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.

# Genetic Studies in Coeliac Disease

**Graham Turner**

**2011**

A thesis submitted to the University of Dublin, Trinity College

for the degree of

Doctor of Philosophy (Ph.D.)

by

Graham Turner,

MSc. (Molecular Medicine),

PgDip (Statistics),

B.A. (Mod. Human Genetics).

Supervisor

Ross McManus, Ph.D.

Department of Clinical Medicine,

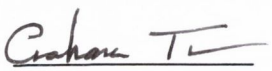
University of Dublin,

Trinity College,

January, 2011.

**DECLARATION**

I declare that this thesis has not been submitted as an exercise for a degree at this or any other University, that it is entirely my own work (save where otherwise acknowledged), and I agree that the Library may lend or copy the thesis upon request.



Graham Turner

TRINITY COLLEGE
LIBRARY DUBLIN
LIBRARY DUBLIN

Thesis 9553

**Summary**

This study is primarily concerned with elucidating genetic contributions to coeliac disease in the Irish population via case control based association studies. Due to the nature of complex diseases, and the large sample sizes required to successfully detect genetic factors of relatively small magnitude, as well as the need for association confirmation in independent populations, there is also a lesser focus on the role of genetic factors in the Italian, British and Dutch populations.

In pursuit of these aims, we employ a broad range of contemporary study designs; a candidate gene approach, bioinformatic examination of the previously associated 5q31-33 coeliac linkage region, and genome wide association study.

650 Irish coeliac patients, and over 1,000 Irish population controls, were genotyped across a range of studies, using PCR, Taqman, Amplifluor, and Illumina GoldenGate technology. Allele, genotype and carrier status frequencies for each individual SNP were compared statistically between patient and control groups using the chi-square or fishers exact test. Haplotype frequencies were estimated from unphased genotype data using computational methods and were statistically compared between groups.

This study has uncovered a number of statistically significant associations with coeliac disease in the Irish population, and has also, in conjunction with collaborative partners in England and the Netherlands, revealed multiple statistically significant associations with coeliac disease across multiple European populations, many of which have subsequently been replicated in other independent populations.

In the Irish sample, the NfKB regulator TNIP1 on chromosome 5 has shown significant association, an observation which we have confirmed in an independent Italian population. The CCR-Δ32 allele has also shown evidence of association with coeliac disease, as has an IL-18 promoter centred haplotype, and a haplotype spanning the TNFSF14 gene.

In a larger context, as part of a consortium conducting a follow-up study into the results of the UK coeliac GWAS, we have been instrumental in discovering 12 novel coeliac susceptibility regions, which harbour genes with key immunological function; IL2/IL21, RGS1, IL-18RAP, IL-12A, LPP, TAGAP, LNK, CCR1/CCR3, TNFAIP3, REL, CD80 and CTLA4. These observations occurred in the Irish, British and Dutch populations, and many have subsequently been replicated.

We also address, to an extent, the potential role for population variation in the context of complex disease, looking at trends in the Irish population from the GWAS follow-up data.

In conclusion, this study has been successful in elucidating multiple variants associated with coeliac disease, both in a standalone context, and as part of a broader research arc.

**Acknowledgements**

I would like to thank Dr. Ross McManus for providing me with the opportunity for such a unique challenge over the last five years, and his support in producing this thesis.

I would also like to thank my fellow members of the Genomic Research group; especially Dr. Anthony Ryan, without whom I would not be where I am today. For his knowledge, friendship, insight, perspective and guidance, I am truly grateful. My thanks too to Dr. Arun Kumar Mankan and Dr. Matthew Lawless, for friendship in times of adversity, and to Dr. Micheal O'Dwyer, Dr. Mary White, Dr. Robert Greely, Dr. Mawish Babar for their clinician's perspective. To Dr. Karen Brophy, Dr. Megan Dring and Dr. Kunal Patel, my thanks.

My thanks also to Dr. Karen Hunt, for her friendship, expertise, keen mind and excellent organisation skills. I cannot think of a better lab partner. To Dr. Peter Olwell, whose good humour kept me laughing, I extend my sincere thanks.

To the doctors, nurses and patients of the coeliac clinic, thank you for your continued support in providing clinical samples for this project, as well as to the Irish Blood Transfusion Service, for providing healthy blood donors for use as controls, and Dr. Joe Mc Partlin for his help in sample preparation, and good humour.

My thanks too to my friends, and family, who have had to put up with talk of genes and SNPs, inflammation and association, for quite some time now. For their patience, support, and good humour, I am forever grateful.

And a very special thanks, of course, to my wife, Lynn, without whom this thesis would never have gotten written.

# Table of Contents


Genetic Studies in Coeliac Disease ..... 1

DECLARATION ..... 2

Summary ..... 3

Acknowledgements ..... 5

Abbreviations ..... 12

Publications from this Thesis ..... 14

Dedication. ..... 16

CHAPTER I ..... 17

Introduction ..... 17

1.1 Coeliac Disease; A Primer ..... 18

1.2 Clinical Aspects of Coeliac Disease ..... 18

1.3 Diagnostic Methods ..... 21

1.4 Incidence and Prevalence ..... 23

1.5 Genetic Components of Coeliac Disease ..... 28

1.6 HLA contribution to Coeliac Disease ..... 29

1.7 Gluten and the Immune Response ..... 32

1.8 Non-HLA genetic factors and Coeliac Disease ..... 34

1.9 Linkage Disequilibrium ..... 34

1.10 Haplotype Analysis ..... 36

1.11 Discovering the Genetic Basis of Complex Diseases ..... 38

1.12 Genome Scans and Coeliac Disease ..... 39

1.13 Candidate Gene Approaches to Complex Disease ..... 43

1.14 International Hap Map Project ..... 45

1.15 GWAS and Complex Disease ..... 46

1.16 Complex Disease and Regulatory Elements ..... 47

1.17 Case-Control Studies ..... 48

1.18 Irish Population ..... 49

CHAPTER II ..... 54

Materials and Methods ..... 54

2.1 Coeliac Patient DNA collection ..... 55

2.2 Control Subject collection ..... 55

2.3 DNA extraction from whole blood ..... 55

2.4 DNA QC and quantification ..... 55

2.5 Whole Genome Amplification ..... 56

2.6 Power Calculations ..... 57
2.7 ECR Browser ..... 57
2.8 Rvista ..... 58
2.9 Selection of SNPs for Haplotype Tagging ..... 61
2.10 Genotyping of Polymorphisms ..... 61
2.10.1 Polymerase Chain Reaction ..... 61
2.10.2 Agarose Gel Electrophoresis ..... 62
2.10.3 Allelic Discrimination using Taqman Genotyping technology ..... 62
2.10.4 Ampiflour Genotyping ..... 64
2.10.5 Illumina GoldenGate Genotyping Platform ..... 66
2.10.6 Confirmation of Genotyping Results ..... 72
2.11 Data Analysis ..... 72
2.11.1 Beadstudio ..... 72
2.11.2 Genetic case-control differentiation analysis of allele, genotype and carrier status frequencies ..... 73
2.11.2 Hardy-Weinberg Equilibrium (HWE) testing ..... 73
2.11.3 Haplotype estimation and differentiation ..... 73
2.11.4 Linkage disequilibrium analysis ..... 74
2.11.5 Cochran-Mantel-Haenszel test for significance in stratified populations ..... 74
CHAPTER III ..... 75
Chromosome 5 & Coeliac Disease ..... 75
3.1 Introduction ..... 76
3.1.1 Linkage to the 5q region ..... 76
3.2 Objectives ..... 81
3.3 Results ..... 81
3.3.1 ECR Browser; Identification of Human/Murine ECRs ..... 81
3.3.2 Tagging of Variability within observed ECRs ..... 82
3.3.3 rVISTA; Transcription Factor Biding Site Characterisation ..... 83
3.3.4 SNP Selection in the 146-151Mb interval of chromosome 5q. ..... 84
3.3.5 TNIP1 ..... 86
3.3.6 STK32A ..... 87
3.3.7 Allele, Genotype and Carrier Status frequencies and disease association testing of genotyped polymorphisms in TNIP1 associated SNPs ..... 88

3.3.8 Allele, Genotype and Carrier Status frequencies and disease association testing of genotyped polymorphisms in TNIP1 associated SNPs in an independent Italian population ..... 90

3.3.9 Cochran-Mantel-Haenszel test for significance of TNIP1 associated SNPs in the combined Irish and Italian populations ..... 92

3.3.10 Allele, Genotype and Carrier Status frequencies and disease association testing of genotyped polymorphisms in STK32A 3' htSNPs ..... 93

3.3.11 Haplotype analysis of TNIP1 associated SNPs ..... 95

3.3.12 Haplotype analysis of TNIP1 associated SNPs in the Italian Population .. ..... 97

3.3.13 Homozygous Haplotypes ..... 98

3.3.14 Patterns of Linkage Disequilibrium ..... 99

3.3.15 Hardy Weinberg Analysis ..... 100

3.4 Discussion ..... 102

CHAPTER IV ..... 107

Coeliac Disease Genomewide Association Study Replication ..... 107

4.1 Introduction ..... 108

4.1.1 Early Genome Wide Association Studies ..... 108

4.1.2 A Genome Wide Association Study in the UK population ..... 109

4a.2 Objectives ..... 111

4a.3 Results ..... 112

4a.3.1 SNP Selection ..... 112

4a.3.2 Allele frequencies and disease association testing of non IL-2/IL-21 LD block genotyped SNPs ..... 113

4a.3.3 Allele, Genotype and Carrier Status frequencies and disease association testing of IL-2/IL-21 LD block genotyped SNPs ..... 114

4a.3.4 LD Structure of the IL-2/IL-21 Region ..... 116

4a.3.5 Allele, Genotype and Carrier Status frequencies and disease association testing of Block 1 and Block 2 IL-2/IL-21 LD block genotyped htSNPs ..... 117

4a.3.6 Haplotype Analysis of the IL-2/IL-21 region in the Irish population ..... 120

4a.3.7 Joint analysis of IL-2/IL-21 SNPs in Irish, British and Dutch populations via the Cochran-Mantel-Haenszel test. ..... 121

4a.3.8 Allele, Genotype and Carrier Status frequencies and disease association testing of selected IL-2/IL-21 LD block genotyped tgSNPs in an Italian population ..... 122

4a.3.9 Hardy Weinberg Analysis ..... 124

4b.1 Introduction ..... 126

4b.1.1 Genome Wide Association Study in the Irish, British and Dutch Populations ....126
4b.2 Objectives ....128
4b.3 Results ....129
4b.3.1 Selection of SNPs for inclusion in study ....129
4b.3.2 Genotyping and Quality Control of the Irish Dataset ....129
4b.3.3 Analysis of the Irish Dataset ....130
4b.3.4 Locations of the Most Associated SNPs in the Irish Dataset ....130
4b.3.5 Allele, Genotype and Carrier Status frequencies and disease association testing of SNPs with P < 0.015 in the Irish population ....131
4b.3.6 CMH Analysis of the Irish, Dutch, British and GWAS datasets ....135
4c.1 Introduction ....137
4c.2 Objectives ....137
4c.3 Results ....137
4c.3.1 Part 3 Replication of Effect in an Independent Italian Population....137
4.4 Discussion ....141
4a.4 IL-2/IL21....141
4b.4 Genome Wide Association Follow up in Multiple Populations ....144
4b.4.1 Overall Significant SNPs in Combined Populations ....144
4b.4.2 SNPs Suggestive of Significance in the Irish Population ....148
4c.4 Replication of significant Golden Gate SNPs in an Italian population....152
CHAPTER V ....154
Candidate Genes & Coeliac Disease ....154
5.1 Introduction ....155
5.1.1 Candidate Gene Selection....155
5.1.2 CCR5 ....155
5.1.3 TNFSF14 ....156
5.1.4 IL-18 ....157
5.1.5 MyoixB....157
5.1.6 Candidate Gene Biased Follow up of Multi-population GWAS....158
5.2 Objectives ....159
5.3 Results ....159
5.3.1 TNFSF14 ....159
5.3.2 IL-18....160
5.3.3 MyoIXB ....161

5.3.4 Allele, Genotype and Carrier Status frequencies and disease association testing of CCR5 Δ32 ..... 161

5.3.5 Allele, Genotype and Carrier Status frequencies and disease association testing of TNFSF14 ..... 163

5.3.6 Allele, Genotype and Carrier Status frequencies and disease association testing of IL-18 SNPs ..... 165

5.3.7 Allele, Genotype and Carrier Status frequencies and disease association testing of MyoIXB SNPs. ..... 166

5.3.8 Candidate Gene Biased Follow up of Multi-population GWAS in the Irish Population ..... 168

5.3.9 Haplotype analysis of TNFSF14 associated SNPs ..... 171

5.3.10 Haplotype analysis of IL-18 associated SNPs ..... 172

5.3.11 Haplotype analysis of MyoIXB associated SNPs ..... 172

5.3.12 Hardy Weinberg Analysis ..... 173

5.4 Discussion ..... 174

5.4.1 Single Candidate Genes ..... 174

5.4.2 Candidate Gene Biased Follow up of Multi-population GWAS ..... 176

5.4.3 Candidate Gene Biased Follow up of Multi-population GWAS ; The Irish Population ..... 178

Chapter VI ..... 181

General Discussion ..... 181

6.1 Autoimmune Disease; Common Factors ..... 182

6.2 Regulatory Variants play a large role in complex disease ..... 184

6.3 General Functional Overview ..... 185

6.4 Pathways, Interactants and Associations ..... 188

6.5 Cyokines IL-18, IL-18RAP, IL-2, IL-21, IL12A ..... 188

6.6 Tolerance ..... 189

3.1 6.7 NfKappaB ..... 190

3.2 6.8 Summary of Interacting Gene Functions ..... 193

3.3 6.9 Population specific variants ..... 194

6.9.1 Common Pathway, different member ..... 194

6.9.2 GWAS follow up; Differences between populations ..... 194

6.9.3 Irish Population ..... 196

6.9.4 5q linkage peak disparity ..... 196

6.10 A Model of Coeliac Disease: common core, with varying modifying factors ..... 197

6.11 Further Studies ..... 199

References .....200

**Abbreviations**

| | |
|---|---|
| 95%CI | 95% Confidence Interval |
| AGA | Anti-Gliadin Antibody |
| AI | Autoimmune Disease |
| AIDP | Acute inflammatory Demyelinating Polyneuropathy |
| APC | Antigen Presenting Cell |
| ASE | Allele Specific Extension |
| ASO | Allele specific Oligionucleotide |
| CD | Coeliac Disease |
| cM | CentiMorgans |
| CMH | Cochran-Mantel-Haenszel |
| CV/MD | Common Variant / Multiple Disease |
| CY3 | A fluorophore (fluorescent part of a molecule) |
| CY5 | A fluorophore |
| DBSNP | Database of SNPs |
| DH | Dermatitis Herpetiformis |
| DM: | Diabetes Mellitus |
| DNA | Deoxyribonucleic acid |
| ECR | Evolutionary Conserved Region |
| EM | Expectations Maximisation |
| EMA | Anti-Endomysial Antibody |
| FAM | A fluorophore |
| Fluor | Fluorophore (the fluorescent part of a molecule) |
| g | grams |
| GWAS | Genome Wide Association Study |
| HLA | Human Leucocyte Antigen |
| hLOD | heterogeneity LOD |
| htSNP | Haplotype Tagging Single Nucleotide Polymorphism |
| HWE | Hardy Weinberg Equilibrium |
| IBD: | Inflammatory Bowel Disease |
| IEL | Intra-Epithelial Lymphocyte |
| IGA | Immuno-globulin A |
| IGG | Immuno-globulin G |
| JIA: | Juvenile Idiopathic Arthritis |
| JOE | A fluorophore |
| KB | Kilobase |
| LD | Linkage Disequilibrium |
| LOD | Logarithm of the Odds |
| LSO | Locus Specific Oligonucleotide |
| MAF | Minor Allele Frequency |
| MB | Megabase |
| MHC | Major Histocompatibility Complex |
| ml | Millilitre |
| MLS | Maximised LOD Score |
| MS: | Multiple Sclerosis |

| | |
|---|---|
| NFKB | Nuclear Factor Kappa B |
| NK | Natural Killer |
| nm | Nanometre |
| NPL | Non Parametric Linkage |
| OR | Odds Ratio |
| PCR | Polymerase Chain Reaction |
| RA | Rheumatoid Arthritis |
| RNA | Ribonucleic acid |
| SAM | SAM Sentrix® Array Matrix |
| Sib pairs | Sibling Pairs |
| SLE: | Systemic Lupus Erythematosus |
| SNP | Single Nucleotide Polymorphism |
| SS: | Sjorgenson Syndrome |
| SSD | Steven's Smith Donnelly |
| SUD | Single-Use DNA (plate) |
| T1D: | Type 1 Diabetes |
| TAE | Tris Acetate EDTA |
| TCR | T-Cell Receptor |
| TF | Transcription Factor |
| TFBS | transcription Factor Binding Site |
| ttG | Tissue Trans Glutaminase |
| UC | Ulcerative Colitis |
| UDG | Uracil-DNA Glycosylase |
| µg | Microgram |
| µl | Microlitre |
| µM | Micromole |
| UTR | Untranslated Region |
| UV | Ultra-violet |
| VIC | A fluorophore |
| WGA | Whole Genome Amplification |

**Publications from this Thesis**

**Original Papers**

Hunt KA, Zhernakova A, **Turner G**, Heap GA, Franke L, Bruinenberg M, Romanos J, Dinesen LC, Ryan AW, Panesar D, Gwilliam R, Takeuchi F, McLaren WM, Holmes GK, Howdle PD, Walters JR, Sanders DS, Playford RJ, Trynka G, Mulder CJ, Mearin ML, Verbeek WH, Trimble V, Stevens FM, O'Morain C, Kennedy NP, Kelleher D, Pennington DJ, Strachan DP, McArdle WL, Mein CA, Wapenaar MC, Deloukas P, McGinnis R, McManus R, Wijmenga C, van Heel DA.

Newly identified genetic risk variants for celiac disease related to the immune response. Nature Genetics. 2008 Apr;40(4):395-402.

Trynka G, Zhernakova A, Romanos J, Franke L, Hunt KA, **Turner G**, Bruinenberg M, Heap GA, Platteel M, Ryan AW, de Kovel C, Holmes GK, Howdle PD, Walters JR, Sanders DS, Mulder CJ, Mearin ML, Verbeek WH, Trimble V, Stevens FM, Kelleher D, Barisani D, Bardella MT, McManus R, van Heel DA, Wijmenga C.

Coeliac disease associated risk variants in TNFAIP3 and REL implicate altered NF-kappaB signalling. Gut 2009;58:1078–1083.

Brophy K, Ryan AW, **Turner G**, Trimble V, Patel KD, O'Morain C, Kennedy NP, Egan B, Close E, Lawlor G, MacMathuna P, Stevens FM, Abuzakouk M, Feighery C, Kelleher D, McManus R.

Evaluation of 6 candidate genes on chromosome 11q23 for coeliac disease susceptibility: a case control study. BMC Medical Genetics. 2010. May 17;11:76.

Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, Zhernakova A, Heap GA, Adány R, Aromaa A, Bardella MT, van den Berg LH, Bockett NA, de la Concha EG, Dema B, Fehrmann RS, Fernández-Arquero M, Fiatal S, Grandone E, Green PM, Groen HJ, Gwilliam R, Houwen RH, Hunt SE, Kaukinen K, Kelleher D, Korponay-Szabo I, Kurppa K, MacMathuna P, Mäki M, Mazzilli MC, McCann OT, Mearin ML, Mein CA, Mirza MM, Mistry V, Mora B, Morley KI, Mulder CJ, Murray JA, Núñez C, Oosterom E, Ophoff RA, Polanco I, Peltonen L, Platteel M, Rybak A, Salomaa V, Schweizer JJ, Sperandeo MP, Tack GJ, **Turner G**, Veldink JH, Verbeek WH, Weersma RK, Wolters VM, Urcelay E, Cukrowska B, Greco L, Neuhausen SL, McManus R, Barisani D, Deloukas P, Barrett JC, Saavalainen P, Wijmenga C, van Heel DA.

Multiple common variants for celiac disease influencing immune gene expression. Nature Genetics. 2010 Mar;42:295-302.

Alexandra Zhernakova,, Eli A. Stahl, Gosia Trynka, Soumya Raychaudhuri, Eleanora Festen, Lude Franke , Rudolf S.N. Fehrmann , Fina A.S. Kurreeman, Brian Thomson , Namrata Gupta , Jihane Romanos, Ross McManus, Anthony W Ryan, **Graham Turner**, Elaine F. Remmers, Luigi Greco, Rene Toes, Elvira Grandone, Maria Cristina Mazzilli, Anna Rybak, Bozena Cukrowska, Yonghong Li, Paul I.W. de Bakker, Peter K. Gregersen, Jane Worthington, Katherine A. Siminovitch , Lars Klareskog, Tom W.J. Huizinga, Cisca Wijmenga, Robert M. Plenge

Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci .Submitted to and accepted by PLOS, awaiting publication.

**Turner G** & McManus R

Genetics of Coeliac Disease ; Review. World Journal of Gastroenterology. In Preparation.

**Turner G**, Ryan AW, Trimble V, O'Morain C, Kennedy NP, Egan B, Close E, Lawlor G,, Stevens FM, Abuzakouk M, Feighery C, Kelleher D, McManus R.

TNIP1 and coeliac disease in the Irish and Italian populations. In Preparation.

**This Thesis is dedicated to my loving wife, Lynn.**

# CHAPTER I

# Introduction

## 1.1 Coeliac Disease; A Primer

Coeliac disease (CD) is an autoimmune inflammatory disorder, primarily localised to the small intestine, although extra-intestinal symptoms are evidenced in some sufferers. CD results from an immune response, in genetically susceptible individuals, to ingested dietary gliadins in wheat (Arentz-Hansen *et al.* 2002) and the corresponding prolamins in rye and barley (Vader *et al.* 2003). It is principally characterised by intestinal villous atrophy with crypt hyperplasia, increased number of intraepithelial lymphocytes and extensive inflammatory infiltrate in the intestinal lamina propria. Circulating disease-specific Immunoglobulin A (IgA) and anti-tissue transglutaminase 2 (tTG2) antibodies are also characteristics of the condition (Maki *et al.* 1997). Gluten-reactive intestinal T-cells can be isolated from virtually all patients with coeliac disease, but not from disease free individuals (Arentz-Hansen *et al.* 2004). As the disease typically goes into remission when triggering cereals are avoided, a gluten-free diet is the standard treatment for this disorder (Chapman *et al.* 1974).

## 1.2 Clinical Aspects of Coeliac Disease

The pathology of coeliac disease principally, though not exclusively, manifests itself in the small intestine. Intestinal mucosa possesses hair like protrusions, villi, which act to increase its surface area; this increase in surface area is integral to the efficient absorption of nutrients in the small intestine. Coeliac disease results in localised auto-immune reactions which lead to atrophy of these intestinal villi, resulting in a flattening of the intestinal mucosa, and corresponding reduction in surface area, with resulting reduction in absorption (Green *et al.* 2005). In 1992 the Marsh Scale, a scale for codifying the degree of mucosal damage, was laid out (Marsh *et al.* 1992). According to this scale, there were five distinct stages of mucosal damage. Marsh stage 0 corresponded to a histologically normal mucosal surface, the other 4 stages are as follows;

**Marsh Lesions**

- Stage 1: Presence of Intra Epithelial Lymphocytes.
- Stage 2: Proliferation of cells in the crypts of Lieberkuhn, resulting in crypt hyperplasia.
- Stage 3: Evidence of characteristic Villous Atrophy.
- Stage 4: Hypoplasia (underdevelopment) of the small bowel.

**Figure 1.1** gives a diagrammatic representation of this spectrum of changes outlined above:

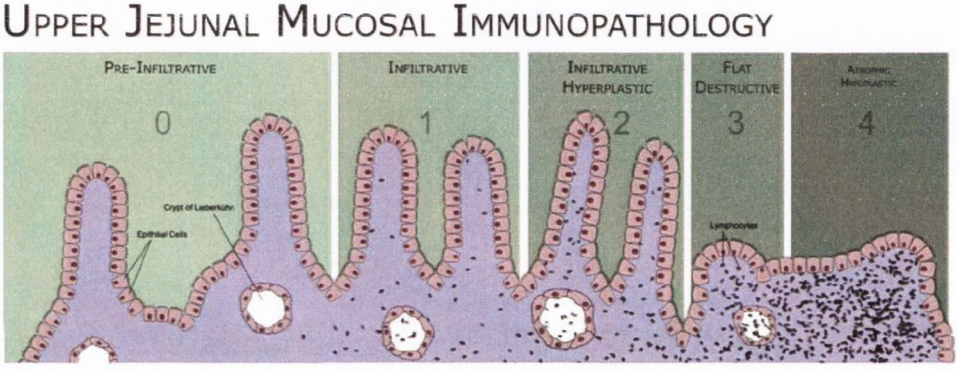



**Figure 1.1 Coeliac Mucsoal Immunopathology. The spectrum of changes that take place in the mucosa of the intestine across the 5 classic Marsh Lesions.** Source: Wikipedian Prolific.

These stages were further refined in 1999 (Oberhuber *et al.* 1999) adding several distinct sub-phenotypes to the broadly classified Stage 3. The additional subtypes, and their characteristics are detailed in **Table 1.1.**

**Table 1.1 Marsh Lesion Subtypes and their characteristics.**

| Marsh Type | IELs/100 Enterocytes | Crypts | Villi | Classification |
|---|---|---|---|---|
| 0 | <40 | Normal | Normal | Normal Mucosa |
| 1 | >40 | Normal | Normal | Infiltrative lesion |
| 2 | >40 | Increased | Normal | Hyperplasic Type |
| 3a | >40 | Increased | Mild Atrophy | Destructive Lesion |
| 3b | >40 | Increased | Marked Atrophy | Destructive Lesion |
| 3c | >40 | Increased | Absent | Destructive Lesion |

Damage to the intestinal wall results in mal-absorption; steatorrhea (diarrhoea due to malabsorption of fat) is common, as is nutrient deficiency, particularly calcium, iron, folic acid and fat soluble vitamins. These complications can, in turn, give rise to a number of consequent problems; weight loss, weakness, anaemia, depression, tiredness, reduced bone density, and iron and folate deficiency (Maki *et al.* 1997). Lactose intolerance can also occur, due to the inability of the damaged villi to adequately absorb the lactose (Di Sabatino & Corazzo, 2009).

In addition to the expected intestinal symptoms, a wide variety of extra-intestinal symptoms have also been observed in a subset of coeliac patients. Dermatitis herpetiformis (DH) is an outbreak of itching blisters on the knees, elbows and buttocks, and is often a symptom of the disease, affecting about 15%-25% of patients with CD (Reunala *et al.* 2001). Increased rates of gastrointestinal tract cancers, specifically intestinal lymphomas, are also seen in CD patients. The risk of cancer falls off with increasing length of gluten free diet following CD diagnosis, and is not significantly increased after 10 years on without gluten. Patients diagnosed with CD before the age of 10 years display no significantly increased risk for malignant lymphoma, and those diagnosed with CD older than 20 years had a considerably higher risk (Askling *et al.* 2002).

Epilepsy with occipital calcifications has also been linked to CD (Gobbi 2005), and osteopenia/osteoporosis appears to be present in virtually all patients who fail to adhere to the gluten-free diet, (due to calcium malabsorption) with a gluten-free diet reducing or normalising the condition (Molteni *et al.* 1990). Failure to adhere to the

gluten-free diet also increases the risk for teenagers of having circulating organs specific autoantibodies, as well as other autoimmune diseases (Toscano *et al.* 2000).

It is worth noting that the full range of clinical manifestations may not always be present, and if they are present, may be so with varying degrees of severity; coeliac disease has a wide range of phenotypes, from severe, dangerous instances of the above symptoms, to clinically silent, asymptomatic instances (Lo *et al.* 2003). Patients presenting with the classical symptoms of so called Coeliac sprue, weight loss and anaemia are becoming increasingly rare as better diagnostic tests detect patients earlier, when only mild symptoms such as indigestion and anaemia are present (Heap & Van Heel, 2009).

### 1.3 Diagnostic Methods

Following the identification of the characteristic Coeliac histological lesion of the small bowel, the gold standard for diagnosis of CD became intestinal biopsy. This is done in order to determine to what degree, if any, the characteristic villus atrophy, IEL and crypt cell hyperplasia are present. The epithelial cells gathered by biopsy are scored according to the Marsh scale (see section 1.2) and the results inform the diagnosis.

Presence of normal mucosa suggests CD is highly unlikely, while Type 1 (Infiltrative lesions) tend to be observed either in patients on a gluten-free diet. Type 2 (Hyperplastic type) are very rare, being seen occasionally in DH, or the family members of patients with CD. Type 3 subtypes (Destructive lesions) are representative of the spectrum of changes seen in symptomatic CD, and observation of these types of lesion would constitute proof positive for the disease (Antonioli *et al.* 2003).

There are several variables which need to be considered when relying on a biopsy-based diagnosis, however. An important element in the interpretation of the observed Marsh Lesions is whether or not the patient has removed gluten from their diet by the time of the biopsy, and if so for how long. Reduction, or elimination, of gluten from the patient's diet prior to biopsy, especially over a period of time, will result in amelioration of intestinal damage, and lessening of the severity of the Marsh Lesions. As such, recent challenge with gluten may be necessary prior to biopsy, to reduce the change of Type II error.

Furthermore, duodenal biopsy has several pitfalls: (1) at least four biopsies are needed to achieve a representative picture; (2) poorly oriented or inadequate biopsies may not be useful for diagnosis; and (3) it is an invasive procedure, both in children and adults. Additionally, the variability of lesions from a single biopsy are quite high, and if biopsies are taken from areas of the intestine which are only mildly affected, a type II error can also occur (Vivas *et al.* 2009).

As a synergistic approach, testing for the presence of anti-gliadin antibodies in serum, was used as an adjunct to biopsy, in order to refine the diagnostic method. However, this test lacks specificity and is only moderately sensitive, making it of limited utility (Kelly *at al.* 1991).

A test for for IgA anti-endomysial antibodies was developed in the early 1980s, and has seen widespread use since the '90s. Evidence shows that the presence of endomysial antibodies is virtually 100% specific for coeliac disease, with a reported sensitivity of 87–89% (Chorzelski *et al.* 1983). Due to its comparatively non-invasive nature, this test has gained popularity as an initial step in Coeliac Diagnosis, and as a method of estimating frequencies of Coeliac Disease in populations (see section 1.4). The EMA test, however, is complex and expensive, and as such, a variety of other auto-antibody tests have been employed, with varying levels of success. For example, when it was determined that the EMA antibody was being generated in response to endogenous enzyme Tissue Transglutaminase , screening for the direct presence of the auto-antigen was utilised in an attempt to devise a cheaper and more straightforward assay.

When utilising antibody tests which assay IgA subtypes, caution must also be exercised, as many Coeliac patients exhibit a reduction in circulating IgA, which may again result in a type II error. It is recommended that IgA levels are checked, to insure sufficient circulating levels. In those patients with insufficient IgA, IgG subtype antibodies may be utilised instead.

As of this writing, the current best practice for diagnosis of Coeliac Disease is as follows; patients presenting with symptoms, gastroenterological or otherwise, which are suggestive of the disease, are given a preliminary serological tTG test. Anyone with tTG over a specified threshold is selected for Ema based testing. If this test is positive, a follow up biopsy is required to secure a definitive diagnosis. Due to the latitude involved in histogical diagnosis, and antibody detection, a synergy of these two approaches seems the most prudent in hoping to obtain an accurate diagnosis

## 1.4 Incidence and Prevalence

Coeliac Disease was first described around the 2nd Century BC, by the Ancient Greek physician Aretaeus of Cappadocia, who identified it as an affliction of the old, and most commonly affecting women (Adams 1856). Samuel Gee gave the first modern-day description of the condition in 1887, building upon Aretaeus' observations in his description of the condition. Gee, however, observed the condition in infants, and considered it a disease of childhood. The 'classical' picture of coeliac disease; occurring in the young, presenting with characteristic abdominal symptoms, diarrhoea, and 'failure to thrive' owes itself to Gee's observations at this time (Gee 1888).

This view of Coeliac Disease, as described by Gee, is one of a rare illness of childhood, affecting individuals of European Caucasian descent. Estimates of incidence of Coeliac Disease as recently as 1985, placed the population frequency at 0.1%. Even at this time however, the classical view of the disease as an affliction of childhood was beginning to be questioned, with the Coeliac Society noting an increase in age of diagnosis amongst its members. **Table 1.2** shows the Coeliac Society data on age of onset of their members as of 1985 (Langman *et al.* 1985).

**Table 1.2 Changing trends in age of diagnosis of celiac disease, 1967-1981, Celiac Society, showing increase in age of diagnosis over time.**

Table *Numbers and percentage of Coeliac Society members according to age and to year of diagnosis*

| | *Age in years* | | | | | |
|---|---|---|---|---|---|---|
| *Year of diagnosis* | *0–4* | *5–9* | *10–24* | *25–39* | *40+* | *Not known* |
| 1967 | 125 | 22 | 22 | 38 | 39 | 246 |
| or earlier | 50·8 | 8·9 | 8·9 | 15·4 | 15·9 | 12·7 |
| 1968–69 | 31 | 7 | 5 | 11 | 13 | 67 |
| | 46·3 | 10·4 | 7·5 | 16·4 | 19·4 | 3·5 |
| 1970–71 | 33 | 9 | 11 | 21 | 21 | 95 |
| | 34·7 | 9·5 | 11·6 | 22·1 | 22·1 | 4·9 |
| 1972–73 | 60 | 18 | 20 | 52 | 42 | 192 |
| | 31·3 | 9·4 | 10·4 | 27·1 | 21·9 | 9·9 |
| 1974–75 | 67 | 22 | 36 | 59 | 70 | 254 |
| | 26·4 | 8·7 | 14·2 | 23·2 | 27·6 | 13·1 |
| 1976–77 | 80 | 19 | 44 | 72 | 101 | 316 |
| | 25·3 | 6·0 | 13·9 | 22·8 | 32·0 | 16·4 |
| 1978–79 | 44 | 18 | 43 | 78 | 102 | 285 |
| | 15·4 | 6·3 | 15·1 | 27·4 | 35·8 | 14·8 |
| 1980–81 | 37 | 8 | 30 | 73 | 93 | 241 |
| | 15·4 | 3·3 | 12·4 | 30·3 | 38·6 | 12·5 |
| Not known | | | | | 236 | 236 |
| | | | | | | 12·2 |
| | 477 | 123 | 211 | 404 | 481 | 236 1932 |
| | 24·7 | 6·4 | 10·9 | 20·9 | 24·9 | 12·2 100·0 |

As can be seen from this table, the demographic amongst Coeliac Society members shifts over the 24 years; in 1967, only 15.9% of the members were diagnosed at age 40+, by 1980-1981, this has increased to 38.6%. At the other end of the spectrum, 50.8% of the members in 1967 were diagnosed between the ages of 0 and 4, a number which has fallen to 15.4% by 1980-1981. A marked increase over this time can also be seen in the 25-39 age of diagnosis, and the 5-9 years age of diagnosis also shows a decrease in incidence.

CD development in adults appears to follows a bimodal distribution, with an initial peak in the fourth decade of life (mostly women) and a second, smaller peak in the

sixth or seventh decade of life (predominantly men) (Swinson & Levi 1980). The reasons for the apparent increase in incidence amongst the older population may be due to improved screening and diagnostic techniques leading to the recognition of coeliac disease where it would previously have gone unnoticed or misdiagnosed. Additionally, individuals with 'silent', or asymptomatic disease, may convert to a more aggressive phenotype with age. Lastly, increased recognition of non-classical symptoms may aid in diagnosis (Ferguson *et al.* 1993).

An exception to the estimated incidence was found in a study carried out in the west of Ireland, where the prevalence of coeliac disease was calculated to be 1 in 300 (Cronin & Shanahan, 2001).This finding led to the perception that coeliac disease was uniquely common in Ireland, especially in the western seaboard region. A model was proposed, relating the prevalence of CD on different continents to the interaction between genetic gradients largely determined by the advance of agriculture, and historical patterns of cereal ingestion (Cronin & Shanahan, 2001).

However, the advent of the non-invasive serological tests capable of detecting coeliac disease, such as those for anti-endomysial antibodies, anti-IGA antibodies and others (see section 1.3) has facilitated improved screening for coeliac disease, and a more accurate estimation of the true disease prevalence. Several studies have been carried out, using these anti-body screening protocols as diagnostic methods.

These studies have revealed the presence of coeliac disease in a wide variety of non-European populations, including African via Ema and AGA (Boudraa *et al.* 1996), Brazillian (Gangolfi *et al.* 2000), American (Hill *et al.* 2000), and Middle Eastern (Al-Tawaty & Elbargathy 1998). There is also evidence from intestinal biopsy to suggest that coeliac disease may be prevalent in Northern China, but that its presence may be masked by symptoms from other conditions common to the region, especially in the case of non classical presenters (Ling-ling Jiang *et al.* 2009).

In addition, studies carried out within European populations have suggested the true prevalence of coeliac disease to be ten-fold higher than previously expected, placing it at 1%.This figure is obtained from sampling large numbers of people from the population and testing for the presence of coeliac disease auto-antibodies (Fasano *et al.* 2003, Maki *et al.* 2003, Bingley *et al.* 2004).

Bingley *et al.* examined a population of 5470 children, aged 7 and a half years, using IGA-EMA and IGA-anti tTG antibody tests. Based on the IGA-EMA, they detected 54

children with significantly elevated antibody levels (1%). Confirmation via biopsy was, however, not subsequently preformed. An additional 137 children were tTG antibody positive but IgA-EMA negative. It should be noted that only 40% of the 54 children reported three or more of the following gastrointestinal symptoms; stomach pain, diarrhoea, constipation or vomiting.

Maki *et al.* 2003 took samples from 3654 students (age range, 7 to 16 years) in 1994 and screened in 2001 for endomysial and tissue transglutaminase antibodies. HLA typing was also performed on stored blood samples. All antibody-positive subjects were asked to undergo small-bowel biopsy in 2001. The subjects were tested for both IgA and IgG EMA and tTG antibodies. The study yielded 50 subjects Positive for IgA-class endomysial and tissue transglutaminase antibodies, 1 Positive for IgA-class endomysial antibodies and negative for IgA-class tissue transglutaminase antibodies 2 Positive for IgA-class tissue transglutaminase antibodies and negative for IgA-class endomysial antibodies 3 Negative for IgA-class endomysial and tissue transglutaminase antibodies but positive for IgG-class endomysial and tissue transglutaminase antibodies. Following anti-body screening, diagnosis was confirmed via biopsy in 37 of the patients, giving a population prevalence of 1.01%.

In the Fasano *et al.* study the serologic criteria for the diagnosis of CD were met in 27 of 2845 not-at-risk adults (1:105) and 4 of 1281 children (1:320). Of these 31 EMA-positive subjects, only 7 (30%) underwent small bowel biopsy. All biopsy specimens revealed a lesion consistent with CD.

Although all 3 of these studies produced similar estimates of coeliac disease in the population, the Finish study (Maki *et al.* 2003) was more rigorous, and insisted on biopsy for a definitive diagnosis. The American based studies were less rigorous, and accepted a positive diagnosis without direct evidence via biopsy. This suggests the population estimate in the American population may be inflated.

Despite this estimated prevalence of 1%, only approximately 0.14% of the UK population is diagnosed with coeliac disease (Van Heel *et al.* 2006). The ratio of diagnosed to undiagnosed cases, in the UK, therefore stands at about 1:7. This ratio of diagnosed to undiagnosed is also consistent with what has been observed in Italy (Catassi *et al.* 1996). One of the chief causes of the misconceptions with regard to the prevalence of coeliac disease is due to the heterogeneity of symptoms seen in patients, as outlined in section 1.2. For a long time, children exhibiting severe, classical symptoms were easily diagnosed, while those children or adults with milder

or extra-intestinal symptoms failed to be properly identified as disease sufferers (Steele, 2011).

Also, asymptomatic, or 'silent' coeliac disease cases are an important aspect of the epidemiology of this disease. Individuals who display no detectable symptoms, or symptoms of a severity which would not lend them-selves to a positive diagnosis of the disease, would seem to out-number those with more overt symptoms. This idea is known as the Coeliac Iceberg, with the diagnosed or 'visible' portion of sufferers being likened to the visible portion of an Iceberg, with the majority, the undiagnosed cases, still lurking 'beneath the surface' as it were (Goddard & Gillett, 2006).

These individuals represent two particular categories of risk; one, classically asymptomatic individuals with complications of the sort outlined in section 1.2, who would benefit from elimination of gluten from the diet, and two; individuals with no detectable complications or disease, who are at risk of converting to an active disease state in later life.

A caveat should be mentioned when relying on these tests as a definitive estimate of the disease frequency however. The sensitivity and specificity of various antibody tests varies considerably, both between tests and within them, depending on the parameters of the experiment.

In dealing with a disease of relatively low prevalence, specificity becomes exceedingly important. While a test with a sensitivity of even 50% will fail to detect only half of the 'true' coeliacs, underestimating the 'true' frequency by 50%, an anti-body test with a specificity of even 99%, means that for every 100 individuals tested, 1 will erroneously be diagnosed with coeliac disease. If the 'true' population frequency were 1%, then a test with a specificity of 99% would increase the estimate to 2%, twice the 'true' frequency. It is useful to consider the statistical ramifications of study design when basing assumption of 'true' incidence on these results. Only EMA antibodies appear to give a high enough specificity to be reliable, and results from studies based on other antibody tests should be treated with due circumspection.

There is a definitive issue here; ascertaining if an individual has a form of coeliac disease which may be actively harmful to them while not presenting classically (osteomalacia, increased cancer risk etc.) is important, as is identifying predisposed individuals who are at risk for developing disease in later life. However, a distinction must be drawn between these two types of individual, and those who are not at risk,

but who have indicated otherwise due to insufficient specificity of an anti-body test. Failure to confirm a putative anti-body based diagnosis with a follow up endoscopy may artificially inflate the true prevalence of the disease.

### 1.5 Genetic Components of Coeliac Disease

Studies have revealed Coeliac Disease to be a complex genetic disorder; a type of genetic illness associated with the effects of multiple genetic factors in combination with environmental elements. Also referred to as polygenic, or multifactorial disease, many such disorders exist and include most, if not all, autoimmune diseases, cardiovascular disease and psychiatric disorders. Due to the profusion of contributory factors, complex disorders display increased incidence in family groups, but do not follow a clear pattern of inheritance in the manner of Mendelian Disorders.

Familial and twin studies have revealed coeliac disease to have a strong genetic component . The concordance of coeliac disease between monozygotic twins is very high, at least 75% (Greco *et al.* 2002), while the concordance rate between dizygotic twins is still high at 11% (Nistico *et al.* 2006). This strong heritability is also evident in family studies (Sollid *et al.* 2002), with the prevalence of CD among first-degree relatives being considerably higher than the prevalence of the disease in the general population; the risk of coeliac disease is approximately 10% for a sibling of an affected individual contrasted with the estimated 1% risk for the general population. The risk of CD in the members of nuclear families with two affected children is approximately three times higher than that when only one child is affected (Dolinsek *et al.* 2004; Gudjonsdottir *et al.* 2004).

By way of a comparison, the concordance rate for Crohn's disease (another inherited complex disease of the gastrointestinal tract) is 33% in monozygotic twins (van Heel *et al.* 2000). Concordance rates in most other autoimmune diseases are similarly low, with 25% in Multiple Sclerosis (Willer *et al.* 2003), 12.3% in Rheumatoid Arthritis (Aho *et al.* 1986) and 35% in Type 1 Diabetes (Olmos *et al.* 1988), underlying the particularly strong genetic component of coeliac diseFor coeliac disease, a sibling relative risk ratio of 30 has been reported (Bevan *et al.* 1999) higher than most other disorders, such as type 1 diabetes, rheumatoid arthritis and Crohn's disease (Lewis *et al.* 2007).

Unusually for a multifactorial disease, CD has a major genetic element which, alone, accounts for some 35% of its heritability. This element is located with the HLA, and is the subject of the following section.

## 1.6 HLA contribution to Coeliac Disease

The Human Leukocyte Anitgen complex is located on the short arm of chromosome 6, and encodes the genes which make up the human Major Histocompatability Complex. Roughly four megabases in size, this region encodes the class I and class II HLA molecules in addition to a number of other genes in the MHC class III region, many of which are involved in immune function (Shiina *et al.* 2004).

HLA class I cell membrane molecules (A, B and C) present cell derived peptides (average 9 amino acids in length) from within endogenous cells, along with viral peptides. HLA class II cell membrane molecules (DP, DOA,DOB, DQ, and DR)are expressed on antigen presenting cells to present exogenous antigens to lymphocytes. This results in T Helper Cell activation, and a B-Cell mediated antibody response. The MHC class III region encodes genes many of which are strongly involved in the immune system, including components of the complement system, TNFa and related genes, and a variety of other genes involved in cell signalling and metabolism (Shiina *et al.* 2004).

Linkage to this region is typically found in autoimmune and inflammatory diseases; type 1 diabetes and rheumatic diseases both evidence strong association with this area (Dausset *et al.* 1976, Stenszky *et al.* 1978). However, regional linkage and causative variant are two very different things. The HLA complex displays very strong linkage disequilibrium (LD, see section 1.9); strong LD of this sort can confound fine mapping of disease associated genes, as it is difficult to disentangle marker from true causative variant.

The initial association with Coeliac Disease and the HLA occurred in the 1970s, with an association to the then utilised 'HLA serotype' method of tissue typing. This method assigns codes to HLA types based on antibody recognition of expressed HLA antigens. The HLA serotypes, HLAB8 (originally named HLA-A8) and HLA-DR3 (originally HLA-DW3) were found to be associated with increased risk of coeliac disease development (Falchuk *et al.* 1972, Keuning *et al.* 1976).

Further studies revealed that the true causal variant was in fact the HLA-DQ2 serotype. HLA-DR3, HLA-B8 and HLA-DQ2 are in strong linkage disequilibrium and are inherited on the same ancestral haplotype (Price *et al.* 1999).

The DQ2 serotype consists in turn of two distinct haplotypes; DQ2.2 and DQ2.5. DQ2.2 does not appear to confer any additional risk with regard to CD, with DQ2.5 appearing to be entirely responsible for the HLA contribution to the disease (Sollid *et al.* 1989).

The HLA-DQ2.5 haplotype contains the specific alleles of the genes which code for the alpha and beta chain subunits of the HLA DQ2 class II molecule. The alleles of the DQ2.5 haplotype are DQA1*0501 (alpha chain) and DQB1*0201 (beta chain), with HLA-DQ2.2 having the alleles DQA1*0201 and DQB1*0202. Thus the DQ2.5 haplotype in itself carries all of the requisite components to generate the disease associated HLA DQ2 molecule on the same chromosome ( i.e. they are encoded in cis) In contrast, the DQ2.2 haplotype supplies only one component- the beta chain encoding allele DQB1*0202- and the remaining allele corresponding to an appropriate alpha chain allele, usually DQA1*0501, or the closely related *0505, needs to be supplied in trans, with each parent thus supplying one chain (Sollid *et al.* 2002).

The HLA-DQ2 molecule's association with CD, encoded as described above, by specific alleles of HLA-DQA1 and HLA-DQB1, is one of the strongest HLA associations among the HLA-associated diseases. In most populations, over 90% of CD patients carry the DQ2 heterodimer encoded by alleles DQA1*0501 and DQB1*0201 (Sollid *et al.* 2000).

The remaining patients tend to possess a related DQ8 serotype, containing the risk allelic combination of DQA1*0301-DQB1*0302, or at least one chain of the DQ2 heterodimer (e.g. DQA1*0501 or DQB1*0201 but not both. A very small proportion of patients are both DQ2 and DQ8 negative, and they have been found to carry one of the two specific alleles coding for DQ2.5 That the risks associated with the different HLA susceptibility alleles differ between populations have been shown (Margaritte-Jeannin *et al.* 2004) as well as a dose dependence of the DQB1*02 allele (Ploski *et al.* 1993, Louka *et al.* 2003, Vader *et al.* 2003).

Therefore, possession of a HLA haplotype which can express either the HLA-DQ2 or HLA-DQ8 heterodimer is required in almost all cases for presentation of the disease -

causing dietary antigens to T cells, which initiates the primary immune response leading to the coeliac phenotype (Falchuk *et al.* 1972, Keuning *et al.* 1976, Tosi *et al.* 1983).

The heterodimers present the peptides to CD4+ T Cells, which in turn release a cytokine cascade which mediates the immune response (van Belzen *et al.* 2004). However, these haplotypes are common within the general population (roughly 30%), making the possession of these variants necessary, but not sufficient to induce the disease (Michalski *et al.* 1996). It is estimated that the HLA-DQ locus accounts for some 35% of heritability, with non-HLA genetic factors making up the majority of heritable risk. The heritability of Coeliac Disease is currently estimated at 50%, and approximately 40% of this genetic contribution has been identified (Dubois *et al.* 2010).

As can be seen in **Figure 1.2,** the alpha and beta chain subunits encoded by the genes on the aforementioned haplotypes, form a heterodimer. This heterodimer is a class II MHC molecule, and its function is to present exogenously derived antigens to T cell receptors. In the case of coeliac disease, antigen presenting cells present gluten peptides derived from the gut lumen to T cells, thus eliciting an immune response.

The heterodimers formed by DQ2.5 and DQ2.2 haplotypes differ in their amino acid sequence by one residue, with respect to their beta chains. This residue alters the peptide binding affinity for gluten-derived peptides (Van de Wal *et al.* 1997). The coeliac disease associated heterodimer, DQ2.5, binds a wider range of gliadin derived epitopes, with a much higher affinity than the non-associated DQ2.2 heterodimer (Vader et al. 2003, Qiao *et al.* 2005). However, when gluten peptides do bind to DQ2.2 they elicit an identical T cell response compared to DQ2.5 binding, indicating that the binding characteristics of HLA-DQ and gluten derived peptides are likely to be the primary mediator of the genetic association between HLA-DQ2.5 and coeliac disease. The increased presentation of gluten epitopes to T cells by DQ2.5 homozygotes explains the observed gene dosage effect, due to the increased density of presenting molecules on the cell's surface (Vader *et al.* 2003).

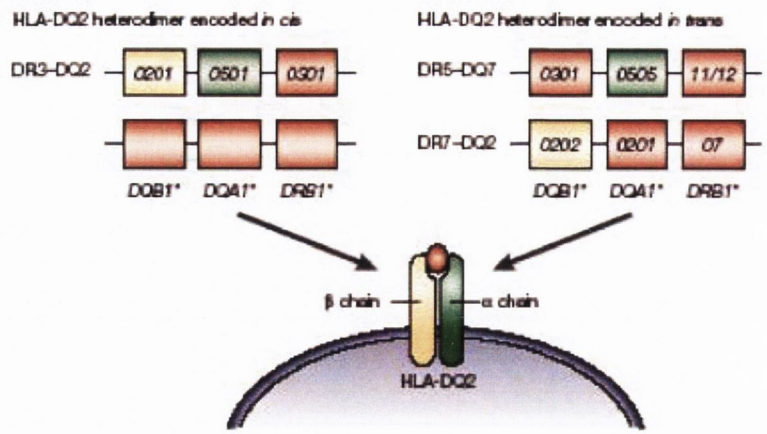



**Figure 1.2 HLA-DQ2 heterodimer formation. The HLA-DQ2 (DQA1*05–DQB1*02) molecule is encoded either in cis (DR3-DQ3 haplotype), or in trans in(DR5–DQ7 and DR7–DQ2 heterozygous).** Source: Sollid et *al.* 2002.

