- ---------- ---------- ----------
#3241 CCCATTGCTT AGCGTTGCCC GACACCTAGT GGCCGTCTGG
helen1 #3281 ---------- ---------- ---------- ----------
consensus... #3281 ---------- ---------- ---------- ----------
#3281 GGAGCCGCTA GCGCGGTGGG AGTGGTTAGC TAACTTCTGG
helen1 #3321 ---------- ---------- ---------- ----------
consensus... #3321 ---------- ---------- ---------- ----------
#3321 ACTATTTGCG GACTTTTTGG TTCTTTGGCT AAAAGTGACC
helen1 #3361 ---------- ---------- ---------- ----------
consensus... #3361 ---------- ---------- ---------- ----------
#3361 TGGAGGCATT GGCTGGCTTT GGGGGACTGG GGATGGCCCC
helen1 #3401 ---------- ---------- ---------- ----------
consensus... #3401 ---------- ---------- ---------- ----------
#3401 GAGAGCGGGC TTTTAAGATG TCTAGGTGCT GGAGGTTAGG
helen1 #3441 ---------- ---------- ---------- ----------
consensus... #3441 ---------- ---------- ---------- ----------
#3441 GTGTCTCCTA ATTTTGAGGT ACATTTCAAG TCTTGGGGGG
helen1 #3481 --G------- ---------- ---------- -T--------
consensus... #3481 ---------- ---------- ---------- ----------
#3481 GCCTCCCTTC CAATCAGCCG CTCCCATTCT CCTAGCCCCG
helen1 #3521 ---------- -----ATG:: :---------- ----------
consensus... #3521 ---------- ---------- ---------- ----------
#3521 CCCCCGCCAC CCCAC:::CT GCCCAGGGAA TGGGGGCGGG
helen1 #3558 -------A-- ---------- ---------- ----------
consensus... #3558 ---------- ---------- ---------- ----------
#3558 ATGAGGGCTG GACCTCCCTT CTCTCCTCCC TCGCCCTCCT
helen1 #3598 ---------- ---------- ---------- ----------
consensus... #3598 ---------- ---------- ---------- ----------
#3598 CCTGTCTCTA CCACGCAAGC CACTCCCCAC GAGCCTGCCC
helen1 #3638 ---------- ---------- ---------- ----------
consensus... #3638 ---------- ---------- ---------- ----------
#3638 TCCCGATGGG GCCCCTCCTA TTCTCCCCCC GCCCTCCCCC
helen1 #3678 ---------- ---:------ ---------- ----------
consensus... #3678 ---------- ---------- ---------- ----------
#3678 TCTCACCCTG TGGTTTTTAT TTCACTTGGC TTCAGCGCCA
helen1 #3717 ---------- ---------- ---------- ----------
consensus... #3717 ---------- ---------- ---------- ----------
#3717 ATGGGCTGAG GTTGGAGTTG GAAGCCACCG CGGACTAAAG
helen1 #3757 ---------- ---------- ---------- ----------
consensus... #3757 ---------- ---------- ---------- ----------
#3757 CTTTGTTTAA ATTCCTGAGA ACTGGAAAGA GTTACAGCCT
helen1 #3797 ---------- ---------- ---------- ----------
consensus... #3797 ---------- ---------- ---------- ----------

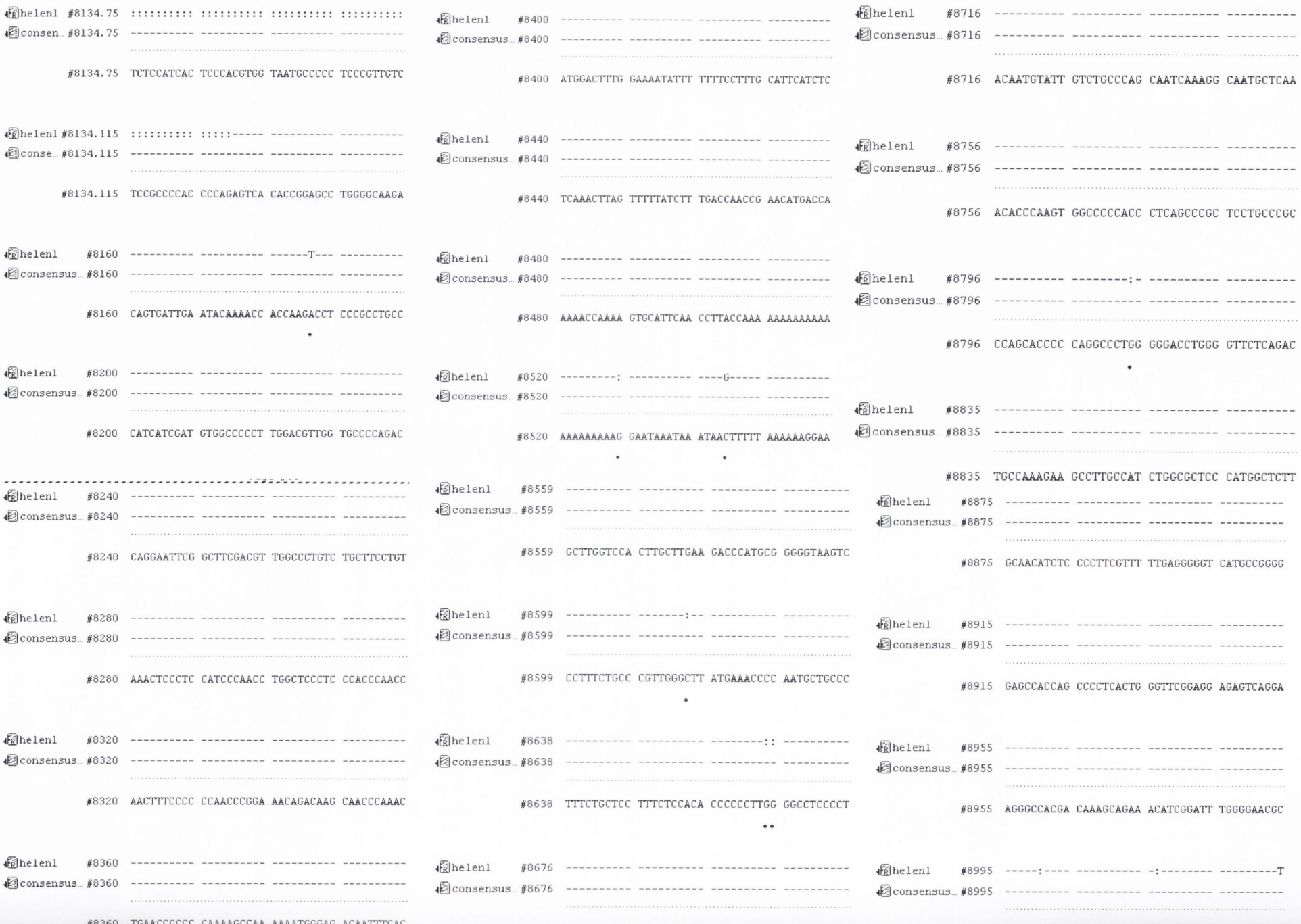

#3837 GGAGCAGGCG AGCTTTTAAG GATTTGAGGA AAGAAGAACG
helenl #3877 ---------- ---------- ---------- ----------
consensus... #3877 ---------- ---------- ---------- ----------
#3877 GGGGGAGGGG CGGGAAGTGA AAAATCCAAG TGTGCCTCTT
helenl #3917 ---------- ---------- ---------- ----------
consensus... #3917 ---------- ---------- ---------- ----------
#3917 AGACCCGGGG GAAAGGTGGT TAAGCTGGGG GTTGCAGTCA
helenl #3957 ---------- ---------- ---------- ----------
consensus... #3957 ---------- ---------- ---------- ----------
#3957 CTACTGACAA CGCCCCTCTT CCGCCTGTCC CAGTCCCACC
helenl #3997 ---------- ---------- ---------- ----------
consensus... #3997 ---------- ---------- ---------- ----------
#3997 AATCACCTGC GTACAGAACG GCCTCAGGTA CCATGACCGA
helenl #4037 ---------- ---------- ---------- ----------
consensus... #4037 ---------- ---------- ---------- ----------
#4037 GACGTGTGGA AACCCGAGCC CTGCCGGATC TGCGTCTGCG
helenl #4077 ---------- ---------- ---------- ----------
consensus... #4077 ---------- ---------- ---------- ----------
#4077 ACAACGGCAA GGTGTTGTGC GATGACGTGA TCTGTGACGA
helenl #4117 ---------- ---------- ---------- ----------
consensus... #4117 ---------- ---------- ---------- ----------
#4117 GACCAAGAAC TGCCCCGGCG CCGAAGTCCC CGAGGGCGAG
consensus... #4157 ---------- ---------- ---------- ----------
#4157 TGCTGTCCCG TCTGCCCCGA CGGCTCAGAG TCACCCACCG
helenl #4197 ---------- ---------- ---------- ----------
consensus... #4197 ---------- ---------- ---------- ----------
#4197 ACCAAGAAAC CACCGGCGTC GAGGGACCCA AGGGAGACAC
helenl #4237 ---------- ---------- ---------- ----------
consensus... #4237 ---------- ---------- ---------- ----------
#4237 TGGCCCCCGA GGCCCAAGGG GACCCGCAGG CCCCCCTGGC
helenl #4277 ---------- ---------- ---------- ----------
consensus... #4277 ---------- ---------- ---------- ----------
#4277 CGAGATGGCA TCCCTGGACA GCCTGGACTT CCCGGACCCC
helenl #4317 ---------- ---------- ---------- ----------
consensus... #4317 ---------- ---------- ---------- ----------
#4317 CCGGACCCCC CGGACCTCCC GGACCCCCTG GCCTCGGAGG
helenl #4357 ---------- ---------- ---------- ----------
consensus... #4357 ---------- ---------- ---------- ----------
#4357 AAACTTTGCT CCCCAGCTGT CTTATGGCTA TGATGAGAAA
helenl #4397 ---------- ---------- ---------- ----------
consensus... #4397 ---------- ---------- ---------- ----------
#4397 TCAACCGGAG GAATTTCCGT GCCTGGCCCC ATGGGTCCCT
helenl #4437 ---------- ---------- ---------- ----------
consensus... #4437 ---------- ---------- ---------- ----------
#4437 CTGGTCCTCG TGGTCTCCCT GGCCCCCCTG GTGCACCTGG
#4477 TCCCCAAGGC TTCCAAGGTC CCCCTGGTGA GCCTGGCGAG
helenl #4517 ---------- ---------- ---------- ----------
consensus... #4517 ---------- ---------- ---------- ----------
#4517 CCTGGAGCTT CAGGTCCCAT GGGTCCCCGA GGTCCCCCAG
helenl #4557 ---------- ---------- ---------- ----------
consensus... #4557 ---------- ---------- ---------- ----------
#4557 GTCCCCCTGG AAAGAATGGA GATGATGGGG AAGCTGGAAA
helenl #4597 ---------- ---------- ---------- ----------
consensus... #4597 ---------- ---------- ---------- ----------
#4597 ACCTGGTCGT CCTGGTGAGC GTGGGCCTCC TGGGCCTCAG
helenl #4637 ---------- ---------- ---------- ----------
consensus... #4637 ---------- ---------- ---------- ----------
#4637 GGTGCTCGAG GATTGCCCGG AACAGCTGGC CTCCCTGGAA
helenl #4677 ---------- ---------- ---------- ----------
consensus... #4677 ---------- ---------- ---------- ----------
#4677 TGAAGGGACA CAGAGGTTTC AGTGGTTTGG ATGGTGCCAA
helenl #4717 ---------- ---------- ---------- ----------
consensus... #4717 ---------- ---------- ---------- ----------
#4717 GGGAGATGCT GGTCCTGCTG GTCCTAAGGG TGAGCCTGGC
helenl #4757 ---------- ---------- ---------- ----------
consensus... #4757 ---------- ---------- ---------- ----------
#4757 AGCCCTGGTG AAAATGGAGC TCCTGGTCAG ATGGGCCCCC


helen1 #4797 ---------- ---------- ---------- ----------
consensus... #4797 ---------- ---------- ---------- ----------
#4797 GTGGCCTGCC TGGTGAGAGA GGTCGCCCTG GAGCCCCTGG
helen1 #4837 ---------- ---------- ---------- ----------
consensus... #4837 ---------- ---------- ---------- ----------
#4837 CCCTGCTGGT GCTCGTGGAA ATGATGGTGC TACTGGTGCT
helen1 #4877 ---------- ---------- ---------- ----------
consensus... #4877 ---------- ---------- ---------- ----------
#4877 GCCGGGCCCC CTGGTCCCAC CGGCCCCGCT GGTCCTCCTG
helen1 #4917 ---------- ---------- ---------- ----------
consensus... #4917 ---------- ---------- ---------- ----------
#4917 GCTTCCCTGG TGCTGTTGGT GCTAAGGGTG AAGCTGGTCC
helen1 #4957 ---------- ---------- ---------- ----------
consensus... #4957 ---------- ---------- ---------- ----------
#4957 CCAAGGGCCC CGAGGCTCTG AAGGTCCCCA GGGTGTGCGT
helen1 #4997 ---------- ---------- ---------- ----------
consensus... #4997 ---------- ---------- ---------- ----------
#4997 GGTGAGCCTG GCCCCCCTGG CCCTGCTGGT GCTGCTGGCC
helen1 #5037 ---------- ---------- ---------- ----------
consensus... #5037 ---------- ---------- ---------- ----------
#5037 CTGCTGGAAA CCCTGGTGCT GATGGACAGC CTGGTGCTAA
helen1 #5077 ---------- ---------- ---------- ----------
consensus... #5077 ---------- ---------- ---------- ----------
#5077 AGGTGCCAAT GGTGCTCCTG GTATTGCTGG TGCTCCTGGC
helen1 #5117 ---------- ---------- ---------- ----------
consensus... #5117 ---------- ---------- ---------- ----------
#5117 TTCCCTGGTG CCCGAGGCCC CTCTGGACCC CAGGGCCCCG
helen1 #5157 ---------- ---------- ---------- ----------
consensus... #5157 ---------- ---------- ---------- ----------
#5157 GCGGCCCTCC TGGTCCCAAG GGTAACAGCG GTGAACCTGG
helen1 #5197 ---------- ---------- ---------- ----------
consensus... #5197 ---------- ---------- ---------- ----------
#5197 TGCTCCTGGC AGCAAAGGAG ACACTGGTGC TAAGGGAGAG
helen1 #5237 ---------- --------C- ---------- ----------
consensus... #5237 ---------- ---------- ---------- ----------
#5237 CCTGGCCCTG TTGGTGTTTA AGGACCCCCT GGCCCTGCTG
helen1 #5277 ---------- ---------- ---------- ----------
consensus... #5277 ---------- ---------- ---------- ----------
#5277 GAGAGGAAGG AAAGCGAGGA GCTCGAGGTG AACCCGGACC
helen1 #5317 ---------- ---------- ---------- ----------
consensus... #5317 ---------- ---------- ---------- ----------
#5317 CACTGGCCTG CCCGGACCCC CTGGCGAGCG TGGTGGACCT
helen1 #5357 ---------- ---------- ---------- ----------
consensus... #5357 ---------- ---------- ---------- ----------
#5357 GGTAGCCGTG GTTTCCCTGG CGCAGATGGT GTTGCTGGTC
helen1 #5397 ---------- ---------- ---------- -------C--
consensus... #5397 ---------- ---------- ---------- ----------
helen1 #5437 ---------- ---------- ---------- ----------
consensus... #5437 ---------- ---------- ---------- ----------
#5437 TGGCCCCAAA GGATCTCCTG GTGAAGCTGG TCGTCCCGGT
helen1 #5477 ---------- ---------- ---------- ----------
consensus... #5477 ---------- ---------- ---------- ----------
#5477 GAAGCTGGTC TGCCTGGTGC CAAGGGTCTG ACTGGAAGCC
helen1 #5517 ---------- ---------- ---------- ----------
consensus... #5517 ---------- ---------- ---------- ----------
#5517 CTGGCAGCCC TGGTCCTGAT GGCAAAACTG GCCCCCCTGG
helen1 #5557 ---------- ---------- ---------- ----------
consensus... #5557 ---------- ---------- ---------- ----------
#5557 TCCCGCCGGT CAAGATGGTC GCCCCGGACC CCCAGGCCCA
helen1 #5597 ---------- ---------- ---------- ----------
consensus... #5597 ---------- ---------- ---------- ----------
#5597 CCTGGTGCCC GTGGTCAGGC TGGTGTGATG GGATTCCCTG
helen1 #5637 ---------- ---------- ---------- ----------
consensus... #5637 ---------- ---------- ---------- ----------
#5637 GACCTAAAGG TGCTGCTGGA GAGCCCGGCA AGGCTGGAGA
helen1 #5677 ---------- ---------- ---------- ----------
consensus... #5677 ---------- ---------- ---------- ----------
#5677 GCGAGGTGTT CCCGGACCCC CTGGCGCTGT CGGTCCTGCT
helen1 #5717 ---------- ---------- ---------- --------C-
consensus... #5717 ---------- ---------- ---------- ----------


#5757 CTGCTGGTCC CGCTGGCGAG AGAGGTGAAC AAGGCCCTGC
helenl #5797 -----C----
consensus #5797
#5797 TGGCT:CCCC GGATTCCAGG GTCTCCCTGG TCCTGCTGGT
helenl #5837
consensus #5837
#5837 CCTCCAGGTG AAGCAGGCAA ACCTGGTGAA CAGGGTGTTC
helenl #5877
consensus #5877
#5877 CTGGAGACCT TGGCGCCCCT GGCCCCTCTG GAGCAAGAGG
helenl #5917
consensus #5917
#5917 CGAGAGAGGT TTCCCTGGCG AGCGTGGTGT GCAAGGTCCC
helenl #5957
consensus #5957
#5957 CCTGGTCCTG CTGGACCCCG AGGGGCCAAC GGTGCTCCCG
helenl #5997
consensus #5997
#5997 GCAACGATGG TGCTAAGGGT GATGCTGGTG CCCCTGGAGC
helenl #6037
consensus #6037
#6037 TCCCGGTAGC CAGGGCGCCC CTGGCCTTCA GGGAATGCCT
#6077 GGTGAACGTG GTGCAGCTGG TCTTCCAGGG CCTAAGGGTG
helenl #6117
consensus #6117
#6117 ACAGAGGTGA TGCTGGTCCC AAAGGTGCTG ATGGCTCTCC
helenl #6157
consensus #6157
#6157 TGGCAAAGAT GGCGTCCGTG GTCTGACCGG CCCCATTGGT
helenl #6197
consensus #6197
#6197 CCTCCTGGCC CTGCTGGTGC CCCTGGTGAC AAGGGTGAAA
helenl #6237
consensus #6237
#6237 GTGGTCCCAG CGGCCCTGCT GGTCCCACTG GAGCTCGTGG
helenl #6277
consensus #6277
#6277 TGCCCCCGGA GACCGTGGTG AGCCTGGTCC CCCCGGCCCT
helenl #6317
consensus #6317
#6317 GCTGGCTTTG CTGGCCCCCC TGGTGCTGAC GGCCAACCTG
helenl #6357 -----C--------
consensus #6357
#6357 GTGCTAAAGG CGAACCTGGT GATGCTGGTG CTAAAGGCGA
#6397 TGCTGGTCCC CCTGGCCCTG CCGGACCCGC TGGACCCCCT
helenl #6437
consensus #6437
#6437 GGCCCCATTG GTAATGTTGG TGCTCCTGGA GCCAAAGGTG
helenl #6477
consensus #6477
#6477 CTCGCGGCAG CGCTGGTCCC CCTGGTGCTA CTGGTTTCCC
helenl #6517
consensus #6517
#6517 TGGTGCTGCT GGCCGAGTCG GTCCTCCTGG CCCCTCTGGA
helenl #6557
consensus #6557
#6557 AATGCTGGAC CCCCTGGCCC TCCTGGTCCT GCTGGCAAAG
helenl #6597
consensus #6597
#6597 AAGGCGGCAA AGGTCCCCGT GGTGAGACTG GCCCTGCTGG
helenl #6637
consensus #6637
#6637 ACGTCCTGGT GAAGTTGGTC CCCCTGGTCC CCCTGGCCCT
helenl #6677
consensus #6677
#6677 GCTGGCGAGA AAGGATCCCC TGGTGCTGAT GGTCCTGCTG


helenl #6717 ---------- ---------- ---------- ----------
consensus... #6717 ---------- ---------- ---------- ----------
#6717 GTGCTCCTGG TACTCCCGGG CCTCAAGGTA TTGCTGGACA
helenl #6757 ---------- ---------- ---------- ----------
consensus... #6757 ---------- ---------- ---------- ----------
#6757 GCGTGGTGTG GTCGGCCTGC CTGGTCAGAG AGGAGAGAGA
helenl #6797 ---------- ---------- ---------- ----------
consensus... #6797 ---------- ---------- ---------- ----------
#6797 GGCTTCCCTG GTCTTCCTGG CCCCTCTGGT GAACCTGGCA
helenl #6837 ---------- ---------- ---------- ----------
consensus... #6837 ---------- ---------- ---------- ----------
#6837 AACAAGGTCC CTCTGGAGCA AGTGGTGAAC GTGGTCCCCC
helenl #6877 C--------- ---------- ---------- ----------
consensus... #6877 ---------- ---------- ---------- ----------
#6877 TGGTCCCATG GGCCCCCCTG GATTGGCTGG ACCCCCTGGT
helenl #6917 ---------- ---------- -----C---- ----------
consensus... #6917 ---------- ---------- ---------- ----------
#6917 GAATCTGGAC GTGAGGGGGC TCCTGGTGCC GAAGGTTCCC
helenl #6957 ---------- ---------- ---------- ----------
consensus... #6957 ---------- ---------- ---------- ----------
#6957 CTGGACGAGA CGGTTCTCCT GGCGCCAAGG GTGACCGTGG
helenl #6997 ---------- ---------- ---------- ----------
consensus... #6997 ---------- ---------- ---------- ----------
#6997 TGAGACCGGC CCCGCTGGAC CCCCTGGTGC TCCTGGTGCT
helenl #7037 ---------- ---------- ---------- ---------G
consensus... #7037 ---------- ---------- ---------- ----------
#7037 CCTGGTGCCC CTGGCCCCGT TGGCCCTGCT GGCAAGAGTA
helenl #7077 G--------- ---------- ---------- ----------
consensus... #7077 ---------- ---------- ---------- ----------
#7077 ATGATCGTGG TGAGACTGGT CCTGCTGGTC CCGCCGGTCC
helenl #7117 C--------C --C------- ---------- ----------
consensus... #7117 ---------- ---------- ---------- ----------
#7117 TGTCGGCCCT GTTGGCGCCC GTGGCCCCGC CGGACCCCAA
helenl #7157 ---------- ---------- ---------- ----------
consensus... #7157 ---------- ---------- ---------- ----------
#7157 GGCCCCCGTG GTGACAAGGG TGAGACAGGC GAACAGGGCG
helenl #7197 ---------- ---------- ---------- ----------
consensus... #7197 ---------- ---------- ---------- ----------
#7197 ACAGAGGCAT AAAGGGTCAC CGTGGCTTCT CTGGCCTCCA
helenl #7237 ---------- ---------- ---------- ----------
consensus... #7237 ---------- ---------- ---------- ----------
#7237 GGGTCCCCCT GGCCCTCCTG GCTCTCCTGG TGAACAAGGT
helenl #7277 ---------- ---------- ---------- ----------
consensus... #7277 ---------- ---------- ---------- ----------
#7277 CCCTCTGGAG CCTCTGGTCC TGCTGGTCCC CGAGGTCCCC
helenl #7317 ---------- ---------- ---------- ----------
consensus... #7317 ---------- ---------- ---------- ----------
helenl #7357 ---------- ---------- ---------- ----------
consensus... #7357 ---------- ---------- ---------- ----------
#7357 TCTCCCTGGC CCCATTGGGC CCCCTGGTCC TCGCGGTCGC
helenl #7397 ---------- ---------- ---------- ----------
consensus... #7397 ---------- ---------- ---------- ----------
#7397 ACTGGTGATG CTGGTCCTGT TGGTCCCCCC GGCCCTCCTG
helenl #7437 ---------- ---------- ---------- ----------
consensus... #7437 ---------- ---------- ---------- ----------
#7437 GACCTCCTGG TCCCCCTGGT CCTCCCAGCG CTGGTTTCGA
helenl #7477 ---------- ---------- ---------- ----------
consensus... #7477 ---------- ---------- ---------- ----------
#7477 CTTCAGCTTC CTGCCCCAGC CACCTCAAGA GAAGGCTCAC
helenl #7517 ---------- ---------- ---------- ----------
consensus... #7517 ---------- ---------- ---------- ----------
#7517 GATGGTGGCC GCTACTACCG GGCTGATGAT GCCAATGTGG
helenl #7557 ---------- ---------- ---------- ----------
consensus... #7557 ---------- ---------- ---------- ----------
#7557 TTCGTGACCG TGACCTCGAG GTGGACACCA CCCTCAAGAG
helenl #7597 ---------- ---------- ---------- ---------A
consensus... #7597 ---------- ---------- ---------- ----------
#7597 CCTGAGCCAG CAGATCGAGA ACATCCGGAG CCCAGAGGGC
helenl #7637 ---------- ---------- ---------- ----------
consensus... #7637 ---------- ---------- ---------- ----------


#7677 TGTGCCACTC TGACTGGAAG AGTGGTGAGG GCCTGCCCTA
helenl #7705.12 :::::::::: :::::::::: :::::::::: ::::::::::
consen... #7705.12 ---------- ---------- ---------- ----------
#7705.12 GCCTCTCCCT CCCTCCTACT CCTGCCATGC CAGGGTCCCC
helenl #7705.52 :::::::::: :::::::::: :::::::::: ::::::::::
consen... #7705.52 ---------- ---------- ---------- ----------
#7705.52 ATGCCCATAT GTGCCCCTAC CATATGGTGC TGGCTGCTCC
helenl #7705.92 :::::::::: :::::::::: :::::::::: ::::::::::
consen... #7705.92 ---------- ---------- ---------- ----------
#7705.92 CCTTCCCTGA CTCCATCTTG CCCTGCCCTA CCACAGGAGA
helenl #7705.132 :--------- ---------- ---------- ----------
conse... #7705.132 ---------- ---------- ---------- ----------
#7705.132 GTACTGGATT GACCCCAACC AAGGCTGCAA CCTGGATGCC
helenl #7745 ---------- ---------- ---------- ----------
consensus... #7745 ---------- ---------- ---------- ----------
#7745 ATCAAAGTCT TCTGCAACAT GGAGACTGGT GAGACCTGCG
helenl #7785 ---------- ---------- ---------- ----------
consensus... #7785 ---------- ---------- ---------- ----------
#7785 TGTACCCCAC TCAGCCCAGT GTGGCCCAGA AGAACTGGTA
helenl #7825 ---------- ---------- ---------- ----------
consensus... #7825 ---------- ---------- ---------- ----------
#7825 CATCAGCAAG AACCCCAAGG ACAAGAGGCA TGTCTGGTTC
consensus... #7865 ---------- ---------- ---------- ----------
#7865 GGCGAGAGCA TGACCGATGG ATTCCAGGTG CGTGAGCTGG
helenl #7891.14 :::::::::: :::::::::: :::::::::: ::::::::::
consen... #7891.14 ---------- ---------- ---------- ----------
#7891.14 ACCTCAGAGC CAGTGTTAGG AGATGGGCTA GCCCAGTGCT
helenl #7891.54 :::::::::: :::::::::: :::::::::: ::::::::::
consen... #7891.54 ---------- ---------- ---------- ----------
#7891.54 CAGAAGGGAC ATGAAGTCCT GGAGTAGGTC TCTGCTAAGG
helenl #7891.94 :::::::::: :::::::::: :::::::::: ::::::::::
consen... #7891.94 ---------- ---------- ---------- ----------
#7891.94 GTGATGGACA GAGCTGGGCT GGGAGGCAGG GGTCTCAGGT
helenl #7891.134 :::::::::: :::::::::: :::::::::: ::::::::::
conse... #7891.134 ---------- ---------- ---------- ----------
#7891.134 CCCTGATAGT GGTTCAGACA CAGGCTGCCG ATGGGCAGGT
helenl #7891.174 :::::::::: :::::::::: :::::::::: ::::::::::
conse... #7891.174 ---------- ---------- ---------- ----------
#7891.174 GGTGCCTCCT CTCGATATAA CGGTGCATTG GGCAGCTCTC
helenl #7891.214 :::::::::: :::::::::: :::::::::: ::::::::::
conse... #7891.214 ---------- ---------- ---------- ----------
#7891.214 TGAGGACCCT GGACAGGGAG GCCAGGCAGG ACTAGAGGTT
helenl #7891.254 :::::::::: :::::::::: :::::::::: ::::::::::
conse... #7891.254 ---------- ---------- ---------- ----------
#7891.254 CCCGCATAGC GGCTCACTCT TCCCTCTCTC CCCTCCCCTG
#7891.294 CAGTTCGAGT ATGGCGGCCA GGGCTCCGAC CCTGCCGATG
helenl #7929 ---------- ---------- ---------- ----------
consensus... #7929 ---------- ---------- ---------- ----------
#7929 TGGCCATCCA GCTGACCTTC CTGCGCCTGA TGTCCACCGA
helenl #7969 ---------- ---------- ---------- ----------
consensus... #7969 ---------- ---------- ---------- ----------
#7969 GGCCTCCCAG AACATCACCT ACCACTGCAA GAACAGCGTG
helenl #8009 ---------- ---------- ---------- ----------
consensus... #8009 ---------- ---------- ---------- ----------
#8009 GCCTACATGG ACCAGCAGAC TGGCAACCTC AAGAAGGCCC
helenl #8049 --------A- ---------- ---------- ----------
consensus... #8049 ---------- ---------- ---------- ----------
#8049 TGCTCCTCCA GGGCTCCAAC GAGATCGAGA TCCGCGCCGA
helenl #8089 ---------- ---------- ---------- ----------
consensus... #8089 ---------- ---------- ---------- ----------
#8089 GGGCAACAGC CGCTTCACCT ACAGCGTCAC TGTCGATGGC
helenl #8129 ------:::: :::::::::: :::::::::: ::::::::::
consensus... #8129 ---------- ---------- ---------- ----------
#8129 TGCACGGTGA GTGCCCAGAA TCCCCAGGCA GGGCCCCACC
helenl #8134.35 :::::::::: :::::::::: :::::::::: ::::::::::
consen... #8134.35 ---------- ---------- ---------- ----------
#8134.35 TCTCCGGCCT TGGGCTTTTT GGCCAGGCCA TAGTGCCCTC


helen1 #8134.75 :::::::::: :::::::::: :::::::::: ::::::::::
consen... #8134.75 ---------- ---------- ---------- ----------
#8134.75 TCTCCATCAC TCCCACGTGG TAATGCCCCC TCCCGTTGTC
helen1 #8134.115 :::::::::: :::::----- ---------- ----------
conse... #8134.115 ---------- ---------- ---------- ----------
#8134.115 TCCGCCCCAC CCCAGAGTCA CACCGGAGCC TGGGGCAAGA
helen1 #8160 ---------- ---------- ------T--- ----------
consensus... #8160 ---------- ---------- ---------- ----------
#8160 CAGTGATTGA ATACAAAACC ACCAAGACCT CCCGCCTGCC
helen1 #8200 ---------- ---------- ---------- ----------
consensus... #8200 ---------- ---------- ---------- ----------
#8200 CATCATCGAT GTGGCCCCCT TGGACGTTGG TGCCCCAGAC
helen1 #8240 ---------- ---------- ---------- ----------
consensus... #8240 ---------- ---------- ---------- ----------
#8240 CAGGAATTCG GCTTCGACGT TGGCCCTGTC TGCTTCCTGT
helen1 #8280 ---------- ---------- ---------- ----------
consensus... #8280 ---------- ---------- ---------- ----------
#8280 AAACTCCCTC CATCCCAACC TGGCTCCCTC CCACCCAACC
helen1 #8320 ---------- ---------- ---------- ----------
consensus... #8320 ---------- ---------- ---------- ----------
#8320 AACTTTCCCC CCAACCCGGA AACAGACAAG CAACCCAAAC
helen1 #8360 ---------- ---------- ---------- ----------
consensus... #8360 ---------- ---------- ---------- ----------
helen1 #8400 ---------- ---------- ---------- ----------
consensus... #8400 ---------- ---------- ---------- ----------
#8400 ATGGACTTTG GAAAATATTT TTTTCCTTTG CATTCATCTC
helen1 #8440 ---------- ---------- ---------- ----------
consensus... #8440 ---------- ---------- ---------- ----------
#8440 TCAAACTTAG TTTTTATCTT TGACCAACCG AACATGACCA
helen1 #8480 ---------- ---------- ---------- ----------
consensus... #8480 ---------- ---------- ---------- ----------
#8480 AAAACCAAAA GTGCATTCAA CCTTACCAAA AAAAAAAAAA
helen1 #8520 ---------: ---------- ----G----- ----------
consensus... #8520 ---------- ---------- ---------- ----------
#8520 AAAAAAAAAG GAATAAATAA ATAACTTTTT AAAAAAGGAA
helen1 #8559 ---------- ---------- ---------- ----------
consensus... #8559 ---------- ---------- ---------- ----------
#8559 GCTTGGTCCA CTTGCTTGAA GACCCATGCG GGGGTAAGTC
helen1 #8599 ---------- -------:-- ---------- ----------
consensus... #8599 ---------- ---------- ---------- ----------
#8599 CCTTTCTGCC CGTTGGGCTT ATGAAACCCC AATGCTGCCC
helen1 #8638 ---------- ---------- --------:: ----------
consensus... #8638 ---------- ---------- ---------- ----------
#8638 TTTCTGCTCC TTTCTCCACA CCCCCCTTGG GGCCTCCCCT
helen1 #8676 ---------- ---------- ---------- ----------
consensus... #8676 ---------- ---------- ---------- ----------
helen1 #8716 ---------- ---------- ---------- ----------
consensus... #8716 ---------- ---------- ---------- ----------
#8716 ACAATGTATT GTCTGCCCAG CAATCAAAGG CAATGCTCAA
helen1 #8756 ---------- ---------- ---------- ----------
consensus... #8756 ---------- ---------- ---------- ----------
#8756 ACACCCAAGT GGCCCCCACC CTCAGCCCGC TCCTGCCCGC
helen1 #8796 ---------- ---------:- ---------- ----------
consensus... #8796 ---------- ---------- ---------- ----------
#8796 CCAGCACCCC CAGGCCCTGG GGGACCTGGG GTTCTCAGAC
helen1 #8835 ---------- ---------- ---------- ----------
consensus... #8835 ---------- ---------- ---------- ----------
#8835 TGCCAAAGAA GCCTTGCCAT CTGGCGCTCC CATGGCTCTT
helen1 #8875 ---------- ---------- ---------- ----------
consensus... #8875 ---------- ---------- ---------- ----------
#8875 GCAACATCTC CCCTTCGTTT TTGAGGGGGT CATGCCGGGG
helen1 #8915 ---------- ---------- ---------- ----------
consensus... #8915 ---------- ---------- ---------- ----------
#8915 GAGCCACCAG CCCCTCACTG GGTTCGGAGG AGAGTCAGGA
helen1 #8955 ---------- ---------- ---------- ----------
consensus... #8955 ---------- ---------- ---------- ----------
#8955 AGGGCCACGA CAAAGCAGAA ACATCGGATT TGGGGAACGC
helen1 #8995 -----:---- ---------- -:--------- ---------T
consensus... #8995 ---------- ---------- ---------- ----------