### 1.7 Gluten and the Immune Response

Although Gee suspected a dietary influence on coeliac disease, it was not until the middle of the 20th century that Wilhelm Dicke identified the ingestion of wheat flour as the environmental trigger for coeliac disease. Dicke observed a marked reduction in the incidence of so called Coeliac sprue during World War II, which corresponded to a reduction in availability of cereals and bread (Dicke *et al.* 1953).

Since then, the immunogenic portion of wheat has been well characterised. Gluten, found in the endosperm of cereals including wheat, barley and rye, consists of various storage proteins. Composed of distinct portions of monomeric, alcohol soluble gliadins (responsible for the elasticity of dough), and polymeric, prolamine rich glutenins (Di Sabatino & Corazza, 2009) gluten proteins have a complex chemistry and are responsible for the baking properties of wheat; water absorption capacity, cohesivity, viscosity, and dough elasticity (Wieser, 2007).

Analysis of the structure of gliadin has identified more than a hundred components that can be grouped into four main types (ω5-, ω1,2-, α/β-, γ-gliadins), with multiple gliadin-dervied epitopes evidencing immunogenicity and toxicity in CD sufferers (Kagnoff , 2007). Storage proteins (prolamines) with a similar amino acid composition to the gliadin fractions of wheat have been identified in barley (hordeins) and rye

(secalines) and show a close relation to the taxonomy and toxic properties of wheat cereal that affect people with coeliac disease (Wieser, 1995).

Although multiple gluten-derived epitopes are immuno stimulatory, some are more active than others. A peptide of 33 aminoacids (residues 57–89) identified from an α-gliadin fraction appears to be immuno dominant, and has properties attributable to its many proline and glutamine rich peptides. The density of proline residues gives the peptide increased resistance to gastrointestinal proteolysis (in people with and without coeliac disease), and causes a left-handed helical conformation, which strengthens binding with HLA-DQ2 and HLA-DQ8 molecules on antigen-presenting cells. Additionally, glutamine residues are a preferred substrate for tissuetransglutaminase- mediated deamidation, which confers an enhanced immunogenicity (Shan *et al.* 2002).

This enhanced immunogenicity is due to the peptide-binding groove structure of DQ2 and DQ8 dimers (Tollefsen *et al.* 2006). The criteria for epitopes binding to DQ2 or DQ8 are well recognised with both DQ2 and DQ8 dimers exhibiting preferences for negatively charged residues at positions within the core peptide-binding groove (Vartdal *et al.* 1996). Such negatively charged residues are not naturally found within gluten derived peptide sequences, however, deamidation of existing glutamine residues within these peptides to glutamate results in suitably negatively charged residues, with a corresponding increase in immunogenicity (Qiao *et al.* 2005).

The endogenous enzyme Tissue Transglutaminase, originally implicated in coeliac disease due to the presence of auto-antibodies, is the catalyst for the deamidation and subsequent increase in epitope affinity for the DQ2 and DQ8 binding sites. For the many gluten derived epitopes that have been identified as putative HLA-DQ2 binding molecules, deamidation by TG2 appears to be a pre-requisite for binding and antigen presentation (van deWal *et al.* 1998).

Tellingly, tTG is highly expressed in the small intestine, particularly in the lamina propria, and up-regulated in inflammation (Fleckenstein *et al.* 2002). A secreted protein, tTG binds with very high affinity to the Extra Cellular Matrix, particularly fibronectin, and has been implicated in cross linking gliadin peptides, TG2, and extra-cellular proteins, acting to localise the gliadin to the lamina propria (Griffin *et al.* 2002, Dieterich *et al.* 2006).

## 1.8 Non-HLA genetic factors and Coeliac Disease

The integral role of the DQ2 heterodimer in Coeliac Disease predisposition has already been discussed, however, the DQ2 heterodimer is present at a frequency of between 20-30% in the Caucasian population. If we accept the frequency of Coeliac Disease at 1%, then only 3-5% of individuals with the predisposing heterodimer go on to develop the disease phenotype. As the concordance between identical twins is 75%, HLA identical siblings is 20%, and non HLA identical twins is 11%. This further confirms the important, but not exclusive role of the HLA in genetic predisposition.

As stated above, it has been estimated, that the contribution from the HLA comprises some 35% of the hereditary component of the disease. The remaining 65% of the hereditary component then, likely lies elsewhere in the genome, including possibly elsewhere in the HLA. Given the nature of the disease, genes which code for proteins involved in the immune system and inflammatory response, the structure and function of the small intestine, and the processing of exogenous antigens would appear to be most likely functional candidates for investigation.

## 1.9 Linkage Disequilibrium

With an estimated 3 billion base pairs in the human genome, discovering a single SNP, or even a group of SNPs, which contributes to disease status may seem an insurmountable task. Fortunately, not all genetic markers act according to the principle of independent assortment, and many SNPs co-segregate with one or more other SNPs.

Linkage Disequilibrium (LD) is the non-random association of alleles at two or more loci (which may or may not be on the same chromosome); some combinations of alleles occur more or less frequently than would be expected from a random formation of haplotypes based on observed allele frequencies. The degree of non-random association between polymorphisms is measured by the LD statistic, which is the difference between observed and expected allelic frequencies (assuming random distributions).

There are two statistics which are commonly used to quantify the degree of LD between two genetic markers, the D' and $r^2$ statistics. D' is a normalised version of D, a statistic which represents the difference in probability of observing two alleles in

the same haplotype, and the probability of observing them independently. D is derived as follows:

$D = f(A1B1)-f(A1)(B1)$

Where f(A1B1) is the observed frequency of the two alleles in the one haplotype, and f(A1)(B1) is the expected value of observing them independently. As is immediately apparent, whether D is positive or negative is irrelevant, only its absolute magnitude is relevant when determining the degree of LD. In order to avoid confusion, D is routinely normalised to D' via dividing it by $D_{Max}$ if positive, and $D_{Min}$ if negative ($D_{max}$ is given by the smaller of p1q2 and p2q1. $D_{min}$ is given by the larger of – p1q1 and – p2q2). This is due to the fact that the extent of D is highly influenced by the frequencies of the constituent alleles.

This results in a value between 0 and 1, with 0 representing completely independent assortment between the two loci, and 1 dictating total linkage, a situation in which the two loci invariably co-segregate. D' as a statistic is unconnected with allele frequency, and can indicate a high degree of linkage even in cases where allele frequencies at different loci on a haplotype are drastically different. For example, two alleles may have a D' = 1, even though one (a) has a frequency of 0.6 and the other (b) has a frequency of 0.4. In this scenario, b is fully in LD with a, however a is not always linked to b, since a is more frequent.

In order to take in account differences in allele frequency, the $r^2$ statistic can be utilised. $r^2$ is derived from r, an expression which represents the correlation coefficient between a pair of loci. r is derived as follows:

$r = D/\sqrt{p1p2q1q2}$

With $r^2$ being:

$r^2 = D^2/(p1p2q1q2)$

This statistic takes into account allele frequency similarity (or lack therof) when reporting the degree of LD. Thus, two alleles with an $r^2 = 1$ are in complete LD and have identical frequencies.

The LD between loci is due to shared ancestry, and the degree of LD observed can be modified by a variety of historical factors; recombination, natural selection,

mutation, drift, non-random mating, and any other factors which influence population equilibrium.

If two or more markers are in strong LD, as evidenced by a D' or r2 value of close to 1, then they exhibit a useful property which can be exploited by a researcher when performing an association study. The genotype at one of the linked loci becomes informative for the genotypes at the other linked loci. Once one such locus is genotyped, the genotypes at the other loci in LD can be inferred with a reasonable expectation of accuracy (depending on the level of LD and accurate knowledge of the populations involved).

This means that rather than having to genotype all existing variation in a genomic region of interest, a subset of SNPs can be selected to act as proxies for markers with which they have a high degree of LD, vastly increasing their informativeness, and the chance of identifying a causal variant in LD with a proxy SNP.

### 1.10 Haplotype Analysis

A region of a chromatid which contains several SNPs, all of which are in LD, is referred to as a Haplotype (so called because it is in effect a series of haploid genotypes from a region of one of the parental chromosomes). Haplotypes are arranged within the human genome in what has come to be referred to as 'Haplotype Blocks', essentially a series of regions of high LD (the blocks) which are separated by short regions of comparatively low LD. Due to allele frequencies, and the high LD between SNPs in a block, all possible haplotypes will not be found within a population, some combinations will simply not exist, or will be present at a statistically negligible frequency.

Putting the concept of LD together with what is known about the structure of the human genome in terms of haplotype blocks, we arrive at the following conclusion; within a haplotype block, assuming sufficient information, a subset of the existing SNPs can be genotyped and used to infer which of the limited set of haplotypes an individual possesses. The genotypes of the remaining polymorphisms at other loci within the haplotype block can then be themselves inferred. The degree to which this inference is accurate is based on the degree of LD between the proxy SNP or SNPs and the SNPs to which they are linked, as well as the accuracy of the information

used to establish the initial population conditions. These proxy SNPs are termed 'Haplotype Tagging SNPs' (htSNPs).

Identifying these htSNPs is typically done as follows; the entire panel of SNPs within a given haplotype are genotyped in a number of controls in order to ascertain the patterns of LD between the SNPs, and which of the possible haplotypes are present within the population. From this, a subset of SNPs which will unambiguously identify individual existing haplotypes can be selected.

**Figure 1.3** Illustrates how this process is carried out. In this instance, genotyping of a control panel for SNPs within the selected Haplotype block indicates only 4 of the possible haplotypes are present at any appreciable frequency within the population. The information provided at only 3 loci is sufficient to establish which of the 4 haplotypes an individual possesses.

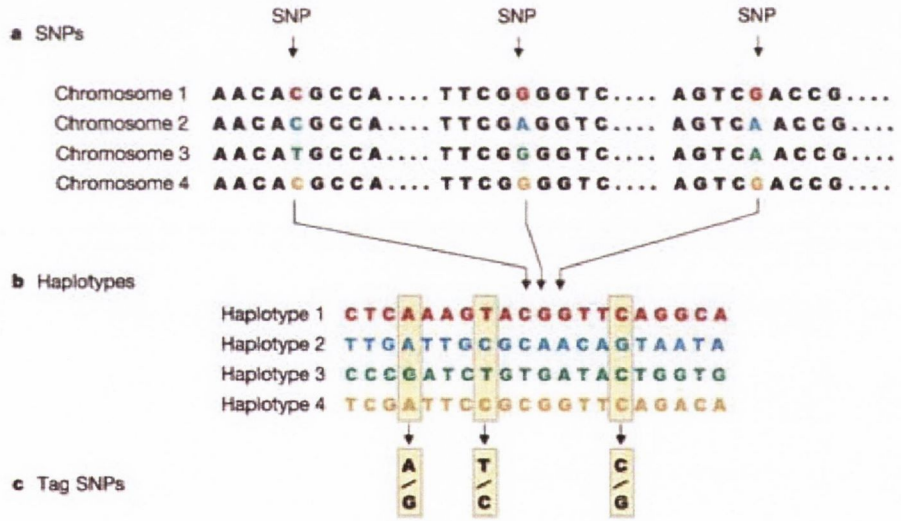



**Figure 1.3 Selection of htSNPs from a limited population pool of Haplotypes. Source: International HapMap Project website.**

When selecting htSNPS for a region, a researcher can either examine an area one haplotype block at a time, or simply ignore underlying block structure, and identify the SNPs which are most informative for the region of interest. For example, an exon of a candidate gene may comprise two distinct haplotype blocks. The first approach would treat each block separately, the second would simply look at the region as a whole.

Both approaches have merits; the block by block approach requires prior knowledge of the haplotype block structure of the region, but produces a panel of SNPs that are better equipped to deal with regions of low LD and/or missing data. The region approach typically generates a smaller panel of SNPs required to tag the area, but will result in increased error in the inference of haplotypes when difficult conditions of the types described above occur (Forton *et al.* 2005).

### 1.11 Discovering the Genetic Basis of Complex Diseases

An estimated 35% of the heritability of coeliac disease has been attributed to the necessary, but not sufficient, contribution of the epitope-presenting DQ2 and DQ8 molecules, leaving the majority of the contributing genetic factors to be elucidated.

In attempting to discern genetic factors contributing to inherited disease two fundamental approaches may be taken; genome-wide studies, and the examination of functional candidate genes (genes which due to their known biological effects are likely to be involved in similarly known, or suspected, mechanisms of disease).

Likewise, there are two methods of tying variants to diseases; family based studies, using linkage analysis to identify a chromosomal region more often inherited by disease sufferers within the family being examined, and studies using cohorts of unrelated individuals, which employ association analysis to identify an inherited allelic variant from a common ancestor which contributes to disease susceptibility (founder effect) (Zlotogora, 1994).

Regardless of which of the two approaches is implemented, searching for causative factors is usually informed by the 'common disease, common variant' hypothesis. This essentially posits that a disease which is common within a population will be caused by a combination of factors which occur with relative frequency within the population. Generally this is taken to mean allelic variants that occur at with at least 5% frequency (Becker, 2004).

There has been much success in elucidating the genetic factors underlying Mendelian disease; however diseases which evidence complex inheritance have been more of a mixed bag. In the complex inheritance model two or more (and in some cases a lot more) genetic factors contribute to varying degrees to the disease state.

In these instances, it can be difficult to identify the multiple loci responsible for disease causation. Regardless of which approach is used, diseases with complex genetic inheritance show a relatively weak relationship between a variant influencing the disease and the disease phenotype when compared to Mendelian Disease. This is due to there being many such factors, each imparting a relatively small overall effect, as opposed to the single, large effect variant present in Mendelian inheritance. Elementary power calculations indicate that in order to have the power to detect effects of lower magnitude, sample sizes must increase; in the case of a disease influenced by many elements of modest effect, having a sample population of sufficient size becomes a key issue.

Quite aside from the possibility of having to hunt down a plethora of low effect loci, a disease phenotype which arises from multiple interactions between disparate factors introduces other complications. A disease-causing variant may be present in a control population at high frequency, harmless without the interactions of other key factors. This high rate of incidence in the general population may make it challenging to prove an association with disease susceptibility.

Epistatic gene interactions can muddy the waters, as control populations can evidence one or more of the disease predisposing variants, but not in conjunction with others required to cause the disease state, leading to the false assumption that these loci are neutral with regard to disease susceptibility if they are analysed in isolation. A disease may also exhibit genetic heterogeneity; a situation where different combinations of predisposing factors is capable of causing the disease. In this instance, the number and nature of predisposing factors can vary from individual to individual, resulting in inconsistent associations, and possibly differences in the presentation of the disease.

Variants between populations may also play a part. Locus or allelic heterogeneity refers to regional or ethnic differences may result in genes, genotypes or alleles contributing differently to the disease phenotype in different populations, making replication of observed associations difficult between populations.

### 1.12 Genome Scans and Coeliac Disease

Genome scans are a method for establishing linkage to disease with a broad chromosomal region (5q31 for example) as opposed to a specific SNP. In the past,

this has largely been achieved by genotyping an array of genetic markers known as microsatellites. These are repeated sequences of a 1-6 base pair motif, for example (CAG)n, where 'n' is the number of tandem repeats. Allelic variability comes from individuals possessing differing numbers of tandem repeats (Buchanan & Thue, 1998). More recently, SNPs have replaced microsatellites as the genetic marker of choice.

Once linkage has been established to a broad genomic region a follow-up study is usually performed in order to refine the observed linkage peak to more specific sub-region. Increased microsatellite marker density around the region, or a fine mapping approach using SNPs, and more recently resequencing of the region, are all common approaches.

Multiple genome scans in coeliac disease have been performed in European populations, and two in the North American population, in order to identify susceptibility genes outside the HLA region. A summary of these scans is detailed in **Table 1.3** including data from a Meta-Analysis involving four of the populations.

**Table 1.3 Coeliac Genome Scans compiled from studies carried out in multiple European populations over a seven year period, showing linkage to chromosomes 5 and 19.**

| Year and Population | Number of Families | Type of Family | Suggestive Linkage $P < 7x10^{-4}$ | *Significant Linkage* $P < 2x10^{-5}$ |
|---|---|---|---|---|
| 1996 Irish | 15 | Multi-Affected | 6p23, 11p11 | |
| 1998, 2000 **Italian** | 39/71/89 | Sib-Pairs | 5q32-34 | |
| **2000, 2001 British** | 16/34 | Multi-Affected | | |
| **2001 Scandinavian** | 70/36 | Sib-Pairs | | |
| **2002 Finnish** | 60/38 | Sib-Pairs | | |
| 2002 Finnish | 9/1 | Large Extended Families | 15q12 | |
| 2002 Northern European | 24 | Multi-Affected | | |
| 2002 North American | 62 | Sib-Pairs | 3p26 | |
| 2002 Dutch | 67/15 | Sib-Pairs | 6q21-22 | *19p13.1* |
| 2004 Finnish | 54 | Sib-Pairs | | |
| 2004 Dutch | 1 | Large Extended Family | 9p21-13 | |
| 2007 North American | 160 | Multi-Affected | | |
| 2003 Meta Analysis of Scandinavian, Italian, Finnish and British (in **bold**) | | | | *5q31-33* |

The table details the number of families that were included in each study, the type of relationship evidenced between affected members of these families, and the broad genomic regions which evidenced linkage to coeliac disease in each study.

As can be observed from the above table, only one of the population studies, in the Dutch population, provided a significant linkage score ($P < 2x10^{-5}$) and there is little in the way of overlap between suggestive linkage peaks across populations.

When assessing the data obtained across these populations, study structure and experimental design must be taken into consideration. There is substantial heterogeneity between the number and types of families analysed in these studies, ranging from only a few, or even one single large pedigree, to analysis of affected sib-pairs.

Also, if related sib-pairs are analysed and not only independent ones, the linkage statistic can be inflated if it is not corrected for. Another reason for discrepancies is the use of different statistical methods. Some genome-wide scans have used non-parametric linkage (NPL), others have analysed the data using several different parameters and LOD scores or heterogeneity LOD scores (HLOD). Testing of different models and using different types of families (multi-affected pedigrees versus sib-pairs), makes the comparison between studies difficult.

The vast majority of these genome scans have failed to provide consistent, replicable linkages when compared with one another, however. The conflicting results may be due in part to the heterogeneity of family structures, and also in part to small sample sizes lacking the statistical power to detect a contributory variant with relatively small effect.

A meta analysis of the Scandinavian, Italian, Finnish and British studies indicated a significant linkage peak in the 5q31-33 region. The raw data consisted of genotypes from 2025 individuals in 442 families; 1056 of whom were affected. The results pointed to chromosome 5q31–33 as being the only significant locus in these families apart from the HLA.

Interestingly, the families from the United Kingdom showed negative linkage scores in this region. This could be a real result, or could be caused by random variation due to the comparatively small number of families and/or locus heterogeneity. These families also have a somewhat different makeup, being large multigeneration pedigrees with many affected individuals. The three remaining data sets consisted of small nuclear families with typically 2–3 affected individuals. It is possible that many unrelated nuclear families favour the detection of a more common susceptibility gene shared by a large portion of CD patients, while large pedigrees with many affected individuals could be the result of a gene variant, which is relatively rare but with stronger penetrance. This would illustrate how study design could influence the outcome of the results.

In 5q, a large fine-mapping study failed to single out a strong susceptibility candidate (Amundsen *et al.* 2007). There were several genes showing nominal association in this Norwegian/Swedish family sample but no single association could explain the large linkage peak found in the same sample (Amundsen *et al.* 2007). Unless, there are regions inadequately covered due to, for example, copy number variation, these findings could suggest that there is too much allelic heterogeneity or too many founder haplotypes to pin a gene down.

Fine-mapping of the chromosome 19 region showed the MYOIXB gene associate with coeliac disease in two different CD cohorts (Van Belzen *et al.* 2003). It is a gene likely to be involved in Rho-dependent signalling pathways and remodelling of the cytoskeleton and tight junction assembly. This seems to fit very well with the possibility of CD being caused by a more permeable intestinal barrier. However, attempts to replicate this finding have been made with little success. So far only one other group has been able to replicate this finding, while numerous others have failed (Amundsen *et al.* 2006, Giordano *et al.* 2006, Hunt *et al.* 2006, Nunez *et al.* 2006, Cirillo *et al.* 2007).

### 1.13 Candidate Gene Approaches to Complex Disease

Functional candidate genes are genes which due to their known biological effects are likely to be involved in similarly known, or suspected, mechanisms of disease. Selecting functional candidate genes as targets for genotyping is a method of refining a genomic search to focus on genes which are more likely to prove to be the causal variant, based on *a priori* knowledge of their function. This may occur when an association has been demonstrated in one gene, and candidate genes are selected due to their known pathway interactions with it, or in the case of a linkage study, as described in the previous section, a genomic region may evidence disease linkage. The size of this region, however, can be megabases in length, and exhaustive genotyping of markers within every gene in the region may represent a very large undertaking. In this case, a subset of candidate genes may be selected from the region for analysis on the basis of their known functions. The difficulty with this approach is, of course, that the selection of appropriate candidates is based on incomplete data with regard to gene function and interaction, as well as assumptions with regard to pathway contribution.

Functional candidate gene approaches have been undertaken with regard to coeliac disease, but with only very limited success. Although some initial findings have been promising, replicating these associations in subsequent populations has proved inconsistent at best.

In the case of chromosome 5, the 5q region which evidenced linkage has been the focus of a several candidate gene driven studies. Association with any has yet to be determined.

Similarly, the linkage peak observed on chromosome 19 in the Dutch family based study was the focus of a candidate gene based follow up study. MYOIXB, encoding the myosin IXB protein, emerged as a candidate gene from studies of the 19p13.1 linkage region in the Dutch population (Monsuur *et al.* 2005). Despite the promising association initially reported, studies in the British, Norwegian, Italian and Irish populations failed to replicate the effect, as mentioned above (Amundsen *et al.* 2006, Hunt *et al.* 2006, Cirillo *et al.* 2007). Therefore, prior to the first GWAS in 2007, despite intensive efforts, no genetic susceptibility loci other than HLA DQ had been definitively identified.

Other potential candidate genes on chromosome 19 include the intercellular adhesion molecule-1 precursor (*ICAM-1*) and *CD209* both located in chromosomal band 19p13.2. A study of 489 CD patients from 257 families recently suggested that a functional variant of the *CD209* promoter is associated with DQ2-negative CD patients (Nunez *et al.* 2006).

A candidate gene approach in an Irish Case-Control study identified an association with a haplotype of the *CTLA4* gene, as well as a weaker association with an extended haplotype encompassing the *CD28-CTLA4- 5' ICOS* genes. *CTLA4* is a gene on chromosome 2q encoding the cytotoxic T lymphocyte antigen 4. CTLA-4 is expressed on T cells and is a receptor for B7 molecules that inhibit T cell activation (Brophy *et al.* 2006) Replication studies of the CTLA4 association have been somewhat inconclusive however, with a British study identifying a different associated haplotype (Hunt *et al.* 2005), a Finnish study finding no significant association (Haimila *et al.* 2004) and a Norwegian/Swedish study suggesting several independent loci in the region were contributing to the effect (Amundsen *et al.* 2004).

CTLA4 acts to maintain tolerance to self-antigens, functioning as a negative regulator of T-cell proliferation through the co-stimulatory signal, and as an inducer of clonal

anergy (Eagar *et al.* 2002). This gene has been under investigation in a number of autoimmune conditions. It is therefore an interesting candidate gene for CD. So far no evidence has come forward showing the involvement of a certain functional genetic variant of this gene in CD. However, there are many reports indicating the involvement of CTLA4 in CD.

The relative risk of CTLA4 in diabetes is estimated to be as low as 1.299, and if the same is true for CD, almost all studies in CD have too little power to be able to detect association. Still, some linkage to the region has been reported in a few studies (Holopainen *et al.* 1999, Naluai *et al.*, 2001, Rioux *et al.* 2004) and most published studies in CD do show signs of association as well. However, they are not completely congruent.

Different studies show association to different genetic variants or haplotypes, while some studies fail to show any association. Other studies with negative results may well remain unpublished. Nevertheless, a negative result is not surprising if the relative risk is low and study samples are small. Different haplotypes showing association could be a result of allelic heterogeneity between and within populations but also of that CTLA4 can have different impacts on disease sub-phenotypes such as disease onset and severity. Also, different markers were used in many instances making comparisons difficult.

If one study includes many silent cases amongst their cases, and another includes classical cases, this could be a reason for the different associations shown. Genotypes at HLA have also been suggested to influence the outcome of susceptibility to alleles in CTLA4 (Naluai *et al.* 2001, Zhernakova *et al.* 2005).

### 1.14 International Hap Map Project

Regardless of approach, the genotyping of a representative control group in order to determine patterns of LD, and which subsets of possible haplotypes exist within a given population is required *a priori*. In order to provide a public database detailing large, control groups for several populations, with SNPs covering the entire human genome, the International HapMap Project was established in 2002 (The International HapMap Consortium, 2005).

Four populations were selected for inclusion: 30 trios from Nigeria (YRI), 30 trios of U.S. residents of northern and western European descent (CEU), 45 unrelated Han Chinese individuals (CHB) and 44 unrelated individuals from Japan (JPT) (The International HapMap Consortium, 2005).

Phase I of the project had a marker coverage of one common SNP per 5,000 base pairs, with more than one million SNPs genotyped in total, and was made publically available in October 2005 (Goldstein & Cavalleri, 2005). Phase II of the project involved increasing marker density threefold, with over 3.1 million SNPs genotyped in these populations, and was released in October 2007. Phase III, released in October of 2009, did not focus on increasing marker density. Rather, it increased diversity, adding a number of additional populations to the database.

This data was made publically available on the HapMap website (see materials and methods for more information). This information greatly increased the ability of researchers to design haplotype tagging experiments for these populations, and as such increased the effectiveness of SNP based association studies. Having such a repository of readily available data on the LD structure of the populations, combined with advances in genotyping technology, also made possible a new era in linkage study; the Genome Wide Association Study.

### 1.15 GWAS and Complex Disease

Family based studies and candidate gene approaches have had limited success in elucidating the genetic underpinnings of coeliac disease, the reasons for which have been outlined above, lack of power being chief amongst them. Testing only a handful of markers at a time is also a slow process; the human genome contains millions of variations.

With the HapMap project providing detailed information regarding the patterns of linkage disequilibrium, and marker frequency, tagging SNPs could be readily identified with sufficient number and density to cover the entire human genome. In conjunction with advances in multiplex genotyping technology, a first generation of genotyping chip sets were produced, capable of simultaneously generating 100's of thousands of genotypes per individual. Subsequent generations of this technology have increased the number of genotypes capable of being generated simultaneously to several million.

WGA (Whole Genome Analysis) studies, then, generally seek to select a panel of markers with which to optimally tag variation (the exact parameters of this will vary based on the constraints and aims of a particular study) rather than seeking to select candidate genes or SNPs. Due to the nature of the genetics of complex disease, multiple variants of relatively small effect, WGA is uniquely suited as a tool for analysis.

The first GWAS study was carried out in 2005, and was successful in determining a disease causing variant in age related macular degeneration (Klein *et al.* 2005). As of this writing, 904 GWAS studies have been carried out, seeking to examine the genetic underpinnings of 165 separate traits (Taken from genome.gov, "A Catalog of Published Genome-Wide Association Studies").

GWAS comes with several distinct advantages; due to the level of genome wide coverage, a more holistic picture can be seen. Gene-gene interactions, or epistatic effects, are more easily detected, such as GAB2 and APOE in Alzheimer disease (Reiman *et al.* 2004). Haplotypes, or combinations of multiple causative SNPs within a single gene also much more evident, such as atrial fibrillation (Gudbjartsson *et al.* 2007).

GWAS also presents its own unique difficulty; interrogation of so many loci via statistical tests comes with a huge bias created by multiple testing, resulting in questions as to the approporiate level of stringency in acceptable levels of statistical significance and requirements for replication of findings (Hunter & Kraft, 2007). Because of this, especially when searching for multiple risk factors of relatively small effect (as are found in many complex diseases, such as coeliac disease) large populations become vital in carrying out a successful GWAS.

### 1.16 Complex Disease and Regulatory Elements

Genome wide association studies have led to a vast increase in the catalogue of genetic associations with complex diseases. Although the HLA coeliac risk alleles are functional, resulting in alterations to protein structure and function, this appears to be the exception rather than the rule. Only a minority of GWAS signals have to date been shown to involve coding variants, and while it is possible that the signals result from their linkage to currently unobserved coding variants, it seems more likely that

the GWAS associations are involved in more subtle regulatory effects on gene expression (Dubois *et al.* 2010).

Over half the variants significantly associated with coeliac disease have been correlated with expression changes in nearby genes (Dubois *et al.* 2010), suggesting that they, or linked surrogates, may play functional roles in controlling expression of these genes. Due to the commonality of factors in many complex diseases, this is likely to be applicable to other common, complex diseases.

These observations are concordant with current theories about complex disease genetics; several, perhaps very many, factors of relatively small effect, combining to produce a disease phenotype (Hindorff *et al.* 2009). Debate is ongoing regarding the source of missing genetic factors in complex diseases; in most cases, only a small proportion of assigned heritability is accounted for. Current studies still use dense genetic mapping to identify associated regions, but this will likely be superseded by DNA sequencing in the near future, allowing a more direct approach towards addressing the problems of causal variants and missing heritability, for example are rare variants important in these relatively common diseases? At the moment, evidence suggests regulatory gene variants are more common causes of disease susceptibility than coding variants.

### 1.17 Case-Control Studies

As mentioned in sections 1.12 and 1.13, genetic association studies take two forms; family based studies and case-control studies, with both study types having their own merits and drawbacks. Family based studies, typically comprising case-sibling pairs or case-parent trios, have the advantage of a common genetic lineage amongst the participants, thus avoiding any issues of population stratification. Additionally, families have a greater tendency to share environmental exposures, which may reduce confounding environmental factors.

However, crucially, sample size when working with family based studies is limited. The importance of sufficient power, and by extension, sufficient sample size, is paramount when assessing complex disease. Multiple variants contributing relatively small effects require a large sample size to detect, thus making the case-control approach the more suitable of the two study designs.

However, there are certain complications which come with choosing a case-control based study. Population stratification (systematic difference in allele frequencies between populations due to different ancestry) and admixture (interbreeding between individuals from two or more previously separated populations) are two possible confounding factors which must be considered.

Additionally, recruitment of an appropriate control group is paramount. Ideally, controls should be matched to the cases in such a way as to ensure they are both drawn from the same population, and that the controls represent the full spectrum of genetic diversity and environmental exposure evidenced in the case group.

There is also the double edged sword of recruitment from multiple centres, such as regional hospitals. On the one hand, this increases sample size, with a corresponding increase in power. On the other hand, inclusion of multiple regional subpopulations may increase the genetic and environmental heterogeneity of subjects, though this largely dependent on there being regional variation in population substructure.

Selection of controls is also a concern; in the case of coeliac disease, the estimated ratio of diagnosed to 'silent' coeliac disease is 1:7. As such, it must be assumed that the control population contains a proportion of 'silent' coeliac disease sufferers (roughly 0.86%).

### 1.18 Irish Population

When examining the Irish population in the course of genetic studies, a special consideration must be made. Broadly speaking, for the purposes of genetic and genomic research, the human race is sub-divided into three categories: African, Han Chinese and Caucasian. Different degrees of variation is present in each of these three populations, reflecting their historical legacy, with the African population evidencing the most variation, the Han Chinese the least, and the Caucasian population falling somewhere in between.

For the purposes of genetic research, the 'Caucasian' population is often taken to encompass a broad variety of sub-populations: European Americans, English, Irish, and much of mainland Europe. It has been assumed previously, that there is little to distinguish one sub-population from another on the genomic level, and that

assumptions and discoveries which hold true for one will hold equally true for another.

However, a recent paper has undertaken the sequencing of the entire genome of an anonymous Irish male, in an effort to ascertain to what degree, if any, systemic differences in the population structure exist between Ireland and mainland Europe/Britain/European Americans (Tong *et al.* 2010).

Although the paper has only sequenced the genome of a single Irish subject, the individual in question had been genotyped in previous association and population structure studies, including analysis of the genetic structure of 2,099 individuals from various Northern European countries. The data generated from these studies indicated that he is representative, genetically, of typical Irish sample. The individual in question is healthy, and is of Irish-Caucasian ancestry to at least three generations.

Comparison of the sequence data indicates that the sample is genetically distinct from other European populations. **Figure 1.4** illustrates principle component analysis of the Irish individual in relation to other Europeans. **Figure 1.5** is non- parametric population structure analysis, which shows the Irish individual relative to other sequenced genomes..

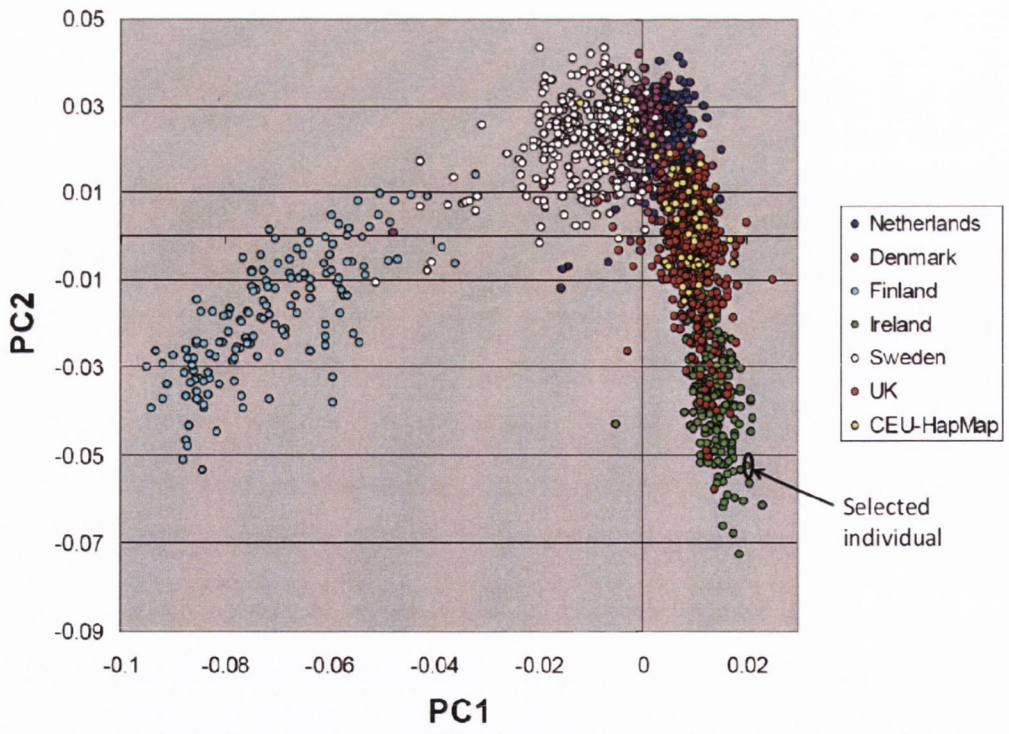



**Figure 1.4 Principal components analysis plot Taken from Tong *et al.* 2010. This figure shows asequenced Irish Individual with respect to other Western Europeans.**

The Irish individual can be seen to cluster near the edge of the European population, suggesting a differing genetic composition.

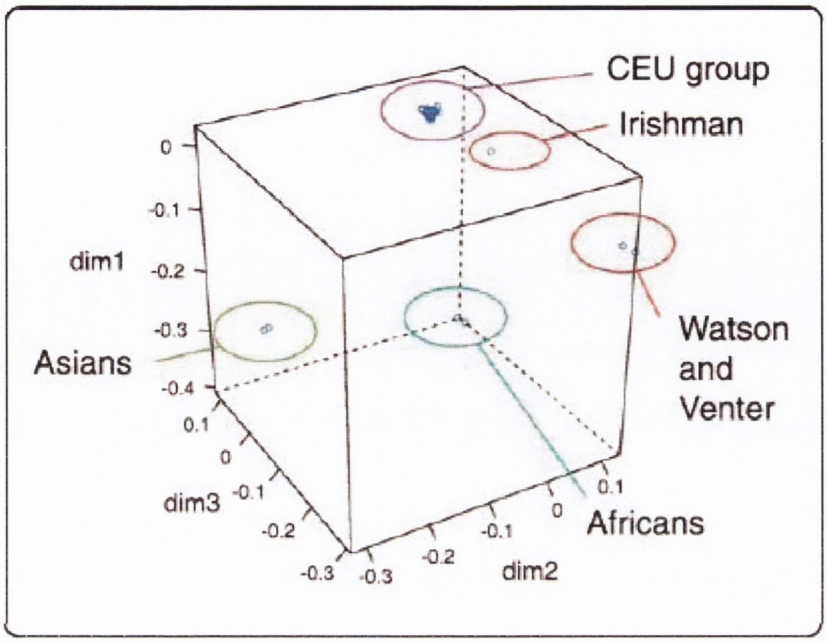



**Figure 1.5 Non-Parametric Analysis. Several distinct groups can be seen, with the European samples forming three distinct subgroups.**

Another important finding which is especially relevant when considering the Irish population in terms of genetic studies, is the heterogeneity of SNPs which were identified in the sequencing of this genome. 3,125,825 SNPs were identified, of which 13% were novel, and may include specific markers for Irish ancestry.

One of these SNPs was a non-synonymous nonsense mutation located in the macrophage-stimulating immune gene MST1, within 280bp of a nonsynonymous coding variant marker (rs3197999) that has been shown in several cohorts to be strongly associated with inflammatory bowel disease and primary sclerosing cholangitis. The implications for novel, auto-immune predisposing SNPs in the Irish population are wide ranging, and must be considered.

As these findings indicate that the Irish population is genetically distinct from other Caucasian sub-groups, and that it contains a suite of unique genetic markers, we must proceed with caution when directly comparing the results of Irish genetic association studies across populations.

Firstly, the linkage patterns may differ at certain loci; differing halpotype block structures and degrees of LD between SNPs may produce disparate results in

association studies. Secondly, we must consider pleiotropy, with differing genetic contributions to disease in the disparate populations.

Taking both differences in LD structure and pleiotropy into consideration, we must be wary when comparing genetic studies across populations, and also when engaged in meta analysis of multiple populations. It should be noted, that despite the documented differences, there is still much commonality between these populations, and a high degree of shared disease contributing factors can safely be assumed to exist. However, we must treat results unique to the Irish population with the consideration they deserve, rather than discarding them as artefacts. This is slightly complicated by the fact that the Irish population is smaller, and has a correspondingly smaller coeliac population, resulting in a smaller sample and reduced power.

### 1.19 Aims of this Study

This study is primarily concerned with elucidating genetic contributions to coeliac disease in the Irish population via case control based association studies. However, due to the nature of complex disease, chiefly that large sample sizes are required to detect genetic effects of relatively small magnitude, and the need for association confirmation in independent populations, samples from the Italian, British and Dutch populations are also the subject of this study's focus, to a lesser degree.

In attempting to ascertain these genetic contributions, we utilise a broad range of contemporary study designs, as outlined in this Chapter; candidate gene approaches, bioinformatic interrogation of public databases in order to identify potential regulatory elements, and genome wide association studies. Replication of observed associations in independent populations are also undertaken.

Due to the linkage observed with 5q31-33 and coeliac disease, via the meta analysis of several independent populations, special focus has been given to this region of the genome. A combination of genomic analysis to identify possible key regulatory sites, in conjunction with a candidate gene approach has been used to identify association of specific SNPs with coeliac disease (Chapter 3; Chromosome 5 & Coeliac Disease).

Utilising the resources of the International Hap Map Project, and taking advantage of the advances in multiplex PCR technology, we have taken part in several GWAS-

follow-up studies in conjunction with laboratories in Holland and Britain, analysing marker variants across the entire genome (Chapter 4; Coeliac Disease Genomewide Association Study Replication).

In keeping with current discoveries and better understanding of immunological pathways, several genes have also been targeted for candidate gene analysis throughout the genome (Chapter 5; Candidate Genes & Coeliac Disease).

Finally, we also seek to consolidate the observations we have made with regard to the genes involved in coeliac disease, both in the Irish population and across multiple independent populations, to determine if underlying pathways, or mechanisms of disease can be discerned from the associations our studies have evidenced (Chapter 6; General Discussion).

# CHAPTER II

# Materials and Methods

## 2.1 Coeliac Patient DNA collection

Blood samples from 650 Irish individuals suffering from coeliac disease were collected from the Department of Gastroenterology, Saint James' Hospital, Dublin; the University College Hospital, Galway; and the Adelaide and Meath Hospital, Tallaght. Patients were diagnosed based on histological markers present in biopsy (see **Section 1.3**) and symptom related response to a gluten free diet.

The Italian case population used in this study was collected by Professor Luigi Greco, and consisted of 400 Italian coeliac patients collected from the greater Naples area of Italy.

## 2.2 Control Subject collection

Early experiments as part of this study use DNA derived from Blood samples taken from 400 population matched healthy volunteer healthcare workers from the Trinity Centre for Molecular Medicine and Saint James Hospital as controls.

Later experiments include as controls, samples procured from the Trinity Biobank. This control population comprises DNA extracted from blood samples taken from healthy volunteers via the Irish Blood Transfusion Board.

The Italian control population used in this study was collected by Professor Luigi Greco, and consisted of 450 Italian women collected from the greater Naples area of Italy.

## 2.3 DNA extraction from whole blood

DNA was extracted from whole blood samples at the Trinity Biobank, using the Gentra automated DNA extraction system (Gentra Systems Inc MN, U.S.A.). At the end of the process, the DNA was stored in the supplied elution buffer (AE) at -20°C, or -80°C for long term storage.

## 2.4 DNA QC and quantification

DNA concentration was assayed using spectrophotometry on a Nanodrop (Thermo Fisher Scientific, DW, USA). 1µl of the extracted DNA was placed on the nandrop

pedestal, and optical density readings at 260nm and 280nm were taken to identify the concentration and purity of the extracted DNA.

In older samples, which had been subject to degradation over time, it was noted that the reading from the nanodrop was not always indicative of DNA quality. This is due to the degradation producing free nucleotides which the spectrophotometric assay identifies as DNA.

In order to assay the amount of double stranded DNA, and obtain a more accurate assessment of DNA quality, we began using the Qubit® Fluorometer based assays for historic samples (>5 years since extraction). Qubit® based assays use fluorescent probes which bind to dsDNA directly, and quantify the DNA by the amount of fluorescence emitted.

A dye Working Solution was prepared in an eppendorf ; 200 µL of buffer and 1 µL of dye for every sample, which was then mixed by vortexing. 190 µL of the working solution was pipette into two assay tubes for standards, and 10 µL of each standard added to an assay tube and mixed by vortexing. 195µL of Working Solution was then alliquoted into assay tubes for samples, and 5µL of each sample added to the assay tube and mixed by vortexing. Samples were incubated for 2 minutes, and the results read on the Qubit® Fluorometer.

### 2.5 Whole Genome Amplification

Due to the Irish Coeliac collection having been compiled over the course of many years, several of the historic samples (>8 years) were found to be of low concentration and/or volume. As it was not practical to re-extract DNA from all of the contributing patients over a short period of time, Whole Genome Amplification was used to increase the amount of DNA available for these samples.

Whole Genome Amplification was carried out on 270 samples, using the Quiagen REPLI-g Midi Kit (Quiagen, West Sussex, UK). Amount of initial DNA used per reaction varied by sample, but was at least 10µg, resulting in a typical yield of 40 µg per reaction.

## 2.6 Power Calculations

The Genetic Power Calculator (Purcell et al. 2003) was used to determine the power of the tests carried out as part of this study. Assumptions were as follows; a disease allele frequency of 10%, disease prevalence of 1%, $\alpha = 0.05$, and a variable estimated genotype relative risk. Effective power was calculated for each case/control group, and the number of cases required to provide 80% power was also calculated.

## 2.7 ECR Browser

http://ecrbrowser.dcode.org/

Evolutionary Conserved Regions (ECRs) are orthologous, intergenic or intronic regions which display high levels of sequence homology between species. ECRs that have been mapped within alignments of the genomes contained in the browser database are presented graphically. The ECRs are colour coded, according to their position in relation to known genes in the reference genome, as shown in **Figure 2.1.**

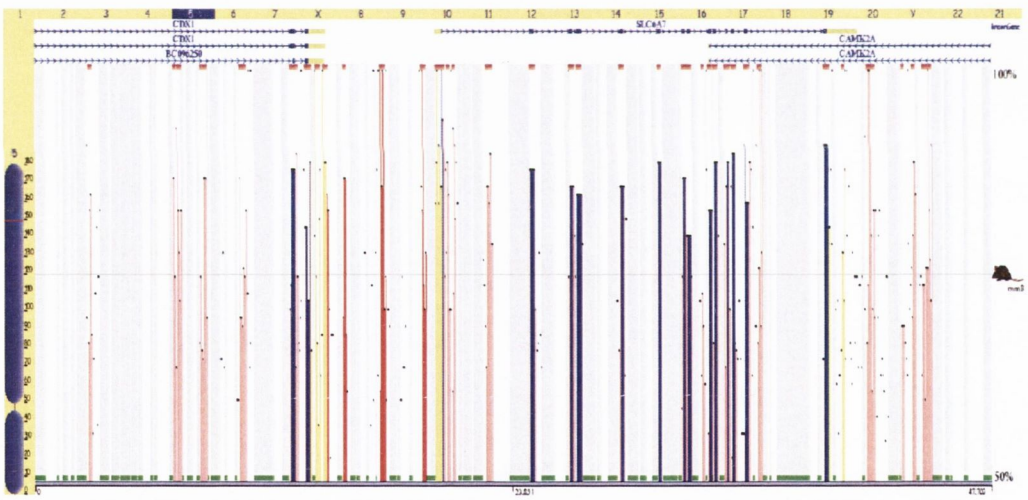


**Figure 2.1 The ECR Browser. Exonic regions are blue, intronic regions are pink, intergenic regions are red, and UTRs are yellow. The Browser recognises ECRs subject to user defined parameters. The default ECR definition is an orthologous region of at least 100 base pairs in length, and 70% sequence homology.**

ECR Browser is a publicly available web-based tool .It functions as a dynamic whole-genome navigation tool for visualising and studying evolutionary relationships between vertebrate and non-vertebrate genomes.

The tool includes several genomes for comparison. At the time we carried out our study, fifteen genomes were available: human, dog, mouse, rat, chicken, frog two

pufferfish (Fugu and Tetraodon), zebrafish, and 6 fruitflies. Since then, the databases have been expanded to include 13 distinct species; (*Homo sapiens*), chimpanzee (*Pan troglodytes*), rhesus monkey (*Macaca mulatta*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), dog (*Canis familiaris*), cow (*Bos taurus*), opossum (*Monodelphis domestica*), chicken (*Gallus gallus*), frog (*Xenopus laevis*), zebrafish (*Danio rerio*), fugu pufferfish (*Takifugu rubripes*), and spotted green pufferfish (*Tetraodon nigrovoridis*).

One or more of these genomes can be compared simultaneously, with the base or 'reference' genome, allowing for viewing of ECRs across multiple species simultaneously.

The 'Grab ECR' feature allows users to rapidly extract sequences that correspond to any ECR, to visualise underlying sequence alignments, and/or use the rVISTA program to identify conserved transcription factor binding sites. In addition to accessing pre-computed alignments for the available genomes, the ECR Browser can also be used as an alignment tool. It allows users to map submitted sequences to specific homologous positions within selected genomes and to create a detailed alignment using the BLASTZ alignment program. The user can retrieve locations of homologous sequences within aligned genomes, and view the extent of their sequence match to the reference genome. (Ovcharenko *et al.* 2004)

### 2.8 Rvista

http://rVISTA2.0.dcode.org/

rVISTA2.0 is a publicly available web tool for identifying putative transcription factor binding sites. rVISTA2.0 utilises BLASTZ to compare orthologous sequences from user defined related organisms. The local alignment methods provide an assessment of evolutionary rearrangement in an attempt to detect TFBSs that have undergone positional changes relative to nearby genes.

rVISTA2.0 analysis proceeds in four main steps:

1. Detect TFBS matches in each individual sequence via the TRANSFAC database.
2. Identify pairs of locally aligned TFBSs
3. Select TFBSs present in regions of high DNA conservation
4. Create a graphical display that dynamically overlays individual or clustered TFBSs with the conservation profile of the genomic locus.

TFBS are identified by cross referencing the selected sequence with either a user-defined consensus sequence or one or more selected TFBS motifs selected from the TRANSFAC library. This includes matrices from vertebrates, plants, nematodes, insects, fungi and bacteria, and contains representatives from 500 vertebrate TF matrices that comprise 400 TF families. Once TFBSs in both sequences are localised, rVISTA2.0 identifies pairs of aligned TFBSs that are interconnected in the local BLASTZ alignment.

Genomic DNA insertions and deletions that occur in the core region of either TFBS disqualify the prediction as rVISTA2.0 requires aligned TFBS predictions to be locally highly conserved (at least 80% sequence identity in a 20 bp sliding window spanning and including the core of the binding site).

Results are ouputed to a web page containing detailed information including positional information on TFBS predictions in both sequences and distribution of aligned and aligned-and conserved TFBSs, as shown in **Figure 2.2.**

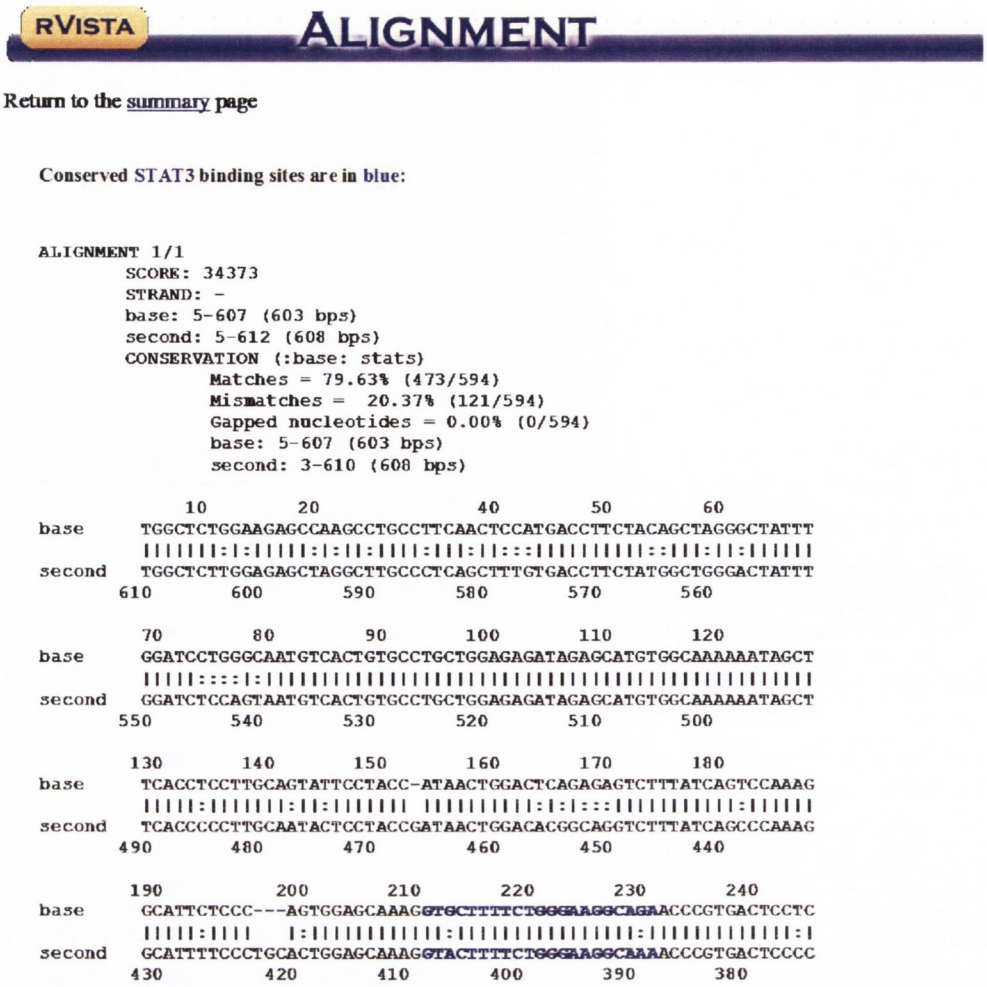



**Figure 2.2 rVISTA The figure shows a comparion of human and mouse orthologous sequences, highlighting in blue a shared consensus binding sequence for the Transcription Factor STAT3.**

Identification of TFBS, especially clustered and of similar function/family, is one way of identifying regions with a potentially regulatory function. Consensus sequences for TF binding are short sequences, typically 5-25bp in length (Moses *et al.* 2003). The extent to which a sequence can depart from the consensus sequence while retaining binding specificity depends largely on two factors; the TF in question, and position specific sequence preference of particular binding sites, essentially which bases in a particular recognition sequence can be altered while still retaining binding specificity. (Moses *et al.* 2003, Stormo 2000).