#9033 ::::TGTTCT GTTCCTTGTG TAACTGTGTT GCTGAAAGAC
helenl #9073 ---------- ---------- ---------- ----------
consensus... #9073 ---------- ---------- ---------- ----------
#9073 TACCTCGTTC TTGTCTTGAT GTGTCACCGG GGCAACTGCC
helenl #9113 ---------- ---------- ---------- ---------T
consensus... #9113 ---------- ---------- ---------- ----------
#9113 TGGGGGCGGG GATGGGGGCA GGGTGGAAGC GGCTCCCCA:
helenl #9153 ---------- ---------- ---------- ----------
consensus... #9153 ---------- ---------- ---------- ----------
#9153 TTTTATACCA AAGGTGCTAC ATCTATGTGA TGGGTGGGGT
helenl #9193 ---------- ---------- ---------- ----------
consensus... #9193 ---------- ---------- ---------- ----------
#9193 GGGGAGGGAA TCACTGGTGC TATAGAAATT GAGATGCCCC
helenl #9233 ---------- ---------- ---------- ----------
consensus... #9233 ---------- ---------- ---------- ----------
#9233 CCCAGGCCAG CAAATGTTCC TTTTTGTTCA AAGTCTATTT
helenl #9273 ---------- ---T------ ----:----- ----------
consensus... #9273 ---------- ---------- ---------- ----------
#9273 TTATTCCTTG ATA:TTTTTC TTTCTTTTTT TTTTTTTTTT
helenl #9312 ---------- ---------- ---------- ----------
consensus... #9312 ---------- ---------- ---------- ----------
#9312 GTGGATGGGG ACTTGTGAAT TTTTCTAAAG GTGCTATTTA
#9352 ACATGGGAGG AGAGCGTGTG CGGCTCCAGC CCAGCCCGCT
helenl #9391 ---------- ---------- ---------- ----------
consensus... #9391 ---------- ---------- ---------- ----------
#9391 GCTCACTTTC CACCCTCTCT CCACCTGCCT CTGGCTTCTC
helenl #9431 ---------- ---------- ---------- ----------
consensus... #9431 ---------- ---------- ---------- ----------
#9431 AGGCCTCTGC TCTCCGACCT CTCTCCTCTG AAACCCTCCT
helenl #9471 ---------- ---------- ---------- ----------
consensus... #9471 ---------- ---------- ---------- ----------
#9471 CCACAGCTGC AGCCCATCCT CCCGGCTCCC TCCTAGTCTG
helenl #9511 ---------- ---------- ---------- ----------
consensus... #9511 ---------- ---------- ---------- ----------
#9511 TCCTGCGTCC TCTGTCCCCG GGTTTCAGAG ACAACTTCCC
helenl #9551 ---------- ---------: :---------- ----------
consensus... #9551 ---------- ---------- ---------- ----------
#9551 AAAGCACAAA GCAGTTTTTC CCCCTAGGGG TGGGAGGAAG
helenl #9589 ---------- ---------- ---------- ----------
consensus... #9589 ---------- ---------- ---------- ----------
#9589 CAAAAGACTC TGTACCTATT TTGTATGTGT ATAATAATTT
helenl #9629 ---------- ---------- ---------- ----------
consensus... #9629 ---------- ---------- ---------- ----------
#9629 GAGATGTTTT TAATTATTTT GATTGCTGGA ATAAAGCATG
#9669 TGGAAATGAC CCAAACATAA ::TCCGCAGT GGCCTCCTAA
helenl #9709 ---------- ---------- ---------- ----------
consensus... #9709 ---------- ---------- ---------- ----------
#9709 TTTCCTTCTT TGGAGTTGGG GGAGGGGTAG ACATGGGGAA
helenl #9749 ---------- ---------- ---------- ----------
consensus... #9749 ---------- ---------- ---------- ----------
#9749 GGGGCCTTGG GGTGATGGGC TTGCCTTCCA TTCCTGCCCT
helenl #9789 ---------- ---------- ---------- ----------
consensus... #9789 ---------- ---------- ---------- ----------
#9789 TTCCCTCCCC ACTATTCTCT TCTAGATCCC TCCATAACCC
helenl #9829 ---------- ---------- ---------- ----------
consensus... #9829 ---------- ---------- ---------- ----------
#9829 CACTCCCCTT TCTCTCACCC TTCTTATACC GCAAACCTTT
helenl #9869 ---------- ---------- ---------- ----------
consensus... #9869 ---------- ---------- ---------- ----------
#9869 CTACTTCCTC TTTCATTTTC TATTCTTGCA ATTTCCTTGC
helenl #9909 ---------- ---------- ---------- ----------
consensus... #9909 ---------- ---------- ---------- ----------
#9909 ACCTTTTCCA AATCCTCTTC TCCCCTGCAA TACCATACAG
helenl #9949 ---------- ---------- ---------- ----------
consensus... #9949 ---------- ---------- ---------- ----------
#9949 GCAATCCACG TGCACAACAC ACACACACAC TCTTCACATC


helen1 #9989 ---------- ---------- ---------- ----------
consensus... #9989 ---------- ---------- ---------- ----------
#9989 TGGGGTTGTC CAAACCTCAT ACCCACTCCC CTTCAAGCCC
helen1 #10029 ---------- ---------- ---------- ----------
consensu... #10029 ---------- ---------- ---------- ----------
#10029 ATCCACTCTC CACCCCCTGG ATGCCCTGCA CTTGGTGGCG
helen1 #10069 ---------- ---------- ---------- ----------
consensu... #10069 ---------- ---------- ---------- ----------
#10069 GTGGGATGCT CATGGATACT GGGAGGGTGA GGGGAGTGGA
helen1 #10109 ---------- ---------- ---------- ----------
consensu... #10109 ---------- ---------- ---------- ----------
#10109 ACCCGTGAGG AGGACCTGGG GGCCTCTCCT TGAACTGACA


helen1 #10149 ---------- ---------- ---------- ----------
consensu... #10149 ---------- ---------- ---------- ----------
#10149 TGAAGGGTCA TCTGGCCTCT GCTCCCTTCT CACCCACGCT
helen1 #10189 ---------- ---------- ----------
consensu... #10189 ---------- ---------- ---------- ----------
#10189 GACCTCCTGC CGAAGGAGCA ACGCAACAGG AGCTTAAGTT
consensu... #10229 ---------- ---------- ---------- ----------
#10229 TAAACCGCTG ATCAGCCTCG ACTGTGCCTT CTAGTTGCCA
consensu... #10269 ---------- ---------- ---------- ----------
#10269 GCCATCTGTT GTTTGCCCCT CCCCCGTGCC TTCCTTGACC
consensu... #10309 ---------- ---------- ---------- ----------
#10309 CTGGAAGGTG CCACTCCCAC TGTCCTTTCC TAATAAAATG

Abdelgany A. W., M. Beeson, D (2003). Allele-specific silencing of a pathogenic mutant acetylcholine receptor subunit by RNA interference. *Hum. Mol. Genet.* **12:** 2637-2644.

Abe E. (2006). Function of BMPs and BMP Antagonists in Adult Bone. *Ann NY Acad Sci* **1068:** 41-53.

Allen M., Hock J., and Burr D. (2004). Periosteum: biology, regulation, and response to osteoporosis therapies. *Bone* **35:** 1003-1012.

Aluigi M., Fogli M., Curti A., Isidori A., Gruppioni E., Chiodoni C., Colombo M., Versura P., D'Errico-Grigioni A., Ferri E., Baccarani M., and Lemoli R. (2006). Nucleofection Is an Efficient Nonviral Transfection Technique for Human Bone Marrow-Derived Mesenchymal Stem Cells. *Stem Cells* **24:** 454-461.

Amar L. D., M. Faucon-Biguet, N. Mallet, J. Vogel, R. (2006). Control of small inhibitory RNA levels and RNA interference by doxycycline induced activation of a minimal RNA polymerase III promoter. *Nucl. Acids Res.* **34:** e37-.

Amarzguiou M., Rossi J., and Kim D. (2005). Approaches for chemically synthesized siRNA and vector-mediated RNAi. *FEBS Letters* **579:** 5974-5981.

Arziman Z., Horn T., and Boutros M. (2005). E-RNAi: a web application to design optimized RNAi constructs. *Nucl. Acids Res.* **33:** W582-588.

Aslan H., Zilberman Y., Arbeli V., Sheyn D., Matan Y., Liebergall M., Li J. Z., Helm G. A., Gazit D., and Gazit Z. (2006). Nucleofection-Based Ex Vivo Nonviral Gene Delivery to Human Stem Cells as a Platform for Tissue Regeneration. *Tissue Engineering* **12:** 877-889.

Aubin J. E. (2001). Regulation of Osteoblast Formation and Function. *Reviews in Endocrine & Metabolic Disorders* **2:** 81-94.

Aza-Blanc P., Cooper C. L., Wagner K., Batalov S., Deveraux Q. L., and Cooke M. P. (2003). Identification of Modulators of TRAIL-Induced Apoptosis via RNAi-Based Phenotypic Screening. *Molecular Cell* **12:** 627-637.

Bakowska J. D. M., M. Camp, SM. Wang, Y. Allen, PD. Breakefield, XO. (2003). Targeted transgene integration into transgenic mouse fibroblasts carrying the full-length human AAVS1 locus mediated by HSV/AAV rep(+) hybrid amplicon vector. *Gene therapy* **10:** 1691-702.

Baksh D. S., L. Tuan, RS. (2004). Adult mesenchymal stem cells: characterization, differentiation, and application in cell and gene therapy. *Journal of Cellular and Molecular Medicine* **8:** 301-16.

Barker D. W. H., Hartung S, Breindl M, Jaenisch R. (1991). Retrovirus-induced insertional mutagenesis: mechanism of collagen mutation in Mov13 mice. *Molecular and Cellular Biology* **11:** 5154-5163.

Barry F. P., and Murphy J. M. (2004). Mesenchymal stem cells: clinical applications and biological characterization. *The International Journal of Biochemistry & Cell Biology* **36:** 568-584.

Bartel D. P. (2004). MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **116:** 281-297.

Bartholomew A., Sturgeon C., Siatskas M., Ferrer K., McIntosh K., Patil S., Hardy W., Devine S., Ucker D., and Deans R. (2002). Mesenchymal stem cells suppress lymphocyte proliferation in vitro and prolong skin graft survival in vivo. *Experimental Hematology* **30:** 42-48.

Baum J., and Brodsky B. (1999). Folding of peptide models of collagen and misfolding in disease. *Current Opinion in Structural Biology* **9:** 122-128.

Bayne E. H., and Allshire R. C. (2005). RNA-directed transcriptional gene silencing in mammals. *Trends in Genetics* **21:** 370-373.

Beck K., Chan V. C., Shenoy N., Kirkpatrick A., Ramshaw J. A. M., and Brodsky B. (2000). Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. *PNAS* **97:** 4273-4278.

Bedalov A. S., R. Dodig, M. Kronenberg, MS. Kapural, B. Bogdanovic, Z. Kream, BE. Woody, CO. Clark, SH. Mack, K. (1995). Regulation of COL1A1 expression in type I collagen producing tissues: identification of a 49 base pair region which is required for transgene expression in bone of transgenic mice. *Journal of Bone and Mineral Research* **10:** 1443-51.

Behlke M. A. (2006). Progress towards in Vivo Use of siRNAs. *Molecular Therapy* **13:** 644-670.

Bennink J. R., and Palmore T. N. (2004). The promise of siRNAs for the treatment of influenza. *Trends in Molecular Medicine* **10:** 571-574.

Berns K., Hijmans E. M., Mullenders J., Brummelkamp T. R., Velds A., Heimerikx M., Kerkhoven R. M., Madiredjo M., Nijkamp W., Weigelt B., Agami R., Ge W., Cavet G., Linsley P. S., Beijersbergen R. L., and Bernards R. (2004). A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428:** 431-437.

Bertoni C., Jarrahian S., Wheeler T. M., Li Y., Olivares E. C., Calos M. P., and Rando T. A. (2006). Enhancement of plasmid-mediated gene therapy for muscular dystrophy by directed plasmid integration. *PNAS* **103:** 419-424.

Bhate M., Wang X., Baum J., and Brodsky B. (2002). Folding and Conformational Consequences of Glycine to Alanine Replacements at Different Positions in a Collagen Model Peptide. *Biochemistry* **41:** 6539-6547.

Birmingham A., Anderson E. M., Reynolds A., Ilsley-Tyree D., Leake D., Fedorov Y., Baskerville S., Maksimova E., Robinson K., Karpilow J., Marshall W. S., and Khvorova A. (2006). 3[prime] UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat Meth* **3:** 199-204.

Blair H. Z., M. Schlesinger, PH. (2002). Mechanisms balancing skeletal matrix synthesis and degradation. *Biochemical Journal* **342:** 329-41.

Bohnsack M. T., Czaplinski K., and Gorlich D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10:** 185-191.

Bonadio J., Saunders T. L., Tsai E., Goldstein S. A., Morris-Wiman J., Brinkley L., Dolan D. F., Altschuler R. A., Hawkins J. E., Jr., Bateman J. F., Mascara T., and Jaenisch R. (1990). Transgenic Mouse Model of the Mild Dominant Form of Osteogenesis Imperfecta. *PNAS* **87:** 7145-7149.

Breault D., Lichtler A., and Rowe D. (1997). COL1A1 Transgene Expression in Stably Transfected Osteoblastic Cells. Relative contributions from the first intron, 3' flanking seqeunce and sequences delivered from the body of the human Col1a1 minigene.

. *J. Biol. Chem.* **272:** 31241-31250.

Bridge A. J., Pebernard S., Ducraux A., Nicoulaz A.-L., and Iggo R. (2003). Induction of an interferon response by RNAi vectors in mammalian cells. *Nat Genet* **34:** 263-264.

Bruder S., Kurth A., Shea M., Hayes W., Jaiswal N., and Kadiyala S. (1998). Bone regeneration by implantation of purified, culture-expanded human mesenchymal stem cells. *Journal of Orthopaedic Research* **16:** 155-162.

Brummelkamp T., Bernards R., and Agami R. (2002). Stable suppression of tumorigenicity by virus-mediated RNA interference. *Cancer Cell* **2:** 243-247.

Brummelkamp T. R., Bernards R., and Agami R. (2002). A System for Stable Expression of Short Interfering RNAs in Mammalian Cells. *Science* **296:** 550-553.

Byers P. H. (2002). Killing the messenger: new insights into nonsense-mediated mRNA decay. *J. Clin. Invest.* **109:** 3-6.

Cabral W., and Marini J. (2004). High Proportion of Mutant Osteoblasts Is Compatible with Normal Skeletal Function in Mosaic Carriers of Osteogenesis Imperfecta. *American Journal of Human Genetics* **47:** 752-760.

Calin, and Croce (2006). Genomics of Chronic Lymphocytic Leukemia MicroRNAs as New Players With Clinical Significance. *Seminars in Oncology* **33:** 167-173.

Calvi L. M., Adams G. B., Weibrecht K. W., Weber J. M., Olson D. P., Knight M. C., Martin R. P., Schipani E., Divieti P., Bringhurst F. R., Milner L. A., Kronenberg H. M., and Scadden D. T. (2003). Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* **425:** 841-846.

Caplen N. J., Fleenor J., Fire A., and Morgan R. A. (2000). dsRNA-mediated gene silencing in cultured Drosophila cells: a tissue culture model for the analysis of RNA interference. *Gene* **252:** 95-105.

Carmell M. A., Zhang L., Conklin D. S., Hannon G. J., and Rosenquist T. A. (2003). Germline transmission of RNAi in mice. *Nat Struct Mol Biol* **10:** 91-92.

Castanotto D., Tommasi S., Li M., Li H., Yanow S., Pfeifer G. P., and Rossi J. J. (2005). Short hairpin RNA-directed cytosine (CpG) methylation of the RASSF1A gene promoter in HeLa cells. *Molecular Therapy* **12:** 179-183.

Chalberg T. W., Genise H. L., Vollrath D., and Calos M. P. (2005). {phi}C31 Integrase Confers Genomic Integration and Long-Term Transgene Expression in Rat Retina. *Invest. Ophthalmol. Vis. Sci.* **46:** 2140-2146.

Chalberg T. W., Portlock J. L., Olivares E. C., Thyagarajan B., Kirby P. J., Hillman R. T., Hoelters J., and Calos M. P. (2006). Integration Specificity of Phage [phi]C31 Integrase in the Human Genome. *Journal of Molecular Biology* **357:** 28-48.

Chamberlain J. R., Schwarze U., Wang P.-R., Hirata R. K., Hankenson K. D., Pace J. M., Underwood R. A., Song K. M., Sussman M., Byers P. H., and Russell D. W. (2004). Gene Targeting in Stem Cells from Individuals with Osteogenesis Imperfecta. *Science* **303:** 1198-1201.

Chan J. A., Krichevsky A. M., and Kosik K. S. (2005). MicroRNA-21 Is an Antiapoptotic Factor in Human Glioblastoma Cells. *Cancer Res* **65:** 6029-6033.

Chang S., Chuang H., Chen Y., Chen J., Chung H., Lu Y., Lin H., Tai C., and Lou J. (2003). Ex vivo gene therapy in autologous bone marrow stromal stem cells for tissue-engineered maxillofacial bone regeneration. *Gene Therapy* **10:** 2013-9.

Chapel A. B., JM. Bensidhoum, M. Fouillard, L. Young, RG. Frick, J. Demarquay, C. Cuvelier, F. Mathieu, E. Trompier, F. Dudoignon, N. Germain, C. Mazurier, C. Aigueperse, J. Borneman, J. Gorin, NC. Gourmelon, P. Thierry, D. (2003). Mesenchymal stem cells home to injured tissues when co-infused with hematopoietic cells to treat a radiation-induced multi-organ failure syndrome. *The Journal of Gene Medicine* **5:** 1028-1038.

Chen L., Le H., Qin R., Kumar M., Du Z., Xia R., and Deng J. (2005). Reversal of the phenotype by K-rasval12 silencing mediated by adenovirus-delivered siRNA in human pancreatic cancer cell line Panc-1. *World of Gastroenterology* **11** 831-838.

Chen W., Liu M., Jiao Y., Yan W., Wei X., Chen J., Fei L., Liu Y., Zuo X., Yang F., Lu Y., and Zheng Z. (2006). Adenovirus-Mediated RNA Interference against Foot-and-Mouth Disease Virus Infection both In Vitro and In Vivo. *J. Virol.* **80:** 3559-3566.

Cheng S., Lou J., Wright N., Lai C., Avioli L., and Riew K. (2001). In vitro and in vivo induction of bone formation using a recombinant adenoviral vector carrying the human BMP-2 gene. *Calcified Tissue International* **68:** 87-94.

Chipman S. D., Sweet H. O., McBride D. J., Jr., Davisson M. T., Marks S. C., Jr., Shuldiner A. R., Wenstrup R. J., Rowe D. W., and Shapiro J. R. (1993). Defective Pro{alpha}2(I) Collagen Synthesis in a Recessive Mutation in Mice: A Model of Human Osteogenesis Imperfecta. *PNAS* **90:** 1701-1705.

Cho H. H., Kim Y. J., Kim S. J., Kim J. H., Bae Y. C., Ba B., and Jung J. S. (2006). Endogenous Wnt Signaling Promotes Proliferation and Suppresses Osteogenic Differentiation in Human Adipose Derived Stromal Cells. *Tissue Engineering* **12:** 111-121.

Cicchillitti L. J., SA. Sala, A. Saitta, B. (2004). B-Myb acts as a repressor of human COL1A1 collagen gene expression by interacting with Sp1 and CBF factors in scleroderma fibroblasts. *Biochen J* **378:** 609-16.

Cogoni C., and Macino G. (1999). Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase. *Nature* **399:** 166-169.

Cohen-Solal L., Zolezzi F., Pignatti P. F., and Mottes M. (1996). Intrafamilial variable expressivity of osteogenesis imperfecta due to mosaicism for a lethal G382R substitution in theCOL1A1gene. *Molecular and Cellular Probes* **10:** 219-225.

Colige A., Sokolov B., Nugent P., Baserga R., and Prockop D. (1993). Use of an antisense oligonucleotide to inhibit expression of a mutated human procollagen gene (COL1A1) in transfected mouse 3T3 cells. *Biochemistry* **32** 7-11.

Constantinou C., Pack M., Young S., and Prockop D. (1990). Phenotypic heterogeneity in osteogenesis imperfecta: the mildly affected mother of a proband with a lethal variant has the same mutation

substituting cysteine for alpha 1-glycine 904 in a type I procollagen gene (COL1A1). *American Journal of Human Genetics* **47:** 679 -679.

Coronel M. F., Musolino P. L., and Villar M. J. (2006). Selective migration and engraftment of bone marrow mesenchymal stem cells in rat lumbar dorsal root ganglia after sciatic nerve constriction. *Neuroscience Letters* **405:** 5-9.

Coumoul X., Shukla V., Li C., Wang R.-H., and Deng C.-X. (2005). Conditional knockdown of Fgfr2 in mice using Cre-LoxP induced RNA interference. *Nucl. Acids Res.* **33:** e102-.

Cowan C. M., Aalami O. O., Shi Y.-Y., Chou Y.-F., Mari C., Thomas R., Quarto N., Nacamuli R. P., Contag C. H., Wu B., and Longaker M. T. (2005). Bone Morphogenetic Protein 2 and Retinoic Acid Accelerate in Vivo Bone Formation, Osteoclast Recruitment, and Bone Turnover. *Tissue Engineering* **11:** 645-658.

Cullen L. M., and Arndt G. M. (2005). Genome-wide screening for gene function using RNAi in mammalian cells. *Immunology and Cell Biology* **83:** 217-223.

Czauderna F., Santel A., Hinz M., Fechtner M., Durieux B., Fisch G., Leenders F., Arnold W., Giese K., Klippel A., and Kaufmann J. (2003). Inducible shRNA expression for application in a prostate cancer mouse model. *Nucl. Acids Res.* **31:** e127-.

Dallas A V. A. (2006). RNAi: a novel antisense technology and its therapeutic potential. *Medical Science Monitor* **12:** RA67-74.

Davidson B. L., and Paulson H. L. (2004). Molecular medicine for the brain: silencing of disease genes with RNA interference. *The Lancet Neurology* **3:** 145-149.

Dawson P. A., and Marini J. C. (2000). Hammerhead ribozymes selectively suppress mutant type I collagen mRNA in osteogenesis imperfecta fibroblasts. *Nucl. Acids Res.* **28:** 4013-4020.

Dayoub H., Dumont R. J., Li J. Z., Dumont A. S., Hankins G. R., Kallmes D. F., and Helm G. A. (2003). Human Mesenchymal Stem Cells Transduced with Recombinant Bone Morphogenetic Protein-9 Adenovirus Promote Osteogenesis in Rodents. *Tissue Engineering* **9:** 347-356.

Dazzi F., Ramasamy R., Glennie S., Jones S. P., and Roberts I. (2006). The role of mesenchymal stem cells in haemopoiesis. *Blood Reviews* **20:** 161-171.

De Bari C., Dell'Accio F., Vandenabeele F., Vermeesch J. R., Raymackers J.-M., and Luyten F. P. (2003). Skeletal muscle repair by adult human mesenchymal stem cells from synovial membrane. *J. Cell Biol.* **160:** 909-918.

De Palma M., Montini E., de Sio F. R. S., Benedicenti F., Gentile A., Medico E., and Naldini L. (2005). Promoter trapping reveals significant differences in integration site selection between MLV and HIV vectors in primary hematopoietic cells. *Blood* **105:** 2307-2315.

Devine M. M., CM. Jang, E. Anderson, PC. Balian, G. (2002). Transplanted bone marrow cells localize to fracture callus in a mouse model. *Journal of Orthopaedic Research* **20:** 1232-1239.

Devine S. M., Bartholomew A. M., Mahmud N., Nelson M., Patil S., Hardy W., Sturgeon C., Hewett T., Chung T., and Stock W. (2001). Mesenchymal stem cells are capable of homing to the bone marrow of non-human primates following systemic infusion. *Experimental Hematology* **29:** 244-255.

Devine S. M., Cobbs C., Jennings M., Bartholomew A., and Hoffman R. (2003). Mesenchymal stem cells distribute to a wide range of tissues following systemic infusion into nonhuman primates. *Blood* **101:** 2999-3001.

Dezawa M., Hoshino M., and Ide C. (2005). Treatment of neurodegenerative diseases using adult bone marrow stromal cell-derived neurons. *Expert Opinion on Biological Therapy* **5:** 427-435.

Di Lullo G. A., Sweeney S. M., Korkko J., Ala-Kokko L., and San Antonio J. D. (2002). Mapping the Ligand-binding Sites and Disease-associated Mutations on the Most Abundant Protein in the Human, Type I Collagen. *J. Biol. Chem.* **277:** 4223-4231.

Dicker A., Le Blanc K., Astrom G., van Harmelen V., Gotherstrom C., Blomqvist L., Arner P., and Ryden M. (2005). Functional studies of mesenchymal stem cells derived from adult human adipose tissue. *Experimental Cell Research* **308:** 283-290.

Dimitriou R., Tsiridis E., and Giannoudis P. V. (2005). Current concepts of molecular aspects of bone healing.

Injury **36:** 1392-1404.

Distler J. H. W., Jungel A., Kurowska-Stolarska M., Michel B. A., Gay R. E., Gay S., and Distler O. (2005). Nucleofection: a new, highly efficient transfection method for primary human keratinocytes. *Experimental Dermatology* **14:** 315-320.

Doench J. G., Petersen C. P., and Sharp P. A. (2003). siRNAs can function as miRNAs. *Genes Dev.* **17:** 438-442.

Downey P. S. M. (2006). Bone biology and the clinical implications for osteoporosis. *Physical Therapy* **86:** 77-91.

Dragoo J. C., JY. Lieberman, JR. Huang, J. Zuk, PA. Zhang, J. Hedrick, MH. Benhaim, P. (2003). Bone induction by BMP-2 transduced stem cells derived from human fat. *Journal of Orthopaedic Research* **21:** 622-629.

Dumont R. D., H. Li, JZ. Dumont, AS. Kallmes, DF. Hankins, GR. Helm, GA. (2002). Ex vivo bone morphogenetic protein-9 gene therapy using human mesenchymal stem cells induces spinal fusion in rodents. *Neurosurgery* **51:** 1239-44;.

Eames B. D. L. F., L. Helms, JA (2003). Molecular ontogeny of the skeleton. *Birth Defects Research Part C: Embryo Today: Reviews* **69:** 93-101.

Echeverri C. J., and Perrimon N. (2006). High-throughput RNAi screening in cultured cells: a user's guide. *Nat Rev Genet* **7:** 373-384.

Edwards M., Wenstrup R., Byers P., and Cohn D. (1992). Recurrence of lethal osteogenesis imperfecta due to parental mosaicism for a mutation in the COL1A2 gene of type I collagen. The mosaic parent exhibits phenotypic features of a mild form of the disease. *Human Mutation* **1:** 45-47.

Edwards M. W., RJ. Byers, PH. Cohn, DH. (1992). Recurrence of lethal osteogenesis imperfecta due to parental mosaicism for a mutation in the COL1A2 gene of type I collagen. The mosiac parent exhibits phenotypic features of a mild form of the disease. *Human Mutation* **1:** 47-54.

Eis P. S., Tam W., Sun L., Chadburn A., Li Z., Gomez M. F., Lund E., and Dahlberg J. E. (2005). Accumulation of miR-155 and BIC RNA in human B cell lymphomas. *PNAS* **102:** 3627-3632.

Elbashir S. M., J. Patkaniowska, A. Lendeckel, W. Tuschl, T. (2001). Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *EMBO* **20:** 6877- 88.

Elbashir S. M., J. Patkaniowska, A. Lendeckel, W. Tuschl, T. (2002). Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *EMBO Journal* **20.:** :6877-88.

Elbashir S. M., Harborth J., Lendeckel W., Yalcin A., Weber K., and Tuschl T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411:** 494-498.

Elbashir S. M., Lendeckel W., and Tuschl T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15:** 188-200.

Eliopoulos N., Stagg J., Lejeune L., Pommey S., and Galipeau J. (2005). Allogeneic marrow stromal cells are immune rejected by MHC class I- and class II-mismatched recipient mice. *Blood* **106:** 4057-4065.

Ellis J., and Yao S. (2005). Retrovirus Silencing and Vector Design: Relevance to Normal and Cancer Stem Cells? *Current Gene Therapy* **5:** 367-373.

EN M.

Farrar G. K., PF. Humphries, P. (2002). On the genetics of retinitis pigmentosa and on mutation-independent approaches to therapeutic intervention. *EMBO Journal* **21:** 857-64.

Fedorov Y., Anderson E. M., Birmingham A., Reynolds A., Karpilow J., Robinson K., Leake D., Marshall W. S., and Khvorova A. (2006). Off-target effects by siRNA can induce toxic phenotype. *RNA* **12:** 1188-1196.

Ferguson C., Alpern E., Miclau T., and Helms J. A. (1999). Does adult fracture repair recapitulate embryonic skeletal formation? *Mechanisms of Development* **87:** 57-66.

Fitzgerald J. L., SR. Bateman, JF. (1999). Proteasomal degradation of unassembled mutant type I collagen pro-alpha1(I) chains. *Journal of Biological Chemistry* **274:** 27392-8.

Flockhart I., Booker M., Kiger A., Boutros M., Armknecht S., Ramadan N., Richardson K., Xu A., Perrimon N., and Mathey-Prevot B. (2006). FlyRNAi: the Drosophila RNAi screening center database. *Nucl. Acids Res.* **34:** D489-494.

Forlino A., Porter F. D., Lee E. J., Westphal H., and Marini J. C. (1999). Use of the Cre/lox Recombination System to Develop a Non-lethal Knock-in Murine Model for Osteogenesis Imperfecta with an alpha 1(I) G349C Substitution. VARIABILITY IN PHENOTYPE IN BrtlIV MICE. *J. Biol. Chem.* **274:** 37923-37931.

Forte A. N., MA. Cipollaro, M. Giordano, A. Cascino, A. Galderisi, U. (2006). An effective method for adenoviral-mediated delivery of small interfering RNA into mesenchymal stem cells. *Journal of Cellular Biochemistry* **9999:** n/a.

Frank P. B. (2003). Biology and clinical applications of mesenchymal stem cells. *Birth Defects Research Part C: Embryo Today: Reviews* **69:** 250-256.

Frank Rauch F. H. G. (2005). Osteogenesis imperfecta, current and future medical treatment. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* **139C:** 31-37.

Friedenstein A. P.-S. I., Petrakova, KV. (1966). Osteogenesis in transplants of bone marrow cells. *Journal of Embryology and Experimental Morphology* **16:** 381-90.

Gafni Y. T., G. Liebergal, M. Pelled, G. Gazit, Z. Gazit, D. (2004). Stem cells as vehicles for orthopedic gene therapy. *Gene Therapy* **11:** 417-26.

Gajko-Galicka A. (2002). Mutations in type I collagen genes resulting in osteogenesis imperfecta in humans. *Acta Biochim Pol.* **49:** 433-41.

Gartner A., Collin L., Lalli G., and William E. Balch C. J. D. a. A. H. (2006). Nucleofection of Primary Neurons. *In* "Methods in Enzymology", pp. 374-388, Academic Press.

Gazit D. T., G. Kelley, P. Wang, E. Jalenak, M. Zilberman, Y. Moutsatsos, I. (1999). Engineered pluripotent mesenchymal cells integrate and differentiate in regenerating bone: a novel cell-mediated gene therapy. *Journal of Gene Medicine* **1:** 121-133.

Glorieux F. H., Rauch F., Plotkin H., Ward L., Travers R., Roughley P., Lalic L., Glorieux D. F., Fassier F., and Bishop N. J. (2000). Type V Osteogenesis Imperfecta: A New Form of Brittle Bone Disease. *Journal of Bone and Mineral Research* **15:** 1650-1658.

Glorieux F. H., Ward L. M., Rauch F., Lalic L., Roughley P. J., and Travers R. (2002). Osteogenesis Imperfecta Type VI: A Form of Brittle Bone Disease with a Mineralization Defect. *Journal of Bone and Mineral Research* **17:** 30-38.