This relaxation from consensus means that 5-25bp sequences which can potentially bind TFs abound throughout the genome. Additionally, even if a site binds a TF, there is no way of knowing that it exerts a functionally relevant effect once it has bound. Thus, given the degeneracy inherent to TF binding identifying putative TFBS purely on

homology to defined consensus sequences will result in the generation of a vast number of possible sites throughout the genome, the majority of which will be non-functional. Identifying a putative TFBS requires that it satisfy a number of more stringent criteria in order to assess its potential utility. A method of identifying regions which are likely to contain regulatory elements is also crucial, as this will reduce the amount of the genome to be scanned to a practical level. Implementation of both of these strategies is crucial in bioinformatic-based location of putative TFBS within regulatory regions.

## 2.9 Selection of SNPs for Haplotype Tagging

When selecting SNPs to haplotype tag genes in this study, the public database HapMap was referred to (see Appendix). We used the freely available Haploview software package to download genotype information for the CEU population, visualize LD blocks, and select a panel of tagging SNPs via the SNP Tagger application.

SNPs were selected as tags based on an $r^2$ value of at least 0.8 with the SNPs with which they were tagging. SNPs with a minor allele frequency of below 0.05 in the CEU population were not selected to be tagged.

An exception to this approach should be noted in the IL-18 and TNFSF14 genes, the tags for which were identified by a previous member of our research group, using public databases available prior to the launch of the HapMap site.

## 2.10 Genotyping of Polymorphisms

### 2.10.1 Polymerase Chain Reaction

The CCR5 Delta-32 polymorphism was genotyped via PCR, separated by electrophoresis, and visualised under a UV light. The Delta 32 polymorphism was amplified using a Primers described in Goulding *et al.* 2005. Primers were supplied by MWG Biotech, Germany. A 10µl PCR reaction consisting of 1x PCR buffer (as supplied by manufacturer), 25µM $MgCl_2$, 200µM of deoxynucleotide triphophates, 12.5µM of both forward and reverse primer, 10ng of extracted DNA, 1 unit of Qiagen HotStar DNA polymerase Taq, were used to genotype each sample in both the subject and control groups.

Thermal cycling conditions of the PCR consisted of an initial enzyme activation step at 95°C for 15 minutes, followed by 35 cycles (a cycle consists of a DNA denaturing step at 94°C for 45 seconds, an annealing step at 62°C for thirty seconds, and an extension step at 72°C for 30 seconds).

**Primers flanking the CCR5-32 mutation**

(sense, 5'-CAAAAAGAAGGTCTTCATTACACC-3')

antisense, 5'-CCTGTGCCTCTTCTTCTCATTTCG-3')

These primers were used to amplify 189-bp (wild-type) and 157-bp (32bp deletion) fragments of the CCR5 gene, respectively. Following amplification the fragments were visualised on a 3% agarose gel.

### 2.10.2 Agarose Gel Electrophoresis

The CCR5 Delta-32 products were separated on 3% agarose gels. Gels were made with 6g of agarose (Melford Laboratories, UK) to 200ml of 1x TAE, which was then boiled. 11µl of ethidium bromide (Sigma Chemical Co, MO, USA) was added, and the gel poured into a gel tray, and left to set (30-40 minutes). 2µl of loading dye was added to each 10µl sample, and 10µl of this was loaded into each well in the gel, along with 10µl of a 100 base pair DNA ladder (New England Biolabs, MA, USA). The gel was then subjected to electrophoresis in 1x TAE buffer for 40 minutes across a voltage of 120 volts. Gels were subsequently visualised under UV light, and the genotype of each sample recorded based on observed fragment size.

### 2.10.3 Allelic Discrimination using Taqman Genotyping technology

In the Italian population, the IL-2/IL-21 SNPs rs4374642, rs13119723, rs1127348, rs6822844 and rs6840978, along with the GWAS replication SNPs rs9851967, rs2816316, rs917997, rs6441961, rs3184504, rs1738074, rs17810546 were all genotyped using Taqman technology. In the Irish population, the TNFSF14 SNPs rs2279627, rs3760746 and rs11878563, along with the IL-18 SNPs rs1946518 and rs187238 and the MyoIXB SNP rs2305764 were also genotyped using this method.

In this technology, primers specific to the region flanking the SNP in question are designed to facilitate specific amplification of the region. Fluorescent probes are also designed, with a probe specific for each of the two alleles. These probes hybridise to

their respective alleles. Each probe has one of two fluorescent markers bound to it in turn, most commonly these are flurophores VIC and FAM. While bound to the probe, these fluorophores are associated with a 'quench' molecule which prevents them from fluorescing.

The polymerase hybridises to the DNA strand being amplified, and moves in a progressive manner along the DNA strand. Upon reaching the bound probe it cleaves it (by virtue of its 5' exonuclease activity), releasing the probe from its binding to the DNA strand and the flurophore from its association with the probe.

Once released, the flurophore emits fluorescence which can be detected by a reader. The amount of fluorescence emitted is proportional to the amount of probe cleaved, which, in turn, is proportional to the amount of amplification.

The two fluorophores (typically FAM and VIC) emit significantly different fluorescence, allowing the two alleles of a SNP to be distinguished from one another. The reader detects and reports the amount of fluorescence, and plots it on a graph for each sample.

One axis of the graph relates to fluorescence emitted by the FAM tagged probe (the X axis), the other to that emitted by the VIC tagged probe (the y axis). A homozygote for the allele that hybridises to the FAM probe will show strong fluorescence on the X axis, and little, if any, fluorescence on the Y axis. The reverse will be true for a homozygote for the other allele. A heterozygote will show fluorescence on both axis, though typically of smaller magnitude than that of a homozyote. **Figure 2.3** shows a graphical overview of this process.

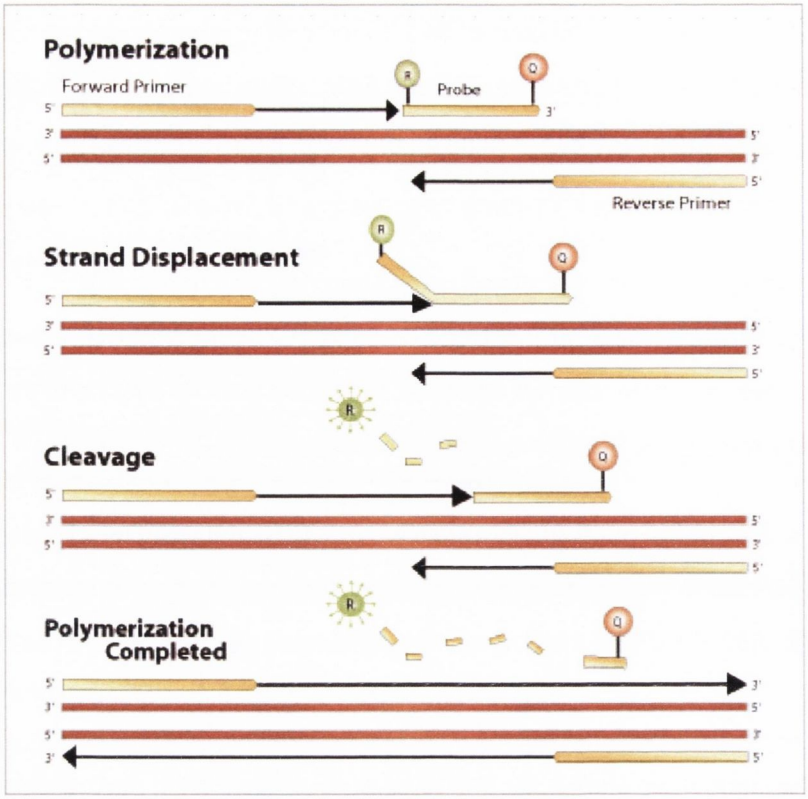



**Figure 2.3 An Overview of Taqman genotyping technology, showing the 4 main stages of the process.**

Genotyping of the SNPs were carried out at a total reaction volume of 5µl. Each reaction consisted of 10ng of DNA, 2.5µl of 2X Universal Master Mix (Applied Biosystems) and 0.125 40x Assay-by-Design/Assay-by-Demand mix.

Fragments were amplified on the ABI 7000 (Applied Biosystems) under the standard conditions associated with the Assays by design and assays by demand; an initial 50°C for 2 minutes, followed by 95°C for 10 minutes, then 40 cycles. A cycle consists of a denaturing step at 92°C for 15s and an annealing/extension step at 60°C for 1 minute.

### 2.10.4 Ampiflour Genotyping

Genotyping of the 5q SNPsrs11558274, rs8177834, rs2017638, rs2233282, rs2233280, rs868531, rs2233292, rs2233289 and rs17728338 near TNIP1 in the Irish population; rs8177834, rs2017638 and rs868531 in the Italian population; and the STK32A haplotype tagging SNPs rs4705038, rs11955126, rs9325041, rs17482628, rs10041912, rs17548933 were carried out commercially by Kbiosciences Ltd, UK.

Kbiosciences utilises the Amplifluor genotyping system: a closed tube system with two Amplifluor primers, and three unlabeled primers. The Amplifluor primers consist of an excited fluorophore, which is linked to an acceptor moiety via an oligonucleotide, resulting in quenching of the fluorescence. A single common reverse primer, and two SNP-specific primers are designed to amplify across the SNP, each with a 5' tail that corresponds to one of the two Amplifluor primers.

As SNP-specific PCR product is generated, the Amplifluor primer hybridizes with its complement on the tail sequence of the newly generated PCR product. With the common reverse primer, this causes the product to amplify further. A FAM or a JOE fluorophore signal is then generated in an allele specific manner. Heterozygotes result in signals from both fluorophores. The Amplifluor primers. and the common reverse primer, are used at x10 excess over the SNP specific primers to ensure high yield of the resulting fluorescent product. **Figure 2.4** shows this graphically.

We used the dbSNP function of the NCBI database to indentify the sequence 50 base pairs either side of the SNP to be genotyped of interest to be genotyped, and sent the sequence to KBioscience to develop an assay for genotyping.

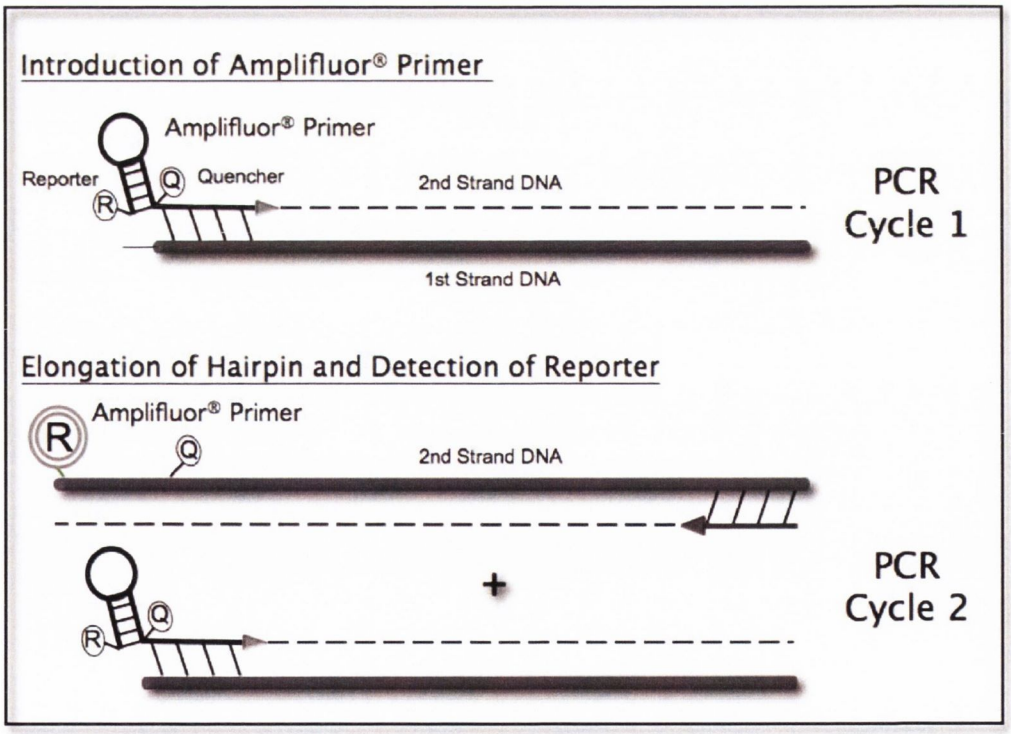



**Figure 2.4 An Overview of Amplifluor genotyping technology, showing the 3 main stages of the process.**

### 2.10.5 Illumina GoldenGate Genotyping Platform

All of the SNPs genotyped as part of the GWAS follow-up described in chapter 4, and those genotyped as part of the candidate gene enriched GWAS follow up described in Chapter 5, were genotyped using the Illumina GoldenGate Platform. The protocol for the massively multiplex 'GoldenGate' genotyping platform is summarised below. For exact experimental protocol, the GoldenGate Manual (Illumina) can be provided.

**Figure 2.5** gives an overview of the workflow for the GoldenGate assay.

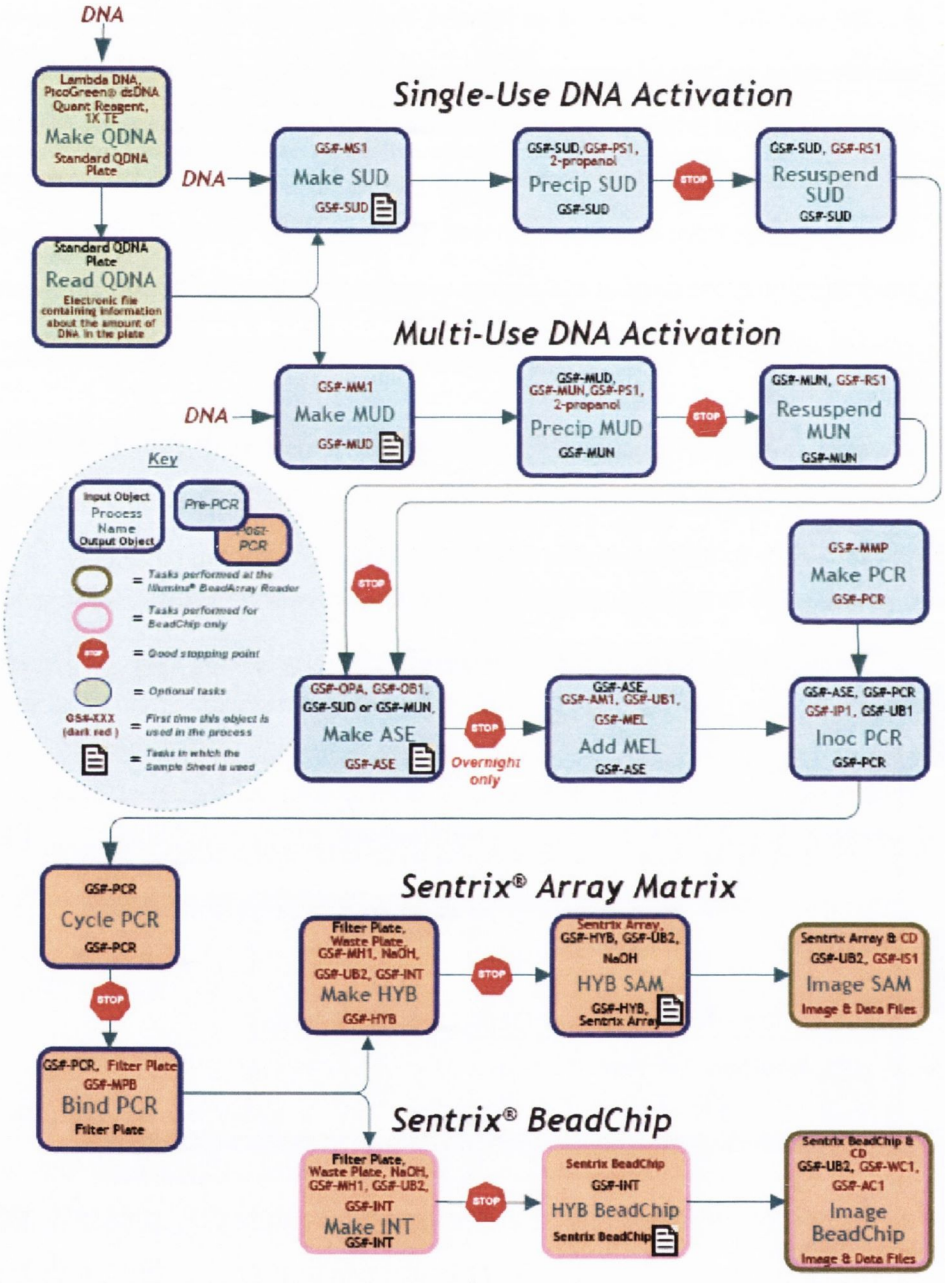



**Figure 2.5 GoldenGate workflow. This diagram gives an overview of the entire, 3 day genotyping process, starting from either SUA DNA plate or MUA DNA plate, and moving on to either the Array Matrix or Bead Chip. We used SUA plates and Bead Cheap in this study.** Source: GoldenGate Manual (Illumina).

*SUD (Single Use DNA) Plate*

The Single-Use and Multi-Use DNA Activation plates are mutually exclusive steps in this protocol (one or the other are used). We utilised the SUD. Creation of the SUD plate follows the steps outlined below.

1) DNA is activated for binding to paramagnetic particles through the addition of a Streptavidin/Biotin tag.
2) Tagged DNA is precipitated from the SUD Plate, and excess DNA activation reagent is removed.
3) The DNA is re-suspended to make the SUD Plate.

Making of the SUD plate is summarised in **Figure 2.6**

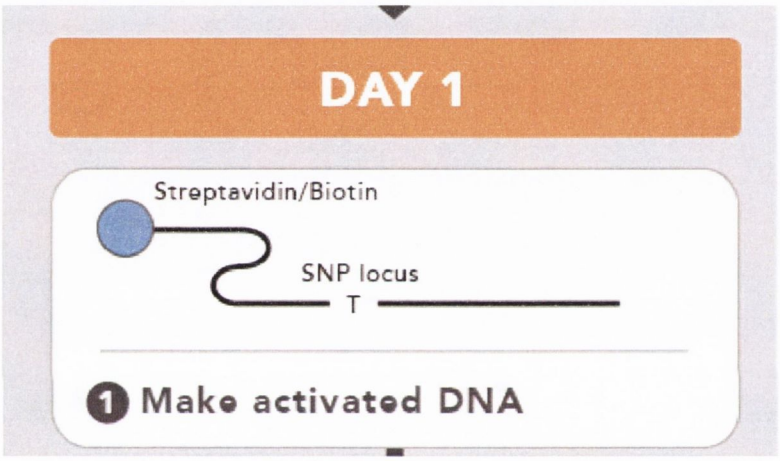



**Figure 2.6 Day 1 GoldenGate Process.**

*ASE (Alelle Specific Extension) Plate*

The ASE plate combines the activated DNA with SNPquery oligonucleotides (detailed below), hybridization reagents, and paramagnetic particles. The plate is then placed in a heat block and the allele specific and locus specific oligos for each SNP locus of interest are allowed to anneal to activated DNA samples. The activated DNA is simultaneously captured by paramagnetic particles. Once this occurs, the ASE plate is ready for the extension and ligation of the hybridized oligos on the bound DNAs. The ASE Plate Process is summarised in **Figure 2.7.**

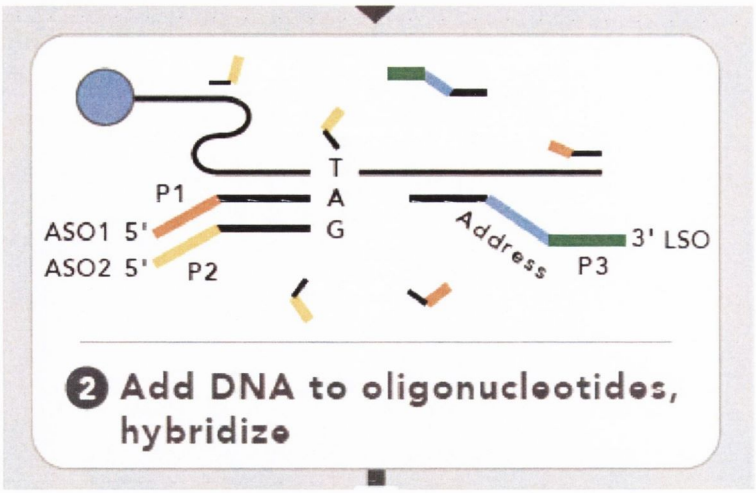



**Figure 2.7 Day 1 ASE Plate Process.**

*Oligonucleotides*

Three oligonucleotdies are designed for each SNP locus in the genotyping panel. Two are specific to the alleles present at that site; they are termed the Allele Specific Oligonucleotides (ASO). One Hybridises several bases downstream of the SNP; it is the Locus Specifc Oligoucleotide (LSO). All three oligonucleotide sequences contain regions of genomic complementarily, and universal PCR primer sites. The LSO also contains a unique address sequence that targets a particular bead type.

During the primer hybridization process, the assay oligonucleotides hybridize to the genomic DNA sample which itself is bound to paramagnetic particles (see making of the SUD plate). Because hybridization occurs prior to any amplification steps, no amplification bias can be introduced into the assay.

*Extension & Ligation in ASE Plate*

1) The ASE Plate is heated, allowing allele and locus specific oligos to anneal to activated DNA samples. Activated DNA is simultaneously captured by paramagnetic particles. In order to assure complete annealing and capture, the plate is heated for two hours.

2) The non-specifically hybridised and excess oligos are washed away from the DNA. Following hybridization, several wash steps are performed, reducing noise by removing excess and mis-hybridized oligonucleotides.

3) Extension and ligation enzymes are added to each sample. Extension and ligation occurs at 45°C. Extension of the appropriate ASO, and ligation of the extended product to the LSO , joins information about the genotype present at the SNP site to the address sequence on the LSO. These joined, full length products provide a template for PCR in the next step. The entire process is summarised in **Figure 2.8.**

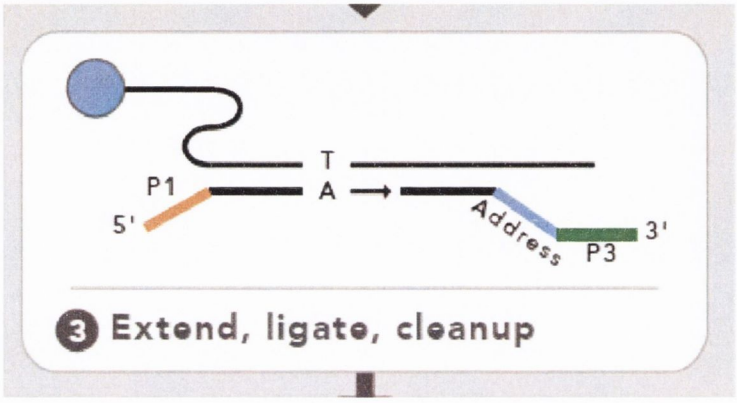



**Figure 2.8 Day 2; Extension Ligation and Cleanup.**

*PCR Plate*

Illumina-recommended DNA Polymerase and Uracil DNA Glycosylase is added to the master mix for PCR, and a 96-sample plate for use in the Inoc PCR process is created.

1) DNA Polymerase and UDG are added to the PCR Master-mix to create the PCR Plate.
2) The template formed by the extension and ligation reaction is transferred to the PCR plate. The PCR reaction uses 3 universal primers: two with fluorescent dyes and one biotinylated primer. Primers are allele specific, with Cy3 and Cy5 corresponding to ASO1s and ASO2s respectively. The Biotinylated primer allows capture of PCR product and elution of the strand containing the fluorescent signal.
3) The PCR plate is thermally cycled. This fluorescently labels and amplifies templates generated in the pre-PCR process. **Figure 2.9** summarises the Universal PCR cycle.

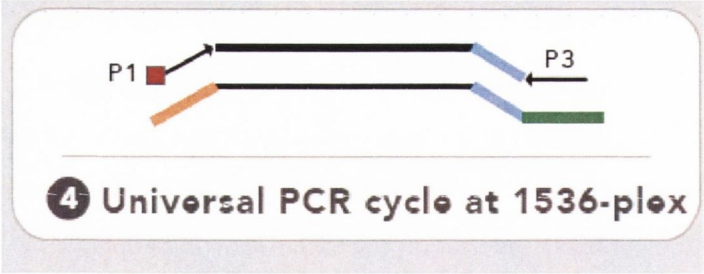



**Figure 2.9 Day 2; Universal PCR Cycle.**

4) Double stranded PCR products are immobilised by binding the biotinylated strand to paramagnetic particles. The solution is transferred to a filter plate and incubated at room temp for one hour so that the PCR product binds to the paramagnetic particles. The single stranded dye labelled DNA is then eluted. This process is summarised in **Figure 2.10.**

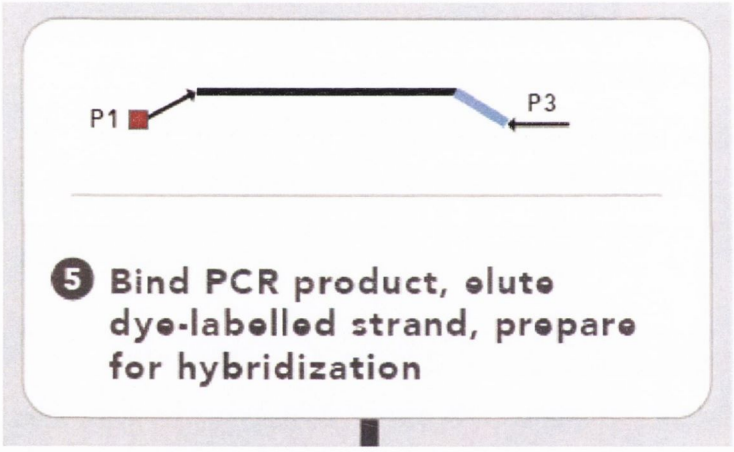



**Figure 2.10 Day 2; PCR Product Binding.**

*Hybridisation Plate*

In this step, the single-stranded, fluor-labeled PCR product from the Filter plate is washed, eluted, into the GS-INT plate, then transferred to a new plate GS- HYB, suitable for mating to the Sentrix® Array Matrices (SAMs).

1) Once the fluor labelled PCR product is transferred to the HYB Plate, it is paired with the SAM, hybridised overnight (60 to 45 degrees) to anneal signal to the array. The LSO address signal targets the oligonueclotide to a specific bead type on the array.

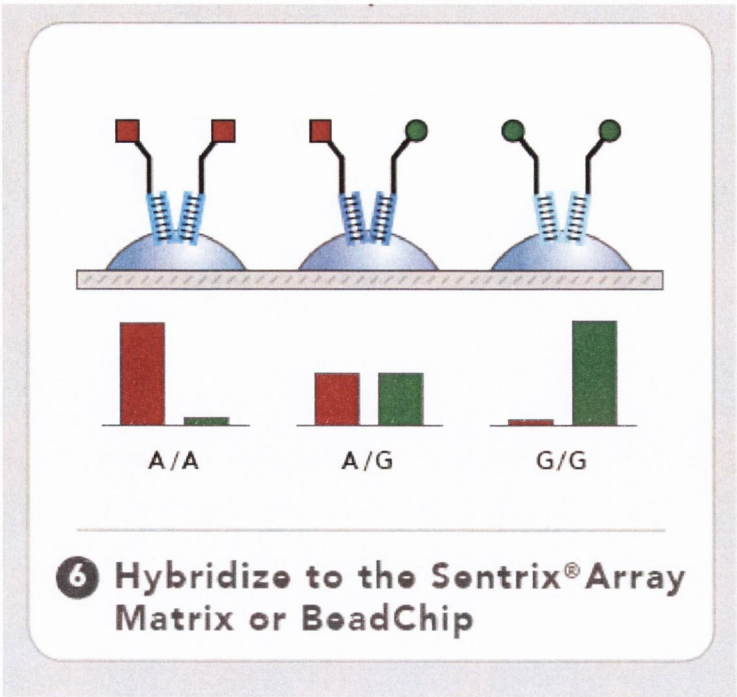



**Figure 2.11 Day 2; Hybridisation.**

The Universal PCR primers P1 and P2 are Cy3- and Cy5-labeled. After downstream-processing, the single- stranded dye-labelled DNAs are hybridized to their complementary bead type through their unique address sequences.

Hybridization of the GoldenGate Assay products onto the Array Matrix allows for the separation of the assay products in solution, onto a solid surface for individual SNP genotype readout. After hybridization, the BeadArray Reader is used to analyze fluorescence signal on the Sentrix Array Matrix or BeadChip 8.

Lasers excite the Cy3 and Cy5 fluorophores of the single stranded PCR products bound to the beads of the fibre optic bundles. Light emissions of these are separately recorded in high resolution images of the fibre optic bundles. Data from these images are analyzed to determine SNP genotypes using Illumina's genotyping software package (Occurs on Day 3, the final day of the protocol). **Figure 2.11**, above, shows product-bead hybridisation, and associated signal.

### 2.10.6 Confirmation of Genotyping Results

In the cases of taqman, ampifluor and PCR gel-visualised-based genotyping, 5% of the sample being genotyped were routinely duplicated to ensure genotyping accuracy.

## 2.11 Data Analysis

### 2.11.1 Beadstudio

Data from the GWAS follow-up studies were viewed graphically via the Illumina software application Beadstudio. This application allows for display of the relative intensities of the fluorescence from the Cy3 and Cy5 fluors of the single stranded PCR products generated via the GoldenGate assay.

The program allows for adjustment of the fields used for allelic calling, and provides various metrics, including goodness of fit for the clustering of the genotype calls. We analyzed the Irish population separately, partially due to the tendency of the WGA samples to cluster in a manner which was internally consistent, but differently centred to other samples. Figures **2.12, 2.13 and 2.14** show three Irish loci, each with varying degrees of clustering spread. Irish clusters were manually adjusted on a SNP by SNP basis, for WGA and genomic samples separately, to ensure the most accurate call rates possible.

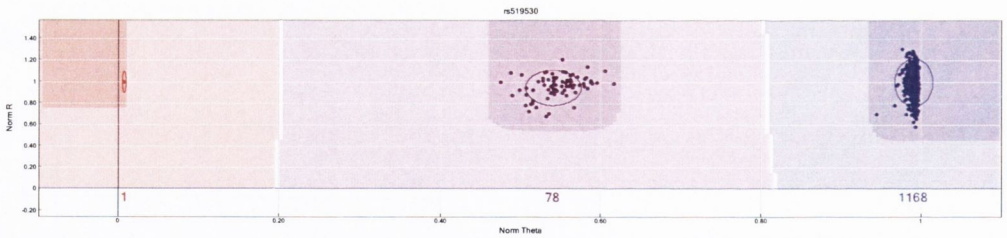



**Figure 2.12 Tightly clustered genotype groups as shown the beadchip software.**

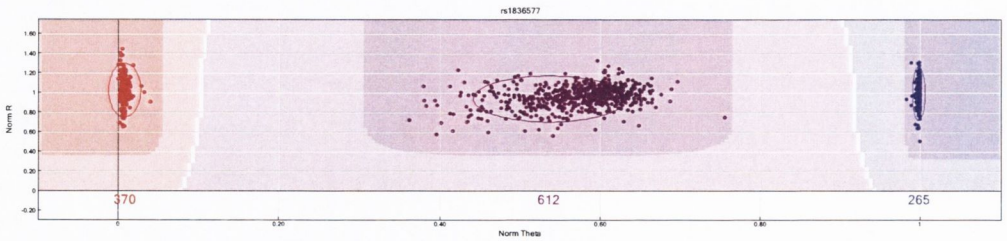



**Figure 2.13 Moderately clustered genotype groups as shown the beadchip software.**

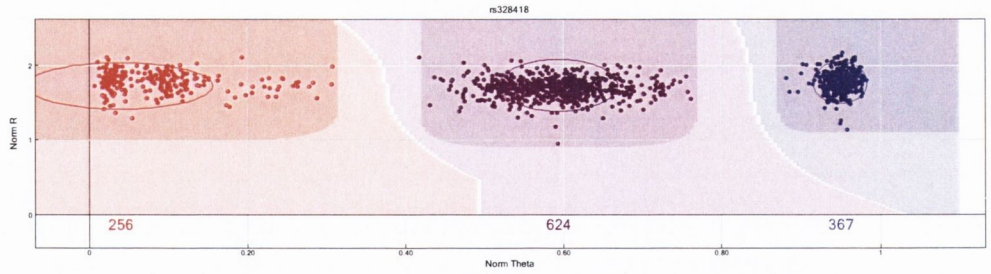



**Figure 2.14 Poorly clustered genotype groups as shown the beadchip software.**

### 2.11.2 Genetic case-control differentiation analysis of allele, genotype and carrier status frequencies

Allelic, genotyping and carrier status frequencies were obtained using PLINK. PLINK is a free, open-source whole genome association analysis toolset. Comparison of frequency differences between patient and control groups was assessed via the $\chi^2$ test, or Fisher's exact test, where appropriate. Results from PLINK were double checked using the Open-Epi webtool.

### 2.11.2 Hardy-Weinberg Equilibrium (HWE) testing

Genotyping results were evaluated for HWE using PLINK (see above).

### 2.11.3 Haplotype estimation and differentiation

Haplotypes were estimated using PHASE program. PHASE implements methods for estimating haplotypes from population genotype data via the Stephens-Smith-Donnelly (SSD) method (Stephens *et al.* 2001; Stephens & Donnelly 2003) was also used to generate haplotypes. This package produced a global p-value to test for any significant difference

In order to establish a P value for each individual haplotype, and to confirm the estimations of frequency, haplotype estimation and hypothesis testing was also performed on HITAGENE.

HITAGENE uses the expectation maximisation (EM) algorithm (Excoffier &Slatkin 1995) to estimate haplotype frequencies. Statistical differences between haplotype frequencies in the case and control groups were examined using p-values obtained by a permutation method as described by Zaykin (Zaykin *et al.* 2002) and carried out using the "haplotype association test" function in HITAGENE.

### 2.11.4 Linkage disequilibrium analysis

Linkage Disequilibrium between SNPs was determined by utilising the Haploview program to view HapMap CEU data for the SNPs in question.

### 2.11.5 Cochran-Mantel-Haenszel test for significance in stratified populations

This test was carried out via PLINK (see above).

# CHAPTER III

# Chromosome 5

# &

# Coeliac Disease

## 3.1 Introduction

### 3.1.1 Linkage to the 5q region

Prior to the emergence of large scale GWAS studies into Coeliac Disease, genome scans using family based association studies were the best indication of promising regions to investigate for association. Many such studies were carried out by European laboratories between 1996 and 2007, and **Table 3.1** summarises the results from these Genome Scans, and subsequent linkage and association studies.

As can be seen, a number of genomic regions have shown potential linkages to CD. However, only two, 19p13.1 and 5q31-33, have shown significant linkage ($P < 2x10^{-5}$) with coeliac disease.

Both of these regions (along with 3q, 10q, 11q, and 15q) were analyzed in an Italian cohort of 89 affected sibpairs, with a set of 50 polymorphic markers, utilising a Maximum Lod Score Method. This population showed evidence for linkage to the 5q region only (Greco *et al.* 2001). This result was reproduced using another cohort of 93 affected Italian sib-pairs from a previous genome screen (Greco *et al.* 1998). The highest MLS value of 2.92 was found for the most centromeric marker tested, the microsatellite repeat marker D5S640 (Greco *et al.* 2001).

A further study was carried out on the 5q31-33 region, adding 12 new microsatellite markers and utilising a cohort of 182 affected Italian sib-pairs (Greco *et al.* 2002). These 24 microsatellite repeat markers spanned a region of 52cM, were evaluated using NPL and MLS statistics, the results of which are summarised in **Table 3.2.** This table refines the maximum linkage peak to the D5S2014 marker (**in bold**), close to the previously identified D5S640 (*in italics*).

The region has also exhibited linkage to coeliac disease in Scandinavian and Finnish families in addition to being associated with a number of multifactorial diseases, including Crohn's Disease, IDDM, asthma and allergy (Percopo *et al.* 2003).

**Table 3.1 Pre GWAS Genome Scans compiled from studies carried out in multiple European populations over a seven year period, showing linkage to chromosomes 5 an 19.**

| **Year and Population** | **Number of Families** | **Type of Family** | **Suggestive Linkage $P < 7x10^{-4}$** | ***Significant Linkage*** $P < 2x10^{-5}$ |
|---|---|---|---|---|
| 1996 Irish | 15 | Multi-Affected | 6p23, 11p11 | |
| 1998, 2000 **Italian** | 39/71/89 | Sib-Pairs | 5q32-34 | |
| **2000, 2001 British** | 16/34 | Multi-Affected | | |
| **2001 Scandinavian** | 70/36 | Sib-Pairs | | |
| **2002 Finnish** | 60/38 | Sib-Pairs | | |
| 2002 Finnish | 9/1 | Large Extended Families | 15q12 | |
| 2002 Northern European | 24 | Multi-Affected | | |
| 2002 North American | 62 | Sib-Pairs | 3p26 | |
| 2002 Dutch | 67/15 | Sib-Pairs | 6q21-22 | *19p13.1* |
| 2004 Finnish | 54 | Sib-Pairs | | |
| 2004 Dutch | 1 | Large Extended Family | 9p21-13 | |
| 2007 North American | 160 | Multi-Affected | | |
| 2003 Meta Analysis of Scandinavian, Italian, Finnish and British (in **bold**) | | | | *5q31-33* |

**Table 3.2 Microsatellite Repeat Markers on 5q31-33.**

| *Location* | *Marker* | *NPL* | *MLS* | *Information* |
|---|---|---|---|---|
| 0 | D5S436 | 1.970 | 1.410 | 0.759 |
| 0.1 | D5S1480 | 2.209 | 1.657 | 0.761 |
| 2.9 | D5S413 | 1.689 | 0.868 | 0.817 |
| 3 | D5S434 | 1.867 | 1.054 | 0.818 |
| 3.1 | D5S2090 | 1.713 | 0.943 | 0.823 |
| 5.4 | D5S2015 | 2.876 | 2.265 | 0.883 |
| *5.5* | *D5S640* | *2.798* | *2.234* | *0.891* |
| 5.6 | D5S2013 | 2.240 | 1.459 | 0.877 |
| **6.2** | **D5S2014** | **3.091** | **2.533** | **0.865** |
| 6.3 | D5S636 | 2.994 | 2.426 | 0.866 |
| 9.1 | D5S673 | 2.376 | 1.486 | 0.879 |
| 9.2 | D5S2077 | 2.032 | 1.123 | 0.879 |
| 9.8 | D5S2007 | 2.484 | 1.617 | 0.879 |
| 9.9 | D5S410 | 2.374 | 1.515 | 0.879 |
| 13.3 | D5S2112 | 2.282 | 1.383 | 0.864 |
| 17.7 | D5S422 | 2.112 | 1.381 | 0.830 |
| 22.7 | D5S2032 | 2.148 | 1.293 | 0.816 |
| 26.4 | D5S671 | 1.596 | 0.778 | 0.798 |
| 28.6 | D5S400 | 1.354 | 0.599 | 0.765 |
| 33 | D5S425 | 1.075 | 0.490 | 0.834 |
| 36.3 | D5S2069 | 1.450 | 0.726 | 0.809 |
| 41.9 | D5S2111 | 1.036 | 0.504 | 0.769 |
| 49.7 | D5S408 | 0.413 | 0.061 | 0.765 |
| 51.7 | D5S2006 | 0.898 | 0.267 | 0.747 |

*Adapted from Greco et al. 2003. Location refers to the distance from the start point along the 52 centimorgans chosen for analysis by Greco et al. Marker delineates the name of the microsatellite repeat, with NPL and MLS representing LD derived statistics. Information refers to the polymorphic information content of the locus, a function of its heterogeneity and the frequency of its alleles.*

Although these Italian studies have examined the 5q region with a greater density of markers than any other, the 2003 meta analysis details the other genome scans which have demonstrated linkage to the area. **Figure 3.1** shows graphically the linkage peaks each of these studies observed, as well as the overall linkage peak when the datasets were merged and subject to a meta analysis. Table 3.3 shows the markers most strongly associated with each population.

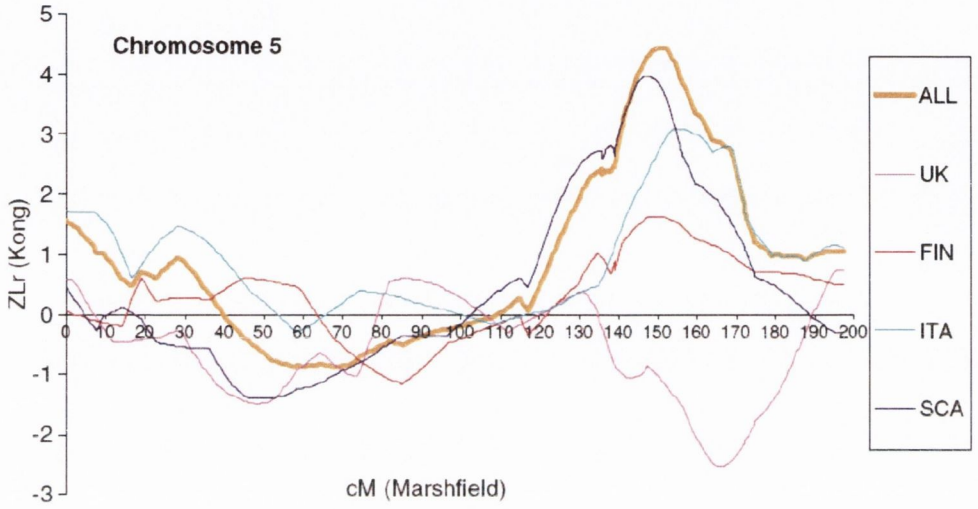



**Figure 3.1 Linkage peaks in 5q; Taken from Nalaui et al, 2008, showing peaks by individual population, and meta-analysis.**

**Table 3.3 Microsatellite Markers with strongest Linkage by study.**

| Study | Microsatellite | Position |
|---|---|---|
| **Italian** | **D5S640** | 149982541-149982629 |
| | **D5S2014** | 149963917-149964076 |
| **Scandinavian** | **D5S816** | 135301510-135301752 |
| **Finnish** | **D5S816** | 135301510-135301752 |
| **British** | **None** | NA |
| **Meta Analysis** | **D5S640** | 149982541-149982629 |

Based on the identification of **D5S640** as the strongest linkage peak in both the well characterised Italian studies, and the meta analysis (Babron *et al.*, 2003) the region around this microsatellite was deemed the most promising for investigation. Due to the linkage peak observed in the Scandinavian and Finnish studies, **D5S816,** it was deemed that an interval encompassing **D5S640,** but extending toward **D5S816** would

be selected for investigation; the Mb region spanning from 146,000,000 to 151,000,000.

There is evidence to suggest that complex diseases owe a significant portion of their genetic predisposition to variants involved in the regulation of gene expression (Hindorf *et al.* 2009, Dubois *et al.* 2010). As such, we decided to attempt to identify SNPs involved in the regulation of gene expression within this interval, and to select a subset of them, based on likely relevant functionality, for genotyping and analysis.

Using the ECR Browser tool (detailed more fully in Chapter 2, Materials and Methods), we investigated the 5q genomic region from 146 Mb to 151 Mb for evidence of Evolutionarily Conserved Regions (ECRs). These are orthologous genomic sections, from two or more organisms, which display a degree of sequence conservation which is higher than would be expected for non functional DNA. This increased level of conservation may be due to the presence of a regulatory element, which as a functional motif is exerting evolutionary pressure to conserve its sequence. Presence of an ECR then, within an intergenic region, UTR or intron, may be indicative of the presence of a functionally active regulatory region, based on the relatively high degree of sequence conservation between the two organisms (Ovcharenko *et al.* 2005).

Additionally, the primacy of the **D5S816** microsatellite in the Scandinavian population studies indicates that there may be a second genetic determinant closer to 135,000Mb. Keeping this in mind, and in response to a 2007 study in Swedish and Norwegian populations (Amundsen *et al.* 2007), a haplotype tagging approach for a gene close to **D5S816,** STK32A, was also undertaken.

## 3.2 Objectives

The aims of this chapter are:

1. Examine the selected 5Mb interval for evidence of non-coding Evolutionarily Conserved Regions which may be indicative of regulatory sites.
2. Determine putative Transcription Factor Binding Sites within these Evolutionarily Conserved Regions.
3. Identify potential variation due to the presence of single nucleotide polymorphisms within these binding sites.
4. Select a subset of these SNPs for analysis, based on their proximity to candidate genes, and/or likelihood of being present within a promoter/enhancer element that responds to known pro-inflammatory transcription factors.
5. While simultaneously applying these criteria, also examine the selected interval for promising candidate genes, based on their role in immune mediation, autoimmunity, or gastrointestinal function, as these are most likely to contribute to coeliac disease.
6. Select a subset of SNPs within/across these candidate genes which are most informative, either in their roles as haplotype tags, or due to known or suspected alteration of protein structure or gene regulation.
7. Carry out statistical tests to determine any association between genotypes or possession of alleles and the disease state.
8. Where haplotype tagging has occurred, also test for haplotype association with disease state.

## 3.3 Results

### 3.3.1 ECR Browser; Identification of Human/Murine ECRs

The ECR Browser has the capability to compare the genomes of 13 distinct species; (*Homo sapiens*), chimpanzee (*Pan troglodytes*), rhesus monkey (*Macaca mulatta*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), dog (*Canis familiaris*), cow (*Bos taurus*), opossum (*Monodelphis domestica*), chicken (*Gallus gallus*), frog (*Xenopus*

*laevis*), zebrafish (*Danio rerio*), fugu pufferfish (*Takifugu rubripes*), and spotted green pufferfish (*Tetraodon nigrovoridis*).

The human genome sequence was compared to that of the mouse across the 5MB of the 5q31-33 region selected for study. The mouse was chosen as an organism for comparison on the basis of its degree of relationship to human. It is expected that the divergence time between the two organisms will allow for appropriate alteration in non-coding sequence, while retaining common gene regulatory structures.

In accordance with the default ECR Browser standards, an ECR was defined as a region of orthologous chromosome at least 100bp in length, which exhibited a homology of 70% or higher with the reference genome. **Figure 3.2** shows the ECR Browser interface displaying a comparison of human and murine orthologous sequences in the 5q region. Within the 146-151Mb genome interval 5250 human-murine ECRs were detected.

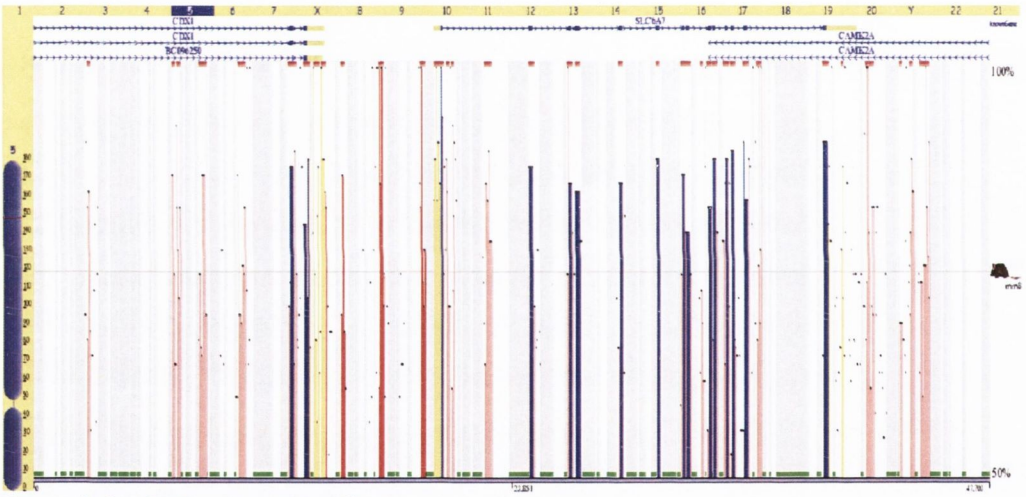


**Figure 3.2 ECR Browser Interface.**

### 3.3.2 Tagging of Variability within observed ECRs

The presence of variability within potential regulatory sites is of interest to us; invariant regulatory sites in humans are not. As such, the next step was to determine how many of the 5250 ECRs within this region contained documented SNPs. The ECR Browser database can be interrogated to provide this information, and of the identified ECRs within this region, 2205 contain one or more SNPs.

### 3.3.3 rVISTA; Transcription Factor Biding Site Characterisation

Having identified SNP containing ECRs, we then used the rVISTA tool to ascertain what, if any, TFBS were likely to be present within the SNP containing ECRs, and whether the SNPs within the ECRs lay within these putative TFBS.

All of the 467 vertebrate Transcription Factor binding motifs from the TRANSFAC library were selected for comparison . Transcription Factor binding motifs which were present in the selected ECR were highlighted, and the selected sequence was inspected visually to determine if the SNP lay within the predicted binding site.`

**Figure 3.3** shows the Rvista interface when highlighting TFBS in an ECR. 167 ECRs evidenced putative TFBS which contained one or more SNPs.

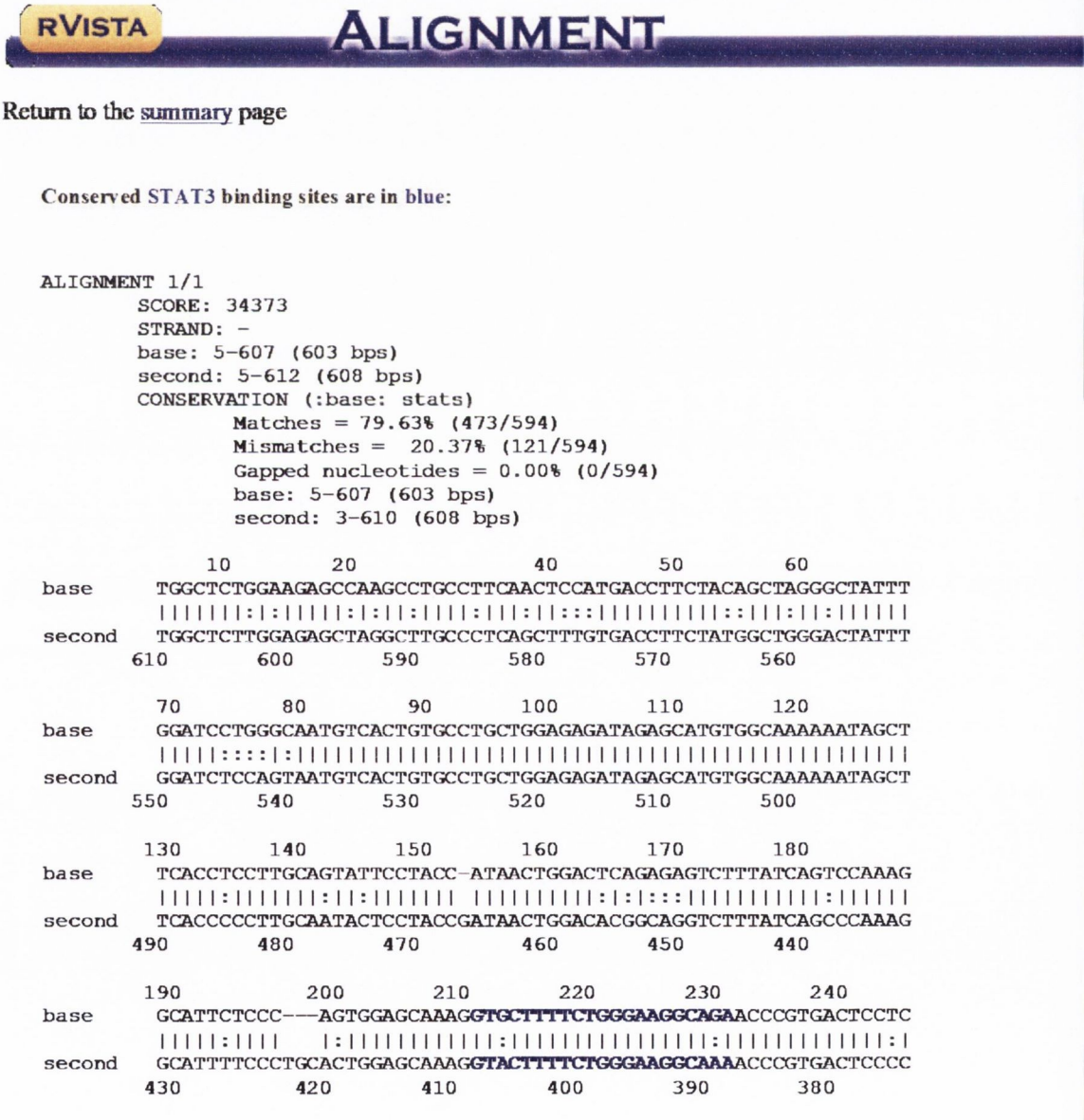



**Figure 3.3 RVISTA TFBS Identification Interface**

### 3.3.4 SNP Selection in the 146-151Mb interval of chromosome 5q.

In determining a subset of the identified SNPs to genotype, a number of different criteria were applied:

1) Known function of the TFBS the SNP is located within, with a preference for those involved in the inflammatory and/or immune responses.
2) Proximity to candidate genes.
3) Location relative to candidate genes (within the 3' or 5' UTR, or other regions likely to contain regulatory elements).
4) Proximity to **D5S640**
5) Regions with a high localisation of TFBS; regulatory elements tend to be composed of multiple, interacting TFBS.

We decided to concentrate our efforts on the 500Kb region from 150240K-150740K, due to it fulfilling multiple criteria simultaneously. The putative SNP containing TFBS in this area had a higher than average proportion of immune and pro-inflammatory roles, and although the interval had a comparable density of TFBS with the rest of the 5Mb examined (13, with 16 per 500kb being average) they were concentrated exclusively around three genes (GPX3, TNIP1 and ANXA6), which was a degree of density unseen elsewhere in the interval. **Table 3.4** summarises.

The functional nature of these TFBS also coincided with promising positions relative to the genes in question (UTR, Intergenic and nearby, and in one case intronic) as can be seen in **Figure 3.4**. The known functions of the 3 genes in question marked TNIP1 as a promising candidate gene, due to its role in the NfKappaB pathway (Cohen *et al.* 2009), and association of its interactant, TNFAIP3 with Coeliac Disease (See Chapter 5), and its genomic position **(150,409,506-150,466,719)** places it close to **D5S640.**

**Table 3.4 SNPs in putative Transcription Factor Binding Sites within ECRs in the 150240K-150740K interval.**

| | Start | stop | ECR Length (bp) | SNP Kind | SNP | % homology | TFBS |
|---|---|---|---|---|---|---|---|
| 1 | 150381400 | 150381538 | 139 | intronic | rs3805435 | 70 | LFA1Q6 |
| 2 | 150385392 | 150385491 | 100 | intronic | rs8177436 | 70 | AP2Q6 |
| 3 | 150389628 | 150389817 | 190 | UTR | rs11558274 | 69.5 | CDXA |
| 4 | 150389917 | 150390244 | 328 | UTR | rs8177834 | 71.3 | CMYB<br>VMYB |
| 5 | 150390325 | 150390603 | 279 | UTR | rs2233311 | 75.3 | ATF3_Q6 |
| 6 | 150403561 | 150403709 | 149 | intronic | rs10051105 | 70 | FOX04 |
| 7 | 150445700 | 150445800 | 100 | intergenic | rs2017638 | 72.2 | NFKB_C<br>NFKB_Q6 |
| 8 | 150446830 | 150447305 | 476 | intergenic | rs2233283 | 75.4 | AP2_Q3<br>MUSCLE_INI_B<br>MYCMAX_B<br>MINI19_B |
| 9 | 150446830 | 150447305 | 476 | intergenic | rs2233282 | 75.4 | STRA13<br>STAF |
| 10 | 150446830 | 150447305 | 476 | intergenic | rs2233280 | 75.4 | SP1 |
| 11 | 150464947 | 150465172 | 226 | intronic | rs868531 | 70.8 | CAAT_C |
| 12 | 150486024 | 150486284 | 261 | intronic | rs2303029 | 86.6 | ATF1_Q6<br>CREB_Q4<br>VJUN |
| 13 | 150495023 | 150495128 | 106 | intronic | rs3811989 | 70.8 | COUP_DR1_Q6<br>DR1_Q3<br>HNF4<br>HNF4_01_B<br>HNF4_DR1_Q3 |

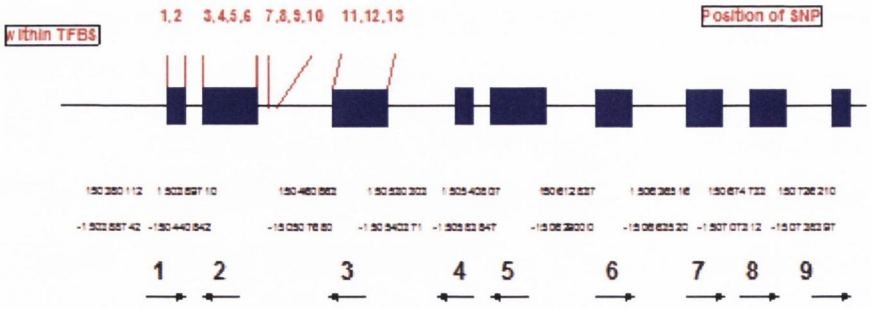



**Figure 3.4 Genes and SNPs within TFBS in ECR.**

## Genes:

1) GPX3 glutathione peroxidase 3 (plasma).
2) TNIP1 TNFAIP3 interacting protein 1
3) ANXA6-annexin A6
4) LOC442141 predicted similar to platelet-derived growth factor receptor like
5) DKFZP434C171-DKFZP434C171 protein
6) GM2A-GM2 ganglioside activator
7) SLC36A3-solute carrier family 36 (proton/amino acid symporter), member 3
8) SLC36A2-solute carrier family 36 (proton/amino acid symporter), member 2
9) LOC391840 similar to Thyroid hormone receptor-associated protein complex 240 kDa component (Trap240) (Activator-recruited cofactor 250 kDa component) (ARC250) (Thyroid hormone receptor associated protein 1)

### 3.3.5 TNIP1

TNIP1 is a strong candidate gene, based on several criteria. Functionally, it codes for a protein (Abin1) which acts to repress NFkB mediated inflammation (Cohen *et al.* 2009). TNIP1's binding partner, TNFAIP3, has shown association with Coeliac Disease (see Chapter 5), so there is already evidence to suggest involvement of this pathway as an underlying mechanism of Coeliac Disease.

Based on TNIP1's strong positional aspect, and its strong candidate gene nature, it was decided to genotype SNPs within TFBS which were likely to be involved in its regulation. **Table 3.5** summarises the selected SNPs within putative TFBS, and the reason for their selection.

**Table 3.5 SNPs suspected to play a regulatory role in TNIP1 Expression.**

| SNP | Positional Criteria | Functional Criteria |
|---|---|---|
| rs11558274 | Within 3' UTR of TNIP1 | Within a putative CDXA binding site. |
| *rs8177834* | Within 3' UTR of TNIP1 | Within a putative CMYB binding site. |
| *rs2017638* | Intergenic; 5' of TNIP1 | Within a putative NFKB binding site. |
| rs2233282 | Intergenic; 5' of TNIP1 | Within a putative STRA13 binding site. |
| rs2233280 | Intergenic; 5' of TNIP1 | Within a putative SP1 binding site. |
| rs868531 | Intronic; ANAX6 | Within a putative CAAT_C binding site. |

Two of its non synonymous SNPs; rs2233292 and rs2233289 were also included for analysis, along with rs17728338, due to its having been linked to other autoimune disease in separate studies.

### 3.3.6 STK32A

While **D5S640** has proven to be the strongest linkage peak in the majority of studies, the Finnish and Scandinavian studies show a linkage peak at **D5S816.** A follow up study in 2007 tested 1404 SNPs across the 16Mb interval 131,895,324–148,053,211 bp in a cohort of Swedish and Norwegian families. **Figure3.5** shows a graph of these SNPs, with chromosomal position on the x axis, and –Log 10 of the P value on the y axis.

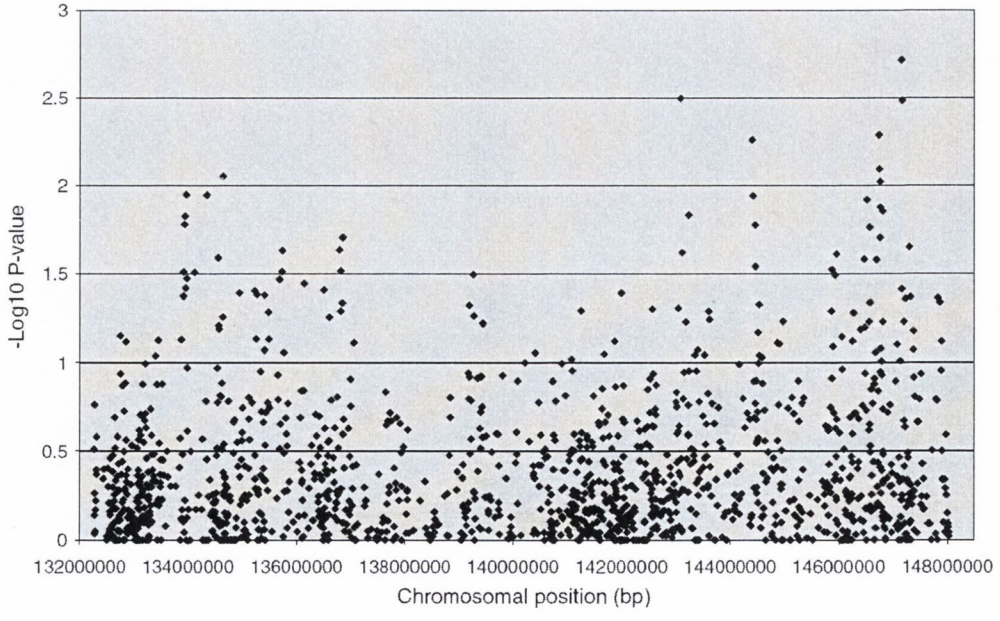



**Figure 3.5 Norwegian and Swedish family based Linkage Study linkage peaks. Source: Adamovic *et al.* 2008.**

59 of these1404 SNPs have an uncorrected p-value of < 0.05, and 6 of them are located within the 3' region of the STK32A gene (146,614,579-146,763,768).

In order to determine whether this linkage peak represents either the coeliac 2 locus, or a second genetic factor predisposing towards the disease, SNPs were chosen to haplotype tag the 3' region of STK32A, from 146,684,633-146,754,589. 6 SNPs were selected, using the data contained in the HapMap project, and the Haploview software package. rs4705038, rs11955126, rs9325041, rs17482628, rs10041912 and rs17548933 were identified as htSNPs which would tag all variations of MAF of 0.05, with $r^2$ cutoff of 0.8 (mean $r^2$ of 0.914).

### 3.3.7 Allele, Genotype and Carrier Status frequencies and disease association testing of genotyped polymorphisms in TNIP1 associated SNPs

All 9 selected SNPs were analyzed for allelic, genotypic and carrier status, and the results, along with Odds Ratios and 95% confidence intervals, are tabulated on tables **3.6** to **3.9.** Both of the selected coding SNPs proved to be invariant in the Irish population, as did 3 of the suspected regulatory SNPs, and genotyping was discontinued once this became apparent.

Significant associations were observed between two of the polymorphisms and disease status. rs8177834 was associated with the disease at the allelic level (Fisher's p = 0.003725), the genotypic level (Fisher's P = 0.0092) and at the level of carrier status, with an association observed between possession of the G allele and the disease state (Fisher's P= 0.04023). It was also observed, that the G allele exhibited a dosage effect, and that homozygosity for the G allele at this locus showed an even stronger association (Fisher's P = 0.009655).

rs2017638 also evidenced association with disease at the carrier level. Homozygosity for the A allele at this locus was associated with the disease state (Fisher's P = 0.02356).

**Table 3.6 Allele Frequencies of TNIP1 SNPs in an Irish Population. rs8177834, a SNP which may be involved in gene regulation shows association with coeliac disease (p = 0.0037).**

| SNP | Alleles | Case Maj | Ctrl Maj | Case Mi | Ctrl Mi | P | Odds Ratio | 95% CI |
|---|---|---|---|---|---|---|---|---|
| rs8177834 | G/A | 0.924 | 0.891 | 0.076 | 0.109 | 0.0037 | 1.49 | 1.03-2.15 |
| rs2017638 | A/G | 0.558 | 0.523 | 0.442 | 0.477 | 0.0667 | 1.15 | 0.94-1.42 |
| rs868531 | G/A | 0.786 | 0.776 | 0.214 | 0.224 | 0.5293 | 1.06 | 0.83-1.36 |
| rs17728338 | G/A | 0.941 | 0.945 | 0.059 | 0.055 | 0.6855 | 1.08 | 0.69-1.68 |

**Table 3.7 Genotype Frequencies of TNIP1 SNPs in an Irish Population rs8177834, a SNP which may be involved in gene regulation shows association with coeliac disease (p = 0.0092).**

| SNP | Alleles | Case AA | Ctrl AA | Case AG | Ctrl AG | Case GG | Ctrl GG | P Fishers |
|---|---|---|---|---|---|---|---|---|
| rs8177834 | G/A | 0.003 | 0.014 | 0.146 | 0.188 | 0.851 | 0.797 | 0.0092 |
| rs2017638 | G/A | 0.325 | 0.270 | 0.464 | 0.504 | 0.210 | 0.226 | 0.07747 |
| rs868531 | G/A | 0.057 | 0.049 | 0.312 | 0.349 | 0.630 | 0.601 | 0.54729 |
| rs17728338 | G/A | 0.008 | 0.002 | 0.101 | 0.106 | 0.891 | 0.891 | 0.68465 |

**Table 3.8 A Dominant/G Recessive Model of TNIP1 SNPs in an Irish Population rs8177834, a SNP which may be involved in gene regulation shows association with coeliac disease (p = 0.0097).**

| SNP | Case (AA)+(AG) | Case (GG) | Ctrl (AA)+(AG) | Ctrl (GG) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| rs8177834 | 0.149 | 0.851 | 0.202 | 0.797 | 0.0097 | 1.45 | 1.10-1.92 |
| rs2017638 | 0.790 | 0.210 | 0.774 | 0.226 | 0.5242 | 1.10 | 0.85-1.41 |
| rs868531 | 0.370 | 0.630 | 0.299 | 0.601 | 0.2798 | 1.13 | 0.91-1.40 |
| rs17728338 | 0.109 | 0.891 | 0.109 | 0.891 | 1 | 1 | 0.72-1.40 |

**Table 3.9 G Dominant/A Recessive Model of TNIP1 SNPs rs8177834 and rs2017638, SNPs which may be involved in gene regulation both show association with coeliac disease (p = 0.0402 and P = 0.0236 respectvely).**

| SNP | Case (GG)+(AG) | Case (AA) | Ctrl (GG)+(AG) | Ctrl (AA) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| rs8177834 | 0.997 | 0.003 | 0.986 | 0.014 | 0.0402 | 4.71 | 0.97-23.1 |
| rs2017638 | 0.675 | 0.325 | 0.730 | 0.270 | 0.0236 | 1.30 | 1.04-1.63 |
| rs868531 | 0.943 | 0.057 | 0.951 | 0.049 | 0.4811 | 1.17 | 0.74-1.85 |
| rs17728338 | 0.992 | 0.008 | 0.998 | 0.002 | 0.1134 | 4.02 | 0.73-22.1 |

**Table 3.10 Irish Case and Control Numbers by SNP for TNIP1**

| SNP | N Case | N Ctrl |
|---|---|---|
| rs8177834 | 576 | 972 |
| rs2017638 | 571 | 967 |
| rs868531 | 573 | 971 |
| rs17728338 | 576 | 933 |

### 3.3.8 Allele, Genotype and Carrier Status frequencies and disease association testing of genotyped polymorphisms in TNIP1 associated SNPs in an independent Italian population

Due to the observed association of rs8177834 and rs2017638 with CD in the Irish population, it was decided that replication of the effect would be attempted in an independent Italian population. rs868531 was also included for analysis, despite its lack of association in the Irish population, as it was deemed possible that as a potential regulatory site it may work in concert with one or both of the other SNPs as part of enhancer or repressor element whose effect would only be apparent on the haplotypic level.

The three SNPs were analyzed for allelic, genotypic and carrier status, and the results, along with Odds Ratios and 95% confidence intervals, are tabulated on tables **3.11** to **3.14.** None of the selected SNPs reached significance in the Italian population.

**Table 3.11 Allele Frequencies of TNIP1 SNPs (Italian) None of the SNPs tested show association with coeliac disease, though this may be due to relatively lower power caused by the reduced sample size.**

| SNP | Alleles | Case Maj | Ctrl Maj | Case Mi | Ctrl Mi | P | Odds Ratio | 95% CI |
|---|---|---|---|---|---|---|---|---|
| rs8177834 | G/A | 0.891 | 0.862 | 0.109 | 0.138 | 0.0919 | 1.30 | 0.96-1.7 |
| rs2017638 | G/A | 0.549 | 0.513 | 0.451 | 0.487 | 0.1685 | 1.16 | 0.95-1.41 |
| rs868531 | G/A | 0.788 | 0.781 | 0.212 | 0.219 | 0.758 | 1.04 | 0.82-1.33 |

**Table 3.12 Genotype Frequencies of TNIP1 SNPs (Italian) ) None of the SNPs tested show association with coeliac disease, though this may be due to relatively lower power caused by the reduced sample size.**

| SNP | Alleles | Case AA | Ctrl AA | Case AG | Ctrl AG | Case GG | Ctrl GG | P |
|---|---|---|---|---|---|---|---|---|
| rs8177834 | G/A | 0.014 | 0.014 | 0.190 | 0.248 | 0.796 | 0.738 | 0.1421 |
| rs2017638 | G/A | 0.303 | 0.261 | 0.491 | 0.503 | 0.205 | 0.235 | 0.1685 |
| rs868531 | G/A | 0.045 | 0.035 | 0.333 | 0.367 | 0.621 | 0.597 | 0.5022 |

**Table 3.13 A Dominant/G Recessive Model of TNIP1 SNPs (Italian) ) None of the SNPs tested show association with coeliac disease, though this may be due to relatively lower power caused by the reduced sample size.**

| SNP | Case (AA)+(AG) | Case (GG) | Ctrl (AA)+(AG) | Ctrl (GG) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| rs8177834 | 0.204 | 0.796 | 0.262 | 0.738 | 0.06353 | 1.39 | 0.99-1.94 |
| rs2017638 | 0.795 | 0.205 | 0.765 | 0.235 | 0.3405 | 1.19 | 0.85-1.68 |
| rs868531 | 0.379 | 0.621 | 0.403 | 0.597 | 0.5089 | 1.11 | 0.83-1.48 |

**Table 3.14 G Dominant/A Recessive Model of TNIP1 SNPs (Italian) ) None of the SNPs tested show association with coeliac disease, though this may be due to relatively lower power caused by the reduced sample size.**

| SNP | Case (GG)+(AG) | Case (AA) | Ctrl (GG)+(AG) | Ctrl (AA) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| rs8177834 | 0.986 | 0.014 | 0.986 | 0.014 | 1 | 1 | 0.30-3.31 |
| rs2017638 | 0.697 | 0.303 | 0.739 | 0.261 | 0.2006 | 1.23 | 0.90-1.68 |
| rs868531 | 0.955 | 0.045 | 0.965 | 0.035 | 0.4682 | 1.30 | 0.63-2.67 |

**Table 3.15 Italian Case and Control Numbers by SNP for TNIP1.**

| SNP | N Case | N Ctrl |
|---|---|---|
| rs8177834 | 357 | 428 |
| rs2017638 | 356 | 421 |
| rs868531 | 354 | 430 |

### 3.3.9 Cochran-Mantel-Haenszel test for significance of TNIP1 associated SNPs in the combined Irish and Italian populations

Having analyzed the TNIP1 SNPs in both populations separately, we pooled the datasets and performed the 2x2xk (Allelic) and IxJxK (Genotypic) Cochran-Mantel-Haenszel tests for stratified datasets. The results can be seen in tables **3.21** and **3.22**.

Analysis of both populations at the allelic level revealed significant association of both rs8177834 and rs2017638 with the disease state (P = 0.0008218 and P = 0.01952 respectively), with the minor alleles of both SNPs showing a protective effect consistent with the observations in the Irish dataset.

Analysis of the combined populations at the genotype level revealed significance only with rs8177834 (P =0.001226) a result also consistent with what had been observed in the Irish population.

**Table 3.16 Cochran-Mantel-Haenszel test for 2x2xK stratified datasets in both the Irish and Italian populations shows associations with both rs8177834 and rs2017638, both of which may be involved in gene regulation (P = 0.00082 and P =0.01952 respectively).**

| SNP | Allele | P | Odds Ratio | 95% CI | Standard Error |
|---|---|---|---|---|---|
| rs8177834 | A | 0.00082 | 0.7142 | 0.586-0.8703 | 0.1009 |
| rs2017638 | G | 0.01952 | 0.8684 | 0.7714-0.9776 | 0.06044 |
| rs868531 | G | 0.4687 | 0.9485 | 0.8223-1.094 | 0.0729 |

**Table 3.17 Cochran-Mantel-Haenszel test for IxJxK stratified datsets in both the Irish and Italian populations shows associations with rs8177834 which may be involved in gene regulation (P = 0.001226).**

| SNP | P |
|---|---|
| rs8177834 | 0.001226 |
| rs2017638 | 0.5404 |
| rs868531 | 0.7602 |

### 3.3.10 Allele, Genotype and Carrier Status frequencies and disease association testing of genotyped polymorphisms in STK32A 3' htSNPs

All 6 selected tgSNPs were analyzed for allelic, genotypic and carrier status, and the results, along with Odds Ratios and 95% confidence intervals, are tabulated on tables **3.18** to **3.21. Table 3.28** shows Haplotype reconstruction, and analysis, of these SNPs.

None of the SNPs evidenced association under any of the tested models.

**Table 3.18 Genotype Frequencies of STK32A 3' tgSNPs in the Irish population reveals no significant association with coealic disease.**

| SNP | Alleles | Case Maj | Ctrl Maj | Case Mi | Ctrl Mi | P | Odds Ratio | 95% CI |
|---|---|---|---|---|---|---|---|---|
| rs4705038 | C/A | 0.868 | 0.840 | 0.132 | 0.160 | 0.11620 | 1.25 | 0.84-1.87 |
| rs11955126 | C/T | 0.833 | 0.817 | 0.167 | 0.183 | 0.43847 | 1.12 | 0.76-1.63 |
| rs9325041 | T/C | 0.592 | 0.581 | 0.408 | 0.419 | 0.71696 | 1.05 | 0.78-1.40 |
| rs17482628 | A/G | 0.667 | 0.693 | 0.333 | 0.307 | 0.30760 | 1.13 | 0.83-1.54 |
| rs10041912 | T/C | 0.869 | 0.845 | 0.131 | 0.155 | 0.23245 | 1.01 | 0.66-1.54 |
| rs17548933 | A/G | 0.884 | 0.868 | 0.116 | 0.132 | 0.36203 | 1.16 | 0.75-1.78 |

**Table 3.19 Allele Frequencies of STK32A 3' tgSNPs in the Irish population reveals no significant association with coealic disease.**

| SNP | Alleles | Case nn | Ctrl nn | Case Nn | Ctrl Nn | Case NN | Ctrl NN | P |
|---|---|---|---|---|---|---|---|---|
| rs4705038 | C/A | 0.014 | 0.065 | 0.234 | 0.275 | 0.751 | 0.702 | 0.11547 |
| rs11955126 | C/T | 0.025 | 0.029 | 0.283 | 0.308 | 0.692 | 0.663 | 0.41934 |
| rs9325041 | T/C | 0.173 | 0.168 | 0.471 | 0.503 | 0.356 | 0.330 | 0.70104 |
| rs17482628 | A/G | 0.096 | 0.097 | 0.473 | 0.419 | 0.431 | 0.483 | 0.28664 |
| rs10041912 | T/C | 0.012 | 0.020 | 0.239 | 0.269 | 0.749 | 0.711 | 0.20606 |
| rs17548933 | A/G | 0.009 | 0.017 | 0.212 | 0.231 | 0.778 | 0.752 | 0.35087 |

**Table 3.20 Major Dominant/Minor Recessive Model of STK32A SNPs in the Irish population reveals no significant association with coealic disease.**

| SNP | Case (NN)+(Nn) | Case (nn) | Ctrl (NN)+(Nn) | Ctrl (nn) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| rs4705038 | 0.986 | 0.014 | 0.022 | 0.978 | 0.4311 | 1.6 | 0.53-4.86 |
| rs11955126 | 0.975 | 0.025 | 0.028 | 0.972 | 0.8227 | 1.15 | 0.46-2.86 |
| rs9325041 | 0.837 | 0.173 | 0.833 | 0.167 | 0.922 | 1.02 | 0.62- 1.67 |
| rs17482628 | 0.905 | 0.095 | 0.903 | 0.097 | 1 | 1.02 | 0.62-1.66 |
| rs10041912 | 0.988 | 0.012 | 0.980 | 0.020 | 0.3935 | 1.66 | 0.51-5.78 |
| rs17548933 | 0.991 | 0.009 | 0.983 | 0.017 | 0.5256 | 1.78 | 0.48-7.21 |

**Table 3.21 Major Recessive/Minor Dominant Model of STK32A SNPs in the Irish population reveals no significant association with coealic disease.**

| SNP | Case (nn)+(Nn) | Case (NN) | Ctrl (nn)+(Nn) | Ctrl (NN) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| rs4705038 | 0.249 | 0.751 | 0.298 | 0.702 | 0.1435 | 1.28 | 0.93-1.76 |
| rs11955126 | 0.308 | 0.692 | 0.337 | 0.663 | 0.4313 | 1.14 | 0.84-1.56 |
| rs9325041 | 0.644 | 0.356 | 0.671 | 0.329 | 0.4849 | 1.13 | 0.83-1.53 |
| rs17482628 | 0.570 | 0.430 | 0.517 | 0.483 | 0.16 | 1.24 | 0.92-1.65 |
| rs10041912 | 0.251 | 0.749 | 0.289 | 0.711 | 0.2512 | 1.22 | 0.89-1.68 |
| rs17548933 | 0.222 | 0.778 | 0.248 | 0.752 | 0.4432 | 1.16 | 0.83-1.61 |

**Table 3.22 Irish Case and Control Numbers by SNP for STK32A.**

| SNP | N Case | N Controls |
|---|---|---|
| rs4705038 | 414 | 356 |
| rs11955126 | 396 | 347 |
| rs9325041 | 393 | 346 |
| rs17482628 | 418 | 329 |
| rs10041912 | 419 | 346 |
| rs17548933 | 419 | 355 |

### 3.3.11 Haplotype analysis of TNIP1 associated SNPs

Haplotype reconstruction is an invaluable tool in the analysis of potential regulatory sequences. Enhancers, repressors, or other regulatory elements are comprised of multiple regions which bind proteins across a single chromosome, and reconstruction and subsequent analysis of haplotypes estimates which variants are present on each chromosome.

Several Haplotype reconstructions were performed for the TNIP1 associated SNPs in the Irish population; a 4 SNP haplotype (**Table 3.23**), a 3 SNP haplotype, utilising all 3 potential regulatory SNPs, (**Table 3.24**) and a 2 SNP Haplotype, both potential regulatory SNPs which evidenced association with disease (**Table 3.25**).

All three reconstructions exhibited significant global P values (0.03, 0,01 and 0,01 respectively) with the GA haplotpye also showing individual significance in the 2 SNP reconstruction (P = 0.03739).

**Table 3.23 TNIP1 4 SNP Haplotype in the Irish population; reconstructed from 3 potential regulatory SNPs, and 1 additional SNP which had shown association in autoimmune disease in a previous study. No single haplotype is significant, although a global P of 0.03 is observed.**

| Haplotype | F(Case) | F(Ctrl) | P | Odds Ratio | 95% CI |
|---|---|---|---|---|---|
| GGGG | 0.350 | 0.361 | 0.66870 | 0.95 | 0.77-1.19 |
| GAGG | 0.333 | 0.294 | 0.09751 | 1.21 | 0.97-1.5 |
| GAGA | 0.115 | 0.103 | 0.48170 | 1.12 | 0.81-1.6 |
| GGGA | 0.069 | 0.080 | 0.35440 | 0.84 | 0.56-1.25 |
| GAAG | 0.0562 | 0.052 | 0.69690 | 1.10 | 0.69-1.73 |

Overall P= 0.03

**Table 3.24 TNIP1 3 SNP Haplotype (rs8177834, rs2017638, rs868531) in the Irish population; reconstructed from 3 potential regulatory SNPs. No single haplotype is significant, although a global P of 0.01 is observed.**

| Haplotype | F(Case) | F(Control) | P | Odds Ratio | 95% CI |
|---|---|---|---|---|---|
| GAG | 0.397 | 0.353 | 0.07869 | 1.21 | 0.98-1.50 |
| GGG | 0.340 | 0.350 | 0.7288 | 0.96 | 0.77-1.120 |
| GAA | 0.112 | 0.102 | 0.5493 | 1.11 | 0.79-1.54 |
| GGA | 0.076 | 0.086 | 0.4713 | 0.87 | 0.59-1.27 |

Overall P= 0.01

**Table 3.25 TNIP1 2 SNP Haplotype (rs8177834, rs2017638) in the Irish population; reconstructed from 2 potential regulatory SNPs. The GA haplotype shows significance (p = 0.03739) in addition to a global P = 0.01.**

| Haplotype | F (Case) | F (Ctrl) | P Value | Odds Ratio | 95% CI |
|---|---|---|---|---|---|
| GA | 0.509 | 0.453 | 0.03739 | 1.25 | 1.01-1.53 |
| GG | 0.416 | 0.438 | 0.39250 | 0.91 | 0.740-1.13 |
| AA | 0.049 | 0.071 | 0.09406 | 0.68 | 0.43-1.07 |
| AG | 0.027 | 0.039 | 0.21060 | 0.68 | 0.36-1.23 |

Overall P= 0.01

### 3.3.12 Haplotype analysis of TNIP1 associated SNPs in the Italian Population

Several Haplotype reconstructions were performed for the TNIP1 associated SNPs in the Italian population; a 3 SNP haplotype, utilising all 3 potential regulatory SNPs (Table 3.26) and a 2 SNP Haplotype, comprising both potential regulatory SNPs which evidenced association with disease in the Irish population, (Table 3.27).

Although the Italian Haplotypes did not evidence significance, this is likely due to reduced power due to the comparatively small sample size, and it is worth noting that the estimated 2 SNP Haplotype frequencies are virtually identical between the 2 populations.

**Table 3.26 3 SNP Italian Haplotype (rs8177834, rs2017638, rs868531); reconstructed from 3 potential regulatory SNPs. No single haplotype is significant, nor is the test significant globally.**

| Haplotype | F(Case) | F(Control) | P | Odds Ratio | 95% CI |
|---|---|---|---|---|---|
| GAG | 0.368 | 0.333 | 0.2767 | 1.18 | 0.88-1.59 |
| GGG | 0.330 | 0.338 | 0.8430 | 0.97 | 0.72-1.31 |
| GAA | 0.123 | 0.102 | 0.3962 | 1.13 | 0.77-1.91 |
| GGA | 0.070 | 0.091 | 0.3232 | 0.77 | 0.45-1.29 |
| AAG | 0.046 | 0.054 | 0.5572 | 0.82 | 0.42 1.58 |
| AGG | 0.044 | 0.053 | 0.5284 | 0.80 | 0.40-1.57 |

Overall P= 0.08

**Table 3.27 2 SNP Haplotype (rs8177834, rs2017638); reconstructed from 2 potential regulatory SNPs. No single haplotype is significant, nor is the test significant globally.**

| Haplotype | F (Case) | F (Control) | P Value | Odds Ratio | 95% CI |
|---|---|---|---|---|---|
| GA | 0.491 | 0.437 | 0.1298 | 1.25 | 0.94-1.65 |
| GG | 0.401 | 0.427 | 0.4673 | 0.90 | 0.67-1.20 |
| AA | 0.057 | 0.071 | 0.3992 | 0.78 | 0.43-1.40 |
| AG | 0.052 | 0.065 | 0.4263 | 0.78 | 0.41- 1.43 |

Overall P= 0.08

**Table 3.28 Haplotype analysis of STK32A 3' tgSNPs(rs11955126, rs17482628, rs4705038, rs10041912, rs9325041, rs17548933). Haplotype reconstruction and analysis of the 6 SNPs used to tag the 3' region of STK32A in the Irish population. No haplotype shows association with coeliac disease, nor is the test significant globally.**

| Haplotype | Fcase | FCtrl | p | OR | 95% |
|---|---|---|---|---|---|
| CGCTTA | 0.330 | 0.307 | 0.4965 | 1.12 | 0.81-1.53 |
| CACTTA | 0.253 | 0.252 | 0.9479 | 1.01 | 0.72- 1.42 |
| CACTCA | 0.221 | 0.227 | 0.8322 | 0.96 | 0.68-1.37 |
| TAACCG | 0.085 | 0.104 | 0.4438 | 0.82 | 0.50-1.36 |

Overall P = 0.76

### 3.3.13 Homozygous Haplotypes

As the GA Haplotype is associated with the disease state, and genotypic analysis has shown the GG genotype to be associated with disease in rs8177834, and the AA genotype to be associated with disease in rs2017638, we investigated the association of two copies of the GA haplotype with disease status in the Irish and Italian populations. The results can be seen in tables **3.29** and **3.30**.

In both populations, possession of two copies of the GA haplotype shows significant association with disease (p = 0.0004272 in the Irish population, and p = 0.03394 in the Italian population) with a notable increase in risk (OR = 1.56, 95% CI = 1.22-1.99 in the Irish population, and OR = 1.45, 95% CI = 1.03-2.05 in the Italian population).

**Table 3.29 GA/GA Haplotypes is the Irish Population. Comparing the proportions of carriers of two copies of the GA haplotype between the Irish case and control populations reveals a significant difference between them. Carrying two copies of the haplotype gives an Odds Ratio of 1.56 (P = 0.0004272).**

| Case | Controls | p | Odds Ratio | 95% CI |
|---|---|---|---|---|
| 0.268 | 0.190 | 0.0004272 | 1.56 | 1.22-1.99 |

**Table 3.30 GA/GA Haplotypes is the Italian Population. Comparing the proportions of carriers of two copies of the GA haplotype between the Italian case and control populations reveals a significant difference between them. Carrying two copies of the haplotype gives an Odds Ratio of 1.451 (P = 0.03394).**

| Case | Controls | p | Odds Ratio | 95% CI |
|---|---|---|---|---|
| 0.247 | 0.185 | 0.03394 | 1.451 | 1.03-2.046 |

### 3.3.14 Patterns of Linkage Disequilibrium

When SNPs evidence an association with disease, there is always the possibility that the causal variant is not the SNP in question, but rather than it is acting as a proxy for another SNP with which it is in LD. As such, we must be wary of taking associations on face value, and examine the known LD structure surrounding associated SNPs.

LD is also important in the context of haplotype tagging SNPs, as it determines which SNPs a tag is informative for, and to what degree.

**Table 3.31** shows the degree of LD which the TNIP1 associated SNPs evidence with nearby SNPs. Table 3.32 does the same for the STK32A tagging SNPs.

**Table 3.31 LD of TNIP1 SNPs, illustrating the LD between tested SNPs and catalogued SNPs in the genomic region.**

| SNP | SNP | $R^2$ | Distance (kb) | Function |
|---|---|---|---|---|
| **rs8177834** | rs9324672 | 0.97 | 10 | TNIP1; Intronic |
| | rs7713223 | 0.72 | 20.7 | TNIP1; Intronic |
| | rs870406 | 0.67 | 9.3 | GPX3; Intronic |
| | rs8177412 | 0.63 | 9.9 | GPX3; 5' UTR |
| | rs8177406 | 0.64 | 10.3 | Near GPX3 |
| | rs8177404 | 0.63 | 10.4 | Near GPX3 |
| | rs1946235 | 0.64 | 10.4 | Near GPX3 |
| | rs1946234 | 0.64 | 10.7 | Near GPX3 |
| | rs2042236 | 0.64 | 11.8 | Near GPX3 |
| **rs2017638** | No SNP with **$R^2 > 0.5$** | | | |
| **rs868531** | No SNP with **$R^2 > 0.5$** | | | |
| **rs17728338** | No SNP with **$R^2 > 0.5$** | | | |

**Table 3.32 LD of STK32A 3' tgSNPs illustrating the LD between tested SNPs and catalogued SNPs in the genomic region.**

| Tagging SNP | SNPs Tagged | $R^2$ |
|---|---|---|
| **rs4705038** | rs2400294 | 0.877 |
| | rs1364371 | 0.935 |
| | rs12514269 | 0.817 |
| | rs10077515 | 0.871 |
| | rs918798** | 0.935 |
| | rs2400295 | 0.935 |
| | rs918797** | 0.871 |
| | rs4705164 | 0.871 |
| | rs4705165 | 1.0 |
| | rs4705166 | 0.86 |
| | rs4705038 | 1.0 |
| | rs11951154 | 0.935 |
| | rs12332485 | 0.817 |
| | rs10079272 | 0.825 |
| | rs6866098 | 0.817 |
| **rs11955126** | rs6884181** | 0.853 |
| | rs11955126 | 1.0 |
| **rs9325041** | rs889829 | 0.892 |
| | rs9325041 | 1.0 |
| **rs17482628** | rs17482628 | 1.0 |
| **rs10041912** | rs10041912 | 1.0 |
| **rs17548933** | rs17548933 | 1.0 |

****denote SNPs within the 3' region of STK32A which evidenced significance in the Swedish and Norwegian populations.**

### 3.3.15 Hardy Weinberg Analysis

We preformed Hardy Weinberg analysis to test for population equilibrium in all tested 5q SNPs, in all assayed populations. Table 3.33 lists the results. Only the cases showed departure from HW Equilibrium in rs17728338, the SNP which showed association with TNIP1 in other autoimmune disease.

**Table 3.33 Hardy Weinberg Analysis of 5q SNPs. Only the Irish case population shows a significant departure from HWE at rs17728338. This departure is only moderately significant, and unlikely to signify selection pressure at this locus.**

| SNP | Case | Control |
|---|---|---|
| rs8177834 Irish Samples | 0.7640 | 0.4124 |
| rs2017638 Irish Samples | 0.1744 | 0.7473 |
| rs868531 Irish Samples | 0.1051 | 0.8533 |
| rs17728338 Irish Samples | **0.0403** | 1.0000 |
| rs8177834 Italian Samples | 0.9148 | 0.9224 |
| rs2017638 Italian Samples | 0.5920 | 0.5403 |
| rs868531 Italian Samples | 1.0000 | 0.1568 |
| rs4705038 | 0.8310 | 0.8447 |
| rs11955126 | 0.8569 | 0.7198 |
| rs9325041 | 0.5997 | 0.5933 |
| rs17482628 | 0.1763 | 0.7922 |
| rs10041912 | 0.5162 | 0.8348 |
| rs17548933 | 0.6367 | 1 |

## 3.4 Discussion

The importance of ECRs as functional genomic regions has been debated since this study was undertaken, with a wide variety of different observations between research groups. Subsequent studies have shown the ability of a moderately conserved ECR to reliably predict the existence of a regulatory element may be less than was envisioned at the time of this study's inception. ECRs certainly have proven to have a predictive function, but it may not be as good an indicator of functionality as early studies suggested.

Investigation of human/rodent genomes utilising the criteria we have applied has had mixed success, with a higher than expected degree of false positives. Deletion of two ECR rich gene desert regions in mouse models , containing 1,243 ECRs (100+ base pairs length, 70% identity) across a 1,511 kilobase region and another 845 kilobase region, produced viable mice with only small differences in levels of gene expression (Nóbrega et al., 2004). Of course, in the case of complex diseases, it is the interplay of multiple, individually minor factors which results in disease, and so this may not be an unexpected result.

Stricter constraints on the definition of an ECR has produced a new classification of conserved region; the ultraconserved region, a region which shows 100% conservation between orthologous regions, and is at least 200bp in length. Identification of regulatory regions via searching for ultraconserved regions between human and mouse, and human and rat, has revealed a number of regions involved in the regulation of developmental genes.

However, even this stricter approach to what constitutes a meaningful ECR predicts functional variants with only reasonable accuracy; an expression study carried out in transgenic mice reported that 45% of identified potential regulatory sequences functioned reproducibly as tissue-specific enhancers of gene expression during embryonic development (Pennacchio *et al.* 2006).

There is clear association with CD shown with the SNPs rs8177834 and rs2017638 in the Irish population, and when the Irish and Italian populations are analysed as a whole, via the Cochran-Mantel-Haenszel test. That an association with these SNPs was not evidenced independently in the Italian population is likely due to a lack of power to detect the effect, caused by the relatively small Italian sample size. Allele

and genotype frequencies are similar, as are haplotype frequencies, and the Allelic model comes close to reaching significance for rs8177834 in the Italian population.

Looking at the carrier status model in the Irish population suggests that the homozygous state for the major allele at both loci carries the additional risk of disease, with possession of only one, or zero, copies of the allele being protective. The same trend is observed for rs8177834 in the Italian population; testing for carrier status effects just falling short of significance (P = 0.06353, OR = 1.39, 95%CI = 0.99-1.94).

In order to understand how the homozygous state may lead to increased risk of disease on the molecular level, we must consider the nature of the SNPs. Both SNPs were selected for study based on the possibility of their being part of a transcription factor mediated regulatory site. rs8177834 lies within a c-MYB binding motif, while rs2017638 is located within an NFKBQ6 motif.

To examine rs8177834 first: the c-MYB recognised consensus binding site general format, aligned with the sequence surrounding rs8177834 is shown below. (Note the motif is described in relation to the opposite strand to the assay used for rs8177834, hence C instead of G is shown as the major allele).

| | | |
|---|---|---|
| Consensus binding motif | 5'-YAAC[GT]G-3' | |
| Sequence flanking SNP | 5'-CAA**C**GG-3' | (SNP Highlighted) |

rs8177834 lies at the centre of a CMYB binding motif. Possession of the major allele (G) is integral to the formation of the TFBS, and consequently, the binding of the CMYB transcription factor. The A allele disrupts the binding motif, and would act to reduce the efficacy of CMYB binding (assuming the binding site is a functional one).

CMYB has a number of recognised roles as a transcription factor. Inducible c-myb knockouts have shown that its expression is essential for the development of multiple hematopoietic lineages, including T cell lines (Sandberg et al. 2005). Its action, in conjunction with GATA-3 is required for Th2 cell maturation in primary human peripheral blood T cells, and the formation of Th2 memory cells (Nakata et al. 2010). cMYB is also required for the survival of CD4(+)CD8(+) double-positive (DP) thymocytes and pro-B cell differentiation (Fahl *et al.* 2009). In addition to T-Cell maturation, c-Myb protein is also thought to have an important regulatory function

in other cell types, especially epithelial cells including blocking of differentiation (Ramsay *et al.* 2005).

The Cmyb binding site is located within 3' UTR of TNIP1, and may represent part of an enhancer/repressor or even post transcriptional regulatory element. Given TNIP1s role in inhibiting the action of NF-kappa-B-dependent gene expression, and the fact that possession of the major allele (which maintains the binding motif) results in increased risk of disease, it would seem to follow that this binding site represents a Cmyb controlled repressor of TNIP1.

rs2017638 is located at the periphery of an NFKBQ6 binding site, and unlike the putative c-MYB binding site, it does not represent an important 'make or break' base pair change. However, it was considered functionally relevant due to the central role NF-KappaB plays in the immune system, and the extreme scarcity of SNPs located within NF-KappaB binding sites within the entire 5Mb region (there were 2).

Consensus binding motif NGGGGAMTTTCCNN

Sequence flanking SNP CAGGGACTTCCCT**A**GGC (SNP Highlighted)

rs201638 is located in the Intergenic region 5' of TNIP1 and, once again possession of the major allele confers additional risk. As TNIP1 negatively regulates NF-kappa-B, it is possible that this site comprises an NF-kappa-B activated repressor of TNIP1 activation, as part of a negative feedback loop involved in the perpetuation of the immune response.

Haplotype reconstruction in the Irish population reveals an association between the haplotype comprising the major alleles of both SNPs and the disease state. Similar analysis in the Italians does not provide significant association, however the frequencies of the reconstructed haplotypes are almost identical, and the lack of significant association is, once again, most likely due to a lack of statistical power.

Increased risk associated with this haplotype suggests that both sites may work in concert as part of a single negative regulator of TNIP1 activation. Individuals with two copies of the risk haplotype are at significantly increased risk of the disease.

TNIP1's binding partner, TNFAIP3, has shown association with coeliac disease (see Chapter 6), so there is already evidence to suggest involvement of this pathway as an underlying mechanism of coeliac disease. TNFAIP3 has also shown association with

numerous other auto-immune diseases (Graham *et al.* 2008, De Jager 2009Nair *et al.* 2009, Prahalad *et al.* 2009, Wang *et al.* 2010).

TNIP1 showed suggestive, but not conclusive, association with coeliac disease in GWAS II (see Chapter 8). Additionally, it has shown association with several AI diseases, including Rheumatoid Arthritis, Psoriasis and SLE in European, Japanese and Chinese Han Chinese patients (Elder 2009, Gateva *et al.* 2009, Kawasaki et al. 2010, Orozco et al. 2010).

Using HapMap Phase II data to examine the pattern of LD in the TNIP1 region via haploview reveals that of the two associated SNPs, only rs8177834 shows strong linkage with other SNPs. rs9324672, located 10kb away within an intron in TNIP1 has an $r^2$ value of 0.97 with rs8177834, effectively allowing the SNPs to act as proxies for one another. We must keep in mind then, that is it is also possible the effect observed is due to the biological activity of the intronic SNP. rs8177834 also shows LD with 8 other nearby SNPs, but none in excess of an $r^2 > 0.8$. None of these SNPs are known to have a functional role, or are associated with any disease.

No association has been observed with STK32A, despite it showing signs of association in the Norwegian/Swedish family study. As **Table 3.4** shows, population analysis in the Scandinavian populations (Swedish, Finnish, Norwegian) in the past has shown strongest association with the microsatellite marker **D5S816** as opposed to **D5S2014**, which has shown association with the majority of other populations. It is possible that two genetic factors on chromosome 5 are contributing towards the disease state, each with different relative strengths in the Scandinavian and more southern European populations. Also for consideration, there may exist a population specific risk factor, in the form of STK32A, which predisposes towards disease in the Scandinavian populations, but which does not exist outside of them.

In conclusion, our studies into the 5q 31-33 region, which has long shown linkage with CD, has revealed an association with 2 SNPs flanking theTNIP1 gene with coeliac disease in an Irish case control population. The result was not replicated in an Italian case control population, most likely due to the small size of the sample (case =357, control = 430), and corresponding lack of power. Analysis of both populations via the CMH test for association in stratified samples revealed an association with both SNPs. A haplotype formed by the two SNPs also showed association, with individuals possessing two copies of the haplotype being at increased risk of disease. The SNPS lie within putative TFBS, which may be indicative of a functional regulator of

transcription. We also investigated the 3' region of STK32A in an Irish case control population, which had shown association with CD in the Norwegian population, but this observation was not replicated.

# CHAPTER IV

# Coeliac Disease Genomewide Association Study Replication

## 4.1 Introduction

### 4.1.1 Early Genome Wide Association Studies

With the advancement of technology to the point where massively multiplex genotyping reactions became cost effective, a new type of study design began to become prevalent. Hundreds of thousands of SNPs could now be genotyped in an individual as part of a single reaction, allowing much greater coverage of a genome. In addition, the human genome project's publication of the HapMap dataset made public a panel of tagging SNPs across the entire genome. The conjunction of these two advances brought about commercially available products which could be used to perform whole genome scans.

The first GWAS was published in 2005, and tested 116,204 SNPs in a population of 96 patients with age-related macular degeneration, and 50 controls. The population participants self classified themselves as "white, not of Hispanic origin." The study revealed association with a non-synonymous coding SNP in complement factor H gene, (CFH), located in a region of chromosome 1 which had shown linkage to Age-related macular degeneration in family studies (Klein *et al.* 2005).

A second study in 2006 tested 97,824 SNPs in a population of 96 Southeast Asian patients with wet age-related macular degeneration, and 130 population matched controls (Dewan *et al.* 2006). This study revealed a polymorphism in the HTRA1 gene which, it transpired, is the major genetic risk factor for this condition.

A variety of other GWAS studies followed, examining case control populations for a variety of complex diseases. Despite larger populations than the studies described above, they were not as successful in terms of the proportion of the attributable heritable risk identified.

This is due to the different underlying nature of the diseases; age related macular degeneration, both wet and dry, are contributed to genetically but a few factors of relatively strong effect. The majority of complex autoimmune diseases, however, seem to be governed by a myriad of genetic elements, each with relatively small effect, which requires additional statistical power to uncover.

### 4.1.2 A Genome Wide Association Study in the UK population

In 2007, Professor David Van Heel (Centre for Gastroenterology, Institute of Cell and Molecular Science, Queen Mary University of London) undertook a genome-wide association study for coeliac disease in the UK population. He tested 310,605 SNPs for association in 778 UK coeliac disease cases, and 1422 controls. The cases were genotyped on the HumanHap300 BeadChip, and the controls on the HumanHap500.

The HumanHap300 BeadChip is capable of genotyping a panel of over 317,000 SNPs. The SNPs in question are selected from Phase I HapMap project data, the SNPs being selected for inclusion in the panel based on their ability to act as haploytpe tag proxies (see section 1.17) and thus accurately and efficiently incorporate the maximum amount of information. SNPs were deemed to be suitable proxies for a second SNP if the $r^2$ between them was at least 0.8, assuming the SNP in question was within 10kb of a known gene or evolutionarily conserved region. Outside of these areas of assumed functionality, the criterion was relaxed to an $r^2$ of 0.7 (Carlson *et al.* 2004).