Glover D. J., Lipps H. J., and Jans D. A. (2005). Towards safe non viral gene expression in humans *Nature Reviews Genetics* **6:** 299-310.

Goncalves M. A. F. V., de Vries A. A. F., Holkers M., van de Watering M. J. M., van der Velde I., van Nierop G. P., Valerio D., and Knaan-Shanzer S. (2006). Human mesenchymal stem cells ectopically expressing full-length dystrophin can complement Duchenne muscular dystrophy myotubes by cell fusion. *Hum. Mol. Genet.* **15:** 213-221.

Goncz K. P., NL. Chow, BL. Davis, BR. Gruenert, DC. (2002). Application of SFHR to gene therapy of monogenic disorders. *Gene Therapy* **9:** 691-4.

Gotkin M. G., Ripley C. R., Lamande S. R., Bateman J. F., and Bienkowski R. S. (2004). Intracellular trafficking and degradation of unassociated pro[alpha]2 chains of collagen type I. *Experimental Cell Research* **296:** 307-316.

Gou D., Narasaraju T., Chintagari N. R., Jin N., Wang P., and Liu L. (2004). Gene silencing in alveolar type II cells using cell-specific promoter in vitro and in vivo. *Nucl. Acids Res.* **32:** e134-.

Grassi G., Forlino A., and Marini J. C. (1997). Cleavage of collagen RNA transcripts by hammerhead ribozymes in vitro is mutation-specific and shows competitive binding effects. *Nucl. Acids Res.* **25:** 3451-3458.

Gregory C. A., Prockop D. J., and Spees J. L. (2005). Non-hematopoietic bone marrow stem cells: Molecular control of expansion and differentiation. *Experimental Cell Research* **306:** 330-335.

Gregory R. I., Chendrimada T. P., Cooch N., and Shiekhattar R. (2005). Human RISC Couples MicroRNA Biogenesis and Posttranscriptional Gene Silencing. *Cell* **123:** 631-640.

Gregory R. I., Yan K.-p., Amuthan G., Chendrimada T., Doratotaj B., Cooch N., and Shiekhattar R. (2004). The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432:** 235-240.

Griesenbach U., Geddes D. M., and Alton E. W. F. W. (2006). Gene therapy progress and prospects: cystic fibrosis. *Gene Ther* **13:** 1061-1067.

Grimm D., and Kay M. A. (2006). Therapeutic short hairpin RNA expression in the liver: viral targets and vectors. *Gene Ther* **13:** 563-575.

Groth A. C., and Calos M. P. (2004). Phage Integrases: Biology and Applications. *Journal of Molecular Biology* **335:** 667-678.

Groth A. C., Fish M., Nusse R., and Calos M. P. (2004). Construction of Transgenic Drosophila by Using the Site-Specific Integrase From Phage {phi}C31. *Genetics* **166:** 1775-1782.

Groth A. C., Olivares E. C., Thyagarajan B., and Calos M. P. (2000). A phage integrase directs efficient site-specific integration in human cells. *PNAS* **97:** 5995-6000.

Gu W., Putral L., Hengst K., Minto K., Saunders N. A., Leggatt G., and McMillan N. A. J. (2006). Inhibition of cervical cancer cell growth in vitro and in vivo with lentiviral-vector delivered short hairpin RNA targeting human papillomavirus E6 and E7 oncogenes. *Cancer Gene Ther.*

Gysin R. W. J., Sheng MH, Kasukawa Y, Miyakoshi N, Chen ST, Peng H, Lau KH, Mohan S, Baylink DJ. (2004). Ex vivo gene therapy with stromal cells transduced with a retroviral vector containing the BMP4 gene completely heals critical size calvarial defect in rats. *Gene Therapy* **9:** 991-9.

Hacein-Bey-Abina S., Von Kalle C., Schmidt M., McCormack M. P., Wulffraat N., Leboulch P., Lim A., Osborne C. S., Pawliuk R., Morillon E., Sorensen R., Forster A., Fraser P., Cohen J. I., de Saint Basile G., Alexander I., Wintergerst U., Frebourg T., Aurias A., Stoppa-Lyonnet D., Romana S., Radford-Weiss I., Gross F., Valensi F., Delabesse E., Macintyre E., Sigaux F., Soulier J., Leiva L. E., Wissler M., Prinz C., Rabbitts T. H., Le Deist F., Fischer A., and Cavazzana-Calvo M. (2003). LMO2-Associated Clonal T Cell Proliferation in Two Patients after Gene Therapy for SCID-X1. *Science* **302:** 415-419.

Haleem-Smith H. D., A. Okafor, C. Tuli, R. Olsen, D. Hall, DJ. Tuan, RS. (2005). Optimization of high-efficiency transfection of adult human mesenchymal stem cells in vitro. *Molecular Biotechnology* **30:** 9 - 20.

Haley B., and Zamore P. D. (2004). Kinetic analysis of the RNAi enzyme complex. *Nat Struct Mol Biol* **11:** 599-606.

Hall T. M. T. (2005). Structure and Function of Argonaute Proteins. *Structure* **13:** 1403-1408.

Hamazaki H., Ujino S., Miyano-Kurosaki N., Shimotohno K., and Takaku H. (2006). Inhibition of hepatitis C virus RNA replication by short hairpin RNA synthesized by T7 RNA polymerase in hepatitis C virus subgenomic replicons. *Biochemical and Biophysical Research Communications* **343:** 988-994.

Hamilton A. V., O. Chappell, L. Baulcombe, D. (1999). Two classes of short interfering RNA in RNA silencing. *EMBO* **21:** 4671-9.

Hamm A., Krott N., Breibach I., Blindt R., and Bosserhoff A. K. (2002). Efficient Transfection Method for Primary Cells. *Tissue Engineering* **8:** 235-245.

Hammond S. M., Bernstein E., Beach D., and Hannon G. J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404:** 293-296.

Han J., Lee Y., Yeom K.-H., Nam J.-W., Heo I., Rhee J.-K., Sohn S. Y., Cho Y., Zhang B.-T., and Kim V. N. (2006). Molecular Basis for the Recognition of Primary microRNAs by the Drosha-DGCR8 Complex. *Cell* **125:** 887-901.

Harada S.-i., and Rodan G. A. (2003). Control of osteoblast function and regulation of bone mass. *Nature* **423:** 349-355.

Hartung S., Jaenisch R., and Breindl M. (1986). Retrovirus insertion inactivates mouse [alpha] 1(I) collagen gene by blocking initiation of transcription. *Nature* **320:** 365-367.

Hayashi S., and McMahon A. P. (2002). Efficient Recombination in Diverse Tissues by a Tamoxifen-Inducible Form of Cre: A Tool for Temporally Regulated Gene Activation/Inactivation in the Mouse. *Developmental Biology* **244:** 305-318.

Heister T., Heid I., Ackermann M., and Fraefel C. (2002). Herpes Simplex Virus Type 1/Adeno-Associated Virus Hybrid Vectors Mediate Site-Specific Integration at the Adeno-Associated Virus Preintegration Site,

AAVS1, on Human Chromosome 19. *J. Virol.* **76:** 7163-7173.

Held P. K., Olivares E. C., Aguilar C. P., Finegold M., Calos M. P., and Grompe M. (2005). In Vivo Correction of Murine Hereditary Tyrosinemia Type I by [phi]C31 Integrase-Mediated Gene Delivery. *Molecular Therapy* **11:** 399-408.

Hibbitt O. C., and Wade-Martins R. (2006). Delivery of Large Genomic DNA Inserts 100 kb Using HSV-1 Amplicons. *Current Gene Therapy* **6:** 325-336.

Hickman-Davis J. M., and Davis I. C. (2006). Transgenic mice. *Paediatric Respiratory Reviews* **7:** 49-53.

Hoelters J. C., M. Drechsel, M. Geissler, C. Gulkan, H. Bocker, W. Schieker, M. Jochum, M. Neth, P. (2005). Nonviral genetic modification mediates effective transgene expression and functional RNA interference in human mesenchymal stem cells. *The Journal of Gene Medicine* **7:** 718-728.

Hollis R. S., SM. Sclimenti, CR. Lin, J. Chen-Tsai, Y. Calos, MP. (2003). integrases for the construction and manipulation of transgenic mammals. *Reproductive Biology and Endocrinology* **1**.

Horwitz E. M., Gordon P. L., Koo W. K. K., Marx J. C., Neel M. D., McNall R. Y., Muul L., and Hofmann T. (2002). Isolated allogeneic bone marrow-derived mesenchymal cells engraft and stimulate growth in children with osteogenesis imperfecta: Implications for cell therapy of bone. *PNAS* **99:** 8932-8937.

Horwitz E. M., Prockop D. J., Fitzpatrick L. A., Koo W. W. K., Gordon P. L., Neel M., Sussman M., Orchard P., Marx J. C., Pyeritz R. E., and Brenner M. K. (1999). Transplantability and therapeutic effects of bone marrow-derived mesenchymal cells in children with osteogenesis imperfecta. *Nat Med* **5:** 309-313.

Horwitz E. M., Prockop D. J., Gordon P. L., Koo W. W. K., Fitzpatrick L. A., Neel M. D., McCarville M. E., Orchard P. J., Pyeritz R. E., and Brenner M. K. (2001). Clinical responses to bone marrow transplantation in children with severe osteogenesis imperfecta. *Blood* **97:** 1227-1231.

Hosono T., Mizuguchi H., Katayama K., Xu Z.-L., Sakurai F., Ishii-Watabe A., Kawabata K., Yamaguchi T., Nakagawa S., Mayumi T., and Hayakawa T. (2004). Adenovirus Vector-Mediated Doxycycline-Inducible RNA Interference. *Human Gene Therapy* **15:** 813-819.

Hou Z., Nguyen Q., Frenkel B., Nilsson S. K., Milne M., van Wijnen A. J., Stein J. L., Quesenberry P., Lian J. B., and Stein G. S. (1999). Osteoblast-specific gene expression after transplantation of marrow cells: Implications for skeletal gene therapy. *PNAS* **96:** 7294-7299.

Inoue S. P., FC.; Koehl, GE.; Piso, P; Schlitt, HJ.; Geissler, EK.; Dahlke, MH (2006). Immunomodulatory Effects of Mesenchymal Stem Cells in a Rat Organ Transplant Model. *Transplantation* **81:** 1589-1595, .

Ishikawa Y. T., N. Murakami, K. Uchiyama, T. Kumaki, S. Tsuchiya, S. Kugoh, H. Oshimura, M. Calos, MP. Sugamura, K. (2006). Phage &phis;C31 integrase-mediated genomic integration of the common cytokine receptor gamma chain in human T-cell lines. *The Journal of Gene Medicine* **8:** 646-653.

Ito M., Kawano K., Miyagishi M., and Taira K. (2005). Genome-wide application of RNAi to the discovery of potential drug targets. *FEBS Letters* **579:** 5988-5995.

Jackson A. L., Bartz S. R., Schelter J., Kobayashi S. V., Burchard J., Mao M., Li B., Cavet G., and Linsley P. S. (2003). Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotech* **21:** 635-637.

Jackson A. L., Burchard J., Leake D., Reynolds A., Schelter J., Guo J., Johnson J. M., Lim L., Karpilow J., Nichols K., Marshall W., Khvorova A., and Linsley P. S. (2006). Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. *RNA* **12:** 1197-1205.

Jackson A. L., Burchard J., Schelter J., Chau B. N., Cleary M., Lim L., and Linsley P. S. (2006). Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* **12:** 1179-1187.

Jaenisch R., Harbers K., Schnieke A., Lohler J., Chumakov I., Jahner D., Grotkopp D., and Hoffmann E. (1983). Germline integration of Moloney murine leukemia virus at the Mov13 locus leads to recessive lethal mutation and early embryonic death. *Cell* **32:** 209-216.

Javazon E. H., Beggs K. J., and Flake A. W. (2004). Mesenchymal stem cells: paradoxes of passaging. *Experimental Hematology* **32:** 414-425.

Jiang W., Ma A., Wang T., Han K., Liu Y., Zhang Y., Zhao X., Dong A., Du Y., Huang X., Wang J., Lei X., and Zheng X. (2006). Intravenous transplantation of mesenchymal stem cells improves cardiac performance

after acute myocardial ischemia in female rats. *Transplant International* **19:** 570-580.

Jones D., Hosalkar H., and Jones S. (2002). The orthopaedic management of osteogenesis imperfecta. *Current Orthopaedics* **16:** 374-388.

Jorgensen C., Gordeladze J., and Noel D. (2004). Tissue engineering through autologous mesenchymal stem cells. *Current Opinion in Biotechnology* **15:** 406-410.

Kahler R. G., M. Lian, J. Stein, GS. van Wijnen, AJ. Westendorf, JJ. (2006). Lymphocyte enhancer-binding factor 1 (Lef1) inhibits terminal differentiation of osteoblasts. *Journal of Cellular Biochemistry* **97:** 969-983.

Karsenty G., and Wagner E. F. (2002). Reaching a Genetic and Molecular Understanding of Skeletal Development. *Developmental Cell* **2:** 389-406.

Kartsogiannis V., and Ng K. W. (2004). Cell lines and primary cell cultures in the study of bone cell biology. *Molecular and Cellular Endocrinology* **228:** 79-102.

Kassis I., Zangi L., Rivkin R., Levdansky L., Samuel S., Marx G., and Gorodetsky R. (2006). Isolation of mesenchymal stem cells from G-CSF-mobilized human peripheral blood using fibrin microbeads. *Bone Marrow Transplant* **37:** 967-976.

Kennerdell J. R., and Carthew R. W. (2000). Heritable gene silencing in Drosophila using double-stranded RNA. *Nat Biotech* **18:** 896-898.

Keravala A. G., AC. Jarrahian, S. Thyagarajan, B. Hoyt, JJ. Kirby, PJ. Calos MP. (2006). A diversity of serine phage integrases mediate site-specific recombination in mammalian cells. *Molecular Genetics and Genomics* **276:** 135-46.

Keravala A. P., JL. Nash, JA. Vitrant, DG. Robbins, PD. Calos, MP. (2006). PhiC31 integrase mediates integration in cultured synovial cells and enhances gene expression in rabbit joints. *The Journal of Gene Medicine* **8:** 1008-1017.

Khillan J. S., Li S., and Prockop D. J. (1994). Partial Rescue of a Lethal Phenotype of Fragile Bones in Transgenic Mice with a Chimeric Antisense Gene Directed Against a Mutated Collagen Gene. *PNAS* **91:** 6298-6302.

Khillan J. S., Olsen A. S., Kontusaari S., Sokolov B., and Prockop D. J. (1991). Transgenic mice that express a mini-gene version of the human gene for type I procollagen (COL1A1) develop a phenotype resembling a lethal form of osteogenesis imperfecta. *J. Biol. Chem.* **266:** 23373-23379.

Khvorova A., Reynolds A., and Jayasena S. D. (2003). Functional siRNAs and miRNAs Exhibit Strand Bias. *Cell* **115:** 209-216.

Kiang A.-S., Palfi A., Ader M., Kenna P. F., Millington-Ward S., Clark G., Kennan A., O'Reilly M., Tam L. C. T., and Aherne A. (2005). Toward a Gene Therapy for Dominant Disease: Validation of an RNA Interference-Based Mutation-Independent Approach. *Molecular Therapy* **12:** 555-561.

Kim D. R., JJ. (2003). Coupling of RNAi-mediated target downregulation with gene replacement. *Antisense Nucleic Acid Drug Dev* **13:** 151-5.

Kim J. K., Gabel H. W., Kamath R. S., Tewari M., Pasquinelli A., Rual J.-F., Kennedy S., Dybbs M., Bertin N., Kaplan J. M., Vidal M., and Ruvkun G. (2005). Functional Genomic Analysis of RNA Interference in C. elegans. *Science* **308:** 1164-1167.

Kim V. N., and Nam J.-W. (2006). Genomics of microRNA. *Trends in Genetics* **22:** 165-173.

Knothe Tate M. L., Adamson J. R., Tami A. E., and Bauer T. W. (2004). The osteocyte. *The International Journal of Biochemistry & Cell Biology* **36:** 1-8.

Ko K., Furukawa K., Takahashi T., Urano T., Sanai Y., Nagino M., Nimura Y., and Furukawa K. (2006). Fundamental study of small interfering RNAs for ganglioside GD3 synthase gene as a therapeutic target of lung cancers. *Oncogene.*

Koc O. D., J. Nieder, M. Gerson, SL. Lazarus, HM. Krivit, W. (2003). Allogeneic mesenchymal stem cell infusion for treatment of metachromatic leukodystrophy (MLD) and Hurler syndrome (MPS-IH). *Bone Marrow Transplantation.* **30:** 215-22.

Kofron M. D., and Laurencin C. T. (2006). Bone tissue engineering by gene delivery. *Advanced Drug Delivery*

Reviews **58:** 555-576.