Under these criteria, eighty percent of all Phase I+II loci which are recorded as possessing a MAF ≥ 0.05 in the CEU population are tagged by one or more of the SNPs on the HumanHap300 BeadChip. As htSNPs are dictated by the patterns of local LD, htSNPs in regions of high LD will tag a greater proportion of the known variation, while less than 80% coverage will be seen in regions of lower LD (See **Figure 4.1**) (Illumina Technical Bulletin).

Due to its design as an aide in indentifying genetic elements in autoimmune conditions, this panel is also further enriched by the addition of 7,300 non-synonymous SNPs, and an increased density of htSNPs in the Major Histocompatibility Complex (MHC) region (Illumina Human Hap300 data sheet).

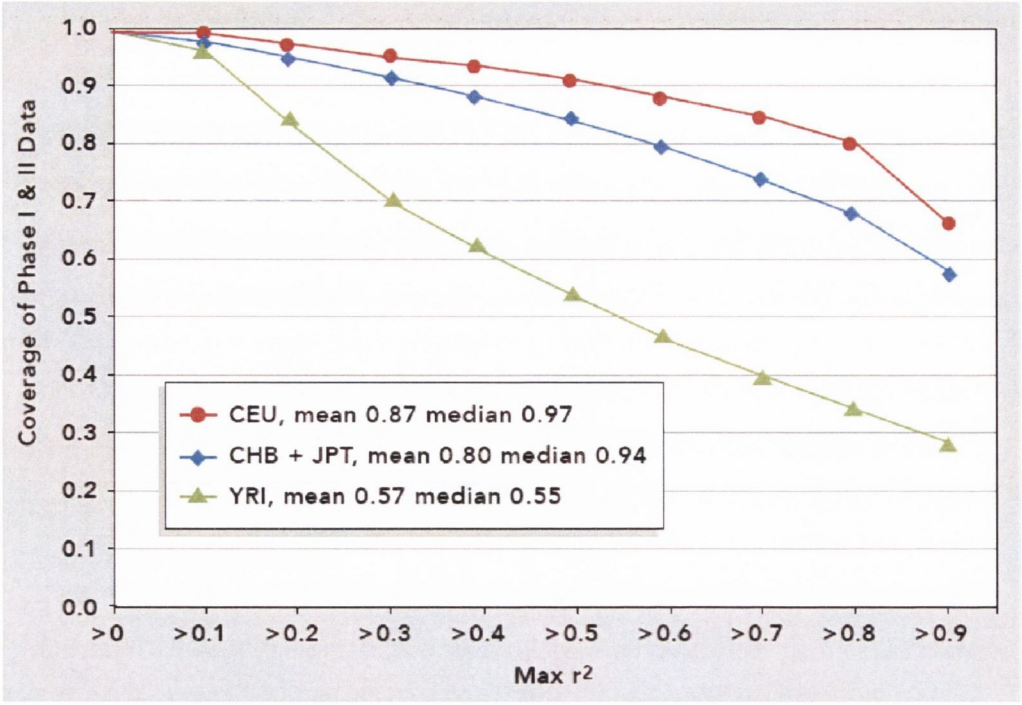



**Figure 4.1 Genome coverage by the HumanHap300 BeadChip SNP panel, showing the $r^2$ of proxy SNPs with SNPs from Phase I an II HapMap SNPs in European (CEU), African (CHB) and Asian (YRI) populations**. Source: Illumina.

HumanHap500 (used for the controls) has a panel selected under similar criteria, but with a larger number of SNPs (500,000).

Outside of the expected HLA association, 56 SNPs were found to have a $P < 10^{-4}$ in the UK population, and following the completion of the GWAS, we worked in conjunction with this British based group, and another group in the Netherlands, led by Professor Cisca Wijmenga, to genotype some of the most significant results from the GWAS in three additional independent populations; an Irish Case Control population, a Dutch case control population, and a second, independent, British Case Control population. The aim of this collaboration was to establish a study with additional power, provided by the vastly increased sample size afforded by the inclusion of the 3 additional collections, with an eye to overcoming the 'noise' generated by the massive degree of multiple testing involved in carrying out a GWAS.

This joint study was comprised of two distinct phases; an initial, small scale attempt at replication of a handful of some of the most strongly associated SNPs, and a later, more comprehensive study involving the genotyping of a much larger set of associated SNPs. There was also a third, adjunct phase, in which replication of the observed effects were attempted in fifth and sixth independent populations.

To reflect the tripartite nature of the follow up, this Chapter is sub-divided into three parts;

**Part 1**: Initial Replication of observed GWAS effects in independent populations. (Sub-Section a).

**Part 2**: Comprehensive follow up of GWAS in indepdendent Irish, British and UK populations, using the Illumina Goldengate Platform(Sub-Section b).

**Part 3**: Replication of observed associations from the Goldengate Platform study in an independent Italian population (Sub-Section c).

### 4a.2 Objectives

The aims of Part One are:

1. To determine a subset from the panel of 56 SNPs ($P < 10^{-4}$) in which to attempt confirmation in an Irish population.
2. Genotyping of selected panels in the Irish population.
3. Carry out statistical tests to determine any association between genotypes or possession of alleles and the disease state.
4. Where appropriate, estimation of haplotypes, and test for haplotype association with disease state.
5. Examination of the LD structure around SNPs which showed consistent association across populations; selection and genotyping of additional SNPs in the Irish population to more fully detail the haplotype structure of these regions.
6. Joint analysis of Irish, British and Dutch SNPs via Cochran-Mantel-Haenszel test.
7. Comparison of allele and genotype frequencies, along with degree of association, of associated SNPs across the 3 main populations.

### 4a.3 Results

#### 4a.3.1 SNP Selection

Working in conjunction with the UK group which carried out the initial study, we aimed to identify a short list of the most significant and relevant of the 56 associated SNPs, which evidenced a $P < 10^{-4}$ in the British population, and attempt to replicate these associations in independent Irish and Italian populations.

Of these 56 SNPs, several associations occurred in, and near, the IL-2/IL-21 gene LD block. 4 SNPs in or near the IL-2/IL-21 linkage block were selected for replication in the Irish population, as were 8 SNPs outside of the IL-2/IL-21 area, in promising candidate regions (see **Table 4.1** for information on these SNPs and regions),

**Table 4.1 Non IL-2/IL-21 Regions for replication in the Irish population based on their associations in the UK GWAS study.**

| SNP | Gene | Function |
|---|---|---|
| rs1021621 | **POU2F1** | Transcription Factor. Associated with AI Disease. |
| rs9290678<br>rs6762743 | **ZNF639**<br>**ZNF639** | Zinc Finger Binding Protein. Associated with oesophageal carcinoma. |
| rs10505604 | **TG** | Associated with Autoimmune Disease. |
| rs4469515<br><br>rs1064891 | **SMARCA2**<br><br>**SMARCA2** | Regulates Villin Expression in GI cells. NfKappaB signalling. |
| rs7159238 | **SEL1L** | ER Processing and folding. Associated with AI disease. |
| rs3809982 | **ALPK2** | Kinase. SNP is a missense coding SNP. |

As the amount of DNA available for the Italian population was limited, it was determined that we would forestall genotyping in this population until such time as a clearer pattern of association emerged from analysis of the Irish population, and the Cochran-Mantel-Haenszel test of the Irish, Dutch and British populations. Once consistent associations were replicated in multiple populations, we would then attempt replication in the Italian sample.

The Irish SNPs and Dutch SNPs tested for replication and haplotype tagging in **Section a** were genotyped in the Centre for Gastroenterology, Institute of Cell and Molecular Science, Queen Mary University of London by Dr Karen Hunt, as part of the collaboration between our groups. The Italian SNPs were genotyped by us in the Institute of Molecular Medicine.

### 4a.3.2 Allele frequencies and disease association testing of non IL-2/IL-21 LD block genotyped SNPs

Outside of the IL-2/IL-21 region, several SNPs located near potential candidate genes were genotyped and analysed in the Irish population. **Table 4.2** shows the allele frequencies, P-values and nearby genes. Table 4.3 gives the sample size for each SNP.

None of the SNPs showed significant association in the Irish population at any level (only allelic data shown).

**Table 4.2 Allele Frequencies in non IL-2/IL-21 SNPs, replication in Irish Population. None of these 8 SNPs show significance in the Irish population.**

| SNP | Alleles | Case Maj | Ctrl Maj | Case Mi | Ctrl Mi | P-Value | OR | 95% CI |
|---|---|---|---|---|---|---|---|---|
| rs1021621 | C/T | 0.547 | 0.553 | 0.453 | 0.447 | 0.79 | 0.98 | 0.81-1.16 |
| rs6762743 | C/T | 0.755 | 0.719 | 0.245 | 0.281 | 0.075 | 1.20 | 0.99-1.47 |
| rs9290678 | A/G | 0.740 | 0.714 | 0.240 | 0.286 | 0.061 | 1.20 | 0.98-1.40 |
| rs10505604 | A/G | 0.758 | 0.768 | 0.242 | 0.232 | 0.61 | 0.95 | 0.77-1.16 |
| rs4469515 | A/G | 0.613 | 0.640 | 0.387 | 0.360 | 0.23 | 0.89 | 0.74-1.07 |
| rs1064891 | C/T | 0.595 | 0.601 | 0.405 | 0.399 | 0.79 | 0.86 | 0.72-1.02 |
| rs7159238 | C/T | 0.541 | 0.542 | 0.459 | 0.458 | 0.96 | 0.99 | 0.84-1.18 |
| rs3809982 | C/T | 0.504 | 0.496 | 0.496 | 0.504 | 0.75 | 1.03 | 0.87-1.23 |

**Table 4.3 Irish Case and Control Numbers by SNP for Non IL-2/IL-21 loci**

| SNP | N Case | N Ctrls |
|---|---|---|
| rs1021621 | 478 | 555 |
| rs6762743 | 481 | 553 |
| rs9290678 | 475 | 558 |
| rs10505604 | 480 | 544 |
| rs4469515 | 482 | 538 |
| rs1064891 | 477 | 545 |
| rs7159238 | 474 | 555 |
| rs3809982 | 477 | 540 |

### 4a.3.3 Allele, Genotype and Carrier Status frequencies and disease association testing of IL-2/IL-21 LD block genotyped SNPs

All 4 selected SNPs in the IL-2/IL-21 region were analyzed for allelic, genotypic and carrier status, and the results, along with Odds Ratios and 95% confidence intervals, are tabulated on tables **4.4** to **4.7. Table 4.8** details the sample size, per SNP.

All showed association at the allelic frequency level, and all except rs6822844 also showed association at the genotypic frequency level. In all cases the major allele was associated with disease status.

Examination of carrier allele status shows an increased risk for those carrying the major allele of rs6822844, (P = 0.04518, OR =2.274, 95%CI 1.017-5.498) and decreased risk for those carrying the minor alleles of rs13151961 ( P = 0.01859, OR = 0.7262, 95%CI 0.554-0.9481) and rs6840978 (P = 0.02493, OR = 0.7485, 95%CI 0.5804-0.9642)

**Table 4.4 Allele Frequencies of IL-2/IL-21 LD block genotyped SNPs in the Irish population. A significant association with coeliac disease is shown in all 4 SNPs.**

| SNP | Alleles | Case Maj | Ctrl Maj | Case Mi | Ctrl Mi | P | OR | 95% CI |
|---|---|---|---|---|---|---|---|---|
| rs13151961 | A/G | 0.852 | 0.806 | 0.148 | 0.194 | 0.0056 | 1.39 | 0.99-1.93 |
| rs13119723 | A/G | 0.872 | 0.838 | 0.128 | 0.162 | 0.030 | 1.32 | 0.93-1.87 |
| rs6822844 | G/T | 0.858 | 0.803 | 0.142 | 0.197 | 0.0013 | 1.48 | 1.06-2.07 |
| rs6840978 | C/T | 0.808 | 0.760 | 0.192 | 0.240 | 0.0083 | 1.33 | 0.98-1.79 |

**Table 4.5 Genotype Frequencies of IL-2/IL-21 LD block genotyped SNPs in the Irish population. A significant association with coeliac disease is shown in all SNPs except rs13119723.**

| SNP | Alleles | Case nn | Case nN | Case NN | Ctrl nn | Ctrl nN | Ctrl NN | P |
|---|---|---|---|---|---|---|---|---|
| rs13151961 | A/G | 0.017 | 0.262 | 0.713 | 0.043 | 0.302 | 0.654 | 0.015 |
| rs13119723 | A/G | 0.021 | 0.215 | 0.765 | 0.041 | 0.241 | 0.718 | 0.084 |
| rs6822844 | G/T | 0.017 | 0.250 | 0.733 | 0.038 | 0.316 | 0.645 | 0.005 |
| rs6840978 | C/T | 0.040 | 0.303 | 0.657 | 0.060 | 0.360 | 0.580 | 0.032 |

**Table 4.6 Major Dominant/Minor Recessive Model of IL-2/IL-21 SNPs in the Irish population. A significant association with coeliac disease is shown in rs6822844.**

| SNP | Case (NN)+(Nn) | Case (nn) | Ctrl (nn)+(Nn) | Ctrl (nn) | P | OR | 95% CI |
|---|---|---|---|---|---|---|---|
| rs13151961 | 0.976 | 0.017 | 0.957 | 0.043 | 0.1205 | 1.74 | 0.87-3.70 |
| rs13119723 | 0.980 | 0.021 | 0.959 | 0.041 | 0.06170 | 2.03 | 0.97-4.82 |
| rs6822844 | 0.983 | 0.017 | 0.962 | 0.038 | 0.04518 | 2.27 | 1.02-5.49 |
| rs6840978 | 0.960 | 0.040 | 0.940 | 0.060 | 0.1586 | 1.52 | 0.85-0.27 |

**Table 4.7 Minor Dominant/Major Recessive Model of IL-2/IL-21 SNPs in the Irish population. A significant association with coeliac disease is shown in rs13151961 and rs6840978.**

| SNP | Case (Nn)+(nn) | Case (NN) | Ctrl (Nn)+(nn) | Ctrl (NN) | P | OR | 95% CI |
|---|---|---|---|---|---|---|---|
| rs13151961 | 0.277 | 0.713 | 0.345 | 0.655 | 0.01859 | 0.73 | 0.55-0.95 |
| rs13119723 | 0.235 | 0.765 | 0.282 | 0.718 | 0.08638 | 0.78 | 0.59-1.04 |
| rs6822844 | 0.267 | 0.733 | 0.316 | 0.642 | 0.08849 | 0.79 | 0.60-1.04 |
| rs6840978 | 0.351 | 0.657 | 0.420 | 0.580 | 0.02493 | 0.75 | 0.58-0.96 |

**Table 4.8 Irish Case and Control Numbers by SNP for IL-2/IL-21 SNPs.**

| SNP | N Case | N Controls |
|---|---|---|
| rs13151961 | 473 | 553 |
| rs13119723 | 480 | 556 |
| rs6822844 | 464 | 547 |
| rs6840978 | 472 | 552 |

### 4a.3.4 LD Structure of the IL-2/IL-21 Region

The 4 SNPs which evidenced association with CD in the British and Irish populations are part of the HumanHap300 BeadChip panel, and as such are haplotype tags. However, examination of the genomic region surrounding these SNPs using the haploview software revealed that these SNPs were insufficient to tag the majority of variation which exists in that region.

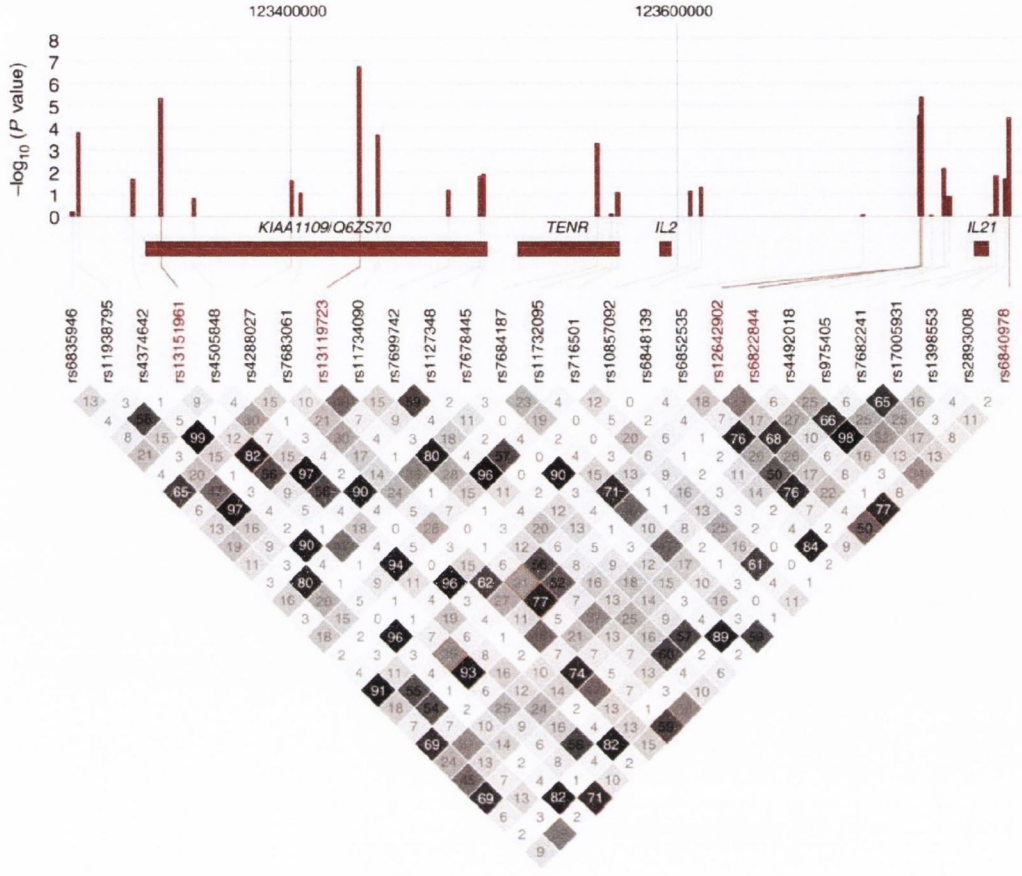
The 480kb region surrounding the SNPs contains two haplotype blocks; a 439kb block and 40kb haplotype block. Using the haploview software and HapMap Phase I data, we determined 10 additional SNPs were required to effectively tag the haplotypes present in the population with a frequency of >5%.

Block one (439kb) is tagged by rs6835946, rs4374642, rs13151961, rs4505848, rs13119723, rs1127348, rs11732095, rs10857092 and rs6822844.

Block 2 (40kb) is tagged by rs4492018, rs975405, rs1398553, rs2893008 and rs6840978. Figure 4.2 shows the patterns of LD in this region.

**Figure 4.2 Patterns of LD in the IL-2/IL-21/KIAA1109/TENR LD Block, based on genotyping data from the Uk and Irish populations.**

Source: van Heel *et al.* 2007.




**4a.3.5 Allele, Genotype and Carrier Status frequencies and disease association testing of Block 1 and Block 2 IL-2/IL-21 LD block genotyped htSNPs**

The remaining htSNPs required to fully characterise the linkage region were genotyped and analyzed for allelic, genotypic and carrier status, and the results, along with Odds Ratios and 95% confidence intervals, are tabulated on tables **4.9** to **4.12. Table 4.13** details the sample size, per SNP.

Two of these SNPs, rs1127348 and rs1398553 showed association at the allelic level (P = 0.0053 and P = 0.015 respectively) with rs1127348 also evidencing association at the genotypic level (P =0.016).

The minor allele appears to confer the additional risk of disease in both rs1127348 and rs1398553.

**Table 4.9 Allele Frequencies in Irish IL-2/IL-21 htSNPs; rs1127348 and rs1398553 show significant associations with coeliac disease (p =0.0053 and p = 0.015).**

| SNP | Alleles | Case Maj | Ctrl Maj | Case Mi | Ctrl Mi | P | OR | 95% CI |
|---|---|---|---|---|---|---|---|---|
| rs6835946 | G/A | 0.686 | 0.665 | 0.314 | 0.335 | 0.32 | 1.10 | 0.85-1.43 |
| rs4374642 | T/C | 0.917 | 0.933 | 0.083 | 0.067 | 0.17 | 1.26 | 0.79-2.00 |
| rs4505848 | A/G | 0.687 | 0.722 | 0.313 | 0.278 | 0.084 | 1.18 | 0.90-1.55 |
| rs1127348 | T/C | 0.783 | 0.831 | 0.217 | 0.169 | 0.0053 | 1.36 | 0.10-1.86 |
| rs11732095 | A/G | 0.915 | 0.923 | 0.085 | 0.077 | 0.50 | 1.11 | 0.71-1.75 |
| rs10857092 | G/A | 0.93 | 0.923 | 0.070 | 0.077 | 0.53 | 1.11 | 0.69-1.78 |
| rs4492018 | G/A | 0.709 | 0.698 | 0.291 | 0.302 | 0.60 | 1.05 | 0.80-1.38 |
| rs975405 | T/C | 0.601 | 0.561 | 0.399 | 0.439 | 0.070 | 1.18 | 0.92-1.51 |
| rs1398553 | C/T | 0.694 | 0.743 | 0.306 | 0.257 | 0.015 | 1.27 | 0.97-1.68 |
| rs2893008 | A/G | 0.927 | 0.935 | 0.073 | 0.065 | 0.50 | 1.13 | 0.70-1.85 |

**Table 4.10 Genotype Frequencies Irish IL-2/IL-21 htSNPs; rs1127348 shows a significant association with coeliac disease (p = 0.016).**

| SNP | Alleles | Case nn | Ctrl nn | Case nN | Ctrl nN | Case NN | Ctrl NN | P-Value |
|---|---|---|---|---|---|---|---|---|
| rs6835946 | G/A | 0.085 | 0.118 | 0.459 | 0.433 | 0.457 | 0.449 | 0.198 |
| rs4374642 | T/C | 0.006 | 0 | 0.153 | 0.133 | 0.841 | 0.867 | 0.113 |
| rs4505848 | A/G | 0.077 | 0.072 | 0.472 | 0.413 | 0.451 | 0.516 | 0.117 |
| rs1127348 | T/C | 0.036 | 0.024 | 0.362 | 0.290 | 0.602 | 0.687 | 0.016 |
| rs11732095 | A/G | 0.011 | 0.011 | 0.150 | 0.133 | 0.840 | 0.856 | 0.743 |
| rs10857092 | G/A | 0.004 | 0.007 | 0.132 | 0.140 | 0.864 | 0.854 | 0.754 |
| rs4492018 | G/A | 0.069 | 0.098 | 0.445 | 0.407 | 0.486 | 0.494 | 0.175 |
| rs975405 | T/C | 0.161 | 0.186 | 0.475 | 0.506 | 0.363 | 0.308 | 0.215 |
| rs1398553 | C/T | 0.090 | 0.064 | 0.431 | 0.387 | 0.479 | 0.549 | 0.051 |
| rs2893008 | A/G | 0.002 | 0 | 0.141 | 0.130 | 0.857 | 0.870 | 0.479 |

**Table 4.11 Major Dominant/Minor Recessive Model of IL-2/IL-21 htSNPs in the Irish population; none of the SNPs show a significant association with coeliac disease.**

| SNP | Case (NN)+(Nn) | Case (nn) | Ctrl (nn)+(Nn) | Ctrl (nn) | P | OR | 95% CI |
|---|---|---|---|---|---|---|---|
| rs6835946 | 0.916 | 0.085 | 0.882 | 0.118 | 0.07615 | 1.45 | 0.96-2.20 |
| rs4374642 | 0.994 | 0.006 | 1.000 | 0.000 | 0.09870 | NA | NA |
| rs4505848 | 0.923 | 0.077 | 0.928 | 0.072 | 0.7454 | 1.08 | 0.68-1.73 |
| rs1127348 | 0.964 | 0.036 | 0.976 | 0.024 | 0.2527 | 1.54 | 0.74-3.20 |
| rs11732095 | 0.990 | 0.011 | 0.989 | 0.011 | 0.9795 | 1.02 | 0.31-3.36 |
| rs10857092 | 0.996 | 0.004 | 0.993 | 0.007 | 0.5677 | 1.74 | 0.32-9.61 |
| rs4492018 | 0.931 | 0.068 | 0.902 | 0.098 | 0.09252 | 1.48 | 0.94-2.33 |
| rs975405 | 0.839 | 0.161 | 0.814 | 0.186 | 0.3011 | 1.19 | 0.86-1.65 |
| rs1398553 | 0.910 | 0.090 | 0.936 | 0.064 | 0.1148 | 1.46 | 0.92-2.33 |
| rs2893008 | 0.998 | 0.002 | 1.000 | 0.000 | 0.4545 | NA | NA |

**Table 4.12 Minor Dominant/Major Recessive Model of IL-2/IL-21 htSNPs in the Irish population; 3 of the SNPs show a significant association with coeliac disease, rs4505848, rs1127348, rs1398553 ( p = 0.04016, p =0.004694 and p = 0.02519).**

| SNP | Case (Nn)+(nn) | Case (NN) | Ctrk (Nn)+(nn) | Ctrl (NN) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| rs6835946 | 0.543 | 0.457 | 0.551 | 0.449 | 0.8087 | 1.03 | 0.81-1.32 |
| rs4374642 | 0.159 | 0.841 | 0.133 | 0.867 | 0.2451 | 1.23 | 0.87-1.17 |
| rs4505848 | 0.549 | 0.451 | 0.484 | 0.516 | 0.04016 | 1.14 | 0.89-1.47 |
| rs1127348 | 0.398 | 0.602 | 0.313 | 0.687 | 0.004694 | 1.45 | 1.12-1.88 |
| rs11732095 | 0.160 | 0.840 | 0.144 | 0.856 | 0.4621 | 1.14 | 0.81-1.60 |
| rs10857092 | 0.136 | 0.864 | 0.147 | 0.853 | 0.6056 | 1.10 | 0.77-1.56 |
| rs4492018 | 0.514 | 0.486 | 0.506 | 0.494 | 0.7917 | 1.03 | 0.81-1.32 |
| rs975405 | 0.637 | 0.363 | 0.692 | 0.308 | 0.06481 | 1.28 | 0.99-1.66 |
| rs1398553 | 0.522 | 0.479 | 0.451 | 0.549 | 0.02519 | 1.33 | 1.03-1.70 |
| rs2893008 | 0.144 | 0.857 | 0.130 | 0.870 | 0.5390 | 1.12 | 0.80-1.60 |

**Table 4.13 Irish Case and Control Numbers by SNP for IL-2/IL-21 Region htSNPs.**

| SNP | N Case | N Controls |
|---|---|---|
| rs6835946 | 473 | 550 |
| rs4374642 | 478 | 555 |
| rs4505848 | 468 | 545 |
| rs1127348 | 472 | 549 |
| rs11732095 | 468 | 550 |
| rs10857092 | 471 | 550 |
| rs4492018 | 467 | 540 |
| rs975405 | 465 | 548 |
| rs1398553 | 466 | 550 |
| rs2893008 | 460 | 552 |

### 4a.3.6 Haplotype Analysis of the IL-2/IL-21 region in the Irish population

Using Phase the htSNPs were used to estimate the haplotypes present in blocks 1 and 2 in the Irish population, which are detailed in **Table 4.14** and **4.15** respectively. In block 1 (439kb) two Haplotypes indicated an association with the disease; ATGAGTAAG and ATAGACAAT (P = 0.0045 and 0.018). ATGAGTAAG shows a protective effect, (OR 0.69), while ATAGACAAT predisposes towards disease.

In block 2, GTTAT predisposes towards CD (P = 0.019) and ACTAC is protective (P = 0.012).

**Table 4.14 Haplotype Estimation ofhalplotype block 1 in the IL-2/IL-21 region (439kb) in the Irish Population. Two haplotypes show a significant association with coeliac disease, and the test shows global significance (p = 0.02).**

| Haplotype | F(Case) | F(Ctrl) | Odds Ratio | 95% CI | P |
|---|---|---|---|---|---|
| GTAAATAGA | 0.317 | 0.335 | 0.92 | 0.71-1.20 | 0.40 |
| ATAGACAAT | 0.20 | 0.16 | 1.31 | 0.95-1.81 | 0.018 |
| ATGAGTAAG | 0.117 | 0.161 | 0.69 | 0.48-0.99 | 0.0045 |
| ACAAATAAT | 0.078 | 0.065 | 1.22 | 0.75-1.97 | 0.26 |
| ATAAATGAT | 0.086 | 0.075 | 1.16 | 0.74-1.83 | 0.33 |
| ATAGATAAT | 0.055 | 0.047 | 1.18 | 0.67-2.07 | 0.43 |
| ATAGATAGT | 0.052 | 0.065 | 0.79 | 0.46-1.34 | 0.20 |
| ATAAATAAT | 0.046 | 0.044 | 1.05 | 0.58-1.90 | 0.85 |
| ATGAATAAG | 0.023 | 0.033 | 0.69 | 0.32-1.48 | 0.20 |

Overall P = 0.02

**Table 4.15 Haplotype Estimation of haplotype block 2 in the IL-2/Il21 region (40kb) in the Irish Population. Two haplotypes show a significant association with coeliac disease, and the test shows global significance (p = 0.04).**

| Haplotype | F(Case) | F(Ctrl) | Odds Ratio | 95% CI | P |
|---|---|---|---|---|---|
| ATCAT | 0.295 | 0.305 | 0.95 | 0.73-1.25 | 0.61 |
| GTTAT | 0.301 | 0.255 | 1.26 | 0.95-1.66 | 0.019 |
| ACTAC | 0.194 | 0.240 | 0.76 | 0.56-1.03 | 0.012 |
| ACTAT | 0.130 | 0.131 | 0.99 | 0.69-1.43 | 0.99 |
| ACTGT | 0.075 | 0.066 | 1.15 | 0.71-1.86 | 0.45 |

Overall P = 0.04

### 4a.3.7 Joint analysis of IL-2/IL-21 SNPs in Irish, British and Dutch populations via the Cochran-Mantel-Haenszel test.

As well as genotyping of rs13151961, rs13119723, rs6822844 and rs6840978 in the British and Irish populations, the SNPs were also replicated by a Dutch group in an independent Dutch population.

The results of genotyping in all three populations for the SNPs that showed significance, as well as the results of Cochran-Mantel-Haenszel analysis of the populations as a whole are listed in **Table 4.16** and **Table 4.17.**

All 4 SNPs are consistently associated in all 3 populations. In addition, CMH analysis reveals highly significant associations (in the order of $10^{-12}$-$10^{-14}$).

**Table 4.16 Combined GWAS results, showing P values for the 4 significant IL-2/IL-21 SNPs in the UK case control population, the Dutch case control population, and the Irish case control population.**

| SNP | UK CASE | UK CTRL | P |
|---|---|---|---|
| rs13151961 | 0.126 | 0.179 | $5.2 \times 10^{-6}$ |
| rs13119723 | 0.101 | 0.158 | $2.0 \times 10^{-7}$ |
| rs6822844 | 0.126 | 0.179 | $4.6 \times 10^{-6}$ |
| rs6840978 | 0.163 | 0.215 | $3.8 \times 10^{-5}$ |
| **SNP** | **DUTCH CASE** | **DUTCH CTRL** | **P** |
| rs13151961 | 0.122 | 0.191 | $2.2 \times 10^{-6}$ |
| rs13119723 | 0.115 | 0.164 | 0.00042 |
| rs6822844 | 0.124 | 0.185 | $2.1 \times 10^{-5}$ |
| rs6840978 | 0.152 | 0.219 | $2.0 \times 10^{-5}$ |
| **SNP** | **IRISH CASE** | **IRISH CTRL** | **P** |
| rs13151961 | 0.148 | 0.194 | 0.0056 |
| rs13119723 | 0.128 | 0.162 | 0.030 |
| rs6822844 | 0.142 | 0.197 | 0.0013 |
| rs6840978 | 0.192 | 0.240 | 0.0083 |

**Table 4.17 Combined GWAS results Odds Ratios, showing P values for the 4 significant IL-2/IL-21 SNPs in combined UK/Dutch and Irish case control populations.**

| SNP | Allele | OR | 95% CI | P |
|---|---|---|---|---|
| rs13151961 | G | 0.65 | 0.58-0.73 | $1.3 \times 10^{-12}$ |
| rs13119723 | G | 0.66 | 0.58-0.74 | $4.8 \times 10^{-11}$ |
| rs6822844 | T | 0.63 | 0.57-0.71 | $1.3 \times 10^{-14}$ |
| rs6840978 | T | 0.70 | 0.63-0.78 | $1.1 \times 10^{-10}$ |

**4a.3.8 Allele, Genotype and Carrier Status frequencies and disease association testing of selected IL-2/IL-21 LD block genotyped tgSNPs in an Italian population**

rs13151961, rs13119723, rs6822844 and rs6840978, along with the supplementary tagging SNPs rs1127348 and rs1398553 all showed association with CD in the Irish population. Of these 6 SNPs, 5 were selected for genotyping to determine if the

effect could be replicated in an independent Italian population; rs4374642 rs13151961, rs13119723 and rs6822844 were selected for their ability to tag all haplotypes which were observed in the Irish population at a frequency of 5% or more. rs6840978 was chosen as the only significant SNP from block 2.

All 5 selected SNPs were analyzed for allelic, genotypic and carrier status, and the results, along with Odds Ratios and 95% confidence intervals, are tabulated on tables **4.18** to **4.21. Table 4.22** details the sample size, per SNP.

Despite the observed association with CD in both the British and Irish populations, only rs6840978 showed association at the genotypic level (P = 0.03633), with the major allele appearing to confer the additional risk, similar to the observed effect in the Irish population.

**Table 4.18 Allele Frequencues of IL-2/IL-21 SNPs in an Italian Population; in contrast to the other tested populations, none of the IL-2/IL21 region SNPs show a significant association with coeliac disease.**

| SNP | Alleles | Case Maj | Ctrl Maj | Case Mi | Ctrl Mi | P-Value | Odds Ratio | 95% CI |
|---|---|---|---|---|---|---|---|---|
| rs4374642 | T/C | 0.922 | 0.912 | 0.078 | 0.088 | 0.4632 | 1.14 | 0.80-1.64 |
| rs13119723 | A/G | 0.993 | 0.993 | 0.007 | 0.007 | 0.9221 | 1.06 | 0.308-3.79 |
| rs1127348 | A/G | 0.831 | 0.828 | 0.169 | 0.172 | 0.9166 | 1.02 | 0.79-1.33 |
| rs6822844 | G/T | 0.911 | 0.896 | 0.089 | 0.104 | 0.336 | 1.199 | 0.86-1.69 |
| rs6840978 | C/T | 0.889 | 0.860 | 0.111 | 0.14 | 0.0877 | 1.303 | 0.96-1.77 |

**Table 4.19 Genotype Frequencies of IL-2/IL-21 SNPs in an Italian population, rs6840978 shows an association with coeliac disease on the genotypic level.**

| SNP | Alleles | Case nn | Ctrl nn | Case nN | Ctrl nN | Case NN | Ctrl NN | P-Value |
|---|---|---|---|---|---|---|---|---|
| rs4374642 | T/C | 0.005 | 0.005 | 0.146 | 0.167 | 0.848 | 0.828 | 0.7035 |
| rs13119723 | A/G | 0.000 | 0.000 | 0.013 | 0.014 | 0.987 | 0.986 | 1 |
| rs1127348 | A/G | 0.024 | 0.021 | 0.292 | 0.301 | 0.684 | 0.678 | 0.9284 |
| rs6822844 | G/T | 0.003 | 0.012 | 0.174 | 0.185 | 0.823 | 0.803 | 0.3527 |
| rs6840978 | C/T | 0.003 | 0.023 | 0.218 | 0.235 | 0.779 | 0.742 | 0.03633 |

**Table 4.20 Minor Dominant/Major Recessive Model of IL-2/IL-21 SNPs in an Italian population; none of the SNPs show association.**

| SNP | Case (Nn)+(nn) | Case (NN) | Ctrl (Nn)+(nn) | Ctrl (NN) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| rs4374642 | 0.848 | 0.152 | 0.827 | 0.173 | 0.4220 | 1.17 | 0.80- 1.71 |
| rs13119723 | 0.987 | 0.013 | 0.986 | 0.014 | 0.9301 | 1.06 | 0.31- 3.81 |
| rs1127348 | 0.684 | 0.316 | 0.678 | 0.322 | 0.8378 | 1.03 | 0.76- 1.39 |
| rs6822844 | 0.823 | 0.177 | 0.803 | 0.197 | 0.4754 | 1.14 | 0.79- 1.64 |
| rs6840978 | 0.779 | 0.221 | 0.742 | 0.258 | 0.2251 | 1.23 | 0.88-1.72 |

**Table 4.21 Major Dominant/Minor Recessive Model of IL-2/IL-21 SNPs in an Italian population; only rs6840978 shows association with coealic disease (p =0.01544).**

| SNP | Case (NN)+(Nn) | Case (nn) | Control (NN)+(Nn) | Control (nn) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| rs4374642 | 0.995 | 0.005 | 0.995 | 0.005 | 0.8893 | 0.86 | 0.09-8.31 |
| rs13119723 | 1.000 | 0.000 | 1.000 | 0.000 | 1 | NA | NA |
| rs1127348 | 0.976 | 0.024 | 0.979 | 0.021 | 0.7887 | 0.88 | 0.33-2.30 |
| rs6822844 | 0.997 | 0.003 | 0.988 | 0.002 | 0.1809 | 4.22 | 0.58-1.08 |
| rs6840978 | 0.997 | 0.003 | 0.976 | 0.024 | 0.01544 | 8.35 | 1.39-183.6 |

**Table 4.22 Italian Case and Control Numbers by SNP for IL-2/IL-21 SNPs.**

| SNP | N Case | N Controls |
|---|---|---|
| rs4374642 | 370 | 429 |
| rs13119723 | 381 | 431 |
| rs1127348 | 377 | 428 |
| rs6822844 | 362 | 432 |
| rs6840978 | 349 | 426 |

### 4a.3.9 Hardy Weinberg Analysis

We preformed Hardy Weinberg analysis to test for population equilibrium in all tested GWAS replication SNPs, in all assayed populations. **Table 4.23** lists the results..

**Table 4.23 Hardy Weinberg Analysis of Selected SNPs; only rs4505848 shows significant departure from Hardy Weinberg Equilibrium (p = 0.034).**

| SNP | Case | Control |
|---|---|---|
| rs1021621 | 0.356 | 0.223 |
| rs6762743 | 0.655 | 0.749 |
| rs9290678 | 0.114 | 0.749 |
| rs10505604 | 0.118 | 0.268 |
| rs4469515 | 0.762 | 0.952 |
| rs1064891 | 0.440 | 0.371 |
| rs7159238 | 0.882 | 0.746 |
| rs3809982 | 0.231 | 0.099 |
| rs13151961 | 0.561 | 0.588 |
| rs13119723 | 0.241 | 0.193 |
| rs6822844 | 0.481 | 0.588 |
| rs6840978 | 0.199 | 0.215 |
| rs6835946 | 0.157 | 0.509 |
| rs4374642 | 0.873 | 0.092 |
| rs4505848 | 0.034 | 0.581 |
| rs1127348 | 0.154 | 0.431 |
| rs11732095 | 0.350 | 0.104 |
| rs10857092 | 0.825 | 0.666 |
| rs4492018 | 0.088 | 0.438 |
| rs975405 | 0.846 | 0.538 |
| rs1398553 | 0.731 | 0.743 |
| rs2893008 | 0.319 | 0.101 |
| rs4374642- Italian | 1 | 0.559 |
| rs13119723- Italian | 1 | 1 |
| rs1127348- Italian | 0.586 | 0.306 |
| rs6822844- Italian | 0.337 | 0.797 |
| rs6840978- Italian | 0.100 | 0.547 |

## 4b.1 Introduction

### 4b.1.1 Genome Wide Association Study in the Irish, British and Dutch Populations

As outlined in the previous section (4a), a UK based GWAS examined 778 coeliac cases and 1,422 population controls from the UK. Several of the SNPs showing evidence of a association ($P < 10^{-4}$) were then genotyped in Dutch and Irish case control samples to provide validation in independent populations. The datasets were then analyzed as a whole using the CMH test.

Despite the large number of SNPs tested, and the added power afforded by the replication in additional populations, only one genomic region was found to be significantly associated with CD after correction for multiple testing; the IL-2/IL-21 region on chromosome 4.

As previously mentioned, it has been estimated that the HLA region accounts for some 35% of the heritability attributed to coeliac disease. The identification of only one additional genomic region associated with the disease may suggest that the contribution from one or more genes in this area is of a similar magnitude. However, estimates suggest that this is not the case.

rs13119723 showed the initial association with the disease state; Haplotype tag markers genotyped in the Irish, Dutch and British populations for the two LD blocks surrounding this SNP allowed more full characterisation of the area, and as such a better picture of the frequency of markers.

Based on the frequencies of these markers, we estimate that the IL2–IL21 region explains a mere < 1% of the increased familial risk of coeliac disease. Rather than the GWAS study revealing a second major determinant, or even a host of minor contributory factors, it has only one region of relatively minor importance to offer.

However, examination of the data from the first GWAS reveals the possibility of multiple other contributing regions which showed trends towards association, but were not significant. The fact that they were not detected could well be due to the manner in which the study was designed and carried out; particularly whether the study was powerful enough to detect the risk effect sizes contributed by these loci. This highlights neatly one of the difficulties in large scale genome analysis with regard to hypothesis testing. Ordinarily, in a single independent test of a locus, significance is recognised as a P value or < 0.05. By definition, this statistic informs that the result

the researcher is seeing, or one more extreme, would only be expected to occur due to random chance only one time in twenty.

However, examination of multiple additional loci requires that a correction must be made to the threshold that is considered significant; after all, under the previous set of parameters, testing 20 loci simultaneously, one would reasonably expect to achieve one result of $P < 0.05$ which would be statistically attributable to random chance.

When the number of loci being tested increases to the levels seen in the GWAS, the number of observed extreme results attributable to chance increase, as do the strictures for correction due to multiple testing. Testing of additional samples, as was done with the Dutch and Irish populations in the IL-2/IL-21 region will increase the power of the test accordingly, to a degree where a contributing locus should evidence significance even after relatively strict correction for multiple testing.

As only a very small proportion of the over 310,000 SNPs initially tested in the GWAS were subsequently typed in additional samples, it is quite possible that one or more contributing loci exist as part of the panel of htSNPs, but did not evidence sufficient degree of association on the strength of the GWAS results alone to merit replication. Examination of the proportion of observed significant results revealed them to be in excess of what would be expected due to random chance, suggesting more contributing loci did exist within the dataset.

This highlights a weakness of the two phase approach; an initial round of genotyping in a base population, and subsequent additional of supplementary populations to confirm the effect is more resource efficient, but may also lack the required power in phase one, thus reducing the value of the study.

In order to avoid this pitfall, the UK-Dutch-Irish consortium decided to alter the study design for the follow up to the GWAS. Instead of genotyping relatively few SNPs in subsequent populations, we resolved to extensively genotype markers in all three populations which had shown weaker evidence of association with disease in the first GWAS.

In selecting SNPs for this follow up study, it was crucial to attempt an approach which was most likely to include causal variants which were included on the GWAS platform, but failed to be flagged as significant due to lack of power. As causal variants would be expected to show a degree of association (just not a sufficient

degree based on power/correction) it was decided to select over 1,000 of the SNPs which showed the highest level of significance in the GWAS. It was thought that by selecting a proportion of the most significant results, (any SNPs with a $P < 0.00275$) that this would give the best chance for inclusion of previously overlooked causal variants. A further group of non-synonymous SNPs showing weaker p values were also included, based on their obvious potential to affect gene function.

In order to maximise power, the selected SNPs were tested in the Dutch, Irish, and a second independent UK case control population. This would allow the 4 populations to be analyzed together via CMH, and increase the chance of identifying susceptibility loci, an approach which has been shown to be more efficient than replication of the entire GWAS SNP list (Skol *et al.* 2006).

We performed genotyping using the Illumina GoldenGate assay (detailed in the Materials and Methods section). Genotyping of the Irish and British case –control collections was carried out in the Genome Centre at Barts and The London School of Medicine and Dentistry in London by myself and Dr. Karen Hunt, and the Dutch sample was genotyped in the Netherlands by Dr. Alexandra Zhernakova.

**4b.2 Objectives**

The aims of Part Two of this Chapter are:

1. To determine a panel of 1500 SNPs to be genotyped in the 3 populations using Illumina 'Golden Gate' multiplex genotyping technology.
2. Genotyping and Quality Control of said SNPs.
3. Analysis of the Irish population data.
4. Comparison of the associations across the populations
5. Joint analysis of the 4 populations via Cochran-Mantel-Haenszel test.

### 4b.3 Results

#### 4b.3.1 Selection of SNPs for inclusion in study

310,605 markers from the GWAS SNP panel passed quality control criteria, which was a Hardy-Weinberg equilibrium of $P > 0.0001$ in controls. As mentioned above, this dataset was examined to select non-HLA SNPs which showed association with disease status with p values using a two-tailed allele based $\chi 2$-test of $P < 0.00275$, or $P < 0.01$ if the SNP in question was a non synonymous coding SNP (nsSNP).

This led to 1,164 non-HLA SNPs initially being selected, comprising 1,088 SNPs with association results of $P < 0.00275$, and 76 nsSNPs with association p values of between $< 0.00275$ and $< 0.01$.

This panel was subsequently subjected to a pre genotyping quality control step; we excluded those SNPs which were unlikely to genotype successfully on the Golden Gate platform on the basis of Illumina GoldenGate assay design criteria (provided by the manufacturer in the form of an index number associated with a SNP) by excluding SNPs which clustered poorly in the original study. Following the removal of unsuitable markers, 1,025 SNPs were selected to be genotyped in the three follow-up populations.

#### 4b.3.2 Genotyping and Quality Control of the Irish Dataset

885 Biobank Control samples were genotyped, along with 481 Irish Coeliac samples, of which 336 were untreated genomic DNA, and 145 samples which had been subjected to whole genome amplification (WGA) due to their being historical samples, and consequently of below the concentration required for the study (250ng per run, at a concentration of 50ng/μl).

Genotyping of the samples in such a broad panel of SNPs allowed for comparison of co-efficient of relationship between samples (using PLINK software) to a degree that had not previously been possible. 8 coeliac samples were revealed to be cryptic duplicates, and a total of 26 coeliac samples showed evidence of previously unknown first degree relationship (co-efficient of relationship of 0.5 or greater) across 10 distinct families. In the control group, two cryptic duplicates (possibly twins) and 9 first degree relatives were identified. All duplicates were removed from the dataset, and in the case of families, related individuals were randomly selected for removal.

The whole genome amplified samples exhibited non standard clustering; although they formed distinct subgroups based on fluorescence, the positions of these subgroups relative to non amplified samples was subject to some skew.

In order to correct for this, clustering was performed on these WGA samples separately, and the samples pooled once genotypes had been determined.

Once the duplicates and family members had been removed, along with samples which failed to amplify, or cluster, the Irish sample was comprised of 438 Coeliacs and 872 controls.

### 4b.3.3 Analysis of the Irish Dataset

In addition to the CMH test performed on all 4 populations, we also analysed the Irish population separately. Population specific effects are a real possibility in complex disease, and although the correction for multiple testing (given assumptions on effect size for each locus) precludes the Irish samples from reaching statistical significance when analyzed individually, it may be useful to evaluate whether the contribution of individual loci may potentially be unique to Irish coeliac patients, or indeed the general population. We report here detailed information on those, non HLA, SNPs which showed an association of $P < 0.015$ at the allelic level in the Irish population.

### 4b.3.4 Locations of the Most Associated SNPs in the Irish Dataset

15 non HLA SNPs showed a P of below the assigned threshold of 0.015. rs9503233 was selected for inclusion despite being over thisthreshold ($P = 0.024$), as the SNP is a non-synonymous SNP within a gene of currently unknown function. **Table 4.24** lists the 16 SNPs, their allelic P-Value, chromosome, genomic position, any nearby gene of functional interest, and the relative location of the SNP to the nearest gene.

**Table 4.24 Characteristic of Most Strongly Associated Irish SNPs from the goldengate GWAS followup study.**

| SNP | Chr | P-Value | Gene | Relative Location |
|---|---|---|---|---|
| rs17810546 | 3 | 1.588e-005 | IL-12A | Intergenic |
| rs6822844 | 4 | 8.684e-005 | IL-2/IL-21 | Intergenic |
| rs1021701 | 3 | 0.001198 | RARB | Intergenic |
| rs2078178 | 12 | 0.000965 | CLEC7A | Intronic to CLEC7A |
| rs6483061 | 11 | 0.01104 | No obvious candidate | Intergenic |
| rs10501723 | 11 | 0.01028 | No obvious candidate | Intergenic |
| rs269243 | 7 | 0.01389 | DENND2A | Non-Syn Coding |
| rs7479104 | 11 | 0.01449 | No obvious candidate | Intergenic |
| rs7852227 | 9 | 0.0007139 | TRPM3 | Intergenic |
| rs13119723 | 4 | 0.0006263 | IL-2/IL-21 | |
| rs9503233 | 6 | 0.02413 | C6orf195 | Non-Syn Coding |
| rs2257212 | 3 | 0.01204 | SLC15A2 | Non-Syn Coding |
| rs13397583 | 2 | 0.01318 | KLHL29 | 5' of KLHL29 |
| rs736191 | 8 | 0.01208 | POP1 | Intergenic |
| rs13151961 | 4 | 0.0007749 | IL-2/IL-21 | Intergenic |
| rs884108 | 1 | 0.01439 | LRRN2 | Intronic to LRRN2 |

### 4b.3.5 Allele, Genotype and Carrier Status frequencies and disease association testing of SNPs with P < 0.015 in the Irish population

All 16 selected SNPs were analyzed for allelic, genotypic and carrier status, and the results, along with Odds Ratios and 95% confidence intervals, are tabulated on tables **4.25-4.28.**

Naturally, all of the genes show significance at the allelic level, as this was a criteria of their inclusion. Looking at the genotypic level, only rs9503233 fails to show any significant association with disease.