Kolfschoten I. G. M., van Leeuwen B., Berns K., Mullenders J., Beijersbergen R. L., Bernards R., Voorhoeve P. M., and Agami R. (2005). A Genetic Screen Identifies PITX1 as a Suppressor of RAS Activity and Tumorigenicity. *Cell* **121:** 849-858.

Krampera M., Glennie S., Dyson J., Scott D., Laylor R., Simpson E., and Dazzi F. (2003). Bone marrow mesenchymal stem cells inhibit the response of naive and memory antigen-specific T cells to their cognate peptide. *Blood* **101:** 3722-3729.

Kratochwil K., von der Mark K., Kollar E. J., Jaenisch R., Mooslehner K., Schwarz M., Haase K., Gmachl I., and Harbers K. (1989). Retrovirus-induced insertional mutation in Mov13 mice affects collagen I expression in a tissue-specific manner. *Cell* **57:** 807-816.

Kratochwil K. G.-T. N., Holzinger I, Harbers K. (1993). Restricted expression of Mov13 mutant alpha 1(I) collagen gene in osteoblasts and its consequences for bone development. *Develpomental Dynamics* **198:** 273-83.

Kraus K. H., and Kirker-Head C. (2006). Mesenchymal Stem Cells and Bone Regeneration. *Veterinary Surgery* **35:** 232-242.

Kumar S., Mahendra G., Nagy T. R., and Ponnazhagan S. (2004). Osteogenic Differentiation of Recombinant Adeno-Associated Virus 2-Transduced Murine Mesenchymal Stem Cells and Development of an Immunocompetent Mouse Model for Ex Vivo Osteoporosis Gene Therapy. *Human Gene Therapy* **15:** 1197-1206.

Kurozumi K., Nakamura K., Tamiya T., Kawano Y., Ishii K., Kobune M., Hirai S., Uchida H., Sasaki K., and Ito Y. (2005). Mesenchymal stem cells that produce neurotrophic factors reduce ischemic damage in the rat middle cerebral artery occlusion model. *Molecular Therapy* **11:** 96-104.

Kuttenkeuler D., and Boutros M. (2004). Genome-wide RNAi as a route to gene function in Drosophila. *Brief Funct Genomic Proteomic* **3:** 168-176.

Laatsch A. R., S. Grewal, T. Beisiegel, U. Joerg, H. (2004). Differential RNA interference: replacement of endogenous with recombinant low density lipoprotein receptor-related protein (LRP). **83:** 113-120.

Landthaler M., Yalcin A., and Tuschl T. (2004). The Human DiGeorge Syndrome Critical Region Gene 8 and Its D. melanogaster Homolog Are Required for miRNA Biogenesis. *Current Biology* **14:** 2162-2167.

Le Blanc K., and Ringden O. (2005). Immunobiology of Human Mesenchymal Stem Cells and Future Use in Hematopoietic Stem Cell Transplantation. *Biology of Blood and Marrow Transplantation* **11:** 321-334.

Le Blanc K., and Ringden O. (2006). Mesenchymal stem cells: properties and role in clinical bone marrow transplantation. *Current Opinion in Immunology* **18:** 1-6.

Le Blanc K. G., C. Ringden, O. Hassan, M. McMahon, R. Horwitz, E. Anneren, G. Axelsson, O. Nunn J, Ewald, U. Norden-Lindeberg, S. Jansson, M. Dalton, A. Astrom, E. Westgren, M. (2005). Fetal mesenchymal stem-cell engraftment in bone after in utero transplantation in a patient with severe osteogenesis imperfecta. *Transplantation* **79:** 1607-14.

Lee K., Majumdar M. K., Buyaner D., Hendricks J. K., Pittenger M. F., and Mosca J. D. (2001). Human Mesenchymal Stem Cells Maintain Transgene Expression during Expansion and Differentiation. *Molecular Therapy* **3:** 857-866.

Lee N. S., Dohjima T., Bauer G., Li H., Li M.-J., Ehsani A., Salvaterra P., and Rossi J. (2002). Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. *Nat Biotech* **20:** 500-505.

Lee S. T., Lee S., Peters D. P., Hoffman G. G., Stacey A., and Greenspan D. S. (1992). Deletion of the pro-alpha 1(I) N-propeptide affects secretion of type I collagen in Chinese hamster lung cells but not in Mov-13 mouse cells. *J. Biol. Chem.* **267:** 24126-24133.

Leung R. K. M., and Whittaker P. A. (2005). RNA interference: From gene silencing to gene-specific therapeutics. *Pharmacology & Therapeutics* **107:** 222-239.

Leuschner P. A., SL. Kueng, S. Martinez, J. (2006). Cleavage of the siRNA passenger strand during RISC assembly in human cells. *EMBO Reports* **7:** 314-20.

Lewis B. P., Burge C. B., and Bartel D. P. (2005). Conserved Seed Pairing, Often Flanked by Adenosines,

Indicates that Thousands of Human Genes are MicroRNA Targets. *Cell* **120:** 15-20.

Liat Ries-Levavi, Ish-Shalom T., Frydman M., Shirley D. L., Barkai C. G., Goldman B., Byers P., and Friedman E. (2004). Genetic and biochemical analyses of Israeli osteogenesis imperfecta patients. *Human Mutation* **23:** 399-400.

Liechty K. W., MacKenzie T. C., Shaaban A. F., Radu A., Moseley A. B., Deans R., Marshak D. R., and Flake A. W. (2000). Human mesenchymal stem cells engraft and demonstrate site-specific differentiation after in utero transplantation in sheep. *Nat Med* **6:** 1282-1286.

Lin X., Yang J., Chen J., Gunasekera A., Fesik S. W., and Shen Y. (2004). Development of a tightly regulated U6 promoter for shRNA expression. *FEBS Letters* **577:** 376-380.

Lindsay R. (2002). Modeling the benefits of pamidronate in children with osteogenesis imperfecta. *J. Clin. Invest.* **110:** 1239-1241.

Liu J., Carmell M. A., Rivas F. V., Marsden C. G., Thomson J. M., Song J.-J., Hammond S. M., Joshua-Tor L., and Hannon G. J. (2004). Argonaute2 Is the Catalytic Engine of Mammalian RNAi. *Science* **305:** 1437-1441.

Liu J., Jeppesen I., Nielsen K., and Jensen T. G. (2006). [phi]c31 integrase induces chromosomal aberrations in primary human fibroblasts. *Gene Ther* **13:** 1188-1190.

Liu X.-d., Ma S.-m., Liu Y., Liu S.-z., and Sehon A. (2004). Short hairpin RNA and retroviral vector-mediated silencing of p53 in mammalian cells. *Biochemical and Biophysical Research Communications* **324:** 1173-1178.

Lohmann J. U., Endl I., and Bosch T. C. G. (1999). Silencing of Developmental Genes in Hydra. *Developmental Biology* **214:** 211-214.

Long X., Olszewski M., Huang W., and Kletzel M. (2005). Neural Cell Differentiation In Vitro from Adult Human Bone Marrow Mesenchymal Stem Cells. *Stem Cells and Development* **14:** 65-69.

Lund E., Guttinger S., Calado A., Dahlberg J. E., and Kutay U. (2004). Nuclear Export of MicroRNA Precursors. *Science* **303:** 95-98.

Ma Q.-w., Sheng H.-q., Yan J.-b., Cheng S., Huang Y., Chen-Tsai Y., Ren Z.-r., Huang S.-z., and Zeng Y.-t. (2006). Identification of pseudo attP sites for phage [phi]C31 integrase in bovine genome. *Biochemical and Biophysical Research Communications* **345:** 984-988.

Machida Y., Okada T., Kurosawa M., Oyama F., Ozawa K., and Nukina N. (2006). rAAV-mediated shRNA ameliorated neuropathology in Huntington disease model mouse. *Biochemical and Biophysical Research Communications* **343:** 190-197.

Mackenzie T. C., and Flake A. W. (2001). Human Mesenchymal Stem Cells Persist, Demonstrate Site-Specific Multipotential Differentiation, and Are Present in Sites of Wound Healing and Tissue Regeneration after Transplantation into Fetal Sheep. *Blood Cells, Molecules, and Diseases* **27:** 601-604.

Makkar R. P., MJ. Lill, M. Frantzen, M. Takizawa, K. Kleisli, T. Zheng, J. Kar, S. McClelan, R. Miyamota, T. Bick-Forrester, J. Fishbein, MC. Shah, PK. Forrester, JS. Sharifi, B. Chen, PS. Qayyum M. (2005). Intramyocardial injection of allogenic bone marrow-derived mesenchymal stem cells without immunosuppression preserves cardiac function in a porcine model of myocardial infarction. *Journal of Cardiovascular Pharmacological Therapy* **10:** 225 - 33.

Makoto Miyagishi H. S. H. M. Y. K. K. T. (2004). Optimization of an siRNA-expression system with an improved hairpin and its significant suppressive effects in mammalian cells. *The Journal of Gene Medicine* **6:** 715-723.

Manoharan M. (2004). RNA interference and chemically modified small interfering RNAs. *Current Opinion in Chemical Biology* **8:** 570-579.

Marieb E. (2003). "Essentials of Human Anatomy and physiology.," Benjamin Cummings, San Francisco.

Marini J. C. (2003). Do Bisphosphonates Make Children's Bones Better or Brittle? *N Engl J Med* **349:** 423-426.

Martinez J., Patkaniowska A., Urlaub H., Luhrmann R., and Tuschl T. (2002). Single-Stranded Antisense siRNAs Guide Target RNA Cleavage in RNAi. *Cell* **110:** 563-574.

Martinez J., and Tuschl T. (2004). RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev.* **18:** 975-980.

Martino S., Marconi P., Tancini B., Dolcetta D., De Angelis M. G. C., Montanucci P., Bregola G., Sandhoff K., Bordignon C., Emiliani C., Manservigi R., and Orlacchio A. (2005). A direct gene transfer strategy via brain internal capsule reverses the biochemical defect in Tay-Sachs disease. *Hum. Mol. Genet.* **14:** 2113-2123.

Matranga C., Tomari Y., Shin C., Bartel D. P., and Zamore P. D. (2005). Passenger-Strand Cleavage Facilitates Assembly of siRNA into Ago2-Containing RNAi Enzyme Complexes. *Cell* **123:** 607-620.

Matsukura S., Jones P. A., and Takai D. (2003). Establishment of conditional vectors for hairpin siRNA knockdowns. *Nucl. Acids Res.* **31:** e77-.

Matthess Y., Kappel S., Spankuch B., Zimmer B., Kaufmann M., and Strebhardt K. (2005). Conditional inhibition of cancer cell proliferation by tetracycline-responsive, H1 promoter-driven silencing of PLK1. **24:** 2973-2980.

Matzk M. B., JA. (2005). RNAi-mediated pathways in the nucleus. *Nature Reviews Genetics* **6:** 24-35.

Meinel L., Fajardo R., Hofmann S., Langer R., Chen J., Snyder B., Vunjak-Novakovic G., and Kaplan D. (2005). Silk implants for the healing of critical size bone defects. *Bone* **37:** 688-698.

Meirelles L. d. S., Chagastelles P. C., and Nardi N. B. (2006). Mesenchymal stem cells reside in virtually all post-natal organs and tissues. *J Cell Sci* **119:** 2204-2213.

Michael M. Z., O' Connor S. M., van Holst Pellekaan N. G., Young G. P., and James R. J. (2003). Reduced Accumulation of Specific MicroRNAs in Colorectal Neoplasia. *Mol Cancer Res* **1:** 882-891.

Miller V., Paulson H., and Gonzalez-Alegre P. (2005). RNA Interference in Neuroscience: Progress and Challenges. *Cellular and Molecular Neurobiology* **25:** 1195-1207.

Millington-Ward S., McMahon, HP, Allen D, Tuohy G, Kiang AS, Palfi A, Kenna PF, Humphries P, Farrar GJ. (2004). RNAi of COL1A1 in mesenchymal progenitor cells. *European Journal of Human Genetics* **12:** 865 -866

Millington-Ward S., Allers C., Tuohy G., Conget P., Allen D., McMahon H. P., Kenna P. F., Humphries P., and Farrar G. J. (2002). Validation in mesenchymal progenitor cells of a mutation-independent ex vivo approach to gene therapy for osteogenesis imperfecta. *Hum. Mol. Genet.* **11:** 2201-2206.

Millington-Ward S., McMahon H. P., and Farrar G. J. (2005). Emerging therapeutic approaches for osteogenesis imperfecta. *Trends in Molecular Medicine* **11:** 299-305.

Millington-Ward S., O'Neill B., Kiang A., Humphries P., Kenna P., and GJ. F. (1999). A mutation-independent therapeutic strategem for osteogenesis imperfecta. *Antisense Nucleic Acid Drug Dev* **9:** 532-537.

Millington-Ward S., O'Neill B., Tuohy G., Al-Jandal N., Kiang A. S., Kenna P. F., Palfi A., Hayden P., Mansergh F., Kennan A., Humphries P., and Farrar G. J. (1997). Strategems in vitro for gene therapies directed to dominant mutations. *Hum. Mol. Genet.* **6:** 1415-1426.

Millington-Ward. S O. N., B. Kiang AS, Humphries P, Kenna PF, Farrar GJ. (1999). A mutation-independent therapeutic strategem for osteogenesis imperfecta. *Antisense Nucleic Acid Drug Dev.* **9:** 537 -42.

Mittal V. (2004). Improving the effficiency of RNA inteference in mammals. *Nature Reviews Genetics* **5:** 355-365.

Miyagishi M. T., K. (2002). U6 promoter-driven siRNAs with four uridine 3' overhangs efficiently suppress targeted gene expression in mammalian cells. *Nat Biotech* **20:** 497-500.

Miyagishi M. T., K. (2004). RNAi expression vectors in mammalian cells. *Methods Molecular Biology* **252:** 483-91.

Moffat J., Grueneberg D. A., Yang X., Kim S. Y., Kloepfer A. M., Hinkle G., Piqani B., Eisenhaure T. M., Luo B., and Grenier J. K. (2006). A Lentiviral RNAi Library for Human and Mouse Genes Applied to an Arrayed Viral High-Content Screen. *Cell* **124:** 1283-1298.

Moore K. A., and Lemischka I. R. (2006). Stem Cells and Their Niches. *Science* **311:** 1880-1885.

Morigi M., Imberti B., Zoja C., Corna D., Tomasoni S., Abbate M., Rottoli D., Angioletti S., Benigni A., Perico N., Alison M., and Remuzzi G. (2004). Mesenchymal Stem Cells Are Renotropic, Helping to Repair the Kidney and Improve Function in Acute Renal Failure. *J Am Soc Nephrol* **15:** 1794-1804.

Morris K. V., Chan S. W. L., Jacobsen S. E., and Looney D. J. (2004). Small Interfering RNA-Induced

Transcriptional Gene Silencing in Human Cells. *Science* **305**: 1289-1292.

Morris K. V., and Rossi J. J. (2006). Lentiviral-mediated delivery of siRNAs for antiviral therapy. *Gene Ther* **13**: 553-558.

Morris K. V., and Rossi J. J. (2006). Lentivirus-Mediated RNA Interference Therapy for Human Immunodeficiency Virus Type 1 Infection. *Human Gene Therapy* **17**: 479-486.

Moser T., Harbers K., and Kratochwil K. (1996). Tissue- and stage-specific activation of an endogenous provirus after transcription through its integration site in the opposite orientation. *Mol. Cell. Biol.* **16**: 384-389.

Myslinski E. A., J. Krol, A. Carbon, P. (2001). An unusually compact external promoter for RNA polymerase III transcription of the human H1RNA gene. *Nucl. Acids Res.* **29**: 2502-2509.

Nagaya N., Fujii T., Iwase T., Ohgushi H., Itoh T., Uematsu M., Yamagishi M., Mori H., Kangawa K., and Kitamura S. (2004). Intravenous administration of mesenchymal stem cells improves cardiac function in rats with acute myocardial infarction through angiogenesis and myogenesis. *Am J Physiol Heart Circ Physiol* **287**: H2670-2676.

Naito Y., Yamada T., Matsumiya T., Ui-Tei K., Saigo K., and Morishita S. (2005). dsCheck: highly sensitive off-target search software for double-stranded RNA-mediated RNA interference. *Nucl. Acids Res.* **33**: W589-591.

Nakayama G K. Y., Koga K, Kusakabe T (2006). Site-specific gene integration in cultured silkworm cells mediated by phiC31 integrase. *Molecular Genetics and Metabolism* **275**: 1-8.

Neth P., Ciccarella M., Egea V., Hoelters J., Jochum M., and Ries C. (2006). Wnt Signaling Regulates the Invasion Capacity of Human Mesenchymal Stem Cells. *Stem Cells* **24**: 1892-1903.

Ngo H., Tschudi C., Gull K., and Ullu E. (1998). Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. *PNAS* **95**: 14687-14692.

Nimmo D. D., Alphey L., Meredith J. M., and Eggleston P. (2006). High efficiency site-specific genetic engineering of the mosquito genome. *Insect Molecular Biology* **15**: 129-136.

Niyibizi C., Patrick Smith Zhibao Mi Charlotte L. Phillips, Robbins, P (2001). Transfer of pro?2(I) cDNA into cells of a murine model of human Osteogenesis Imperfecta restores synthesis of type I collagen comprised of ?1(I) and ?2(I) heterotrimers in vitro and in vivo. *Journal of Cellular Biochemistry* **83**: 84-91.

Niyibizi C., Wang S., Mi Z., and Robbins P. (2004). Gene therapy approaches for osteogenesis imperfecta. *Gene therapy* **11**: 408-416.

Olivares E. C., Hollis R. P., Chalberg T. W., Meuse L., Kay M. A., and Calos M. P. (2002). Site-specific genomic integration produces therapeutic Factor IX levels in mice. *Nat Biotech* **20**: 1124-1128.

Olsen A. S., Geddis A. E., and Prockop D. J. (1991). High levels of expression of a minigene version of the human pro alpha 1 (I) collagen gene in stably transfected mouse fibroblasts. Effects of deleting putative regulatory sequences in the first intron. *J. Biol. Chem.* **266**: 1117-1121.

Olsen B., Reginato A., and Wang W. (2000). Bone Development *Annual Review of Cell and Developmental Biology* **16**: 191-220.

O'Neill B., Millington-Ward S., O'Reilly M., Tuohy G., Kiang A.-S., Kenna P. F., Humphries P., and Farrar G. J. (2000). Ribozyme-Based Therapeutic Approaches for Autosomal Dominant Retinitis Pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **41**: 2863-2869.

Ong S. P., Choo B. G. H., Chu J. J. H., and Ng M. L. (2006). Expression of vector-based small interfering RNA against West Nile virus effectively inhibits virus replication. *Antiviral Research* **In Press, Corrected Proof**.

Ong S. T., Li F., Du J., Tan Y. W., and Wang S. (2005). Hybrid Cytomegalovirus Enhancer H1 Promoter-Based Plasmid and Baculovirus Vectors Mediate Effective RNA Interference. *Human Gene Therapy* **16**: 1404-1412.

Orlic D. (2003). Adult Bone Marrow Stem Cells Regenerate Myocardium in Ischemic Heart Disease. *Ann NY Acad Sci* **996**: 152-157.

Ortiz-Urda S., Thyagarajan B., Keene D. R., Lin Q., Calos M. P., and Khavari P. A. (2003). C31 Integrase-

Mediated Nonviral Genetic Correction of Junctional Epidermolysis Bullosa. *Human Gene Therapy* **14:** 923-928.

Ortiz-Urda S., Thyagarajan B., Keene D. R., Lin Q., Fang M., Calos M. P., and Khavari P. A. (2002). Stable nonviral genetic correction of inherited human skin disease. *Nat Med* **8:** 1166-1170.

Pace J. M., Atkinson M., Willing M., Wallis G., and Byers P. H. (2001). Deletions and duplications of Gly-Xaa-Yaa triplet repeats in the triple helical domains of type I collagen chains disrupt helix formation and result in several types of osteogenesis imperfecta. *Human Mutation* **18:** 319-326.

Pace J. M., Chitayat D., Atkinson M., Wilcox W. R., Schwarze U., and Byers P. H. (2002). A single amino acid substitution (D1441Y) in the carboxyl-terminal propeptide of the pro{alpha}1(I) chain of type I collagen results in a lethal variant of osteogenesis imperfecta with features of dense bone diseases. *J Med Genet* **39:** 23-29.

Paddison P. C., AA. Bernstein, E. Hannon, G.J. Conklin, D.S. (2002). Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16:** 948-958.

Palfi A. A., M. Kiang, AS. Millington-Ward, S. Clark, G. O'Reilly, M. McMahon, HP. Kenna, PF. Humphries, P. Farrar, GJ. (2006). RNAi-based suppression and replacement of <I>rds</I>-peripherin in retinal organotypic culture. *Human Mutation* **27:** 260-268.

Panagiotis M. (2005). Classification of non-union. *Injury* **36:** S30-S37.

Park C. W., Chen Z., Kren B. T., and Steer C. J. (2004). Double-stranded siRNA targeted to the huntingtin gene does not induce DNA methylation. *Biochemical and Biophysical Research Communications* **323:** 275-280.

Paul C. P., Good P. D., Winer I., and Engelke D. R. (2002). Effective expression of small interfering RNA in human cells. *Nat Biotech* **20:** 505-508.

Paule M. W., RJ. (2000). Transcription by RNA polymerases I and III. *Nucl. Acids Res.* **28:** 1283-1298.

Pavlin D., Lichtler A. C., Bedalov A., Kream B. E., Harrison J. R., Thomas H. F., Gronowicz G. A., Clark S. H., Woody C. O., and Rowe D. W. (1992). Differential utilization of regulatory domains within the alpha 1(I) collagen promoter in osseous and fibroblastic cells. *J. Cell Biol.* **116:** 227-236.

Peace B. E., Florer J. B., Witte D., Smicun Y., Toudjarska I., Wu G., Kilpatrick M. W., Tsipouras P., and Wenstrup R. J. (2005). Endogenously Expressed Multimeric Self-Cleaving Hammerhead Ribozymes Ablate Mutant Collagen in Cellulo. *Molecular Therapy* **12:** 128-136.

Pedro Gonzalez-Alegre V. M. M. B. L. D. H. L. P. (2003). Toward therapy for DYT1 dystonia: Allele-specific silencing of mutant TorsinA. *Annals of Neurology* **53:** 781-787.

Peister A. M., JA. Wang, M. Tucker, HA. Prockop, DJ. (2004). Stable transfection of MSCs by electroporation. *Gene Therapy* **11::** 224-8.

Peng J., Zhao Y., Mai J., Pang W. K., Wei X., Zhang P., and Xu Y. (2005). Inhibition of hepatitis B virus replication by various RNAi constructs and their pharmacodynamic properties. *J Gen Virol* **86:** 3227-3234.

Pereira R. F., O'Hara M. D., Laptev A. V., Halford K. W., Pollard M. D., Class R., Simon D., Livezey K., and Prockop D. J. (1998). Marrow stromal cells as a source of progenitor cells for nonhematopoietic tissues in transgenic mice with a phenotype of osteogenesis imperfecta. *PNAS* **95:** 1142-1147.

Pereira R. F. H., K. W. O'Hara, M. D., Leeper D. B., Sokolov B. P., Pollard M. D., Bagasra O., and Prockop D. J. (1995). Cultured Adherent Cells from Marrow can Serve as Long-Lasting Precursor Cells for Bone, Cartilage, and Lung in Irradiated Mice. *PNAS* **92:** 4857-4861.

Pereira R. H., EL. Halford, KW. Prockop, DJ. (1995). Bone fragility in transgenic mice expressing a mutated gene for type I procollagen (COL1A1) parallels the age-dependent phenotype of human osteogenesis imperfecta. *Journal of Bone and Mineral Research* **10:** 1837-43.

Pereira R. K., JS. Helminen, HJ. Hume, EL. Prockop, DJ. (1993). Transgenic mice expressing a partially deleted gene for type I procollagen (COL1A1). A breeding line with a phenotype of spontaneous fractures and decreased bone collagen and mineral. *Journal of Clincal Investigation* **91:** :709-16.

Persengiev S. P., Zhu X., and Green M. R. (2004). Nonspecific, concentration-dependent stimulation and

repression of mammalian gene expression by small interfering RNAs (siRNAs). *RNA* **10:** 12-18.

Persikov A. V., Ramshaw J. A. M., Kirkpatrick A., and Brodsky B. (2002). Peptide investigations of pairwise interactions in the collagen triple-helix. *Journal of Molecular Biology* **316:** 385-394.

Piersanti S., Sacchetti B., Funari A., Cesare S. D., Bonci D., Cherubini G., Peschle C., Riminucci M., Bianco P., and Saggio I. (2006). Lentiviral Transduction of Human Postnatal Skeletal (Stromal, Mesenchymal) Stem Cells: In Vivo Transplantation and Gene Silencing. *Calcified Tissue International* **V78:** 372-384.

Pittenger M. F., Mackay A. M., Beck S. C., Jaiswal R. K., Douglas R., Mosca J. D., Moorman M. A., Simonetti D. W., Craig S., and Marshak D. R. (1999). Multilineage Potential of Adult Human Mesenchymal Stem Cells. *Science* **284:** 143-147.

Pittenger M. F., and Martin B. J. (2004). Mesenchymal Stem Cells and Their Potential as Cardiac Therapeutics. *Circ Res* **95:** 9-20.

Pochampally R. R., Horwitz E. M., DiGirolamo C. M., Stokes D. S., and Prockop D. J. (2005). Correction of a mineralization defect by overexpression of a wild-type cDNA for COL1A1 in marrow stromal cells (MSCs) from a patient with osteogenesis imperfecta: a strategy for rescuing mutations that produce dominant-negative protein defects. *Gene Ther* **12:** 1119-1125.

Primorac D., Rowe D., Mottes M., Barisic I., Anticevic D., Mirandola S., Gomez L., Kalajzic I., Kusec V., and FH. G. (2001). Osteogenesis imperfecta at the beginning of bone and joint decade. *Croat Med J* **42:** 393-415.

Prockop D. J. (1998). What holds us together? Why do some of us fall apart? What can we do about it? *Matrix Biology* **16:** 519-528.

Prockop D. J. (2003). Further proof of the plasticity of adult stem cells and their role in tissue repair. *J. Cell Biol.* **160:** 807-809.

Pulkkinen L., and Uitto J. (1999). Mutation analysis and molecular genetics of epidermolysis bullosa. *Matrix Biology* **18:** 29-42.

Puthenveedu M. A., and Linstedt A. D. (2004). Gene replacement reveals that p115/SNARE interactions are essential for Golgi biogenesis. *PNAS* **101:** 1253-1256.

Quenneville S. P., Chapdelaine P., Rousseau J., Beaulieu J., Caron N. J., Skuk D., Mills P., Olivares E. C., Calos M. P., and Tremblay J. P. (2004). Nucleofection of muscle-derived stem cells and myoblasts with [phi]C31 integrase: stable expression of a full-length-dystrophin fusion gene by human myoblasts. *Molecular Therapy* **10:** 679-687.

Rahkonen O., Su M., Hakovirta H., Koskivirta I., Hormuzdi S. G., Vuorio E., Bornstein P., and Penttinen R. (2004). Mice With a Deletion in the First Intron of the Col1a1 Gene Develop Age-Dependent Aortic Dissection and Rupture. *Circ Res* **94:** 83-90.

Ralph G. S., Mazarakis N. D., and Azzouz M. (2005). Therapeutic gene silencing in neurological disorders, using interfering RNA. *Journal of Molecular Medicine* **83:** 413-419.

Raoul C., Barker S. D., and Aebischer P. (2005). Viral-based modelling and correction of neurodegenerative diseases by RNA interference. *Gene Ther* **13:** 487-495.

Raper S. E., Chirmule N., Lee F. S., Wivel N. A., Bagg A., Gao G.-p., Wilson J. M., and Batshaw M. L. (2003). Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. *Molecular Genetics and Metabolism* **80:** 148-158.

Rauch F., and Glorieux F. H. (2004). Osteogenesis imperfecta. *The Lancet* **363:** 1377-1385.

Reiser J., Zhang X.-Y., Hemenway C. S., Mondal D., Pradhan L., and La Russa V. F. (2005). Potential of mesenchymal stem cells in gene therapy approaches for inherited and acquired diseases. *Expert Opinion on Biological Therapy* **5:** 1571-1584.

Reynolds A., Anderson E. M., Vermeulen A., Fedorov Y., Robinson K., Leake D., Karpilow J., Marshall W. S., and Khvorova A. (2006). Induction of the interferon response by siRNA is cell type- and duplex length-dependent. *RNA* **12:** 988-993.

Reynolds A., Leake D., Boese Q., Scaringe S., Marshall W. S., and Khvorova A. (2004). Rational siRNA design

for RNA interference. *Nat Biotech* **22:** 326-330.

Rodriguez A., Griffiths-Jones S., Ashurst J. L., and Bradley A. (2004). Identification of Mammalian microRNA Host Genes and Transcription Units. *Genome Res.* **14:** 1902-1910.

Rojas M., Xu J., Woods C. R., Mora A. L., Spears W., Roman J., and Brigham K. L. (2005). Bone Marrow-Derived Mesenchymal Stem Cells in Repair of the Injured Lung. *Am. J. Respir. Cell Mol. Biol.* **33:** 145-152.

Romano N. M., G. (1992). Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. *Molecular Microbiology* **6:** 3343-53.

Ross F. P., and Christiano A. M. (2006). Nothing but skin and bone. *J. Clin. Invest.* **116:** 1140-1149.

Rossert J., Terraz C., and Dupont S. (2000). Regulation of type I collagen genes expression. *Nephrol. Dial. Transplant.* **15:** 66-68.

Roughley P. R., F. Glorieux, FH. (2003). Osteogenesis imperfecta--clinical and molecular diversity. *European Cell & Materials* **30:** 41-7;.

Rubinson D. A., Dillon C. P., Kwiatkowski A. V., Sievers C., Yang L., Kopinja J., Zhang M., McManus M. T., Gertler F. B., Scott M. L., and Van Parijs L. (2003). A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nat Genet* **33:** 401-406.

Ryan J., Barry F., Murphy J., and Mahon B. (2005). Mesenchymal stem cells avoid allogeneic rejection. *J Inflammlation* **2.**

Saban J., Zussman M. A., Havey R., Patwardhan A. G., Schneider G. B., and King D. (1996). Heterozygous oim mice exhibit a mild form of osteogenesis imperfecta. *Bone* **19:** 575-579.

Satake K. L., J. Lenke, LG. (2003). Migration of Mesenchymal Stem Cells Through Cerebrospinal Fluid into Injured Spinal Cord Tissue. *Spine* **29:** 1971-9.

Saxena S., Jonsson Z. O., and Dutta A. (2003). Small RNAs with Imperfect Match to Endogenous mRNA Repress Translation: IMPLICATIONS FOR OFF-TARGET ACTIVITY OF SMALL INHIBITORY RNA IN MAMMALIAN CELLS. *J. Biol. Chem.* **278:** 44312-44319.

Schnieke A., Dziadek M., Bateman J., Mascara T., Harbers K., Gelinas R., and Jaenisch R. (1987). Introduction of the Human proalpha 1(I) Collagen Gene into proalpha 1(I)-deficient Mov-13 Mouse Cells Leads to Formation of Functional Mouse-Human Hybrid Type I Collagen. *PNAS* **84:** 764-768.

Schnieke A., Harbers K., and Jaenisch R. (1983). Embryonic lethal mutation in mice induced by retrovirus insertion into the [alpha]1(I) collagen gene. *Nature* **304:** 315-320.

Schwarz D. D., H. Kennington, L. Moore JT. Schnelter, J. Zamore PD. (2006). Designing siRNA that distinguish between genes that differ by a single nucloetide. *PLoS Genetics* **2:** 140 - 146.

Schwarz D. S., Hutvagner G., Du T., Xu Z., Aronin N., and Zamore P. D. (2003). Asymmetry in the Assembly of the RNAi Enzyme Complex. *Cell* **115:** 199-208.

Sekiya I., Larson B. L., Smith J. R., Pochampally R., Cui J.-G., and Prockop D. J. (2002). Expansion of Human Adult Stem Cells from Bone Marrow Stroma: Conditions that Maximize the Yields of Early Progenitors and Evaluate Their Quality. *Stem Cells* **20:** 530-541.

Sethe S., Scutt A., and Stolzing A. (2006). Aging of mesenchymal stem cells. *Ageing Research Reviews* **5:** 91-116.

Sevignani C., Calin G., Siracusa L., and Croce C. (2006). Mammalian microRNAs: a small world for fine-tuning gene expression. *Mammalian Genome* **17:** 189-202.

Shen C., Buck A. K., Liu X., Winkler M., and Reske S. N. (2003). Gene silencing by adenovirus-delivered siRNA. *FEBS Letters* **539:** 111-114.

Shirley D. M., D. Jordan, G. McQuaid, S. Li, G. (2005). Systemic recruitment of osteoblastic cells in fracture healing. *Journal of Orthopaedic Research* **23:** 1013-1021.

Siemen H., Nix M., Endl E., Koch P., Itskovitz-Eldor J., and Brustle O. (2005). Nucleofection of Human Embryonic Stem Cells. *Stem Cells and Development* **14:** 378-383.

Sillence D., Senn A, and DM. D. (1979). Genetic heterogeneity in osteogenesis imperfecta. *Journal of Med*

Genetics. **16:** 101-16.

Silva J. M., Li M. Z., Chang K., Ge W., Golding M. C., Rickles R. J., Siolas D., Hu G., Paddison P. J., Schlabach M. R., Sheth N., Bradshaw J., Burchard J., Kulkarni A., Cavet G., Sachidanandam R., McCombie W. R., Cleary M. A., Elledge S. J., and Hannon G. J. (2005). Second-generation shRNA libraries covering the mouse and human genomes. *Nat Genet* **37:** 1281-1288.