**Table 4.25 Allelic Frequencies of the Most Strongly Associated Irish SNPs from the goldengate GWAS followup study.**

| SNP | Case Maj | Control Maj | Case Mi | Control Mi | P-Value | OR | 95% CI |
|---|---|---|---|---|---|---|---|
| rs17810546 | 0.811 | 0.874 | 0.189 | 0.126 | 1.588e-005 | 1.62 | 1.29-2.02 |
| rs6822844 | 0.861 | 0.798 | 0.139 | 0.202 | 8.684e-005 | 0.64 | 0.51-0.79 |
| rs1021701 | 0.701 | 0.76 | 0.299 | 0.24 | 0.001198 | 1.35 | 1.13-1.62 |
| rs2078178 | 0.701 | 0.76 | 0.299 | 0.24 | 0.000965 | 1.36 | 1.13-1.63 |
| rs6483061 | 0.528 | 0.476 | 0.472 | 0.524 | 0.01104 | 1.24 | 1.05-1.45 |
| rs10501723 | 0.531 | 0.479 | 0.469 | 0.521 | 0.01028 | 1.24 | 1.05-1.46 |
| rs269243 | 0.914 | 0.941 | 0.086 | 0.059 | 0.01389 | 1.48 | 1.08-2.01 |
| rs7479104 | 0.528 | 0.477 | 0.472 | 0.523 | 0.01449 | 1.226 | 1.04-1.44 |
| rs7852227 | 0.683 | 0.746 | 0.317 | 0.254 | 0.0007139 | 1.36 | 1.14-1.63 |
| rs13119723 | 0.855 | 0.801 | 0.145 | 0.199 | 0.0006263 | 0.68 | 0.54-0.85 |
| rs9503233 | 0.774 | 0.733 | 0.226 | 0.267 | 0.02413 | 0.80 | 0.66-0.97 |
| rs2257212 | 0.572 | 0.521 | 0.428 | 0.479 | 0.01204 | 0.81 | 0.69-0.96 |
| rs13397583 | 0.699 | 0.651 | 0.301 | 0.349 | 0.01318 | 0.80 | 0.67-0.95 |
| rs736191 | 0.558 | 0.506 | 0.442 | 0.494 | 0.01208 | 0.81 | 0.69-0.96 |
| rs13151961 | 0.856 | 0.803 | 0.144 | 0.197 | 0.0007749 | 0.68 | 0.55-0.85 |
| rs884108 | 0.709 | 0.753 | 0.291 | 0.247 | 0.01439 | 1.26 | 1.05-1.51 |

**Table 4.26 Genotypic Frequencies of the Most Strongly Associated Irish SNPs from the goldengate GWAS followup study.**

| SNP | Case nn | Ctrl nn | Case nN | Ctrl nN | Case NN | Ctrl NN | P-Value |
|---|---|---|---|---|---|---|---|
| rs17810546 | 0.039 | 0.013 | 0.301 | 0.227 | 0.659 | 0.760 | 5.005e-005 |
| rs6822844 | 0.016 | 0.041 | 0.246 | 0.323 | 0.738 | 0.636 | 0.0004263 |
| rs1021701 | 0.093 | 0.050 | 0.414 | 0.380 | 0.493 | 0.570 | 0.002859 |
| rs2078178 | 0.098 | 0.054 | 0.404 | 0.373 | 0.498 | 0.573 | 0.00234 |
| rs6483061 | 0.295 | 0.218 | 0.468 | 0.516 | 0.237 | 0.266 | 0.009533 |
| rs10501723 | 0.299 | 0.220 | 0.466 | 0.517 | 0.235 | 0.263 | 0.00755 |
| rs269243 | 0.007 | 0.003 | 0.158 | 0.112 | 0.835 | 0.885 | 0.03631 |
| rs7479104 | 0.295 | 0.219 | 0.467 | 0.517 | 0.238 | 0.264 | 0.01035 |
| rs7852227 | 0.094 | 0.059 | 0.447 | 0.390 | 0.459 | 0.551 | 0.002608 |
| rs13119723 | 0.016 | 0.042 | 0.258 | 0.314 | 0.726 | 0.643 | 0.002285 |
| rs9503233 | 0.052 | 0.079 | 0.347 | 0.375 | 0.601 | 0.546 | 0.07789 |
| rs2257212 | 0.203 | 0.234 | 0.450 | 0.492 | 0.347 | 0.274 | 0.02332 |
| rs13397583 | 0.100 | 0.122 | 0.402 | 0.456 | 0.498 | 0.422 | 0.03252 |
| rs736191 | 0.189 | 0.235 | 0.505 | 0.517 | 0.306 | 0.248 | 0.03786 |
| rs13151961 | 0.016 | 0.041 | 0.256 | 0.312 | 0.728 | 0.647 | 0.002873 |
| rs884108 | 0.064 | 0.068 | 0.454 | 0.358 | 0.482 | 0.575 | 0.002957 |

**Table 4.27 Minor Dominant/Major Recessive Model of the Most Strongly Associated Irish SNPs from the goldengate GWAS followup study.**

| SNP | Case (NN) | Ctrl (NN) | Case (Nn)+(nn) | Ctrl (Nn)+(nn) | P | OR | 95% CI |
|---|---|---|---|---|---|---|---|
| rs17810546 | 0.340 | 0.239 | 0.660 | 0.761 | 0.0001177 | 1.64 | 1.27-2.10 |
| rs6822844 | 0.260 | 0.363 | 0.740 | 0.637 | 0.000245 | 0.62 | 0.48-0.80 |
| rs1021701 | 0.507 | 0.430 | 0.493 | 0.570 | 0.008459 | 1.36 | 1.08-1.72 |
| rs2078178 | 0.502 | 0.427 | 0.498 | 0.573 | 0.00942 | 1.36 | 1.08-1.71 |
| rs6483061 | 0.763 | 0.734 | 0.237 | 0.266 | 0.2632 | 1.16 | 0.89-1.52 |
| rs10501723 | 0.765 | 0.737 | 0.235 | 0.263 | 0.2812 | 1.16 | 0.89-1.52 |
| rs269243 | 0.164 | 0.116 | 0.836 | 0.884 | 0.01554 | 1.50 | 1.08-2.08 |
| rs7479104 | 0.760 | 0.736 | 0.240 | 0.264 | 0.3467 | 1.15 | 0.88-1.50 |
| rs7852227 | 0.541 | 0.448 | 0.459 | 0.552 | 0.001876 | 1.45 | 1.15-1.82 |
| rs13119723 | 0.274 | 0.357 | 0.726 | 0.643 | 0.002764 | 0.68 | 0.53-0.87 |
| rs9503233 | 0.399 | 0.454 | 0.601 | 0.546 | 0.0671 | 0.8 | 0.63-1.01 |
| rs2257212 | 0.653 | 0.726 | 0.347 | 0.274 | 0.007189 | 0.71 | 0.56-0.91 |
| rs13397583 | 0.502 | 0.578 | 0.498 | 0.422 | 0.009623 | 0.74 | 0.59-0.93 |
| rs736191 | 0.694 | 0.753 | 0.306 | 0.247 | 0.02887 | 0.75 | 0.58-0.97 |
| rs13151961 | 0.272 | 0.353 | 0.728 | 0.647 | 0.002981 | 0.68 | 0.53-0.87 |
| rs884108 | 0.518 | 0.425 | 0.482 | 0.575 | 0.001466 | 1.45 | 1.15-1.83 |

**Table 4.28 Major Dominant/Minor Recessive Model of the Most Strongly Associated Irish SNPs from the goldengate GWAS followup study.**

| SNP | Case (nn) | Ctrl (nn) | Case (NN)+(Nn) | Ctrl (NN)+(Nn) | P | OR | 95% CI |
|---|---|---|---|---|---|---|---|
| rs17810546 | 0.039 | 0.013 | 0.961 | 0.987 | 0.001982 | 3.16 | 1.47-7.02 |
| rs6822844 | 0.016 | 0.040 | 0.984 | 0.960 | 0.01921 | 0.39 | 0.16-0.85 |
| rs1021701 | 0.091 | 0.050 | 0.909 | 0.950 | 0.004394 | 1.89 | 1.21-2.95 |
| rs2078178 | 0.098 | 0.054 | 0.902 | 0.946 | 0.002802 | 1.91 | 1.24-2.94 |
| rs6483061 | 0.295 | 0.218 | 0.705 | 0.782 | 0.002299 | 1.50 | 1.15-1.94 |
| rs10501723 | 0.299 | 0.220 | 0.701 | 0.780 | 0.001773 | 1.51 | 1.16-1.96 |
| rs269243 | 0.007 | 0.003 | 0.993 | 0.997 | 0.4079 | 1.99 | 0.34-11.6 |
| rs7479104 | 0.295 | 0.219 | 0.705 | 0.781 | 0.003319 | 1.49 | 1.15-1.94 |
| rs7852227 | 0.094 | 0.059 | 0.906 | 0.941 | 0.02987 | 1.63 | 1.06-2.49 |
| rs13119723 | 0.016 | 0.042 | 0.984 | 0.958 | 0.01378 | 0.37 | 0.15-0.80 |
| rs9503233 | 0.052 | 0.079 | 0.948 | 0.921 | 0.08545 | 0.65 | 0.39-1.04 |
| rs2257212 | 0.203 | 0.234 | 0.797 | 0.766 | 0.2322 | 0.84 | 6.29-1.10 |
| rs13397583 | 0.100 | 0.120 | 0.900 | 0.880 | 0.2711 | 0.81 | 0.55-1.17 |
| rs736191 | 0.189 | 0.116 | 0.811 | 0.884 | 0.06595 | 0.76 | 0.57-1.01 |
| rs13151961 | 0.016 | 0.041 | 0.984 | 0.959 | 0.01532 | 0.38 | 0.15-0.82 |
| rs884108 | 0.060 | 0.068 | 0.940 | 0.932 | 0.7979 | 0.94 | 0.58-1.49 |

### 4b.3.6 CMH Analysis of the Irish, Dutch, British and GWAS datasets

Following genotyping of the Dutch (508 coeliacs, 888 controls) and British (416 coeliacs, 957 controls) populations in the Sanger centre and the Netherlands respectively, data from the three newly genotyped populations was combined with the marker data from the original GWAS, and analysed via the Cochran-Mantel-Haenszel test for 2x2xK stratified datasets. The results are listed on **Table 4.29**

SNPs are ranked in order of overall P value, lowest to highest. HLA SNPs are included for reference. The table is also notated with the nearest gene or genes of functional interest to the SNP, and the relative location of that SNP to the gene in questions.

**Table 4.29 CMH Analysis of the Irish, Dutch, British and GWAS datasets from the goldengate study; a list of all of the SNPs which showed global significance.**

| SNP | Chr | P UKGWAS | P Follow UP | P Overall | Gene |
|---|---|---|---|---|---|
| rs2187668 | 6 | 1.43E-167 | 0 | 0 | HLA |
| rs7775397 | 6 | 7.82E-151 | 0 | 0 | HLA |
| rs2395182 | 6 | 2.43E-17 | 1.23E-38 | 2.74E-54 | HLA |
| rs2856705 | 6 | 4.84E-08 | 7.42E-21 | 3.03E-27 | HLA |
| rs7775228 | 6 | 1.27E-05 | 1.29E-13 | 9.77E-18 | HLA |
| rs6822844 | 4 | 4.80E-06 | 9.84E-09 | 2.82E-13 | IL2/IL21 |
| rs13119723 | 4 | 2.51E-07 | 1.23E-07 | 5.94E-13 | IL2/IL21 |
| rs13151961 | 4 | 5.48E-06 | 4.32E-08 | 1.53E-12 | IL2/IL21 |
| rs2816316 | 1 | 0.001159 | 5.11E-09 | 2.58E-11 | RGS1 |
| rs6840978 | 4 | 4.79E-05 | 2.32E-07 | 5.53E-11 | IL2/IL21 |
| rs917997 | 2 | 9.06E-05 | 1.97E-06 | 8.49E-10 | IL-18RAP |
| rs17810546 | 3 | 0.000339 | 7.77E-07 | 1.07E-09 | IL-12A SCHIP1 |
| rs12642902 | 4 | 4.59E-05 | 7.35E-06 | 2.12E-09 | IL2/IL21 |
| rs13015714 | 2 | 0.00041 | 2.66E-06 | 4.37E-09 | IL-18RAP |
| rs1464510 | 3 | 7.65E-05 | 1.21E-05 | 5.33E-09 | LPP |
| rs11938795 | 4 | 0.000257 | 1.07E-05 | 1.30E-08 | IL2/IL21 |
| rs13076312 | 3 | 0.000337 | 1.21E-05 | 1.78E-08 | LPP |
| rs1559810 | 3 | 0.000103 | 3.28E-05 | 1.99E-08 | LPP |
| rs9851967 | 3 | 0.000904 | 7.13E-06 | 2.45E-08 | LPP |
| rs11734090 | 4 | 0.000337 | 2.95E-05 | 4.76E-08 | IL2/IL21 |
| rs9811792 | 3 | 0.002775 | 5.42E-06 | 5.24E-08 | IL-12A SCHIP1 |
| rs1738074 | 6 | 0.001564 | 1.19E-05 | 6.71E-08 | TAGAP |
| rs653178 | 12 | 0.000649 | 2.98E-05 | 8.00E-08 | LNK, ATXN2 |
| rs7684187 | 4 | 0.000911 | 3.11E-05 | 1.13E-07 | IL2/IL21 |
| rs3184504 | 12 | 0.001075 | 3.22E-05 | 1.33E-07 | LNK<br>ATXN2 |
| rs6441961 | 3 | 0.00284 | 3.21E-05 | 3.14E-07 | CCR1<br>CCR3 |

### 4c.1 Introduction

There was a degree of difficulty in replicating the observed association with the IL-2/IL-21 region in the Italian population. Disparate allelic frequencies suggest that this may be due to a different pattern of LD in the IL-2/IL-21 linkage region between the Italian, and other European populations.

However, this difference in LD may be unique to the IL-2/IL-21 region. In an attempt to replicate the effects observed via the Goldengate assay, and to a lesser extent to determine whether different patterns of LD prevail in the Italian population, we undertook to genotype the Italian case control sample for loci which provided significant in the combined population analysis.

### 4c.2 Objectives

The aims of Part Three of this Chapter are:

1. To select one SNP per genomic region which showed significance in the GoldenGate GWAS followup.
2. Genotyping of said SNPs in an idependent Italian population.
3. Analysis of the Italian population data, and comparison with the results seen in the Irish, Dutch and British populations.

### 4c.3 Results

#### 4c.3.1 Part 3 Replication of Effect in an Independent Italian Population

rs9851967 (LPP), rs2816316 (RGS1), rs917997 (IL-18RAP), rs6441961 (CCR1/CCR5), rs3184504 (LNK), rs1738074 (TAGAP) and rs17810546 (IL-12A) were selected to be genotyped in the Italian population. The eighth region, IL-2/IL-21 had already investigated, and had evidenced a different pattern of association than had been seen in the other population.

**Tables 4.30-4.34** show the analysis of the Italian samples. All loci show significance, except for rs2816316 near RGS1.

**Table 4.30 Allele Frequencies of the GWAS follow-up in the Italian Population; all SNPs which evidenced global association in the CMH test also show significane in the Italian population, except for rs2816316.**

| SNP | Alleles | Case Maj | Ctrl Maj | Case Mi | Ctrl Mi | P-Value | OR | 95% CI |
|---|---|---|---|---|---|---|---|---|
| rs9851967 | C/T | 0.588 | 0.502 | 0.413 | 0.498 | 0.0009 | 1.41 | 1.14-1.73 |
| rs2816316 | A/C | 0.854 | 0.826 | 0.146 | 0.174 | 0.1459 | 1.23 | 0.94-1.62 |
| rs917997 | C/T | 0.742 | 0.786 | 0.258 | 0.214 | 0.0430 | 1.27 | 1.01-1.61 |
| rs6441961 | C/T | 0.582 | 0.643 | 0.418 | 0.357 | 0.0145 | 1.29 | 1.05-1.59 |
| rs3184504 | T/C | 0.568 | 0.517 | 0.432 | 0.483 | 0.0482 | 1.23 | 1.07-1.61 |
| rs1738074 | C/T | 0.538 | 0.610 | 0.462 | 0.390 | 0.0044 | 1.31 | 1.01-1.61 |
| rs17810546 | A/G | 0.901 | 0.938 | 0.099 | 0.062 | 0.0081 | 1.67 | 1.14-2.4 |

**Table 4.31 Genotype Frequencies of the GWAS follow-up in the Italian Population; four SNPs, rs9851967, rs6441961, rs1738074 and rs17810546 show association with coeliac disease (p =0.007, p = 0.0439, p = 0.0091 and p = 0.0178).**

| SNP | Alleles | Case nn | Ctrl nn | Case nN | Ctrl nN | Case NN | Ctrl NN | P-Value |
|---|---|---|---|---|---|---|---|---|
| rs9851967 | C/T | 0.139 | 0.247 | 0.545 | 0.502 | 0.315 | 0.251 | 0.0007 |
| rs2816316 | A/C | 0.012 | 0.035 | 0.269 | 0.279 | 0.719 | 0.686 | 0.1026 |
| rs917997 | C/T | 0.061 | 0.046 | 0.393 | 0.336 | 0.546 | 0.618 | 0.1139 |
| rs6441961 | C/T | 0.172 | 0.132 | 0.493 | 0.450 | 0.335 | 0.418 | 0.0439 |
| rs3184504 | T/C | 0.181 | 0.277 | 0.501 | 0.480 | 0.318 | 0.243 | 0.0946 |
| rs1738074 | C/T | 0.229 | 0.147 | 0.465 | 0.487 | 0.305 | 0.367 | 0.0091 |
| rs17810546 | A/G | 0.009 | 0.005 | 0.180 | 0.115 | 0.811 | 0.881 | 0.0178 |

**Table 4.32 Major Dominant/Minor Recessive Model of the GWAS follow up in the Italian Population four SNPs, rs9851967, rss816316, rs3184504 and rs1738074 and show association with coeliac disease (p =0.0002106, p = 0.03857, p = 0.003339).**

| SNP | Case (NN)+(Nn) | Case (nn) | Ctrl (nn)+(Nn) | Ctrl (nn) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| rs9851967 | 0.861 | 0.139 | 0.753 | 0.247 | 0.0002106 | 2.01 | 1.38-2.9 |
| rs2816316 | 0.988 | 0.012 | 0.965 | 0.035 | 0.03857 | 3.06 | 1.06-10.8 |
| rs917997 | 0.939 | 0.061 | 0.954 | 0.046 | 0.3474 | 0.74 | 0.39-1.39 |
| rs6441961 | 0.828 | 0.172 | 0.868 | 0.132 | 0.1234 | 0.73 | 0.49-1.09 |
| rs3184504 | 0.819 | 0.181 | 0.723 | 0.277 | 0.003339 | 0.58 | 0.40-0.83 |
| rs1738074 | 0.771 | 0.229 | 0.853 | 0.147 | 0.2460 | 2.01 | 1.38-2.94 |
| rs17810546 | 0.991 | 0.009 | 0.995 | 0.005 | 0.5467 | 3.06 | 1.06-10.8 |

**Table 4.33 Italian Case and Control Numbers by SNP for GoldenGate replication study.**

| SNP | N Case | N Controls |
|---|---|---|
| rs9851967 | 343 | 426 |
| rs2816316 | 345 | 433 |
| rs917997 | 359 | 434 |
| rs6441961 | 361 | 424 |
| rs3184504 | 359 | 433 |
| rs1738074 | 357 | 409 |
| rs17810546 | 355 | 428 |

**Table 4.34 Minor Dominant/Major Recessive Model of the GWAS follow up in the Italian Population; 3 SNPs show association with coeliac disease, rs917997, rs6441961 and rs17810546 ( p =0.04238, p =0.01805 and p =0.007131).**

| SNP | Case (Nn)+(nn) | Case (NN) | Ctrl (Nn)+(nn) | Ctrl (NN) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| rs9851967 | 0.685 | 0.315 | 0.749 | 0.251 | 0.05146 | 0.73 | 0.53-1.0 |
| rs2816316 | 0.281 | 0.719 | 0.314 | 0.686 | 0.3207 | 0.85 | 0.63-1.17 |
| rs917997 | 0.454 | 0.546 | 0.382 | 0.618 | 0.04238 | 1.34 | 1.01-1.78 |
| rs6441961 | 0.665 | 0.335 | 0.582 | 0.418 | 0.01805 | 1.42 | 1.06-1.90 |
| rs3184504 | 0.682 | 0.318 | 0.757 | 0.243 | 0.2162 | 0.82 | 0.61-1.12 |
| rs1738074 | 0.695 | 0.305 | 0.633 | 0.367 | 0.07356 | 1.32 | 0.97-1.78 |
| rs17810546 | 0.189 | 0.811 | 0.119 | 0.881 | 0.007131 | 1.72 | 1.16-2.56 |

## 4.4 Discussion

### 4a.4 IL-2/IL21

Initial investigation of the IL-2/IL-21 LD region which showed association in the British population, revealed a similar pattern of association in the Irish case control population. rs13151961, rs13119723, rs6822844, rs6840978 all showed association at the allelic level, with the major allele being associated with an increased risk of disease.

Genotyping of additional haplotype tag markers for the 2 LD blocks surrounding these SNPs, and haplotype reconstruction was undertaken. Each of the two blocks evidenced one common haplotype which conferred an increased risk of disease and was significantly associated with CD, and one haplotype which conferred a protective effect, and was also significantly associated with CD. Comparison of the reconstructed haplotypes in the Irish population shows the same pattern as that of the UK population, in terms of haplotype frequencies, and the specific haplotypes associated with either increased or decreased risk. Joint analysis of the Irish, Dutch and UK populations via CMH showed significant association overall with all 4 SNPs, as shown below in **Table 4.35**

**Table 4.35 CMH analysis of IL-2/IL-21 SNPs across UK, Dutch and Irish populations. All selected SNPs show significant associations with coeliac disease.**

| SNP | Allele | OR | 95% CI | P |
|---|---|---|---|---|
| rs13151961 | G | 0.65 | 0.58-0.73 | $1.3 \times 10^{-12}$ |
| rs13119723 | G | 0.66 | 0.58-0.74 | $4.8 \times 10^{-11}$ |
| rs6822844 | T | 0.63 | 0.57-0.71 | $1.3 \times 10^{-14}$ |
| rs6840978 | T | 0.70 | 0.63-0.78 | $1.1 \times 10^{-10}$ |

The association of this region with Coeliac Disease has also been confirmed in both independent Scandinavian (Amundsen *et al.* 2010) and Italian populations (Romanos *et al.* 2009), although the Italian population showed some notable differences in allelic frequencies, which may reflect structural differences between the populations, which will be discussed below.

The LD region spanned by these markers has shown association with a wide variety of other autoimmune diseases. rs6822844 has shown association with Rheumatoid Arthritis in a Caucasian population (Hollis-Moffatt *et al.* 2010), and rs6840978,

rs6822844 and rs13119723 have all shown association with Ulcerative Colitis (Glas *et al.* 2009). A GWAS searching for Type I Diabetes susceptibility loci has also indicated an association with the LD block, (Barrett *et al.* 2009).

Although the LD blocks which are tagged by these SNPs comprise four genes (KIAA1109, TENR and IL-2 and IL-21), there are clear functional candidates which mark them out as the likely source of association.

**KIAA1109:** This gene's exact function is unknown, however it is thought to function in the regulation of epithelial growth and differentiation. Its product is a protein which is similar in structure to a Chinese hamster protein associated with spermatocyte and adipocyte differentiation. The C-terminus of the protein is also similar to a Caenorhabditis elegans protein that plays a role in lipid storage (Kuo *et al.* 2006)

**TENR** :(Testis nuclear RNA-binding protein) plays a role in spermatogenesis, and binds to RNA, but not DNA (Schumacher *et al.* 1995)

Due to the fact that neither of these genes currently have a well defined function, they do not make good prospects for candidate genes. IL-2 and IL-21, on the other hand, are well characterised in terms of function. They are heavily involved in the immune response, and have shown associations with multiple autoimmune conditions. It is likely that SNPs influencing the expression, or structure of one, or both, of these genes are the causative factor, or factors, responsible for the observed association.

**IL-2** : This gene encodes a secreted cytokine that is crucial for the proliferation of T and B cell proliferation and differentiation. An orthologous gene in mice, when subjected to targeted disruption, results in an ulcerative colitis-like disease, suggesting a key role for IL-2 in the intestinal immune response (Spencer *et al.* 1998).

IL-2 is also involved in regulation of auto-reactive mature T-Cells released from the thymus via its involvement with T-Regulation cells. T-Regs exert their effect through cell to cell contact, via the T-Cell Receptor, and activate a suppression of cell proliferation. The primary method of modulation appears to be control of IL-2 production in the responder cell (Horwitz *et al.* 2003) and development of these cells requires the action of IL-2 (Thornton *et al.* 1998) Peturbations in IL-2 production could result in a wide variety of autoimmune predisposing scenarios. Reduction in IL-2 levels could result in a failure to develop, or reduction in, the population of T-Regs,

while increased IL-2 production in the responder cell might abrogate the suppressive effect of the T-Regs.

As illustrated by its role in autoregulation, IL-2 has pleiotrophic effects on different T cell lineages, and its influence on CD is likely to be complex. For example while IL-2 had originally been considered the archetypical Th1 cytokine and thus a suspect in chronic inflammatory conditions, mice deficient in IL-2 production develop chronic inflammatory conditions probably as a result of a lack of IL-2 dependent T reg function. IL-2 is also involved in the thymus during Th1/Th2 lineage determination Rodriguez-Galán *et al.* 2005). IL-2 has been also been implicated in a host of other autoimmune conditions (See Chapter 6 for more details).

**IL-21:** Like IL-2, this also encodes a secreted cytokine with immunoregulatory activity. It plays a role in the maturation of Natural Killer cells, positive regulation of interferon-gamma biosynthetic process (Th1) and T cell proliferation, as well as mediating transition from innate to adaptive immunity. There is evidence to suggest it is also involved in B cell activation, expansion, and plasma cell generation (Kasaian *et al.* 2002, Kuchen *et al.* 2007).

In a similar fashion to IL-2, IL-21 has also been associated with a number of autoimmune conditions (See Chapter 6 for more details).

Our attempts to replicate the association observed with the IL-2/IL-21/KIAA1109/ TENR region in an Italian population showed an association only with rs6840978. In addition to only one of the five loci tested showing association, the observed allele frequencies at four of the five loci were markedly different from those observed in the Irish population. The SNPs are compared in **Table 4.36**.

**Table 4.36 IL-2/IL-21 SNP frequency comparison; Italians vs. Irish.**

| SNP | Alleles | Irish Case Maj | Italian Case Maj | Irish Control Maj | Italian Control Maj |
|---|---|---|---|---|---|
| rs4374642 | T/C | 0.917 | 0.922 | 0.933 | 0.912 |
| rs13119723 | A/G | 0.872 | 0.993 | 0.838 | 0.993 |
| rs1127348 | T/C | 0.783 | 0.831 | 0.831 | 0.828 |
| rs6822844 | G/T | 0.858 | 0.911 | 0.803 | 0.896 |
| rs6840978 | C/T | 0.808 | 0.889 | 0.760 | 0.860 |

The association with rs6840978 was also observed in a second Italian case control study. Although none of the other SNPs were tested in this study, the case and control major allele frequency in this Italian population were very similar to those we observed (0.919 in the cases, and 0.891 in the controls) suggesting that is typical of the distribution of this SNP in the Italian population.

This may represent evidence of a different underlying haplotype structure in the Italian population, or a difference in relative effects of the loci contributing to CD across populations.

In any event, the IL-2, IL-21 LD region has consistently shown an association across multiple, independent populations, marking it as the first non-HLA risk factor to be definitively associated with coeliac disease. The functionally relevant nature of the genes in this region make it an exciting and highly significant find.

## 4b.4 Genome Wide Association Follow up in Multiple Populations

### 4b.4.1 Overall Significant SNPs in Combined Populations

Following on from the confirmation of effect in the Irish population, and to a lesser extent the Italian, we aimed to discover other contributory loci which may have been missed in the initial UK GWAS.

Through genotyping 1,025 loci selected from some of the most significant UK GWAS results across 3 populations, and subsequent CHM analysis, we observed association between CD and 26 SNPs. 6 of these SNPs lay within the already characterised HLA region, and of the other 21 SNPs, they were located, broadly, within 8 distinct regions.

These regions are described in **Table 4.37** as being near named genes which either map, or are adjacent, to the same strong LD block as associated SNPs, or the SNPs physically map within the gene.

**Figure 4.3** is a QQ plot (Quartile-quartile plot) of the frequency distribution of observed significant associations, against the expected distribution frequency for the null hypothesis. Deviation of the plot from the line is indicative of the presence of true associations. Part (a) shows our results, part (b) shows conditions expected under the Null hypothesis (Source: Hunt *et al.* 2008).

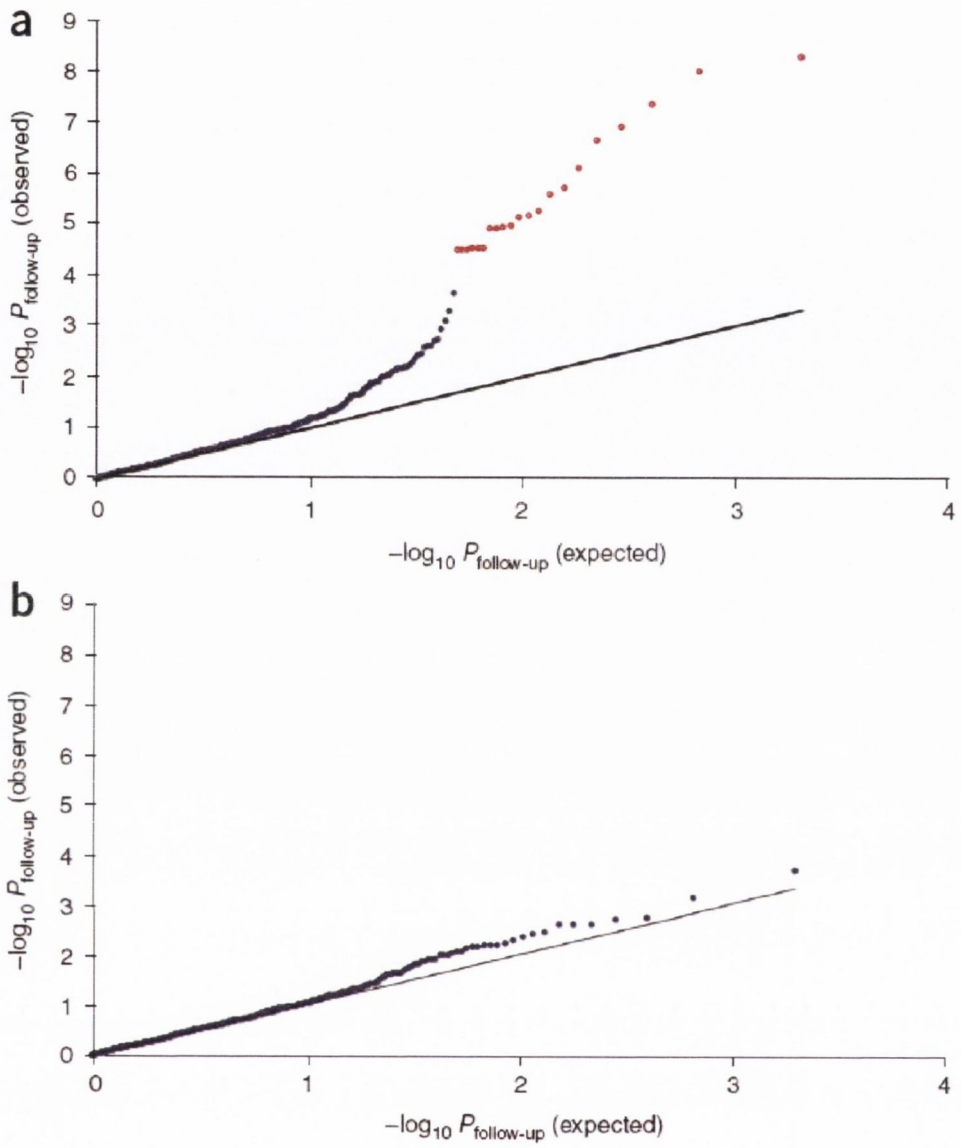



**Figure 4.3 Quartile-Quartile Plot of Observed vs. Expected significant observations.**

**Table 4.37 Genomic Regions showing association with coeliac disease in the GWAS follow-up study.**

| Region | Gene |
|---|---|
| 1 | IL2/IL21 |
| 2 | RGS1 |
| 3 | IL-18RAP |
| 4 | IL-12A/ SCHIP1 |
| 5 | LPP |
| 6 | TAGAP |
| 7 | LNK, ATXN2 |
| 8 | CCR1/CCR3 |

IL-2 and IL-21 have already been discussed above, leaving the likelihood of the genes within the other 7 regions as candidate genes to be considered.

**RGS1**: A cytosolic membrane bound G-protein, which acts to attenuate the signalling of other G-proteins. It has shown the ability to regulate chemokine receptors in germinal center B lymphocytes resulting in a change in homing patterns (Le *et al.* 2005, Han *et al.* 2006). The RGS1 region and has shown association with juvenile idiopathic arthritis (Hinks *et al.* 2010), type I diabetes (Smyth *et al.* 2008) and also recently multiple sclerosis (International Multiple Sclerosis Genetics Consortium, 2010). The association with CD has been confirmed in a Norwegian-Swedish case control population (Amundsen *et al.* 2009), as well as an Italian case control population (Romanos *et al.* 2009).

**IL18RAP:** Along with IL18R1, this SNP encodes the two proteins that form the IL-18 receptor. Their binding to IL-18 is required for high affinity binding of IL-18, and subsequent IL-18 signalling via NFkB (Debets *et al.* 2000). Juvenile idiopathic arthritis (de Jager *et al.* 2009), asthma, atopy, (Reijmerink *et al.*2008) and Crohn's disease along with ulcerative colitis (Zhernakova *et al.* 2008) have all shown association with the IL-18RAP region, and the association with coeliac disease has been replicated in

Spanish (Dema *et al.* 2009), Finnish, Hungarian and Italian populations (Koskinen *et al.* 2009).

**IL-12A**: codes for a subunit of the IL-12 cytokine. IL-12 is a potent mediator of the immune system, regulates apoptosis of Th1 T-Cells, and influences intrathymic T cell development in conjunction with IL-2 and Il-18 (see IL-18 in Section **5.3**). The Il-12 cytokines family has also been shown to play a prominent role in inflammatory, autoimmune disorders and infectious disorders (Gee *et al.* 2009), and IL-12 has been linked with a host of autoimmune diseases.

**LPP:** has a critical role in cytoskeletal remodeling, responding to vascular injury, and inducing migration of smooth muscle cells (Gorenne *et al.* 2006, Jin *et al.* 2007). LPP has shown association with juvenile idiopathic arthritis (Hinks *et al.* 2010), and CD in an independent Swedish-Norwegian family cohort (Amundsen *et al.* 2010).

**TAGAP:** codes for a protein that induces T-Cell activation; although its exact nature and mode of action is unknown, it is believed it may be a Rho GTPase-activating protein (Mao *et al.* 2004). In addition to its association with CD, it has also shown an association with rheumatoid arthritis, type I diabetes (Eyre *et al.* 2010) and Crohn's disease (*Franke et al.* 2010). It has also seen its effect replicated in Italian and Swedish-Norwegian cohorts (Romanos *et al.* 2009, Amundsen *et al.* 2010).

**LNK**: codes for an adaptor protein, which mediates the downstream activity of a host of growth factor and cytokine receptors. An inhibitor of cytokine signalling, LNK affects its negative regulation through induction of TNF alpha (Fitau *et al.* 2006). The gene region is associated with JIA (Hinks *et al.* 2010), multiple sclerosis (Alcina *et al.* 2010), and rheumatoid arthritis (Stahl *et al.* 2010).

Investigations in an African population have revealed that LNK shows a response to bacterial lipopolysaccharide and muramyl dipeptides. Carriers of the A allele in rs3184504 show stronger activation of the NOD2 recognition pathway, suggesting a role in protection against bacterial infection (Zhernakova *et al.* 2010). It has also seen its effect replicated in Italian and Swedish-Norwegian cohorts (Romanos *et al.* 2009, Amundsen *et al.* 2010).

**CCR1**: codes for a transmembrane chemokine receptor which is capable of transducing a cheomkine binding interaction into downstream NFkB activation (Lee & Wong 2009). In addition to mediating the inflammatory response, CCR1 also mediates cells migration, and has been shown to direct mast cells to sites of allergic

mucosal inflammation (Juremalm *et al.* 2002). As regards inflammatory disorders, infiltrative neutrophils in the lungs of patients with inflammatory lung disease have been shown to display a novel arrangement of cytokine receptors, including CCR1 (Hartl *et al.* 2008). As infiltrated neutrophils play a role in damage to the lamina propria in CD, this may indicate a role for CCR1 in the mediation of localised inflammation.

**CCR3**:'s product is a chemokine receptor, similar to CCR1. CCR3 has a binding affinity for Mucosae-associated Epithelial Chemokine (MEC). Constitutively expressed by epithelial cells, with up-regulation in mucosal sites such as the colon, MEC induces immune cell migration through CCR3, and likely plays a role in the recruitment of eosinophils and, potentially, of rare $CCR3^{+}$ T cells into mucosal tissues (Pan *et al.* 2000). Indeed, up-regulation of CCR3 appears to be a characteristic of mucosal inflammation in colonic epithelial cells of ulcerative colitis patients (Manousou *et al.* 2010). IL-2 mediated expression of CCR3 has been shown to result in apoptosis of B cells, in a manner that attenuates the humoral reponse (Jinquan *et al.* 2003). In terms of autoimmune disease, CCR3 is associated with JIA, displays elevated circulating levels of protein in allergy (Azazi *et al.* 2009), and is associated with asthma and atopy (Al-Abdulhadi *et al.* 2010). Its role in CD has been confirmed in Spanish, Italian and Swedish-Norwegian populations, in the same papers as mentioned earlier.

In conclusion, this multi-population approach, genotyping a larger proportion of SNPs in all three populations than the two phase IL-2/IL-21 replication, has proved to be more successful. Seven novel coeliac susceptibility regions have been identified, all associated with genes which play significant roles in the immune system, and the IL-2/IL-21 association has been further confirmed.

### 4b.4.2 SNPs Suggestive of Significance in the Irish Population

Population specific variants may play an important role in coeliac disease, however the nature of GWAS studies do not lend themselves to their identification. Even modest correction for multiple testing will inflate the level of significance required to confidently indicate association to the point where examination of a single population will not result in any definitely significant SNPs, due to the number of SNP loci involved in the experiment, and to the relatively small sample size. One of the reasons multiple populations are used in GWAS studies is to achieve enough power to overcome the correction for multiple testing.

On that basis, we cannot confidently ascribe association to the SNPs which showed significance prior to correction in the Irish population. However, it is worth examining which SNPs did show population specific associations, as it may be indicative of underlying population specific genetic factors.

The fact that SNPs may evidence association in one population and not another may be due to one of several factors. It could, of course, simply be that the association is by chance, as the number of SNPs being examined in a large scale association study of this sort will generate a high degree of false positives. However, there are other potential causes which must also be considered.

We compared the allele frequencies of the significant Irish SNPs in both controls and cases across all three populations. A large difference in control and case allele frequency may be indicative of differing haplotype substructures between the populations. To put it another way, LD patterns may differ from one European population to another. This can result in a different suite of markers being in high LD with causative variants in separate populations. As the strong likelihood is that the SNPs in this study associated with CD do not represent the actual causative variants, but rather are in LD with one or more variants which are the real factors, it should be kept in mind that we are effectively testing proxies. An observation of association may be due to the tag effectively tagging a causative variant, which it does not tag in a population that does not show association.

Different patterns of LD between the populations may result in a SNP being an effective proxy for a causative variant in one population, but not so in another. In this way, although the causative variants which underlie the disease may be represented in different populations, the strength and pattern of their association with disease causing SNPs may not be the same.

Otherwise, if the control allele frequencies are similar, but the case alleles are different, this may suggest a similar haplotype structure across the populations, but a population specific causative variant at work in the population.

There is also the issue of power. All three populations reported above are of differing sizes, and thus the power to detect effects will vary across populations.

Lastly, coeliac disease is a complex disorder which owes its pathology to numerous interacting genetic factors of relatively small effect. Although the HLA association has

proved necessary across all populations tested thus far, the relative contributions of more minor, modifying genes, may vary from population to population.

We set an arbitrary threshold of P = 0.015 at the allelic level for inclusion for discussion in this chapter. Several of the most strongly associated SNPs were also associated in the other 2 populations, and showed overall significance when analyzed via the CMH test (rs6822844, rs13119723 and rs13397583 in the IL-2/IL-21 region, and rs17810546 near IL-12A). Several of the other SNPs are in, or near candidate genes which we will discuss briefly below.

**RARB**: retinoic acid receptor, beta binds retinoic acid, and subsequently directs cellular signalling in embryonic morphogenesis, cell growth and differentiation (Marinelli *et al.* 2007).

**CLEC7A:** C-type lectin domain family 7 member A, is a non-toll-like pattern recognition receptor, which is primarily expressed on myeloid cells (Reid *et al.* 2009). It plays a major role in the recognition of, and subsequent reaction to, fungus (Cunha *et al.* 2010)and mycobacterium (Lee *et al.* 2009, van de Veerdonk *et al.* 2010). CLEC7A recognises the Beta glucans in the yeast cell walls (Kankkunen *et al.* 2010), and induces the inflammatory response via NkKappaB (Gringhuis *et al.* 2010).

Increase in CELC7A expression has also been associated with Inflammatory Bowel Disease (de Vries *et al.* 2009) and Psoriasis (de Koning *et al.* 2010). The receptor also binds T-lymphocytes at a separate binding site, suggesting that CLEC7A can recognize both endogenous and exogenous ligands (Willment *et al.* 2001).

**DENND2A:** DENND2A DENN/MADD domain containing 2A. Although a validated gene, its function is currently unknown. The MADD protein, with which it shares homology, however, is. MADD is involved in apoptosis; lack of endogenous MADD expression results in large scale apoptosis of cells through direct and indirect means (Li *et al.* 2010).

rs269243, the SNP in DENND2A which shows association with CD is a missense causing SNP, causing a change from Proline to Histidine at position 156 of the protein. It has no currently known clinical significance.

**TRPM3:** transient receptor potential cation channel, subfamily M, member 3 codes for a pore-forming channel subunit. It is subject to alternate splicing, with different isoforms resulting in channels with different selectivity for transit through the pore

(Oberwinkler & Phillipp 2007). Although the exact roles of these alternately spliced variants are currently poorly understood, another member of the subfamily, TRPV4, plays a central role in modulating epithelial barrier function (Harteneck & Reiter 2007) and it is possible that one or more of the alternate transcripts of TRPM3 may play a similar role.

**SLC15A2:** (solute carrier family 15 (H+/peptide transporter), member 2) is highly expressed by human myeloid cells, and plays a role in NOD2-dependent NF-kappaB activation. SLC15A2 occurs in phagosome membranes, and is involved in transporting bacteria-derived peptides from phagosomes (Charrière *et al.* 2010).

Analysis of SLC15A2 has shown that bacterial dipeptides are suitable substrates for SLC15A2, and that the protein has a key role in microbial recognition by NLR proteins, especially in the lung (Swaan *et al.* 2008).

SLC15A2 is expressed in kidney, brain, lung, and mammary gland (Pinsonneault *et al.* 2004) but is closely related to the SLC15A1 gene, whose protein is the pre-eminent transporter for di- and tripeptides at the intestinal brush border epithelium (Geissler *et al.* 2010). Both bind similar substrates, although SLC15A1 exhibits markedly reduced binding affinity for most of the substrates (Döring *et al.* 2002).

rs2257212, the SNP in SLC15A2 which shows association with CD, is a missense causing SNP, which causes a Leucine to Phenylanaline change at position 319. The SNP has no currently recognised clinical significance.

**POP1:** processing of precursor 1, ribonuclease P/MRP subunit, Pop1 is a ribonuclease that functions in the nucleus in pre-RNA processing (Welting et al. 2004) and has been identified as an autoantigen in patients suffering from connective tissue diseases (Lygerou *et al.* 1996).

**LRRN2:** Leucine rich repeat neuronal 2. It is unknown what this gene's exact function is, but the protein for which it codes exhibits sequence homology with other proteins which function as cell-adhesion molecules or as signal transduction receptors (Almeida *et al.* 1998).

C6orf195 and KLHL29 have no known relevant functions. rs9503233 is a missense causing SNP in C6orf195, causing a Proline to Serine substitution at position 43 of the protein.

Examination of SNPs which are suggestive of association in the Irish population reveals several excellent candidate genes. Unlike the genes which evidenced association across all 3 populations, whose functions were largely immune based, many of these are of currently unknown or only suspected function (though genes with which they show homology have functionally relevant roles). However, two of these genes. SLC15A2 and CLEC7A are involved in the recognition of, and subsequent immune reaction to, microbial antigens, which may be significant as a contributory factor in the development of active CD in the Irish population. It is also worth noting that 3 non-synonymous SNPs suggest association with CD in the Irish population (rs269243 in DENND2A, rs2257212 in SLC15A2 and rs9503233 in C6orf195).

### 4c.4 Replication of significant Golden Gate SNPs in an Italian population

Replication of the significant results in the Italian population showed more similarity than the attempted replication of the IL-2/IL-21 locus, with all of the SNPs showing association, except rs2816316, near RGS1. Another study in a separate Italian population showed similar allelic frequencies (Romanos *et al.* 2009). A comparison of the two Italian populations is shown below (**Table 4.38**)

We tested rs9851967 for association near LPP, while they tested rs1464510. Both were significant. IL18RAP and CCR3 failed to show significance in their population, though RGS1 did. This may be due to differing haplotype structures between Southern and Northern Italy, or simply due to lack of power.

The broad agreement between the Italian populations and the joint analysis of the 3 populations in terms of which LD blocks show significance may indicate that haplotype substructure differences in the Italian population are confined to specific gene areas, such as the IL-2/IL-21 LD block, rather than being genome-wide.

The replication of all 7 of these novel coeliac loci between two independent Italian populations, would seem to confirm them as true disease susceptibility loci. That the two populations tested are not in complete agreement as to the significance of all 7 loci suggests that there may be a different pattern of LD between northern and southern Italy, perhaps correlated with the observed change in the frequency of HLA predisposing allelic variants between Northern and Southern Europe. Alternatively, as the allele frequencies in both case and controls are very similar in the two

populations (except for rs6441961 near CCR3 which shows a 4.6% difference in the case population) this may simply be due to lack of power.

**Table 4.38 Comparison of GoldenGate replications across two independent Italian populations.**

| SNP | Gene | Alleles | Our Case | Their Case | Our ctrl Maj | Their Control | P-Value |
|---|---|---|---|---|---|---|---|
| rs9851967 | LPP | C/T | 0.413 | NA | 0.498 | NA | 0.0009 |
| rs2816316 | RGS1 | A/C | 0.146 | 0.128 | 0.174 | 0.166 | 0.1459 |
| rs917997 | IL18RAP | C/T | 0.258 | 0.244 | 0.214 | 0.227 | 0.0430 |
| rs6441961 | CCR3 | C/T | 0.418 | 0.372 | 0.357 | 0.368 | 0.0145 |
| rs3184504 | LNK | T/C | 0.432 | 0.446 | 0.4827 | 0.506 | 0.0482 |
| rs1738074 | TAGAP | C/T | 0.462 | 0.454 | 0.390 | 0.412 | 0.0044 |
| rs17810546 | IL12A | A/G | 0.099 | 0.109 | 0.062 | 0.067 | 0.0081 |

# CHAPTER V

# Candidate Genes

# &

# Coeliac Disease

## 5.1 Introduction

### 5.1.1 Candidate Gene Selection

As discussed in Chapter 1, proximity to established linkage peaks is one, but not the only, commonly used strategy for selection of genes for investigation.. A candidate gene approach, in which the gene is selected for study based on the known or suspected function of its product, or pathway in which it plays a part, is also a favoured approach.

Several genes have may be considered as suitable candidates for analysis based on functional criteria and previous disease associations, and are listed below.

### 5.1.2 CCR5

The CCR5 gene codes for a transmembrane protein which is involved in homing of macrophages and T-lymphocytes, as well as the proliferation and differentiation of T-Cells (Oppermann 2004).

CCR5 contains a mutation which is relatively common in European populations (5-14%); a 32 base pair deletion which results in premature termination of transcription, and a resulting truncated, non-functional protein. Individuals who carry the delta 32 mutation are healthy, suggesting that the receptor may not play a unique, essential role in the immune response (Sabeti *et al.* 2005). However due to the role CCR5 plays in T-Cell development and homing, individuals homozygous for the mutation have evidenced unusual reactions to certain diseases.

CCR5-Δ32 individuals have shown to be at higher risk of West Nile virus (Glass *et al.* 2006) indicating that not all of the functions of CCR5 may be redundant. Additionally, CCR5 acts as a co-receptor mediating the entry of certain strains of HIV into immune cells; individuals lacking a functional CCR5 gene have demonstrated resistance to HIV infection (Li *et al.* 1997).

The unusually high incidence of this variant in European populations suggests that it may have had an historical significance in protecting against the effects of Smallpox or Bubonic plague (Sabeti *et al.* 2005).

CCR5-Δ32 carrier status has evidenced association with numerous autoimmune diseases, including rheumatoid arthritis, systemic lupus erythematosus (Martens *et al.* 2009), multiple sclerosis (Shahbazi *et al.* 2009) and juvenile idiopathic arthritis

(Hinks *et al.* 2010). However, in some cases possession of the Δ32 mutation is associated with increased risk of disease (MS and possibly SLE), and in other instances it is associated with decreased risk (RA and JIA).

These apparently contradictory findings may be explained by another property of CCR5; that of immune modulation. Studies in CCR5 (-/-) mice revealed an expected reduction in macrophage function and clearance of infection, but also displayed a delayed-type hypersensitivity reaction and increased humoral responses to T cell-dependent antigenic challenge (Zhou *et al.* 1998). The mouse model for the AI disease Sjögren's syndrome has also shown a lack of functional CCR5 contributes to the pro-inflammatory environment, indicating that CCR5 plays a part in down-modulating the T cell-dependent immune response (Wildenberg *et al.*, 2008).

Expression of CCR5 on the surface of $CD4^{+}CD25^{+}$ T Regulatory cells has also been shown to be a key requirement for these cells to successfully mitigate graft versus host disease, in addition to CCR5 inhibiting T-cell expansion in lymphoid tissues (Wysocki *et al.* 2005).

Due to its ability to both stimulate T-Cell function and modulate the immune response, as well as the relatively common prevalence of the Δ32 mutation, and its association to other AI diseases, it was deemed an excellent candidate for investigation.

### 5.1.3 TNFSF14

TNFSF14 (also known as LIGHT) is a type II membrane protein, with the ability to stimulate the immune response, especially the destructive Th1 type of activity associated with intestinal lesions in coeliac disease (Jin *et al.* 2010). In addition to acting as a powerful costimulatory factor for T-Cells (Wan *et al.* 2002), TNFSF14 signals through the NF-Kappa B pathway, and has been shown to up-regulate CD86 and Stat 3 (Zou *et al.* 2005, Nadiminty *et al.* 2007). Also, crucially, TNFSF14 has the ability to promote T cell proliferation and differentiation, even in a T reg-rich environment, a property which may mark it as key in the development of autoimmune inflammatory diseases (Wang *et al.* 2010).

TNFSF14 also plays a specific role in the regulation of the intestinal immune response. Constitutive, up-regulated, expression of TNFSF14 on T cells leads to lymphocyte activation, inflammation, and tissue destruction, in model animals, all of

which are most severe in the intestine, with mucosal T cells showing increased Th1 cytokine activity (Shaikh *et al.* 2001).

In healthy humans, cell surface TNFSF14 is constitutively expressed on mucosal T and NK cells, and a subpopulation of gut-homing CD4(+) T cells. Indeed, TNFSF14 has also been shown to regulate IFN gamma production in the intestine, with different patterns of expression between the large and small intestines (Brown *et al.* 2010).

Upregulation of TNFSF14 has been associated with active Crohn's Disease, with inflammatory bowel disease patients indicating elevated expression in biopsies from inflamed sites (Wang *et al.* 2005). TNFSF14 appears to exert its influence in this context via the driving of the Th1 immune response.

Nor is Crohn's the only autoimmune disease in which TNFSF14 has been implicated; in rheumatoid arthritis TNFSF14 has been shown to induce $CD14^{+}$monocytes to develop into cells with osteoclast like activity, which then participate in inflammatory bone destruction (Ishida *et al.* 2009). In addition, TNFRSF14, a major ligand of TNFSF14, has evidenced association with Coeliac (Dubois *et al.* 2010) involvement of this pathway with the underlying mechanisms of the disease.

### 5.1.4 IL-18

A pro-inflammatory cytokine, IL-18 signalling leads to up-regulation of the Th1 immune response, and production of IFN-γ (Koutoulaki *et al.* 2010). During positive selection in the thymus, IL-18 acts in conjunction with either IL-2, or Il-12, to skew T cell development towards either the Th1 or Th2 cells types, based on the relative levels of the other two cytokines, with IL-2 and IL-18 predisposing towards the Th2 subset, while IL-12 and IL-18 induce development towards the Th1 subtype (Rodriguez-Galá *et al.* 2005). With regards to disease association, IL-18 has shown association with rheumatoid arthritis, allergic inflammation, systemic lupus erythematosus, atopic eczema, behcet's disease and cardiovascular disease. Its extensive implication in auto immune disease, coupled with its pro-inflammatory nature, and ability to bias towards the destructive Th1 immune response made it an interesting candidate gene for investigation.