Siolas D., Lerner C., Burchard J., Ge W., Linsley P. S., Paddison P. J., Hannon G. J., and Cleary M. A. (2005). Synthetic shRNAs as potent RNAi triggers. *Nat Biotech* **23:** 227-231.

Smicun Y. K., MW. Florer, J Toudjarska, I. Wu, G Wenstrup, RJ. Tsipouras, P (2003). Enhanced intracellular availability and survival of hammerhead ribozymes increases target ablation in a cellular model of osteogenesis imperfecta. *Gene Therapy* **10:** 2005 - 2012.

Sokolov B. M., PK. Khillan, JS. Prockop, DJ. (1993). Tissue- and development-specific expression in transgenic mice of a type I procollagen (COL1A1) minigene construct with 2.3 kb of the promoter region and 2 kb of the 3'-flanking region. Specificity is independent of the putative regulatory sequences in the first intron. *Biochemistry* **7:** 9242-9.

Sokolov B. P., Ala-Kokko L., Dhulipala R., Arita M., Khillan J. S., and Prockop D. J. (1995). Tissue-specific Expression of the Gene for Type I Procollagen (COL1A1) in Transgenic Mice. *J. Biol. Chem.* **270:** 9622-9629.

Sommerfeldt, and Rubin (2001). Biology of bone and how it orchestrates the form and function of the skeleton. *European Spine Journal* **10:** S86-S95.

Song J., Pang S., Lu Y., Yokoyama K. K., Zheng J.-Y., and Chiu R. (2004). Gene Silencing in Androgen-Responsive Prostate Cancer Cells from the Tissue-Specific Prostate-Specific Antigen Promoter. *Cancer Res* **64:** 7661-7663.

Song L., Chau L., Sakamoto Y., Nakashima J., Koide M., and Tuan R. S. (2004). Electric Field-Induced Molecular Vibration for Noninvasive, High-Efficiency DNA Transfection. *Molecular Therapy* **9:** 607-616.

Sontheimer E. J. (2005). Assemly and functioning of RNAi complexes. *Nature Reviews Molecular Cell Biology* **6:** 127-138.

Stacey A. M., R. Jaenisch R. (1987). Rescue of type I collagen-deficient phenotype by retroviral-vector-mediated transfer of human pro alpha 1(I) collagen gene into Mov-13 cells. *Journal of Virology* **61:** 2549-54.

Stark G. R., Kerr I. M., Williams B. R. G., Silverman R. H., and Schreiber R. D. (1998). HOW CELLS RESPOND TO INTERFERONS. *Annual Review of Biochemistry* **67:** 227-264.

Stefanovic B., and Brenner D. A. (2003). 5' Stem-Loop of Collagen alpha 1(I) mRNA Inhibits Translation in Vitro but Is Required for Triple Helical Collagen Synthesis in Vivo. *J. Biol. Chem.* **278:** 927-933.

Sudres M., Norol F., Trenado A., Gregoire S., Charlotte F., Levacher B., Lataillade J.-J., Bourin P., Holy X., Vernant J.-P., Klatzmann D., and Cohen J. L. (2006). Bone Marrow Mesenchymal Stem Cells Suppress Lymphocyte Proliferation In Vitro but Fail to Prevent Graft-versus-Host Disease in Mice. *J Immunol* **176:** 7761-7767.

Sugimoto A. (2004). High-throughput RNAi in Caenorhabditis elegans: genome-wide screens and functional genomics. *Differentiation* **72:** 81-91.

Taguchi K., Ogawa R., Migita M., Hanawa H., Ito H., and Orimo H. (2005). The role of bone marrow-derived cells in bone fracture repair in a green fluorescent protein chimeric mouse model. *Biochemical and Biophysical Research Communications* **331:** 31-36.

Takeshita F., and Ochiya T. (2006). Therapeutic potential of RNA interference against cancer. *Cancer Science* **97:** 689-696.

Thyagarajan B., Olivares E. C., Hollis R. P., Ginsburg D. S., and Calos M. P. (2001). Site-Specific Genomic Integration in Mammalian Cells Mediated by Phage {phi}C31 Integrase. *Mol. Cell. Biol.* **21:** 3926-3934.

Tiscornia G., Singer O., Ikawa M., and Verma I. M. (2003). A general method for gene knockdown in mice by using lentiviral vectors expressing small interfering RNA. *PNAS* **100:** 1844-1848.

Tiscornia G., Tergaonkar V., Galimi F., and Verma I. M. (2004). From The Cover: CRE recombinase-inducible

RNA interference mediated by lentiviral vectors. *PNAS* **101**: 7347-7351.

Toudjarska I K. M., Niu J, Wenstrup RJ, Tsipouras P. (2001). Delivery of a hammerhead ribozyme specifically downregulates mutant type I collagen mRNA in a murine model of osteogenesis imperfecta. *Antisense Nucleic Acid Drug Dev* **11**: 341-6.

Tromp G. K., H. Stacey, A. Shikata, H. Baldwin, CT. Jaenisch, R. Prockop, DJ. (1988). Structure of a full-length cDNA clone for the prepro alpha 1(I) chain of human type I procollagen. *Biochemistry* **253**: 919-22.

Tsuda H. W., T. Yamashita, T. Hamada, H. (2005). Enhanced osteoinduction by mesenchymal stem cells transfected with a fiber-mutant adenoviral BMP2 gene. *The Journal of Gene Medicine* **7**: 1322-1334.

Tsutsumida H., Swanson B. J., Singh P. K., Caffrey T. C., Kitajima S., Goto M., Yonezawa S., and Hollingsworth M. A. (2006). RNA Interference Suppression of MUC1 Reduces the Growth Rate and Metastatic Phenotype of Human Pancreatic Cancer Cells. *Clin Cancer Res* **12**: 2976-2987.

Valencia-Sanchez M. A., Liu J., Hannon G. J., and Parker R. (2006). Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.* **20**: 515-524.

van de Wetering M. O., I. Muncan, V. Pon Fong, MT. Brantjes, H. van Leenen, D. Holstege, FC. Brummelkamp, TR. Agami, R. Clevers, H. (2003). van de Wetering M, Oving I, Muncan V, Pon Fong MT, Brantjes H, van Leenen D, Holstege FC, Brummelkamp TR, Agami R, Clevers H. *EMBO Reports* **4**.

van der Krol A. R., Mur L. A., Beld M., Mol J. N. M., and Stuitje A. R. (1990). Flavonoid Genes in Petunia: Addition of a Limited Number of Gene Copies May Lead to a Suppression of Gene Expression. *Plant Cell* **2**: 291-299.

Ventura A., Meissner A., Dillon C. P., McManus M., Sharp P. A., Van Parijs L., Jaenisch R., and Jacks T. (2004). Cre-lox-regulated conditional RNA interference from transgenes. *PNAS* **101**: 10380-10385.

Venturi G., Tedeschi E., Mottes M., Valli M., Camilot M., Viglio S., Antoniazzi F., and Tato L. (2006). Osteogenesis imperfecta: clinical, biochemical and molecular findings. *Clinical Genetics* **70**: 131-139.

Vermeulen A., Behlen L., Reynolds A., Wolfson A., Marshall W. S., Karpilow J. O. N., and Khvorova A. (2005). The contributions of dsRNA structure to Dicer specificity and efficiency. *RNA* **11**: 674-682.

Wagner E. F., and Karsenty G. (2001). Genetic control of skeletal development. *Current Opinion in Genetics & Development* **11**: 527-532.

Wallis G., Starman B., Zinn A., and Byers P. (1990). Variable expression of osteogenesis imperfecta in a nuclear family is explained by somatic mosaicism for a lethal point mutation in the alpha 1(I) gene (COL1A1) of type I collagen in a parent. *American Journal of Human Genetics* **46**: 1034-40.

Wallis G. S., BJ. Zinn, AB. Byers, PH. (1990). Variable expression of osteogenesis imperfecta in a nuclear family is explained by somatic mosaicism for a lethal point mutation in the alpha 1(I) gene (COL1A1) of type I collagen in a parent. *American Journal of Human Genetics* **46**.

Wang Q., and Marini J. (1996). Antisense Oligodeoxynucleotides Selectively Suppress Expression of the Mutant alpha 2(I) Collagen Allele in Type IV Osteogenesis Imperfecta Fibroblasts . A Molecular Approach to Therapeutics of Dominant Negative Disorders. *J. Clin. Invest.* **97**: 448-454.

Wang X., Li F., and Niyibizi C. (2006). Progenitors Systemically Transplanted into Neonatal Mice Localize to Areas of Active Bone Formation in vivo: Implications of Cell Therapy for Skeletal Diseases. *Stem Cells*: 2005-0430.

Wang X., Skelley L., Cade R., and Sun Z. (2006). AAV delivery of mineralocorticoid receptor shRNA prevents progression of cold-induced hypertension and attenuates renal damage. *Gene Ther* **13**: 1097-1103.

Wang Y. (2006). HSV-1 Amplicon Vectors Are an Efficient Gene Transfer System for Skeletal Muscle Cells. *Current Gene Therapy* **6**: 371-381.

Wang Z., Goh J., De S. D., Ge Z., Ouyang H., Chong J. S. W., Low S. L., and Lee E. H. (2006). Efficacy of Bone Marrow Derived Stem Cells in Strengthening Osteoporotic Bone in a Rabbit Model. *Tissue Engineering* **12**: 1753-1761.

Ward L. M., Rauch F., Travers R., Chabot G., Azouz E. M., Lalic L., Roughley P. J., and Glorieux F. H. (2002). Osteogenesis imperfecta type VII: an autosomal recessive form of brittle bone disease. *Bone* **31**: 12-18.

Wassenegger M. (2005). The Role of the RNAi Machinery in Heterochromatin Formation. *Cell* **122:** 13-16.

Weber M., Roschger P., Fratzl-Zelman N., Schoberl T., Rauch F., Glorieux F. H., Fratzl P., and Klaushofer K. (2006). Pamidronate does not adversely affect bone intrinsic material properties in children with osteogenesis imperfecta. *Bone* **39:** 616-622.

Westbrook T. F., Martin E. S., Schlabach M. R., Leng Y., Liang A. C., Feng B., Zhao J. J., Roberts T. M., Mandel G., and Hannon G. J. (2005). A Genetic Screen for Candidate Tumor Suppressors Identifies REST. *Cell* **121:** 837-848.

Willing M., Deschenes S., Slayton R., and Roberts E. (1996). Premature chain termination is a unifying mechanism for COL1A1 null alleles in osteogenesis imperfecta type I cell strains. *American Journal of Human Genetics* **59:** 799-809.

Wiznerowicz M., and Trono D. (2003). Conditional Suppression of Cellular Genes: Lentivirus Vector-Mediated Drug-Inducible RNA Interference. *J. Virol.* **77:** 8957-8951.

Wu H. B., JF. Schnieke, A. Sharpe, A. Barker, D. Mascara, T. Eyre, D. Bruns, R. Krimpenfort, P. Berns, A, (1990). Human-mouse interspecies collagen I heterotrimer is functional during embryonic development of Mov13 mutant mouse embryos. *Molecular and Cellular Biology* **10:** 1452-60.

Wu X., Li Y., Crise B., and Burgess S. M. (2003). Transcription Start Regions in the Human Genome Are Favored Targets for MLV Integration. *Science* **300:** 1749-1751.

Xia H., Mao Q., Paulson H. L., and Davidson B. L. (2002). siRNA-mediated gene silencing in vitro and in vivo. *Nat Biotech* **20:** 1006-1010.

Xia X., Zhou H., Huang Y., and Xu Z. (2006). Allele-specific RNAi selectively silences mutant SOD1 and achieves significant therapeutic benefit in vivo. *Neurobiology of Disease* **In Press, Corrected Proof.**

Xia X. Z., H. Samper, E. Melov, S. Xu, Z. (2006). Pol II-expressed shRNA knocks down Sod2 gene expression and causes phenotypes of the gene knockout in mice. *PLoS Genetics* **2.**

Xu Y. M.-S., SH. Yang, X. Zhang, J. Oreffo, RO. (2006). The use of small interfering RNAs to inhibit adipocyte differentiation in human preadipocytes and fetal-femur-derived mesenchymal cells. *Experimental Cell Research* **312:** 1856-64.

Yamada T., and Morishita S. (2005). Accelerated off-target search algorithm for siRNA. *Bioinformatics* **21:** 1316-1324.

Yang F., Zhang Y., Cao Y.-L., Wang S.-H., and Liu L. (2005). Establishment and Utilization of a Tetracycline-controlled Inducible RNA Interfering System to Repress Gene Expression in Chronic Myelogenous Leukemia Cells. *Acta Biochimica et Biophysica Sinica* **37:** 851-856.

Yang J.-P., Fan W., Rogers C., Chatterton J. E., Bliesath J., Liu G., Ke N., Wang C.-Y., Rhoades K., Wong-Staal F., and Li Q.-X. (2006). A novel RNAi library based on partially randomized consensus sequences of nuclear receptors: Identifying the receptors involved in amyloid [beta] degradation. *Genomics* **In Press, Corrected Proof.**

Yao S., Sukonnik T., Kean T., Bharadwaj R. R., Pasceri P., and Ellis J. (2004). Retrovirus Silencing, Variegation, Extinction, and Memory Are Controlled by a Dynamic Interplay of Multiple Epigenetic Modifications. *Molecular Therapy* **10:** 27-36.

Yen C. C., Yang S. H., Lin C. Y., and Chen C. M. (2006). Stem cells in the lung parenchyma and prospects for lung injury therapy. *European Journal of Clinical Investigation* **36:** 310-319.

Yin J., Ma Z., Selliah N., Shivers D. K., Cron R. Q., and Finkel T. H. (2006). Effective gene suppression using small interfering RNA in hard-to-transfect human T cells. *Journal of Immunological Methods* **312:** 1-11.

Yoshikawa H. M., A. (2005). Bone tissue engineering with porous hydroxyapatite ceramics. *Journal of Artifical Organs* **8:** 131.136.

Young H., Steele T., Bray R., Hudson J., Floyd J., Hawkins K., Thomas K., Austin T., Edwards C., Cuzzourt J., Duenzl M., Lucas P., and Black A. J. (2001). Human reserve pluripotent mesenchymal stem cells are present in the connective tissues of skeletal muscle and dermis derived from fetal, adult, and geriatric donors. *The Anatomical Record* **264:** 51-62.

Young L. S., PF. Onion, D. Mautner V. (2006). Viral gene therapy strategies: from basic science to clinical application. *The Journal of Pathology* **208:** 299-318.

Yu J.-Y., DeRuiter S. L., and Turner D. L. (2002). RNA interference by expression of short-interfering RNAs and hairpin RNAs in mammalian cells. *PNAS* **99:** 6047-6052.

Yuan J., Wang X., Zhang Y., Hu X., Deng X., Fei J., and Li N. (2006). shRNA Transcribed by RNA Pol II Promoter Induce RNA Interference in Mammalian Cell. *Molecular Biology Reports* **33:** 43-49.

Zamore P. D., and Haley B. (2005). Ribo-gnome: The Big World of Small RNAs. *Science* **309:** 1519-1524.

Zeng Y., Yi R., and Cullen B. R. (2003). MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *PNAS* **100:** 9779-9784.

Zern M. O., I. Duan, L. Pomerantz, R. Liu, SL. Strayer, DS. (1999). A novel SV40-based vector successfully transduces and expresses an alpha 1-antitrypsin ribozyme in a human hepatoma-derived cell line. *Gene Therapy* **6:** 114-20.

Zhang H. F., Y. Wright, C. Gannon, M (2005). Efficient recombination in pancreatic islets by a tamoxifen-inducible Cre-recombinase. *genesis* **42:** 210-217.

Zhang J., Niu C., Ye L., Huang H., He X., Tong W.-G., Ross J., Haug J., Johnson T., Feng J. Q., Harris S., Wiedemann L. M., Mishina Y., and Li L. (2003). Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425:** 836-841.

Zhang X., Mitsuru A., Igura K., Takahashi K., Ichinose S., Yamaguchi S., and Takahashi T. A. (2006). Mesenchymal progenitor cells derived from chorionic villi of human placenta for cartilage tissue engineering. *Biochemical and Biophysical Research Communications* **340:** 944-952.

Zhang X. L., TA. Chen, ST. Peng, H. Wergedal, JE. Guttierez, GG. Sheng, MH. Lau, KH. Baylink, DJ. (2004). Local ex vivo gene therapy with bone marrow stromal cells expressing human BMP4 promotes endosteal bone formation in mice. *Journal of Gene Medicine* **6:** 4-15.

Zhao L.-J., Jian H., and Zhu H. (2003). Specific gene inhibition by adenovirus-mediated expression of small interfering RNA. *Gene* **316:** 137-141.

Zheng L., Bidere N., Staudt D., Cubre A., Orenstein J., Chan F. K., and Lenardo M. (2006). Competitive Control of Independent Programs of Tumor Necrosis Factor Receptor-Induced Cell Death by TRADD and RIP1. *Mol. Cell. Biol.* **26:** 3505-3513.

Ziady A. G., and Davis P. B. (2006). Current prospects for gene therapy of cystic fibrosis. *Current Opinion in Pharmacology* **6:** 515-521.