### 5.1.5 MyoixB

Polymorphisms in the MYOIXB gene have shown association with coeliac disease, ulcerative colitis and Crohn's cisease (Monsuur *et al.* 2005, Núñez *et al.* 2007,

Cooney *et al.* 2009), leading to the speculation that, as the gene product is known to be an actin associated molecular motor (Post *et. al* 2002), it may be involved in remodelling of actin cytoskeleton in the brush border epithelium, potentially resulting in increased gut permeability and consequently increased traffic of gluten epitopes across the gut.

A MyoIXb based haplotype showed association in a Dutch population (Monsuur *et al.* 2005), prompting us to attempt a replication of the effect in the Irish population.

### 5.1.6 Candidate Gene Biased Follow up of Multi-population GWAS

The previous chapter has detailed the study design for the Genome Wide Association Study in multiple populations, whereby the 1020 most strongly associated single nucleotide polymorphisms from the British GWAS were tested in three case–control cohorts (Irish, British and Dutch).

This approach resulted in the identification of eight novel non-HLA loci associated with CD. Examination of the nature of the SNPs involved revealed the following characteristics:

1) **Immune Function is an indicator**; 7 out of the 8 SNPs were located in genes with known Immune function. 4 of these are cytokines, or their receptors.
2) **Underlying Mechanisms of Autoimmunity**: 4 of the loci are in genes which have evidenced association with other autoimmune conditions.
3) **Hierarchy is no guarantee**; only 3 of the 8 loci were located in the top-100 associated signals in the original British GWAS. 3 of the SNPs ranked below the top-500, and 1 was ranked in 1004$^{th}$ place.

With these observations in mind, we elected to join a Dutch led replication study; one in which the SNP selection would be biased towards the inclusion of markers near or within suitable candidate genes; in this case the criteria for suitability being known immune function, or previously established association with an autoimmune condition.

300 SNPs were selected purely on their P values in the initial GWAS (between p.0.000275 and p = 0.004). 55 SNPs were selected for inclusion based on their

mapping to gene regions of known immune function (also a P <0.05 in the original GWAS). Finally, 103 were selected for inclusion based on their previous association of p <0.05 with type 1 diabetes (T1D), rheumatoid arthritis (RA) or Crohn's disease in the WTCCC study (original GWAS P <0.009).

## 5.2 Objectives

The aims of this chapter are:

1. Genotyping of the CCR5 Delta 32 deletion in an Irish Case Control population.
2. Genotyping of previously associated SNPs in and around TNFSF14.
3. Genotyping the IL-18 promoter SNPs that make up the functional haplotype.
4. Selection of a panel of candidate genes for inclusion in a large scale Illumina study.
5. Carry out statistical tests to determine any association between genotypes or possession of alleles and the disease state.
6. Where haplotype tagging has occurred, also test for haplotype association with disease state.

## 5.3 Results

### 5.3.1 TNFSF14

Previous work in our group had determined a panel of tagging SNPs which represented all TNFSF14 haplotypes with a Frequency > 4%. These 3 SNPs; rs2291668, rs3760746 and rs11878563 had all been genotyped in the Irish population as part of a previous study, along with rs2279627 (a SNP located in the 3' region of TNFSF14) and rs344560 (the only non-synonymous SNP in TNFSF14 with a MAF > 5%). Figure 5.1 shows the located of these SNPs within the TNFSF14 gene.

Three of these SNPs showed a minor level of significance, however, at the time the study was undertaken, the sample size of both our case and control populations were significantly smaller (case = 370, control =335).

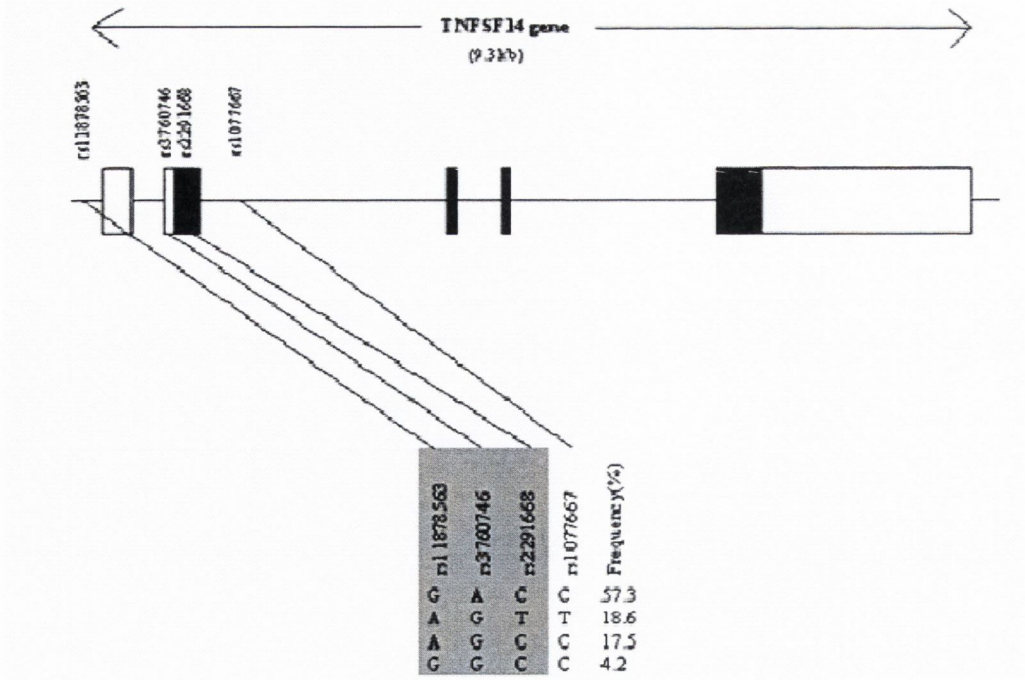



**Figure 5.1 The structure of the TNFSF14 gene and location of haplotype tags.**

With the increase in sample size of both collections over time, and corresponding increase in power, it was determined to further genotype these variants in the new samples.

### 5.3.2 IL-18

rs1946518 and rs187238 are located in the UTR region of IL-18, being at positions -607 and -137 relative to the initiation of transcription respectively. SNPs in IL-18 have evidenced association to a number of autoimmune diseases, including atopic asthma, Behcet's disease, cardiovascular disease, and rheumatoid arthritis (Tiret *et al.* 2005, Imboden *et al.* 2006, Jang *et al.* 2006,).

rs1946518 and rs187238 are believed to influence the regulation of the IL-18 gene by altering regulatory elements in the promoter. The rs187238G to C substitution (IL18-137) disrupts a histone 4 transcription factor-1 (H4TF-1) binding site while the C to A transition at rs1946518 (IL18-607) alters a consensus cyclic adenosine monophosphate (cAMP)-responsive element protein-binding site. Both of these SNPs have been shown to cause altered transcription factor binding and mRNA expression (Giedraitis *et al.* 2002). **Figure 5.2** shows the located of these SNPs in realtion to the IL-18 gene.

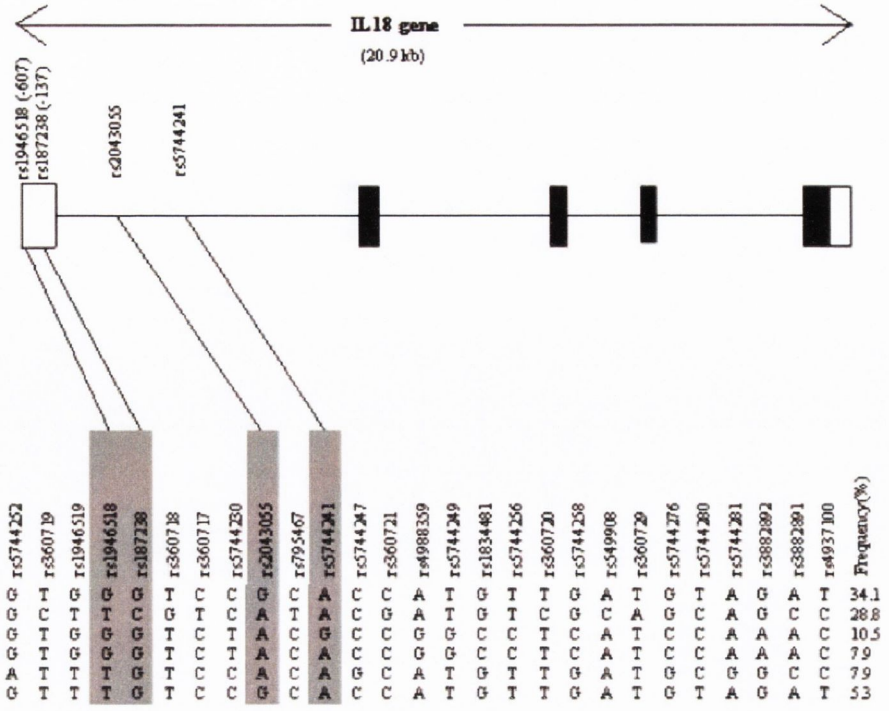



**Figure 5.2 Structure of the IL-18 gene, and location of the promoter proximal polymorphisms.**

### 5.3.3 MyoIXB

A study in the Dutch population revealed a MyoixB haplotype associated with CD (Monsuur *et al.* 2005). The tagging SNPs which comprised the associated haplotype were rs2305767, rs1457092, rs2305765 and rs2305764. We resolved to genotype the tagging SNPs, and test the corresponding haplotype for significance in the Irish population.

### 5.3.4 Allele, Genotype and Carrier Status frequencies and disease association testing of CCR5 Δ32

CCR5 Δ32 (rs333) was analyzed for allelic, genotypic and carrier status, and the results, along with Odds Ratios and 95% confidence intervals, are tabulated on tables **5.1** to **5.5.** An association with disease is observed at both the allelic and genotypic levels, (P = 0.00816 and P = 0.008 respectively). The In/In genotype shows a protective effect (OR = 0.733, 95%CI = 0.57-0.93) with the In/Del genotype predisposing towards disease (OR = 1.32, 95%CI = 1.03-1.68). Analysis of carrier

status reveals a significant association with the minor allele (Deletion) and disease status (P = 0.012, OR = 1.36, 95%CI = 1.07-1.74).

**Figure 5.3** shows a photograph of the agarose gel based genoptying used in identifying this polymorphism.

**Table 5.1 Allele Frequencies of CCR5 Δ32 (rs333). The deletion polymorphism is significantly more frequent in the cases, than the controls (p =0.00816).**

| SNP | Alleles | Case Maj | Ctrl Maj | Case Mi | Ctrl Mi | P-Value | Odds Ratio | 95% CI |
|---|---|---|---|---|---|---|---|---|
| Delta-32 | In/Del | 0.858 | 0.890 | 0.142 | 0.110 | 0.00816 | 1.34 | 0.98-1.83 |

**Table 5.2 Genotype Frequencies of CCR5 Δ32; there is a significant association between genotype status with regard to the deletion polymorphism and coeliac disease.**

| SNP | Alleles | Case D/D | Ctrl D/D | Case I/D | Ctrl I/D | Case I/I | Ctrl I/I | P-Value |
|---|---|---|---|---|---|---|---|---|
| rs333 | In/Del | 0.013 | 0.005 | 0.258 | 0.209 | 0.729 | 0.786 | 0.008 |

**Table 5.3 Insert Dominant/Deletion Recessive Model of CCR5 Δ32.**

| SNP | Case (NN)+(Nn) | Case (nn) | Ctrl (NN)+(Nn) | Ctrl (nn) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| Delta-32 | 0.987 | 0.013 | 0.995 | 0.005 | 0.11255 | 0.65 | 0.23-1.89 |

**Table 5.4 Insert Recessive/Deletion Dominant Model of CCR5 Δ32.**

| SNP | Case (Nn)+(nn) | Case (NN) | Ctrl (Nn)+(nn) | Control (NN) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| Delta-32 | 0.272 | 0.729 | 0.785 | 0.215 | 0.01230 | 1.36 | 1.07-1.74 |

**Table 5.5 Irish Case and Control numbers in CCR5 Δ32.**

| N Case | N Controls |
|---|---|
| 549 | 960 |

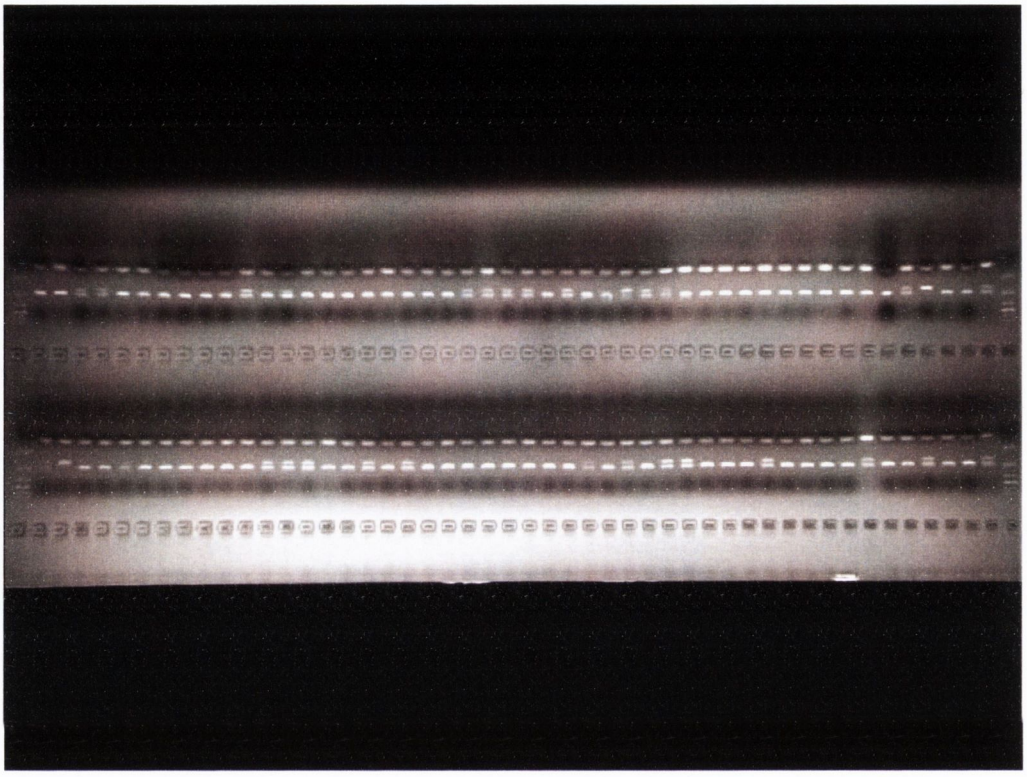


**Figure 5.3 Agarose Gel of CCR5 Δ32 genotyping.**

### 5.3.5 Allele, Genotype and Carrier Status frequencies and disease association testing of TNFSF14

rs2279627, rs3760746 and rs11878563were analyzed for allelic, genotypic and carrier status, and the results, along with Odds Ratios and 95% confidence intervals, are tabulated on tables **5.6** to **5.9.** All 3 SNPs showed association at the allelic and genotypic frequency level. **Table 5.10** details the sample size, per SNP.

In rs2279627 possession of the minor G allele confers an increased risk of disease (P = 0.004, OR = 1.39, 95%CI = 1.11-1.76) with the CC genotype appearing to be protective (OR = 0.72, 95%CI = 0.57-0.89)

Conversely, in rs3760746, it is possession of the major T allele which is associated with increased risk of disease (P = 0.037, OR = 1.36, 95%CI = 1.02-1.81), with the CC genotype being protective (OR = 0.74, 95%CI 0.55-0.98).

rs11878563 also evidences association with its major G allele and disease. On a genotypic level it displays the strongest association of the 3 SNPs (P = 0.00167, OR = 1.28, 95%CI = 1.022-1.62). On the carrier status level, possession of the G allele is associated with, and increases risk of, disease (P = 0.0051, OR = 1.51, 95%CI = 1.13-2.01).

**Table 5.6 Allele Frequencies of TNFSF14 SNPs; all three of the previously identified tagging SNPs show association with coeliac disease on the allelic level.**

| SNP | Alleles | Case Maj | Ctrl Maj | Case Mi | Ctrl Mi | P | Odds Ratio | 95% CI |
|---|---|---|---|---|---|---|---|---|
| rs2279627 | C/G | 0.723 | 0.765 | 0.277 | 0.235 | 0.01464 | 1.25 | 0.96-1.62 |
| rs3760746 | T/C | 0.569 | 0.520 | 0.431 | 0.480 | 0.03104 | 1.22 | 0.96-1.53 |
| rs11878563 | G/A | 0.591 | 0.528 | 0.409 | 0.472 | 0.00167 | 1.29 | 1.02-1.61 |

**Table 5.7 Allele Frequencies of TNFSF14 SNPs; all three of the previously identified tagging SNPs show association with coeliac disease on the genotypic level.**

| SNP | Alleles | Case nn | Ctrl nn | Case Nn | Ctrl Nn | Case NN | Ctrl NN | P-Value |
|---|---|---|---|---|---|---|---|---|
| rs2279627 | C/G | 0.064 | 0.062 | 0.426 | 0.346 | 0.510 | 0.593 | 0.02434 |
| rs3760746 | T/C | 0.188 | 0.240 | 0.486 | 0.478 | 0.326 | 0.280 | 0.03279 |
| rs11878563 | G/A | 0.168 | 0.234 | 0.482 | 0.468 | 0.350 | 0.290 | 0.00264 |

**Table 5.8 Major Dominant/Minor Recessive Model of TNFSF14 SNPs.**

| SNP | Case (NN)+(Nn) | Case (nn) | Control (nn)+(Nn) | Control (nn) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| rs2279627 | 0.936 | 0.064 | 0.938 | 0.062 | 0.88477 | 0.97 | 0.61-1.55 |
| rs3760746 | 0.812 | 0.184 | 0.761 | 0.240 | 0.03685 | 1.36 | 1.02-1.81 |
| rs11878563 | 0.832 | 0.168 | 0.766 | 0.234 | 0.00506 | 1.51 | 1.13-2.01 |

**Table 5.9 Minor Dominant/Major Recessive Model of TNFSF14 SNPs.**

| SNP | Case (Nn)+(nn) | Case (NN) | Ctrl (Nn)+(nn) | Ctrl (NN) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| rs2279627 | 0.490 | 0.510 | 0.407 | 0.593 | 0.0041 | 1.39 | 1.11-1.76 |
| rs3760746 | 0.674 | 0.326 | 0.720 | 0.280 | 0.0864 | 0.80 | 0.62-1.03 |
| rs11878563 | 0.650 | 0.350 | 0.710 | 0.290 | 0.0249 | 0.76 | 0.59-0.97 |

**Table 5.10 Irish Case and Control numbers in TNFSF14**

| SNP | N Case | N Controls |
|---|---|---|
| rs2279627 | 502 | 729 |
| rs3760746 | 484 | 693 |
| rs11878563 | 506 | 732 |

### 5.3.6 Allele, Genotype and Carrier Status frequencies and disease association testing of IL-18 SNPs.

Both the -607 and -137 promoter polymorphisms were analyzed for allelic, genotypic and carrier status, and the results, along with Odds Ratios and 95% confidence intervals, are tabulated on tables **5.11** to **5.14. Table 5.15** details the sample size, per SNP.Neither of the SNPs demonstrated significant association with coeliac disease at any level.

**Table 5.11 Allele Frequencies of IL-18 SNPs; although both promoter SNPs showed association with CD in a previous study in our lab, adding additional cases and controls proved the association to be spurious.**

| SNP | Alleles | Case Maj | Ctrl Maj | Case Mi | Ctrl Mi | P-Value | Odds Ratio | 95% CI |
|---|---|---|---|---|---|---|---|---|
| -137 | G/C | 0.732 | 0.726 | 0.268 | 0.274 | 0.75804 | 1.03 | 0.89-1.32 |
| -607 | G/T | 0.622 | 0.588 | 0.378 | 0.412 | 0.11676 | 1.16 | 0.92-1.45 |

**Table 5.12 Allele Frequencies of IL-18 SNPs; although both promoter SNPs showed association with CD in a previous study in our lab, adding additional cases and controls proved the association to be spurious.**

| SNP | Alleles | Case nn | Ctrl nn | Case nN | Ctrl nN | Case NN | Ctrl NN | P-Value |
|---|---|---|---|---|---|---|---|---|
| -137 | G/C | 0.073 | 0.081 | 0.390 | 0.385 | 0.537 | 0.534 | 0.75859 |
| -607 | G/T | 0.134 | 0.173 | 0.488 | 0.479 | 0.379 | 0.348 | 0.0976 |

**Table 5.13 Major Dominant/Minor Recessive Model of IL-18 SNPs.**

| SNP | Case (NN)+(Nn) | Case (nn) | Ctrl (NN)+(Nn) | Ctrl (nn) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| -137 | 0.927 | 0.073 | 0.919 | 0.081 | 0.56743 | 1.13 | 0.74-1.72 |
| -607 | 0.866 | 0.134 | 0.828 | 0.173 | 0.06414 | 1.35 | 0.98-1.85 |

**Table 5.14 Major Recessive/Minor Dominant Model of IL-18 SNPs.**

| SNP | Case (nn)+(Nn) | Case (NN) | Ctrl (nn)+(Nn) | Ctrl (NN) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| -137 | 0.463 | 0.537 | 0.466 | 0.534 | 0.89626 | 0.99 | 0.79-1.23 |
| -607 | 0.621 | 0.379 | 0.652 | 0.348 | 0.26819 | 0.88 | 0.70-1.11 |

**Table 5.15 Irish Case and Control numbers in IL-18**

| SNP | N Case | N Controls |
|---|---|---|
| -137 | 495 | 860 |
| -607 | 478 | 864 |

### 5.3.7 Allele, Genotype and Carrier Status frequencies and disease association testing of MyoIXB SNPs.

The 4 tgSNPs which showed association in the Dutch population (rs2305767, rs1457092, rs2305765 and rs2305764) were analyzed for allelic, genotypic and carrier status, and the results, along with Odds Ratios and 95% confidence intervals,

are tabulated on tables **5.16** to **5.19**. None of the 4 SNPs analyzed showed association at any level.

**Table 5.16 Allele Frequencies of MyoIXB SNPs; these SNPs showed association with coeliac disease in a Dutch case control study. The association has not been replicated, either in our Irish population, or other independent populations.**

| SNP | Alleles | Case Maj | Ctrl Maj | Case Mi | Ctrl Mi | P-Value | Odds Ratio | 95% CI |
|---|---|---|---|---|---|---|---|---|
| rs2305767 | A/G | 0.576 | 0.594 | 0.424 | 0.406 | 0.52608 | 1.07 | 0.80-1.44 |
| rs1457092 | C/A | 0.630 | 0.616 | 0.370 | 0.384 | 0.57702 | 1.06 | 0.79-1.43 |
| rs2305765 | C/T | 0.588 | 0.567 | 0.412 | 0.433 | 0.44045 | 1.09 | 0.82-1.46 |
| rs2305764 | A/G | 0.585 | 0.574 | 0.416 | 0.426 | 0.73016 | 1.04 | 0.77-1.42 |

**Table 5.17 Genotype Frequencies of MyoIXB SNPs; these SNPs showed association with coeliac disease in a Dutch case control study. The association has not been replicated, either in our Irish population, or other independent populations.**

| SNP | Alleles | Case nn | Case nN | Case NN | Ctrl nn | Ctrl nN | Ctrl NN | P-Value |
|---|---|---|---|---|---|---|---|---|
| rs2305767 | A/G | 0.160 | 0.528 | 0.312 | 0.164 | 0.485 | 0.351 | 0.53209 |
| rs1457092 | C/A | 0.133 | 0.473 | 0.394 | 0.154 | 0.460 | 0.386 | 0.58179 |
| rs2305765 | C/T | 0.154 | 0.515 | 0.331 | 0.192 | 0.483 | 0.326 | 0.41666 |
| rs2305764 | A/G | 0.157 | 0.516 | 0.327 | 0.176 | 0.497 | 0.325 | 0.73362 |

**Table 5.18 Major Dominant/Minor Recessive Model of MyoIXB SNPs.**

| SNP | Case (NN)+(Nn) | Case (nn) | Ctrl (NN)+(Nn) | Ctrl (nn) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| rs2305767 | 0.840 | 0.160 | 0.836 | 0.164 | 0.88478 | 1.03 | 0.70-1.52 |
| rs1457092 | 0.867 | 0.133 | 0.846 | 0.154 | 0.39878 | 1.20 | 0.79-1.18 |
| rs2305765 | 0.846 | 0.154 | 0.808 | 0.192 | 0.17413 | 1.30 | 0.89-1.91 |
| rs2305764 | 0.846 | 0.157 | 0.824 | 0.176 | 0.96433 | 1.17 | 0.78-1.76 |

**Table 5.19 Major Recessive/Minor Dominant Model of MyoIXB SNPs.**

| SNP | Case (GG)+(AG) | Case (AA) | Control (GG)+(AG) | Control (AA) | P | OR | 95% |
|---|---|---|---|---|---|---|---|
| rs2305767 | 0.688 | 0.312 | 0.649 | 0.351 | 0.26241 | 1.19 | 0.88-1.61 |
| rs1457092 | 0.606 | 0.394 | 0.614 | 0.386 | 0.88478 | 0.97 | 0.72- 1.30 |
| rs2305765 | 0.669 | 0.331 | 0.674 | 0.326 | 0.87946 | 0.98 | 0.72-1.33 |
| rs2305764 | 0.674 | 0.327 | 0.675 | 0.325 | 0.96433 | 0.99 | 0.72-1.38 |

**Table 5.20 Irish Case and Control Numbers for MyoIXB.**

| SNP | N Case | N Controls |
|---|---|---|
| rs2305767 | 413 | 342 |
| rs1457092 | 414 | 350 |
| rs2305765 | 402 | 344 |
| rs2305764 | 343 | 323 |

### 5.3.8 Candidate Gene Biased Follow up of Multi-population GWAS in the Irish Population

As part of this study, we genotyped 444 Irish Coeliacs and 960 Irish controls in the functional-gene-enriched panel on the Goldengate platform, as part of a larger study involving independent British (724 cases, 1398 controls) and Dutch (514 cases, 900 controls) populations. The three populations were analyzed separately, and as a whole, using the CMH test.

**Table 5.21** shows the SNPs which evidenced association in the Irish population, at a threshold of $P < 0.01$ before correction for multiple testing.

**Table 5.22** lists the SNPs which proved globally significant ($P = <10^{-4}$) after analysis via the CMH test and their P value in the Irish population.

The following tables (5.a1-5.a5) can be found in Appendix 1.

**Table 5.a1** compares the P values of SNPs which evidenced the strongest association in the Irish population, with the P values of the same SNPs in the other 2 populations. Only rs2327832, near to the TNFAIP3 gene on chromosome 6 and rs2797825, near

VAV2 on chromosome 9, also show significant associations in other populations. rs2327832 is significant in the UK population, but not the Dutch, whereas VAV2 shows significance in the Dutch population, but not the UK.

**Table 5.a2** examines the SNPs which showed global significance following CMH testing, and lists the P value of each SNP by population. 10 of the 13 SNPs showed significance in the Dutch population, with only rs2327832 (near TNFAIP3), rs4439920 and rs9533935 showing no evidence of association.

5 SNPs showed association in the British population; rs7559479 and rs3817465 near IL-18RAP, rs2327832 near TNFAIP3 and rs4439920 and rs4324288 (near ITGA4).

The Irish population shows association with the fewest SNPs, with only 4 evidencing significance; rs2327832 (near TNIP1), rs842639 and rs842647 (near REL) and rs2279001 (near MyoIXb).

We examined the frequencies in the cases and control samples across all three populations in regard to the allele frequencies of their associated SNPs. **Tables 5.a3-5.a5** show these comparisons.

**Table 5.21 Highly Significant SNPs in the Irish Population; these are the SNPs from a Dutch led, candidate gene enriched GWAS follow up study which, although not significant when corrected for multiple testing, were the most significant in the Irish population. They may indicative of population specific variants unique to the Irish.**

| Gene | Chr | SNP | P | Source | SNP Type |
|---|---|---|---|---|---|
| SLC15A2 | 3 | rs2141615 | 0.0002139 | Immune | Intronic SLC15A2 |
| TGFA | 2 | rs377122 | 0.000958 | Immune | Intronic TGFA |
| TGFA | 2 | rs432203 | 0.001402 | Immune | Intronic TGFA |
| HMHB1 | 5 | rs325238 | 0.00191 | Top | Near to HMHB1 |
| TNFSF14 | 19 | rs2279627 | 0.003535 | Candidate | 3'UTR TNFSF14 |
| MACROD2 | 20 | rs6135590 | 0.003941 | Top | Intronic to MACROD2 |
| LOC401014 | 2 | rs4662248 | 0.004781 | Top | Intergenic, near |
| ACVR2A | 2 | rs1385363 | 0.005126 | Top | Intergenic |
| TNFAIP3 | 6 | rs2327832 | 0.005157 | Disease | Intergenic |
| VAV2 | 9 | rs2797825 | 0.008587 | Top | Intronic to VAV2 |
| loc235965 | 7 | rs11760312 | 0.008884 | Disease | Within Gene |
| TRIM60 | 4 | rs7684974 | 0.008912 | Top | Intergenic |
| DAO | 12 | rs3825251 | 0.009045 | Top | Intronic DAO |
| TMEM132B | 12 | rs10083119 | 0.009608 | Top | Intergenic |

**Table 5.22 CHM Result overall; these SNPs are SNPs from a Dutch led, candidate gene enriched GWAS follow up study which included Dutch, UK and Irish case control grops. The SNPs listed below are those which, when analysed via the CMH test, remained significant after multiple testing.**

| Gene | Chr | SNP | P CHM | P in Irish | |
|---|---|---|---|---|---|
| IL-18RAP | 2 | rs7559479 | 1.266e-006 | 0.4862 | Immune |
| IL-18RAP | 2 | rs3817465 | 3.304e-006 | 0.8642 | Immune |
| TNFAIP3 | 6 | rs2327832 | 6.607e-006 | 0.005157 | Disease |
| REL | 2 | rs842639 | 8.005e-006 | 0.04439 | Disease |
| REL | 2 | rs842647 | 9.256e-006 | 0.03128 | IBD Association |
| | 2 | rs4439920 | 1.184e-005 | 0.06171 | Proxy |
| CD80 | 3 | rs1599796 | 4.07e-005 | 0.33 | Immune |
| ITGA4 | 2 | rs4324288 | 4.474e-005 | 0.1902 | Immune |
| MyoIXb | 19 | rs2279001 | 4.599e-005 | 0.02181 | GWAS Rep |
| | 2 | rs1160542 | 6.684e-005 | 0.2377 | Disease |
| TAGAP | 6 | rs212402 | 8.101e-005 | 0.199 | Disease |
| | 13 | rs9533935 | 9.132e-005 | 0.4862 | Disease |
| | 2 | rs4851274 | 9.762e-005 | 0.8642 | Disease |

### 5.3.9 Haplotype analysis of TNFSF14 associated SNPs

Haplotype reconstruction was performed for the 3 TNFSF14 associated SNPs, and the results are shown on Table 5.23. One Haplotype, CCA (rs2279627, rs3760746, rs11878563 respectively) evidences association with CD (P = 0.036, OR = 1.29, 95%CI = 1.02-1.63).

**Table 5.23 TNFSF14 Haplotype (rs2279627, rs3760746, rs11878563); haploypte conconstruction and analysis reveals a single haplotype (CCA) which is significantly more prevelant in cases than controls (p =0.03577) conferring an additional risk (OR = 1.29).**

| Haplotype | F(Case) | F(Ctrl) | p | Odds Ratio | 95% CI |
|---|---|---|---|---|---|
| CCA | 0.464 | 0.401 | 0.03577 | 1.29 | 1.02-1.63 |
| CTG | 0.290 | 0.276 | 0.6084 | 1.07 | 0.83-1.38 |
| GTG | 0.270 | 0.232 | 0.1355 | 1.22 | 0.94-1.60 |

**Overall P = 0.1**

### 5.3.10 Haplotype analysis of IL-18 associated SNPs

Haplotype reconstruction is especially relevant in the case of the -607 and -137 IL-18 promoter polymorphisms, as it has been suggested that they interact directly with DNA binding molecules.

**Table 5.24** Shows the results of haplotype reconstruction for the two IL-18 SNPs. The rare CG Haplotype shows association with disease (P = 0.0044, OR = 4.4, CI = 1.63-11.87).

**Table 5.24 IL-18 Haplotpye (IL-137, IL-607). Haplotype consconstruction and analysis of the 2 IL-18 promotor SNPs shows a rare haplotype (CG) which is associated with the disease state (p =0.0044). Due to the extreme rarity of the haplotype is difficult to ascertain whether this is a real effect, although this rare haplotype has been reported to influence IL-18 expression.**

| Haplotype | F(Case) | F(Ctrl) | p | Odds Ratio | 95% CI |
|---|---|---|---|---|---|
| GG | 0.591 | 0.586 | 0.8902 | 1.02 | 0.82-1.29 |
| CT | 0.238 | 0.264 | 0.3068 | 0.87 | 0.67-1.13 |
| GT | 0.142 | 0.144 | 0.9279 | 0.99 | 0.72-1.36 |
| CG | 0.028 | 0.007 | 0.0044 | 4.40 | 1.63-11.87 |

**Overall P = 0.01**

### 5.3.11 Haplotype analysis of MyoIXB associated SNPs

Haplotype reconstruction was performed on the 4 MyoIXB tgSNPs. The results are shown on **Table 5.25.** None of the Haplotypes showed a significant association with disease in the Irish population.

**Table 5.25 MyoIXB Haplotypes. Haplotype reconstruction and analysis using the marker SNPs selected in MyoIXB revealed no one haplotype associated with disease, but did show an overall p value of 0.03.**

| Haplotype | F(Case) | F(Control) | P | Odds Ratio | 95% CI |
|---|---|---|---|---|---|
| GCCG | 0.431 | 0.402 | 0.4495 | 1.13 | 0.83-1.53 |
| AATA | 0.353 | 0.372 | 0.6159 | 0.92 | 0.67-1.26 |
| ACCG | 0.156 | 0.161 | 0.8194 | 0.96 | 0.63-1.45 |
| ACTA | 0.041 | 0.051 | 0.4769 | 0.79 | 0.38-1.64 |

**Overall P= 0.03**

### 5.3.12 Hardy Weinberg Analysis

**Hardy Weinberg Analysis**

**Table 5.26 Hary Weinberg Analysis by SNP. The Delta-32 mutation in CCR5 shows a significant departure from Hardy Weinberg Equilbirum; this is to be expected, as this locus is known to have undergone selection in the European population relatively recently.**

| SNP | Case | Control |
|---|---|---|
| Delta-32 | 0.1529 | 0.0298 |
| rs2279627 | 0.1814 | 0.3050 |
| rs3760746 | 0.8558 | 0.3267 |
| rs11878563 | 1.0000 | 0.2500 |
| rs2305767 | 0.1105 | 1.0000 |
| rs1457092 | 0.8274 | 0.6475 |
| rs2305765 | 0.2217 | 0.7477 |
| rs2305764 | 0.2650 | 0.8232 |
| -137 | 0.9119 | 0.3399 |
| -607 | 0.4400 | 0.7752 |

## 5.4 Discussion

### 5.4.1 Single Candidate Genes

The CCR5 Δ32 allele was associated with CD in the Irish population, with possession of the non functional allele predisposing towards disease. Given the documented role of CCR5 in the homing of immune cells, reduction in the amount of functional protein on the cell surface might impair normal immune function. The hypersensitivity and increased humoral response seen in CCR5 (-/-) mice in response to T cell-dependent antigenic challenge could possibly mirror the similar symptoms observed in Coeliac Disease under conditions of similar antigenic challenge (gluten).

One appealing mechanism for CCR5 Δ32 to influence an autoimmune disease would be through an impact on resolution of immune responses. In fact, there is evidence that expression of CCR5 on the surface of $CD4^{+}CD25^{+}$ T-reg cells is important to their role of suppressing such responses (Zhang *et al.* 2009). Reduced cell surface expression of CCR5 then, might plausibly impair T-reg homing to sites of inflammation, and thereby strengthen and perpetuate an existing immune response.

The SNPs which tag TNFSF14 were also associated with coeliac disease, as was the CCA haplotype, spanning the gene, suggesting a functional variant within the gene contributing to this condition. Recently, the receptor for TNFSF14, TNFRSF14, has also been associated with coeliac disease in a new multi-population GWAS, strengthening the evidence for this gene's involvement, or at a minimum this pathway (Dubois *et al.* 2010). The TNFSF14 tagging SNPs were also included in the candidate gene-biased follow up study, but failed to show association with either the Dutch or British populations.

Neither of the IL-18 SNPs were linked with coeliac disease on any level individually. Haplotype analysis, however, reveals an association with the rare CG haplotype. These SNPs are in strong LD and it may be that the rare haplotype represents a particular configuration of variable transcription factor response elements that has functional consequences. In support of this, the same rare haplotype was also noted to be a risk factor for rheumatoid arthritis in a Scottish/German study (Gracie *et al.* 2005) and a study in multiple sclerosis sufferers shows that the common haplotype generates somewhat greater levels of IL18 mRNA than other haplotypes tested (Giedraitis *et al.* 2001).

MYOIXB showed no association, either at the individual SNP, or haplotype level. **Table 5.27** compares the frequencies of the haplotypes in the Irish population, with the frequencies of the haplotypes from the study in which the association was observed (Monsuur *et al.* 2005).

**Table 5.27 MYOIXB haplotype frequencies; a comparison; this table compares the frequency of haplotypes in the MYOIXB region between Irish and Ducth populations. The Dutch control population shows a markedly different haplotpye frequency than the other populations, supporting the idea that the original Dutch association with coeliac disease may have been due to stratification in the Dutch control population.**

| Haplotype | F(Case) Irish | F(Case) Dutch | F(Ctrl) Irish | F (Ctrl) Dutch |
|---|---|---|---|---|
| GCCG | 0.431 | 0.373 | 0.402 | 0.455 |
| AATA | 0.353 | 0.387 | 0.372 | 0.302 |
| ACCG | 0.156 | 0.155 | 0.161 | 0.174 |
| ACTA | 0.041 | 0.056 | 0.051 | 0.054 |

The haplotype which showed association with disease status in the Dutch population was the GCCG haplotype. As can be readily seen from the table above, the observed effect is most likely due to stratification, or a similar population based effect in the Dutch controls.

From this single candidate gene bases approach, we have identified various levels of association in the Irish population. The association of the TNFSF14 tagging SNPs, and resulting haplotypes may point to a causative variant somewhere within the gene. In IL-18 the observed rare haplotype, which has been shown to influence other autoimmune conditions may be indicative of a rare contributory factor of relatively large effect. The association with of the CCR5 Δ32 mutation with CD would also suggest this variant plays a functional role in CD, though its exact contribution is currently unknown. Investigation of MYOIXB haplotypes confirms the association observed there is limited to the Dutch population, and likely due to population stratification in the Dutch control group.

### 5.4.2 Candidate Gene Biased Follow up of Multi-population GWAS

#### 5.4.2.1 Overall Significant SNPs in Combined Populations

Results of the Goldengate assay which were enriched for functional genes were quite positive, identifying 6/8 genetic regions which showed overall significance. As TAGAP also appeared in the first Goldengate GWAS replication in Chapter 4, we will not mention it here a second time.

**IL18RAP:** encodes an accessory subunit of the heterodimeric receptor for IL18. This protein enhances the IL18 binding activity of IL18R1 (IL1RRP), a ligand binding subunit of IL18 receptor. The coexpression of IL18R1 and this protein is required for the activation of NF-kappaB and MAPK8 (JNK) in response to IL18. It has been implicated in Crohn's disease and ulcerative colitis (Zhernakova *et al.* 2008), type 1 diabetes (Smyth *et al.* 2008), and asthma (Reijmerink *et al.* 2008).

**TNFAIP3:** (which codes for the protein also referred to as A20) appears to be an important susceptibility factor in autoimmune diseases, having previously been associated with type 1 diabetes (Fung *et al.* 2009) psoriasis (Nair *et al.* 2009) and SLE (Graham et al. 2008; Musone *et al.* 2008). Here we confirm an association with coeliac disease. TNFAIP3/AP20 inhibits NFkB by regulating the duration of TNF induced NFkB activation( Düwel *et al.* 2009). TNFAIP3 deficient cells show elevated NFkB activity and are resistant to Fas ligand induced cell death (Daniel *et al.* 2004). TNFAIP3 knock-out mice suffer from lethal levels of inflammation and haploinsufficient mice express increased proinflammatory cytokines (Vereecke *et al.* 2009). We have also found evidence for association of its binding partner, TNIP1, with coeliac disease in the Irish and Italian populations (see chapter 3).

**REL:** along with both TNIP1 and TNFAIP3, REL is a component of the NFkB inflammatory signalling pathway. It encodes c-Rel, which is a pro-inflammatory constituent of the Rel family, which includes the archetypical pro-inflammatory NFkB subunit, p65. Over-expression of c-Rel has been shown to promote allergic inflammation of the lungs, and conversely, c-Rel knockout mice are resistant to the development of allergic conditions (Donovan *et al.* 1999; Hilliard *et al.* 2002). In terms of autoimmune conditions, Rel has been associated with RA and Atopic Dermatitis (Gregersen *et al.* 2009).

Thus, three genes directly involved in the NFkB pathway have evidenced susceptibility to coeliac disease, reemphasising the importance of this pathway in inflammatory conditions.

**CD80**:(Cluster of Differentiation 80) codes for a cell surface protein expressed on the surface of antigen presenting cells, which interacts with both CD28 and CTLA4 on the surface of T-Cells (Habib *et al.* 2007), providing a co-stimulatory signal, which acts both to initiate and perpetuate the $CD4^+$ T-Cell proliferative response. (Vasilevko *et al.* 2002). Its prime role with relation to T-Cell proliferation suggests mechanisms by which it might drive the inflammatory process in coeliac disease, and CD80 expression has already shown to be correlated with immune cell infiltration of the intestinal lamina propria in ulcerative colitis (Scarpa *et al.* 2004).

**CTLA4:** had shown suggestions of significance in multiple populations prior to this study (See section 1.13). Coding for an integral membrane protein, ctla4 is expressed on the surface of activated T-Cells, and acts to inhibit T cell activity, and dampen the immune response (Kavanagh *et al.* 2008). CTLA4 is constitutively expressed by T-regulatory cells (Manzotti *et al.* 2002) and mice knocked out for CTLA4 develop lethal lymphoproliferative disorders (Singh *et al.* 2007) emphasising the importance of its inhibitory role to normal immune function.

CTLA4 has been implicated in a wide variety of autoimmune diseases, including diabetes (Douroudis et al. 2009), Grave's disease (Yeşilkaya *et al.* 2008) SLE (Barreto *et al.* 2009), rheumatoid arthritis (Daha *et al.* 2009) and Behcet's Disease (Park *et al.* 2009)

The genes which show association from this study are, like those which showed association in the previous GWAS follow-up study, heavily involved in immune regulation. Two, REL and TNFAIP3, are an intrinsic regulatory part of the NfKB pro-inflammatory pathway, a crucial mediator of inflammation and immune response. CTLA4 and CD80 are both cell surface proteins, involved in the regulation and activation of T Cells. Our investigations in other projects also show common pathways with some of the genes reported here; IL-18 (see section 5.4.1) and IL-18RAP are close interactants, and TNIP1 (see section 3.4) and TNFAIP3 are also binding partners. With this new information, common pathways are beginning to emerge as players in the molecular story of coeliac disease.

### 5.4.3 Candidate Gene Biased Follow up of Multi-population GWAS ; The Irish Population

As mentioned in Section 4b.4.2, population specific variants may play an important role in coeliac disease, although GWAS studies do not lend themselves to their identification. In the same manner as we examined potential population specific variants in that section for the British-Dutch-Irish GWAS follow-up, we will do so here for the candidate-gene-enriched follow-up study.

If we examine tables **5.a1** and **5.a2** (*Appendix i*) it becomes apparent that, by and large, the allele frequencies in the genes which reached overall significance, and those which were most significant in the Irish population, are relatively constant across the control populations. Notable exceptions are the TGFA and VAV2 associated SNPs, which show some disparity in allele frequency between the Irish and other populations.

Examination of tables **5.a3** and **5.a4** shows a rather different story, however. Though the allele frequencies of the cases are broadly similar across the populations in the SNPs which proved the most significant overall, there is a substantial difference in allele frequency between the Irish and other populations at the IL-18RAP associated loci, both in cases and controls. Notably, the Irish population is the only one of the 3 not to show association with these SNPs; this may represent a predisposing factor which is absent from, or comparatively minor in, the Irish population.

Looking at **Table5.a5,** the allele frequencies of the Irish population differ quite markedly from those of the other 2 populations at almost every locus. These SNPs are those which were most strongly associated with CD in the Irish population; might it then be possible that with a uniform allele frequency in the control groups, some of these SNPs might represent predisposing factors unique to, or of greater magnitude in, the Irish population.

**SLC15A2**: The most significant of the SNPs in the Irish population in this study, and the British-Dutch-Irish GWAS follow-up, the function of SLC15A2 is covered in section 4b.4.2. We examined the LD between the intronic SNP detected here, and the non-synonymous SNP detected in the previous Chapter (rs2257212), and discovered them to be in almost perfect LD ($r^2 = 0.99$). As such rs2141615 can be considered a proxy for rs2257212.

**TGF alpha:** Both of the SNPs showing association are in high LDs ($r^2$ = 0.90).Transforming growth factor alpha codes for a protein which has two forms: an inactive membrane bound form, and an active, free form. Interaction with epidermal growth factor receptor activates its intrinsic kinase, and releases mature TGF alpha.

Over-expression of TGF alpha in mouse models has evidenced epithelial hyperplasia in several organs, including liver and intestine, without disruption of normal tissue architecture (Luetteke & Lee, 1990). Normally, TGF alpha expression in the small intestine is restricted to the villous epithelium, and it has different patterns of expression throughout the GI tract (Thomas *et al.* 1992) where it acts to maintain gastric mucosal integrity, via binding to a common cell-surface receptor (EGF/TGF-alpha receptor) (Orsini *et al.* 1993) and has the ability to confer changes in cytoskeletal organization, with changes in cell adhesion and migration resulting from coexpression of TGFalpha with one of its ligands CD9 in epithelial and carcinoma cells (Imhof *et al.* 2008).

Mucin overproduction is a hallmark of chronic inflammatory airway diseases, production of mucin leading to airway mucus obstruction and contributes to morbidity and mortality. The bacterium *Pseudomonas. aeruginosa*, an important human respiratory pathogen causing cystic fibrosis, utilizes reactive oxygen species (ROS) to up-regulate MUC5AC mucin expression via TGF-alpha (Yan et al. 2008). Though not immediately relevant to CD, this does highlight a mechanism by which a pathogen regulates gene expression in epithelial cells via TGF alpha.

**HMHB1**: Human minor histocompatibility antigen inducted of cytotoxic T lymphocyte against leukemia cells. Thus far, expression of HMHB1 has been observed only in B-ALL cells and Epstein-Barr virus-transformed B cells, (Dolstra *et al.* 1999) however a genetic factor leading to its inappropriate expression may influence CD patholoy.

**VAV2 :** a member of the Vav family, the gene product is a guanine nucleotide exchange factor which interacts with Ras-related GTPases, as well as the EGF receptor. Vav2 is expressed across a wide range of tissue types. (Tamás *et al.* 2001) and has been shown to play opposite roles in influencing the activity of both T and B Cells. Binding of Vav2 to the TCR suppresses transcription in activated T-Cells (Tartare-Deckert *et al.* 2001), however its binding to B-Cells appears to exert the opposite effect (Doody *et al.* 2000).

Vav2s interaction with Ras at the cell membrane also plays a role in maintaining the architecture of epithelial cells (Duan *et al.* 2010), possibly through the regulation of cell-cell junctions. Vav2 also plays an important role in host-pathogen interaction; it appears some bacteria have the ability to avoid internalisation via the action of surface expressed adhesions, by inducing cytoskeletal rearrangements in the host cells. Vav2 has been shown to have a direct role in Cav1-mediated cytoskeletal re-arrangement, which leads to the prevention of bacterial uptake, in both Neisseria gonorrhoeae and enteropathogenic Escherichia coli (Boettcher *et al.* 2010).

With regards to autoimmune disease, VAV2 has been shown to have a role in MS (Nischwitz *et al.* 2010).

**ACVR2A**: A cell surface receptor, ACVR2A acts to regulates viability and cell adhesion, retarding cell growth, and promoting cell adhesion(Simon *et al.* 2009).

**TNFAIP3**: Also showed association globally, and was discussed in section 4.4.1.

**TNFSF14**: The most significant SNP from the candidate gene TNFSF14 study was included in this panel, primarily to test for significance in the Dutch and British populations, which was not observed.

MACROD2, LOC401014, TRIM60, DAO,TMEM132B and loc235965 are all genes of currently unknown function.

Examination of the SNPs which are indicative of association in the Irish population, once again shows a trend in function away from those observed in the overall analysis. Intestinal and epithelial function, as well as recognition and reaction to microbial species once again feature heavily (as seen in section 4b.4.2). TGF alpha and VAV2 are both involved in recognition of bacteria, and subsequent epithelial cell response. HMHB1 and ACVR2A both show evidence of being involved in the immune response; HMHB1 is particularly interesting, as it is a promoter of T-Cell cytotoxicity, which is only, usually, expressed in very limited subtypes of cells. Genetic variants, or variants in conjunction with pathogen interactions (such as Epstein Barr virus, detailed above) may result in inappropriate expression of this gene, and a corresponding up-regulated cytotoxic immune response.

# Chapter VI

# General Discussion

## 6.1 Autoimmune Disease; Common Factors

The common variant/multiple disease (CV/MD) hypothesis of common, complex disorders, states that common alleles which contribute to a disease given a certain combination of interacting genes and environmental factors, may result in different, but possibly related, clinical outcomes in differing genetic backgrounds and environmental conditions.

This is shown dramatically by **Table 6.1**, which shows the high number of genes associated with coeliac disease from the first and second GWAS' which have also shown association with other autoimmune diseases. This table contains disease associations not only with the genes from this study, but also from the second coeliac disease GWAS (Dubois *et al.* 2010), of which we were also a minor part. This is inclused to give an holistic overview of the interacting genes in autoimmunity.

Thus, certain molecular pathways appear to underlie a variety of dysfunctional immune processes, giving rise to diseases sharing common mechanisms, for example inflammation. While the majority of associated SNPs have the same effects in different diseases, it should be noted some genetic variants can have opposite effects under certain conditions

An example of this is the rs917997 SNP, which is associated with altered expression of the *IL18RAP* gene. This SNP is protective in type 1 diabetes, but increases risk of coeliac disease; this pattern is also seen with variants of the *TAGAP* gene (181). A more complicated scenario arises in the case of the *TNFAIP3* gene region which is associated with rheumatoid arthritis, type 1 diabetes, coeliac disease, and systemic lupus erythematosis; however the genetic variants most strongly associated with each disease are largely consistent.

**Table 6.1 Legend**
Diseases in **bold** are associated with the genes via genetic linkage, and/or association studies.
*Italics* denote a correlation established by an expression study.

References listed correspond to references in *Appendix ii.*

**Table 6.1 Shared Disease Susceptibility Loci in Coeliac Disease and other AIs.**

| Gene | Autoimmune Disease Association |
|---|---|
| IL-2 | **SLE**[106], **SS**[106], **Behcet's Disease**[106], **T1D**[106], **Psoriasis**[107], **RA**[108], **IBD**[109], **Allergy**[110], **Asthma**[110], **Atopic Eczema**[110],**Systemic Sclerosis**[111], **MS**[112] |
| IL-21 | **RA**[113], **UC**[114], **Asthma**[115], **SLE**[116], **Psoriasis**[107], **SS**[106], **Behcet's Disease**[106] |
| IL-12 | **MS**[117], *AIDP*[118], *Parkinson's*[119], *SS*[120], **Asthma**[121], *SLE*[122], *Crohn's*[123] |
| IL-18RAP | **Crohn's**[124], **UC**[124], **Asthma**[125] |
| ICOS | *SLE*[126], **MS**[127], **RA**[128], *IBD*[129], **Spondyloarthropathies**[130], *Inflammatory Neuropathies*[131]. |
| CIITA | **Addision's Disease**[132], **MS**[133]. |
| LNK | **T1D**[134], **SLE**[135] |
| TNFAIP3 | **RA**[136], **SLE**[137], **MS**[138], **JIA**[139], **Psoriasis**[140], **Crohn's**[141], **T1D**[141]. |
| TNFRSF14 | **RA**[142], *SLE*[143], *Atopic Dermatitus*[144], *Allergic Asthma*[144], **T1D**[145]. |
| BACH-2 | **Diabetes**[134] |
| TAGAP | **T1D**[134], **RA**[146] |
| CD80 | *UC*[147] |
| CD28 | **RA**[146], *SLE*[148], **MS**[149], **Behcet's Disease**[150]. |
| CTLA4 | **Diabtes**[151], **Grave's**[152], **SLE**[153], **RA**[154], **Behcet's Disease**[155]. |
| REL | **RA**[156], *Atopic Dermatitis*[157]. |
| PTPN2 | **Crohn's**[158] |
| ETS1 | **SLE**[159] |
| CLEC16A | **MS**[160], **T1D**[160], **UC**[161], **Addison's disease**[132] |
| ICOSLG | *SLE*[162], IBD[163], *Peripheral nervous system*[164] |
| CCR1 | *Inflammatory Lung Disease*[165] |
| CCR2 | *Inflammatory Lung Diseases*[165], **Atopic Asthma**[166], *Multiple Sclerosis*[167], *RA*[168], *DM*[169]. |
| CCR3 | **Asthma**[170] |
| CCR5 | **RA**[171], **SLE**[171],**JIA**[172], **MS**[173], *Inflammatory Lung Diseases*[165], *Inflammatory Neurodegeneration*[174], **Behçet's disease**[175], *Grave's Disease*[176], |
| CCR9 | *Allergic Asthma*[177] |
| CCRL2 | *RA*[178] |

**Abbreviations Used**

**SLE:** Systemic Lupus Erythematosus
**SS:** Sjorgenson Syndrome
**T1D:** Type 1 Diabetes
**RA:** Rheumatoid Arthritis
**IBD:** Inflammatory Bowel Disease
**MS:** Multiple Sclerosis
**UC:** Ulcerative Colitis
**AIDP:** Acute inflammatory Demyelinating Polyneuropathy
**JIA:** Juvenille Idiopathic Arthritis
**DM:** Diabetes Mellitus

### 6.2 Regulatory Variants play a large role in complex disease

Genome wide association studies have led to a vast increase in the catalogue of genetic associations with complex diseases. Although the HLA coeliac risk alleles are functional, resulting in alterations to protein structure and function, this appears to be the exception rather than the rule. Only a minority of GWAS signals have to date been shown to involve coding variants, and while it is possible that the signals result from their linkage to currently unobserved coding variants, it seems more likely that the GWAS associations are involved in more subtle regulatory effects on gene expression (Dubois *et al.* 2010).

This observation is in accordance with what has been noted in previous GWAS studies. An analysis of 151 of the GWAS studies which have been carried out thus far, has revealed 43% of reported associated SNPs to be intergenic, and 45% to be intronic, suggesting a crucial role for noncoding SNPs in common, complex diseases (Hindorff *et al.* 2009). This, of course, must be balanced with the observation that the vast majority of SNPs selected to form a GWAS panel are non coding, as the selection process is biased towards information content, rather than functionality.

However, over half the variants significantly associated with coeliac disease have been correlated with expression changes in nearby genes, (Dubois *et al.* 2010), suggesting that they, or linked surrogates, may play functional roles in controlling expression of these genes. Due to the commonality of factors in many complex diseases, this is likely to be applicable to other common, complex diseases.

These observations are concordant with current theories about complex disease genetics; several, perhaps very many, factors of relatively small effect, combining to produce a disease phenotype. Debate is ongoing regarding the source of missing

genetic factors in complex diseases – for the majority, only a small proportion of heritability is accounted for. Current studies still use dense genetic mapping to identify associated regions, and this will be superceded by DNA sequencing in the near future. This will allow us to more directly address the issue of causal variants and missing heritability; for example, are rare variants important in these relatively common diseases? At the moment, evidence suggests regulatory gene variants are more common causes of disease susceptibility than coding variants

## 6.3 General Functional Overview

The genes which have shown association with coeliac disease, both from this study, and the second coeliac GWAS have a wide variety of immunological functions. The following table, **Table 6.2**, gives a brief summary of their salient qualities with regard to immunological function. Gene ontology tags are taken directly from the gene ontology database. References listed correspond to references in *Appendix ii.*

**Table 6.2 Immunological Roles of coeliac genes.**

| Gene | Gene Ontology Tags | Other Immunological Funtions |
|---|---|---|
| *IL-21* | • **Cell Maturation (NK Cells).**<br>• **Positive regulation of interferon-gamma biosynthetic process (Th1).**<br>• **Positive regulation of T cell proliferation,** | • Mediates transition from innate to adaptive immunity[63]<br>• B cell activation, expansion, and plasma cell generation .[64] |
| *IL-2* | • **Natural killer cell activation.**<br>• **Negative regulation of lymphocyte proliferation.**<br>• **Positive regulation of regulatory T cell differentiation.**<br>• **T cell differentiation.**<br>• **Anti-apoptosis.**<br>• **Negative regulation of B cell apoptosis**<br>• **Regulation of T cell homeostatic prolife**<br>• **Positive regulation of B cell proliferation.**<br>• **Positive regulation of immunoglobulin secretion.**<br>• **Positive regulation of isotype switching to IgG isotypes ration.**<br>• **Negative regulation of inflammatory response.** | • Regulates activated T and B cell apoptosis [65]<br>• Integral in mediating tolerance.[66]<br>• Stimulates Plasma Cells[67]<br>• Integral to Thymocyte Development[68] |
| *IL-12* | • **Positive regulation of interferon-gamma production.**<br>• **Positive regulation of lymphocyte proliferation.**<br>• **Positive regulation of natural killer cell activation**<br>• **Positive regulation of natural killer cell mediated cytotoxicity.**<br>• **Positive regulation of T cell mediated cytotoxicity.**<br>• **Positive regulation of T cell** | • Regulates apoptosis of TH1 T-Cells[69]<br>• Influences intrathymic T cell development[70]<br>• Prominent Role in AI disorders.[71] |

| | | |
|---|---|---|
| | **proliferation.** | |
| SOCS1 Suppressor of Cytokine Signalling 1 | | • Negative regulation of cytokine initiated signal transduction including IFN gamma pathway inhibition [72] |
| *IL-18RAP IL18R1* | | • IL18R1 and IL18RAP encode the two proteins that form the IL-18 receptor. IL-18 signalling leads to up-regulation of Th1 response and IFN-g production.<br>• Skews towards Th1 or Th2 based on immunological context[73]. |
| *ICOS* | | • Mediates B:T cell interactions, determining survival of activated cells[74]<br>• Induces anergic T cells via tolerogenic dendritic cells[75]<br>• Role in Humoral Immunity[74]<br>• Inducible T cell co-stimulator: Initiates Th1 and Th2 responses by acting as a co-stimulatory molecule to promote T cell activation[74] |
| *CIITA* | • **Inflammatory Response**<br>• **Positive regulation of MHC class II biosynthetic process.** | • Critical for antigen presentation in thymic selection process[75]<br>• Activates MHC Class II gene expression in antigen presenting cells, necessary for T cell activation [76] |
| SH2B3/LNK | • **Embryonic hemopoiesis.** | • Negative Regulation through TNF alpha inhibition[77]<br>• Negative Regulation of cytokine signalling including TNF alpha [77]<br>• Negativeiy regulates Jak2 and c-Kit signalling[78,79] |
| TNFAIP3 | • **Anti-apoptosis**<br>• **Apoptosis** | • Negative Regulation of NF-kappaB signalling cascade[80]<br>• Negative Regulation of Apoptosis[81] |
| *TNFRSF14* | • **Negative regulation of alpha-beta T cell proliferation.** | • Acts as a Bi-directional. Ligand dependent switch, either promoting or inhibiting T-Cell activation[82] |
| MAP3K7 | • **Activation of NF-kappaB-inducing kinase activity.**<br>• **Positive regulation of interleukin-2 production.**<br>• **Positive regulation of T cell activation.**<br>• **Positive regulation of T cell cytokine production.**<br>• **Negative regulation of apoptosis[91].**<br>• **T cell receptor signaling pathway.**<br>• **Transforming growth factor beta receptor signaling pathway.** | |
| RUNX3 | • **Negative regulation of epithelial cell proliferation**<br>• **Induction of apoptosis.** | • Expression correlated with CD8+ mature thymocytes.[83] |
| BACH-2 | | • Regulates IL-2 expression[84] |
| TAGAP | | • Activates T-Lymphocytes[85] |
| CD80 | • **Positive regulation of alpha-beta T cell proliferation.**<br>• **Positive regulation of granulocyte macrophage colony-stimulating factor biosynthetic process.**<br>• **Positive regulation of interleukin-2 biosynthetic process.**<br>• **Positive regulation of T-helper 1 cell differentiation.** | • Initiation and maintenance of CD4+ T-cell proliferation[86] |

| | | |
|---|---|---|
| | • **T cell activation.**<br>• **T cell costimulation.**<br>• **Coreceptor activity.** | |
| CD28 | • **Positive regulation of T cell proliferation.**<br>• **Regulatory T cell differentiation.**<br>• **Positive regulation of anti-apoptosis.**<br>• **Humoral immune response.** | • Triggers NfKappaB activation[87]<br>• CD28 costimulation is essential for human T regulatory expansion and function.[88]<br>• Mediates Cell cycle progression in primary human T-Cells[89] |
| *CTLA4* | • **Negative regulation of regulatory T cell differentiation.**<br>• **Negative regulation of T cell proliferation.**<br>• **Tolerance induction.**<br>• **Negative regulation of immune response.** | • Negative Regulation of regulatory and activated effector CD4+ T cells. [90] |
| REL | • **Positive regulation of I-kappaB kinase/NF-kappaB cascade.** | • Anti-apoptotic activities of Rel and RelA required during B-cell maturation[91].<br>• Crucial for lymphocyte proliferation s[92] |
| PTPN2 | | • Induces caspase-1-dependent apoptosis [93]<br>• Inhibits T-Cell proliferation through mediation of the NO pathway [94] |
| THEMIS | • **Negative T cell selection.**<br>• **Positive T cell selection.**<br>• **T cell receptor signaling pathway.** | • Controls thymocyte selection through regulation of TCR signaling.[95]<br>• Important for late thymocyte development[96] |
| *ETS1* | • **Induction of apoptosis.** | • Negative Regulator of plasmacytic differentiation[97]<br>• Transactivates the human GM-CSF promoter[98] |
| *RGS1* | | • Regulates chemokine receptor in germinal center B lymphocytes. [99] |
| PTPRK | | • Regulates CD4+ T cell development via ERK[100] |
| ICOSLG | • **Positive regulation of activated T cell proliferation.**<br>• **T cell activation.**<br>• **B cell activation.** | • Essential for T-cell activation and function[101]<br>• Differentiation and Plasma Cells[102] |
| *CCR1* | • **Inflammatory Response.**<br>• **Response to wounding.** | • Via NKkB and ERK[103,104] |
| CCR2 | • **Inflammatory response.**<br>• **Response to wounding.**<br>• **Cellular defense response.** | • Via ERK[103] |
| CCR3 | • **Cellular defense response.**<br>• **Inflammatory response.** | • Induces Apoptosis[105] |
| ITGA4 | • **Leukocyte cell-cell adhesion.** | |
| *CCL2* | | |
| *CCR5* | • **Inflammatory response.**<br>• **Cellular defense response.** | |
| *CCR9* | • **Cellular defense response.** | |
| *CCR4* | • **Inflammatory response.** | |

## 6.4 Pathways, Interactants and Associations

As the list of genes associated with coeliac disease has grown over the last 5 years, a clearer picture of the molecular players involved in coeliac disease has begun to emerge. Several genes which have been identified as associated in the Irish population are closely interacting partners, or share pathways with, genes revealed as associated in the Goldengate and candidate gene enriched GWAS follow-up studies reported in this thesis. There is also well documented evidence that many of the candidates from these two GWAS studies' follow-ups interact with one another, share pathways, or are functionally related.

Not only does this give us a much clearer picture of the modifying factors which are an essential adjunct to the HLA predisposition in generating the coeliac phenotype, it also informs further pathway specific approaches which we can capitalise on, and use to expand this knowledge.

## 6.5 Cyokines IL-18, IL-18RAP, IL-2, IL-21, IL12A

IL-2, IL-12, IL-18 and IL-21 signalling pathways are important determinants of T cell activity, and their activities have been extensively reviewed. This makes them excellent candidates for involvement in a chronic autoimmune disease, like coeliac disease.

It should be noted that they have pleiotrophic effects on different T cell lineages, and that their influence is likely to be complex. For example, while IL-2 had originally been considered the archetypical Th1 predisposing cytokine, and thus a suspect in chronic inflammation, mice deficient in IL-2 production develop chronic inflammatory conditions, probably as a result of a lack of IL-2 dependent T reg function.

IL-12 and IL-18 (implicated in coeliac disease through the association of the IL-18 receptor associated protein, IL18RAP and weakly through a haplotype association in this study); and IL-21 and IL-18 act synergistically in the induction of IFNγ expression , which is produced in large quantities in the coeliac lesion (El-Mezayen *et al.* 2004).

IL-21, produced by $CD4^+$ lamina propria lymphocytes, also contributes to increased production of IFNγ and T-bet, a transcription factor which promotes Th1 cell fate Jenner *et al.* 2009). It also promotes cytotoxic $CD8^+$ cell (Lametschwandtner *et al.* 2004) and prevents T reg suppression of $CD4^+$ T cell activity (Koch *et al.* 2009).

These cytokines also influence the proliferation of cytotoxic T cells such as $CD8^+$ T cells and Natural Killer cells (Brady *et al.* 2010) which function to destroy infected or damaged host cells via a cytotoxic mechanism. Some intra-epithelial T lymphocytes (IELs) from coeliac patients have been shown to demonstrate aberrant expression of natural killer (NK) lineage receptors and can perform NK-like functions including T cell receptor-independent killing of enterocytes in active coeliac disease (Jabri *et al.* 2000, Meresse *et al.* 2006).

### 6.6 Tolerance

The cytokines identified as part of this study are also involved in tolerance and thymic selection. The two areas of tolerance to self, central and peripheral, are cornerstones of the immune system. Any dysfunction of these mechanisms might easily fit into a model of autoimmunity such as coeliac disease. Central tolerance occurs in the thymus, and acts to remove cross reactive T and B cells before they are released into the periphery. Auto-reactive immune cells which escape removal at this stage are mediated by T regulatory cells, which are essential for peripheral tolerance (Takahashi *et al.* 1998). Maturation of Thymocytes into T Lymphocytes occurs in the Thymus and involves three stages: beta-selection, positive selection, and negative selection, resulting in T cells capable of responding to foreign antigens, but not reacting to 'self' antigens.

IL-18, in concert with either IL-2 or IL-12, determines the lineage of cells developing in the thymus; IL-2 and IL-18 predispose towards a Th2 subset, with IL-12 and IL-18 inducing a development towards Th1 (Rodriguez-Galán *et al.* 2005).

THEMIS, which has been in the latest GWAS (Dubois *et al.* 2010) plays a role in thymic selection, and adds weight to the conjecture that this process may contribute to coeliac disease. THEMIS is a critical component of the T cell developmental program, functioning to sustain and/or integrate signals required for proper lineage commitment and maturation (Lesourne *et al.* 2009). Its specific role is evidenced by its expression; high in thymocytes between the pre-T cell antigen receptor (pre-TCR) and positive-selection checkpoints and low in mature T cells. THEMIS-deficient thymocytes evidence defective positive selection, resulting in fewer mature thymocytes. Negative selection is also impaired in THEMIS-deficient mice.

Negative selection is not an exact process, and some auto-reactive mature T-Cells are released from the thymus. These cells are regulated via T regs, and other peripheral tolerance mechanisms (Yarilin *et al.* 2004)T regs exert their effect through cell to cell contact, via the T-Cell Receptor, and activate a suppression of cell proliferation. The primary method of modulation appears to be control of IL-2 production in the responder cell (Thornton *et al.* 1998) as development of T reg cells requires the action of IL-2 (Horwitz *et al.* 2003). Peturbations in IL-2 production could result in a wide variety of autoimmune predisposing scenarios; reduction in IL-2 levels could result in a failure to develop, or a reduction in, the population of T regs, while increased IL-2 production in the responder cell might abrogate the suppressive effect of the T Regs.

Several other candidate genes could also be viewed as playing a role in tolerance to self; for example, as T regs exert their suppressive effect via cell contact, failure to home to the site of inflammation would also render them ineffective, highlighting a role for cell surface markers and homing molecules, such as CCR1, CCR3 and CCR5, in the modulation of inflammation (Shevach *et al.* 2001).

Chemokines and their receptors fall into this category, and have been identified as susceptibility variants in coeliac disease. Stimulatory molecules such as CD28, ICOS and TNFSF14 could also be regarded in this light, as they positively and negatively regulate T-Cell activation. With specific reference to T regs, for example, CD28 co-stimulation is required to expand and maintain their suppressive function in vivo, and any reduction of cell surface CD-28 would inhibit the efficacy of the modulatory response.

### 6.7 NfKappaB

Several of the identified coeliac susceptibility genes play a role in signal transduction pathways leading to NFkB activation, suggesting this is an important molecular player in this condition. IL-2, IL-21, IL-18 (and by association IL-18RAP), IL-12, TNIP1, TNFAIP3 and REL signal through, or are components of, the NFkB pathway, to a greater or lesser extent.

NFkB is a transcription regulator that is activated by various intra- and extra-cellular stimuli such as cytokines and bacterial or viral products. It plays a key role in the inflammatory, immune, stress response, and anti- apoptosis processes, and

inappropriate NFkB activity has been associated with a number of inflammatory diseases

IL-2, IL-12, IL-18, IL-18RAP, TNFAIP3, TNIP1 and REL all play essential roles in the NFkB pathway, as mentioned in the discussion sections of chapters 3-5. Many cytokines including IL-2, IL-12, IL-18 and IL-21 either induce or are upregulated by NFkB. Interestingly IL-21 is specifically up regulated by NFkB molecules containing c-Rel (Trynka *et al.* 2009). As discussed in section 5.4, c-Rel is an interesting candidate given that its expression pattern is restricted to the immune compartment and its demonstrated role in allergy.

In addition to the genes we have found to be associated with the NfKappaB pathway, Dubois *et al.* 2010, as part of the second coeliac GWAS, also identified multiple other associations in regions which contain genes which are also key players in the NFkB pathway.

**Figure 6.1** gives an overview of all of the currently known risk factors which interact closely with this important inflammatory pathway.

Genes detailed in this study are coloured in blue, with genes evidencing association from Dubois *et al.* coloured green. MyD88 has not shown association in either study, but is included due to its function in mediating IL-2 and IL-21's interactions with NFkB.

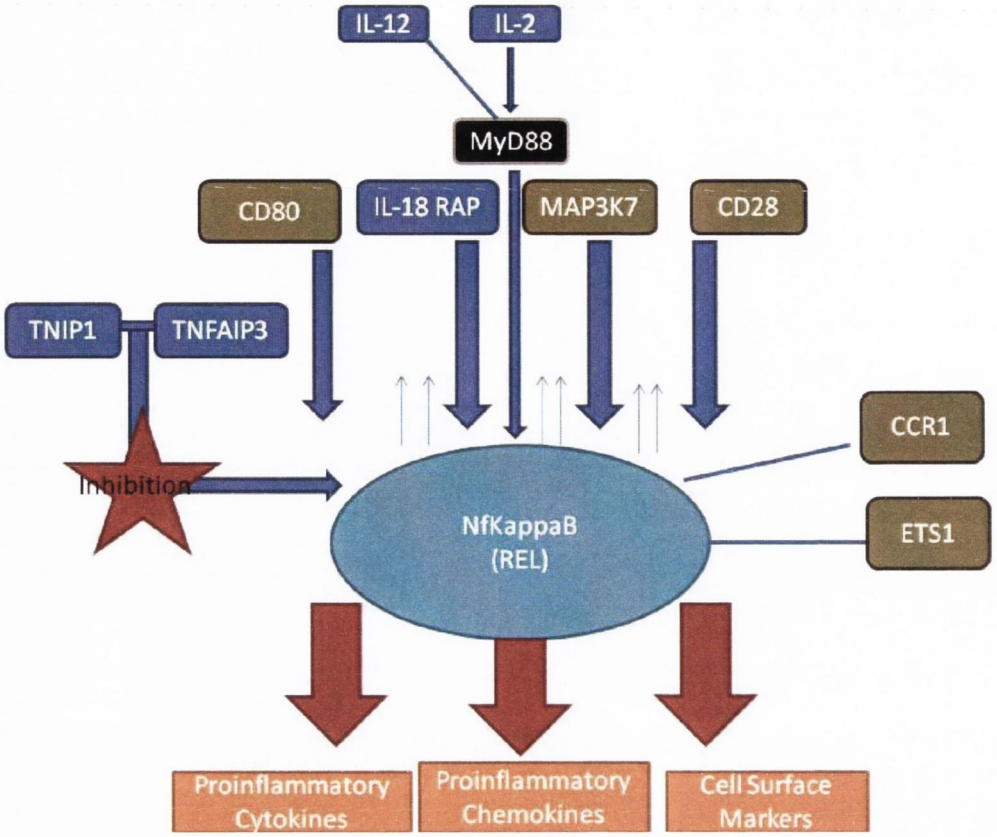



**Figure 6.1 Associated Genes and their interactions with NfKappaB.**

| | |
|---|---|
| REL | Positive regulation of I-kappaB kinase/NF-kappaB cascade |
| TNFAIP3 | Negative regulation of I-kappaB kinase/NF-kappaB cascade |
| CD80/CD28 | The interaction of the antigen presenting cell (APC) and T cells causeing NF-kB activation in both cell types. NF-Kb activation is triggered in T cells by the engagement of the T cell receptor and the CD28 receptor with their ligands, MHC class II, and the costimulatory molecules CD80 and CD86 presented by APCs |
| IL-18RAP | Required for the high affinity binding of interleukin 18 (IL-18) to its receptor complex. Together with IL18R1 it mediates IL-18-dependent activation of NF-kappa-B and JNK. |
| IL-2 and IL-12 | Both activate NF-kappaB. |
| MAP3K7 | IL-1 activates NfKappaB through MAP3K7. |
| ETS1 | Constitutive ETS1, and inducible NFkappaB and AP1, cooperate to coordinate expression of a number of cytokines involved in differentiation, proliferation and activation of the haematopoietic system in activated T-Cells. |
| CCR1 | Induces NfKappaB expression in response to binding ligands. |

### 6.8 Summary of Interacting Gene Functions

Emerging from these observations is a picture of the centrality of the T cell response in coeliac disease, with key signalling and regulatory pathways showing evidence of perturbation. These include both pathways which would be construed as promoting cell activation, such as NFkB, and systems understood to suppress activation after an immune stimulus has passed, such as has been discussed in the context of T reg cells.

Of course, whether the observed variants cause these pro or anti-inflammatory systems to be more or less active is currently unknown. While the functionality of many of the candidate genes is well understood, some observations to date are counter-intuitive to a straightforward interpretation.

For example, the observation that the associated SNP in the IL18RAP gene is linked to reduced expression of the gene; an obvious interpretation of the IL-18 association would be that this pathway would be more active, not less, given that it is a key driver of IFNg expression, which is an observed feature of the coeliac lesion. This situation is analogous to one in Crohn's disease, another gastrointestinally localised auto immune disease, in which nonfunctional NOD2/CARD15 variants failed to activate their normal target, NFkB. Bizarrely, it seemed, a down-regulation of a pro-inflammatory signal was giving rise to an inflammatory disease (Fritz *et al.* 2011).

This has been at least partially explained by the recent discovery that NOD2 is an essential component of the autophagy process, required for the elimination of engulfed bacteria by macrophages and other cells which has emerged as a central mechanism in Crohn's disease (Fritz *et al.* 2011).

The complex pathology of coeliac disease is indicated by the number of susceptibility gene regions identified, but the actual genes and variants still need to be confirmed, their functional effects identified, and the consequences of these for the various cells involved in generating the coeliac lesion in response to gluten elucidated. Additionally, known functions of genes are not necessarily all the functions of genes, and even when a susceptibility locus is identified, pleiotropic effects of a gene product, or assembly of multiple potential proteins from modular domains (such as occurs in many interleukins) must also be considered.

## 6.9 Population specific variants

In the course of our investigations, there has been evidence to suggest that the relative contributions of genetic elements predisposing to CD may vary between even reasonably homologous populations, in this case different European populations.

The differences in associations of SNPs between populations may be due, simply, to a corresponding difference in underlying LD due to population stratification, leading to different proxies for causative variants across different populations. However, it is also possible that a proportion of the attributable heritability varies on a population specific basis. In this section, we will explore this possibility.

### 6.9.1 Common Pathway, different member

We observed an association with CD in the Irish population for the rare CG Haplotype located in the promoter region of IL-18. However, this association was not observed in either the Dutch or British populations as part of the candidate gene enriched GWAS follow up study.

Similarly, TNFSF14 has shown association in the Irish population, but not the British or Dutch. Taken with the lack of association of IL-18RAP with CD in the Irish population, this may point to a perturbation of a common pathway across populations, but with differing aspects of the pathway being disrupted depending on the specific population.

Pleiotropy of this kind is not unusual in Mendelian diseases, and founder effects could cause mutations in different genes which interact together as part of a common pathway. We have already observed many of the regions associated with CD harbour genes whose products are known to interact.

### 6.9.2 GWAS follow up; Differences between populations

Initial GWAS follow up studies showed the pattern of LD in the IL-2/IL-21 blocks to be highly similar between the Irish and UK populations. However, investigation of several of the associated SNPs in an Italian population showed pronounced differences in allele frequencies in this region. That this difference is a factor of the Italian population, rather than the specific sample with which we were working, is

borne out by a second study which showed association with rs6840978 in a separate Italian case control study. As was previously mentioned (section 4c.4), the frequencies of the cases and controls in this study were virtually identical to those we observed in our study at the rs6840978 locus.

As can be seen, the associations are not consistent across the 4 populations. The IL-12A/SCHIP1 region fails to show significance in the UK population, and the TAGAP region is only borderline. Neither the Dutch, nor the Irish population shows an association with LPP, and SH2B3, RGS1, CCR3 and IL-18RAP all show no association in the Irish population. Of the SNPs tested in the Italian population, RGS1 showed no association.

Examination of the allelic frequencies at the loci involved show marked differences between populations which are not associated with CD at those loci, and those which are, both in the case and the control groups.

This would seem to endorse the hypothesis that a certain proportion of risk apportioning loci in complex disorders are population specific. It has already been noted that the HLA risk associated alleles for CD display genetic heterogeneity across Europe, along a north to south gradient.

In Northern Europe, the DQ2.5 molecule is more frequently carried in trans, with the cis configuration becoming more frequent as one travels south. The HLA-DQ8 variant is also more frequently encountered in Southern European populations, with a frequency of 6-8% in CD patients in Southern Europe. With an observed variation of this nature in the occurrence of the most fundamental determinant of CD, it is quite possible that founder effects, population stratification, bottleneck events and local selection pressures on immune modulating genes (such as those suggested to have acted on CCR5) may have resulted in a large degree of heterogeneity in modifying factors between populations.

In the case of inflammatory bowel disease, founder effects have also been suggested for observed differences in the NOD2 risk factor, with allele frequencies having been observed to vary widely between European populations.

If we accept the hypothesis that risk factors vary by population, then it is of interest to us to examine the population specific associations observed in the GWAS follow up studies with respect to the Irish population.

### 6.9.3 Irish Population

Examination of the Irish population alone in the two GWAS follow-up studies (UK, Dutch and Irish as well as candidate gene biased) reveals additional possible mechanisms which may be contributing to the disease state.

A microbial initiating event has long been suspected as a catalyst for the 'activation' or 'stepping-up' of silent coeliac disease. The question of why some predisposed individuals can take decades to evidence symptomatic coeliac disease is one which currently lacks an answer, however, the influence of an infectious microbe has been suggested as a possible answer (De Palma *et al.* 2010).

The genes which suggest significance in the Irish population are primarily concerned with the integrity of the intestinal, mucosal or epithelial surface, and with the interaction with fungi, and various bacteria, including cytoskeletal remodelling of the epithelial cell surface.

Outside of the Irish population, LNK has shown a response to bacterial lipopolysaccharide and muramyl dipeptides, stimulating the NOD2 recognition pathway (Zhernakova *et al.* 2010). This suggests that bacterial response elements may be a common theme in coeliac susceptibility, however the specific elements may vary.

The identification of a seemingly novel SNP in the macrophage attrantant protein unique to Ireland by Tong *et al.* 2010 supports the concept of population specific immunolgical variants. In this case the novel SNP is >300bp from a known gene shown in multiple cohorts to be strongly associated with inflammatory bowel disease. If an Irish population specific variant exists in a causative gene for one complex, autoimmune, intestinal inflammatory disorder, why not another?

### 6.9.4 5q linkage peak disparity

The linkage peaks suggesting the location of the putative Coeliac3 locus on chromosome 5q varies between the Scandinavian populations, and the other European populations were examined. We have found evidence which suggests a role for TNIP1 as the Coeliac3 locus, whereas a Scandinavian study has shown evidence for association with the STK32A gene.

Investigation of the 3' region of the STK32A gene which showed association in the Scandinavian population, including markers for the most significant SNPs in that study, failed to show significance in the Irish population.

STK32A (at position 146.6-146.7Mb on chromosome 5) lies towards the observed Scandinavian linkage peak at **D5S816 (**135.3-135.3Mb), while TNIP1 (at 150.4Mb) is closer to **D5S2014** (149.9-149.9Mb), the peak associated with CD in the other mapped populations. This may point to there being multiple loci on 5q which predispose towards Coeliac Disease and have different relative contributions between Scandanavia and other parts of Europe. It is even possible, that the loci might be population specific, and confer no additional risk in other populations.

### 6.10 A Model of Coeliac Disease: common core, with varying modifying factors

Looking at the SNPs which have shown association, and the populations in which they have proved to be associated, and those in which they have not, a picture of the underlying genetic and immunological nature of (Gregersen *et al.* 2009) begins to emerge.

The fact that a wide spectrum of autoimmune diseases appear to share common causative variants, suggests that there are a number of key immunological pathways, the disruption of which results in a strong tendency towards the autoimmune state. The nature, and relative magnitude, of the contributions of these variants is not constant between the diseases; different diseases presumably also require one or more characteristic genetic determinants specific to its characters (such as the HLA contribution to CD, and autophagy in Crohn's). However, it seems clear that these common variants must play an essential role in contributing to the activities of these principle factors, in the induction of autoimmune disease through the provision of common pathogenic mechanisms.

The heterogeneity in effect which has been observed for some, but not all, of the SNPs associated with CD, suggests that there may be a two tier modifying system; a core of common variants, and a second tier.

This 'core' of contributory factors would display the following characteristics: a relatively consistent magnitude and frequency across all CD populations, and actions necessary, in conjunction with a predisposing HLA, to bring about the disease state.

The factors observed via the GWAS follow up studies would be constituents of such a core.

The 'second tier' then, would be composed of population variable elements, which act to further disrupt the immune system towards the autoimmune state. These variants may be present in only a subsection of populations, but may have analogues in others. These analogues may be risk factors of similar molecular character (a binding partner, or member of a shared immune pathway) which would act to ultimately bring about the same immunological event, but with a different genetic source. Alternatively, they may simply be mutations in the same gene or regulatory site, but at a different locus, and as such contributing to a lesser, or greater, degree towards autoimmunity.

This hypothesis is supported by the fact that a SNP associated in one or more populations, but not in another, shows a markedly different allele frequency in both the case and control samples at the locus in the population which does not evidence association, suggesting a fundamental population difference at that locus.

If this two tier model is correct, it has implications for further investigation of CD susceptibility loci. The GWAS based approach has shown a great deal of success in identifying the shared, or 'core' variants, but the study design prevents the identification of population specific variants, as the correction for multiple testing required excludes analysis of a single population from reaching significance. Additional diagnostic methods, such as the Immunochip (see next section) may shed more light on this issue.

The relatively small degree of heritability associated with the SNPs which have been found to be associated with CD may be due to the fact that tagging SNPs, rather than the true causative variants are being examined. However, a large degree of heritability may also be ascribed to population specific variants, which will not be elucidated via further GWAS based investigation.

### 6.11 Further Studies

A second coeliac GWAS has already been carried out, based in part on data from this study (Dubois *et al.* 2010) in which we played a minor, confirmatory role. In addition to confirming the associations laid out in section 4b.4, it has elucidated an additional 13 associated regions, 11 of which contain obviously relevant candidate genes.

With the elucidation of several pathways involved in Coeliac Disease, and also common to autoimmune diseases in general, and with the prospect of population specific determinants contributing significant heritability to CD, further GWAS based studies may begin to yield diminishing returns.

We are currently using the immunochip platform as a means of fine mapping our Irish case control population for a host of immune related genes, with the aim of uncovering the specific function variants underlying the observed associations. Now that there are signposts pointing the way towards which pathways are critical to disease pathology (NfKB, cytokines, tolerance) future studies can be directed toward exhaustively exploring their constituents.

Investigation into the role of a microbial initiating factor in coeliac disease may also be valuable; although the majority of evidence for this mechanism comes from uncorrected Irish data, LNK has also shown activity in this area. A focus on common microbes which have a molecular similarity to the immunogenic gliadin fraction may be useful. Giardia, for example, is a water borne parasite, whose infection mirrors many of the symptoms of coeliac disease.

Now that a better picture of the molecular underpinnings of the condition has been found, population based fine mapping of the linkage blocks, or candidate genes, surrounding the associated loci may provide a more promising avenue of approach. As sequencing technology advances, making this approach easier, faster and cheaper, the next generation of studies will likely be involved in exhaustively detailing the variation surrounding the associated markers.

# References

Adamovic S, Amundsen SS, Lie BA, Hellqvist A, Gudjónsdóttir AH, Ek J, Nilsson S, Wahlström J, Ascher H, Sollid LM, Naluai AT. Fine mapping study in Scandinavian families suggests association between coeliac disease and haplotypes in chromosome region 5q32. Tissue Antigens. 2008 Jan;71(1):27-34. Epub 2007 Oct 30.

Adams F, translator (1856). "On The Cœliac Affection". The extant works of Aretaeus, The Cappadocian. London: Sydenham Society.

Aho, K., M. Koskenvuo, et al. (1986). "Occurence of rheumatoid arthritis in a nationwide series of twins." J Rheumatol 13: 899-902.

Akobeng AK, Thomas AG (June 2008). "Systematic review: tolerable amount of gluten for people with coeliac disease". Aliment. Pharmacol. Ther. 27 (11): 1044–52.

Al-Abdulhadi SA, Al-Rabia MW. Linkage and haplotype analysis for chemokine receptors clustered on chromosome 3p21.3 and transmitted in family pedigrees with asthma and atopy.

Alcina A, Vandenbroeck K, Otaegui D, Saiz A, Gonzalez JR, Fernandez O, Cavanillas ML, Cénit MC, Arroyo R, Alloza I, García-Barcina M, Antigüedad A, Leyva L, Izquierdo G, Lucas M, Fedetz M, Pinto-Medel MJ, Olascoaga J, Blanco Y, Comabella M, Montalban X, Urcelay E, Matesanz F. The autoimmune disease-associated KIF5A, CD226 and SH2B3 gene variants confer susceptibility for multiple sclerosis. Genes Immun. 2010 Jul;11(5):439-45. Epub 2010 May 27.

Almeida A, Zhu XX, Vogt N, Tyagi R, Muleris M, Dutrillaux AM, Dutrillaux B, Ross D, Malfoy B, Hanash S.

al-Tawaty AI, Elbargathy SM. Coeliac disease in north-eastern Libya. Ann Trop Paediatr. 1998 Mar;18(1):27-30.

Amundsen SS, Adamovic S, Hellqvist A et al: A comprehensive screen for SNP associations on chromosome region 5q31–33 in Swedish/Norwegian coeliac disease families. Eur J Hum Genet 2007.

Amundsen SS, Monsuur AJ, Wapenaar MC, Lie BA, Ek J, Gudjónsdóttir AH, Ascher H, Wijmenga C, Sollid LM. Association analysis of MYOIXB gene polymorphisms with coeliac disease in a Swedish/Norwegian cohort.Hum Immunol. 2006 Apr-May;67(4-5):341-5. Epub 2006 May 3

Amundsen SS, Naluai AT, Ascher H, Ek J, Gudjónsdóttir AH, Wahlström J, Lie BA, Sollid LM. Genetic analysis of the CD28/CTLA4/ICOS (COELIAC3) region in Coeliac disease. Tissue Antigens. 2004 Nov;64(5):593-9.

Amundsen SS, Rundberg J, Adamovic S, Gudjónsdóttir AH, Ascher H, Ek J, Nilsson S, Lie BA, Naluai AT, Sollid LM. Four novel coeliac disease regions replicated in an association study of a Swedish-Norwegian family cohort. Genes Immun. 2010 Jan;11(1):79-86. Epub 2009 Aug 20.

Ann Saudi Med. 2010 Mar-Apr;30(2):115-22. GAC1, a new member of the leucine-rich repeat superfamily on chromosome band 1q32.1, is amplified and overexpressed in malignant gliomas. Oncogene. 1998 Jun 11;16(23):2997-3002.

Antonioli DA. Coeliac disease: a progress report. Mod Pathol. 2003 Apr;16(4):342-6.

Ardlie KG, Kruglyak L, Seielstad M: Patterns of linkage disequilibrium in the human genome. Nat Rev Genet 2002; 3: 299– 309.

Arentz-Hansen H, Fleckenstein B, Molberg Ø, Scott H, Koning F, Jung G, Roepstorff P, Lundin KE, Sollid LM. The molecular basis for oat intolerance in patients with coeliac disease. PLoS Med. 2004 Oct;1(1):e1.

Arentz-Hansen H, McAdam SN, Molberg O, et al. Coeliac lesion T cells recognize epitopes that cluster in regions of gliadins rich in proline residues.

Gastroenterology2002;123:803–9

Askling J, Linet M, Gridley G, Halstensen TS, Ekstrom K, Ekbom A: Cancer incidence in a population-based cohort of individuals hospitalized with coeliac disease or dermatitis herpetiformis. Gastroenterology 2002; 123: 1428– 1435.

Azazi EA, Bakir SM, Mohtady HA, Almonem AA. Circulating chemokine eotaxin and chemokine receptor CCR3 in allergic patients. association loci for human diseases and traits. Proc. Natl Acad. Sci. 106, 9362–9367 (2009).

Babron MC, Nilsson S, Adamovic S et al. Meta and pooled analysis of European coeliac disease data. Eur J Hum Genet 2003; 11: 828–834.

Barreto M, Santos E, Ferreira R, Fesel C, Fontes MF, Pereira C, Martins B, Andreia R, Viana JF, Crespo F, Vasconcelos C, Ferreira C, Vicente AM. Evidence for CTLA4 as a susceptibility gene for systemic lupus erythematosus. Eur J Hum Genet. 2004 Aug;12(8):620-6.

Becker KG. The common variants/multiple disease hypothesis of common complex genetic disorders. Med Hypotheses. 2004;62(2):309-17.

Bevan S, Popat S, Braegger CP, Busch A, O'Donoghue D, Falth-Magnusson K, Ferguson A, Godkin A, Hogberg L, Holmes G, Hosie KB, Howdle PD, Jenkins H, Jewell D, Johnston S, Kennedy NP, Kerr G, Kumar P, Logan RF, Love AH, Marsh M, Mulder CJ, Sjoberg K, Stenhammer L, Walker-Smith J, Marossy AM, Houlston RS. Contribution of the MHC region to the familial risk of coeliac disease. J Med Genet. 1999 Sep;36(9):687-90.

Bingley PJ, Williams AJ, Norcross AJ, et al. Undiagnosed coeliac disease at age seven: population based prospective birth cohort study. BMJ (Clinical Research ed) 2004;328(7435):322–3.

Boettcher JP, Kirchner M, Churin Y, Kaushansky A, Pompaiah M, Thorn H, Brinkmann V, Macbeath G, Meyer TF. Tyrosine-phosphorylated caveolin-1 blocks bacterial uptake by inducing Vav2-RhoA-mediated cytoskeletal rearrangements. PLoS Biol. 2010 Aug 24;8(8). pii: e1000457.

Boudraa G, Hachelaf W, Benbouabdellah M, Belkadi M, Benmansour FZ, Touhami M. Acta Paediatr Suppl. 1996 May;412:58-60. Prevalence of coeliac disease in diabetic children and their first- degree relatives in west Algeria: screening with serological markers.

Brady J, Carotta S, Thong RP, Chan CJ, Hayakawa Y, Smyth MJ, Nutt SL. The interactions of multiple cytokines control NK cell maturation. J Immunol. 2010 Dec 1;185(11):6679-88. Epub 2010 Oct 25.

Brophy K, Ryan AW, Thornton JM, Abuzakouk M, Fitzgerald AP, McLoughlin RM, O'morain C, Kennedy NP, Stevens FM, Feighery C, Kelleher D, McManus R. Haplotypes in the CTLA4 region are associated with Coeliac disease in the Irish population. Genes Immun. 2006 Jan;7(1):19-26.

Brown GR, Lane GW, Whittington BJ. Disparate role of LIGHT in organ-specific donor T cells activation and effector molecules in MHC class II disparate GVHD. J Clin Immunol. 2010 Jan;30(1):178-84. Epub 2009 Oct 14.

Buchanan, F. C. and Thue, T. D. 1998. Intrabreed polymorphic information content of microsatellites in cattle and sheep. Can. J. Anim. Sci. 78: 425–428.

C J R Goddard, H R Gillett. Complications of coeliac disease: are all patients at risk? Postgrad Med J 2006;82:705-712 doi:10.1136/pgmj.2006.048876

Carlson, C.S., Eberle, M.A. Rieder, M.J., Qian, Y., Kruglyak, L., Mickerson, D.A. (2004). Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am. J. Hum. Genet. 74, 106-120.

Catassi, C., E. Fabiani, et al. (1996). "The coeliac iceberg in Italy. A multicentre

antigliadin antibodies screening for coeliac disease in school-age subjects." Acta Paediatr Suppl 412: 29-35.

Chapman BL, Henry K, Paice F, Coghill NF, Stewart JS. Measuring the response of the Coeliac mucosa in adult Coeliac disease to treatment with a gluten-free diet. Gut 1974;15(11):870–4.

Charrière GM, Ip WE, Dejardin S, Boyer L, Sokolovska A, Cappillino MP, Cherayil BJ, Podolsky DK, Kobayashi KS, Silverman N, Lacy-Hulbert A, Stuart LM. Identification of Drosophila Yin and PEPT2 as evolutionarily conserved phagosome-associated muramyl dipeptide transporters. J Biol Chem. 2010 Jun 25;285(26):20147-54. Epub 2010 Apr 20.

Chorzelski TP, Sulej T, Tchorzewska H et al. (1983) IgA class endomysium antibodies in dermatitis herpetiformis and coeliac disease. Ann N Y Acad Sci 420, 325–334.

Cirillo G, Di Domenico MR, Corsi I, Gagliardo T, Del Giudice EM, Perrone L, Tolone C. Dig Liver Dis. Do MYOIXB genetic variants predispose to Coeliac disease? An association study in a cohort of South Italian children. Dig Liver Dis 2007 Mar;39(3):228-31. Epub 2007 Jan 30.

Cohen S, Ciechanover A, Kravtsova-Ivantsiv Y, Lapid D, Lahav-Baratz S. Biochem Biophys Res Commun. ABIN-1 negatively regulates NF-kappaB by inhibiting processing of the p105 precursor. 2009 Nov 13;389(2):205-10. Epub 2009 Aug 18.

Cooney R, Cummings JR, Pathan S, Beckly J, Geremia A, Hancock L, Guo C, Morris A, Jewell DP. Association between genetic variants in myosin IXB and Crohn's disease. Inflamm Bowel Dis. 2009 Jul;15(7):1014-21.

Cronin, Cornelius C. Shanahan, Fergus. Why is Coeliac Disease So Common in Ireland? Perspectives in Biology and Medicine - Volume 44, Number 3, Summer 2001, pp. 342-352

Cunha C, Di Ianni M, Bozza S, Giovannini G, Zagarella S, Zelante T, D'Angelo C, Pierini A, Pitzurra L,

Cytokines in the thymus: production and biological effects. Yarilin, A. A., and I. M. Belyakov. 2004. Curr. Med. Chem. 11:447.

Daha NA, Kurreeman FA, Marques RB, Stoeken-Rijsbergen G, Verduijn W, Huizinga TW, Toes RE. Confirmation of STAT4, IL2/IL21, and CTLA4 polymorphisms in rheumatoid arthritis. Arthritis Rheum. 2009 May;60(5):1255-60.

Daniel S, Arvelo MB, Patel VI, Longo CR, Shrikhande G, Shukri T, Mahiou J, Sun DW, Mottley C, Grey ST, Ferran C: A20 protects endothelial cells from TNF-, Fas-, and NK-mediated cell death by inhibiting caspase 8 activation. Blood 104 : 2376 –2384,2004

Dausset J, Degos L, Hors J: The association of the HL-A antigens with diseases. Clin Immunol Immunopathol 1974; 3: 127– 149. 49 Bluestone R: HL-A antigens in clinical medicine. Dis Mon 1976; 23: 1– 27.

De Jager PL, Chibnik LB, Cui J, Reischl J, Lehr S, Simon KC, Aubin C, Bauer D, Heubach JF, Sandbrink R, Tyblova M, Lelkova P; Steering committee of the BENEFIT study; Steering committee of the BEYOND study; Steering committee of the LTF study; Steering committee of the CCR1 study, Havrdova E, Pohl C, Horakova D, Ascherio A, Hafler DA, Karlson EW. Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. Lancet Neurol. 2009 Dec;8(12):1111-9. Epub 2009 Oct 29.

De Jager W, Vastert SJ, Beekman JM, Wulffraat NM, Kuis W, Coffer PJ, Prakken BJ. Defective phosphorylation of interleukin-18 receptor beta causes impaired natural killer cell function in systemic-onset juvenile idiopathic

arthritis. Arthritis Rheum. 2009 Sep;60(9):2782-93.

De Koning HD, Rodijk-Olthuis D, van Vlijmen-Willems IM, Joosten LA, Netea MG, Schalkwijk J, Zeeuwen PL. A comprehensive analysis of pattern recognition receptors in normal and inflamed human epidermis: upregulation of dectin-1 in psoriasis. J Invest Dermatol. 2010 Nov;130(11):2611-20. Epub 2010 Jul 15.

Debets R, Timans JC, Churakowa T, Zurawski S, de Waal Malefyt R, Moore KW, Abrams JS, O'Garra A, Bazan JF, Kastelein RA. IL-18 receptors, their role in ligand binding and function: anti-IL-1RAcPL antibody, a potent antagonist of IL-18. J Immunol. 2000 Nov 1;165(9):4950-6

Dectin-1 Y238X polymorphism associates with susceptibility to invasive aspergillosis in hematopoietic transplantation through impairment of both recipient- and donor-dependent mechanisms of antifungal immunity. Blood. 2010 Dec 9;116(24):5394-402. Epub 2010 Aug 31.

Dema B, Martínez A, Fernández-Arquero M, Maluenda C, Polanco I, de la Concha EG, Urcelay E, Núñez C. J. Association of IL18RAP and CCR3 with coeliac disease in the Spanish population. Med Genet. 2009 Sep;46(9):617-9. Epub 2009 Jun 18.

Dewan A, Liu M, Hartman S, Zhang SS, Liu DT, Zhao C, Tam PO, Chan WM, Lam DS, Snyder M, Barnstable C, Pang CP, Hoh J. HTRA1 promoter polymorphism in wet age-related macular degeneration. Science. 2006 Nov 10;314(5801):989-92. Epub 2006 Oct 19.

Di Sabatino and Corazza. Coeliac disease. Lancet 2009; 373 (9673):: 1480–93

Dicke WK, Weijers HA, Van De Kamer JH. Coeliac disease. II. The presence in wheat of a factor having a deleterious effect in cases of coeliac disease. Acta Paediatrica 1953;42(1):34–42.

Dieterich W, Esslinger B, Trapp D, et al. Cross linking to tissue transglutaminase and collagen favours gliadin toxicity in eliac disease. Gut 2006;55(4):478–84.

Dolinsek J, Urlep D, Karell K, Partanen J, Micetić-Turk D. The prevalence of coeliac disease among family members of coeliac disease patients. Wien Klin Wochenschr. 2004;116 Suppl 2:8-12.

Dolstra H, Fredrix H, Maas F, Coulie PG, Brasseur F, Mensink E, Adema GJ, de Witte TM, Figdor CG, van de Wiel-van Kemenade E. A human minor histocompatibility antigen specific for B cell acute lymphoblastic leukemia. J Exp Med. 1999 Jan 18;189(2):301-8.

Donovan CE, Mark DA, He HZ, Liou HC, Kobzik L, Wang Y, De Sanctis GT, Perkins DL, Finn PW. NF-kappa B/Rel transcription factors: c-Rel promotes airway hyperresponsiveness and allergic pulmonary inflammation. J Immunol. 1999 Dec 15;163(12):6827-33.

Doody GM, Billadeau DD, Clayton E, Hutchings A, Berland R, McAdam S, Leibson PJ, Turner M. Vav-2 controls NFAT-dependent transcription in B- but not T-lymphocytes. EMBO J. 2000 Nov 15;19(22):6173-84.

Döring F, Martini C, Walter J, Daniel H. Importance of a small N-terminal region in mammalian peptide transporters for substrate affinity and function. J Membr Biol. 2002 Mar 15;186(2):55-62.

Douroudis K, Laine AP, Heinonen M, Hermann R, Lipponen K, Veijola R, Simell O, Knip M, Uibo R, Ilonen J, Kisand K. Association of CTLA4 but not ICOS polymorphisms with type 1 diabetes in two populations with different disease rates. Hum Immunol. 2009 Jul;70(7):536-9. Epub 2009 Apr 17.

Duan L, Chen G, Virmani S, Ying G, Raja SM, Chung BM, Rainey MA, Dimri M, Ortega-Cava CF, Zhao X, Clubb RJ, Tu C, Reddi AL, Naramura M, Band V, Band H. Distinct roles for Rho versus Rac/Cdc42 GTPases downstream of

Vav2 in regulating mammary epithelial acinar architecture. J Biol Chem. 2010 Jan 8;285(2):1555-68. Epub 2009 Oct 13.

Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, Zhernakova A, Heap GA, Adány R, Aromaa A, Bardella MT, van den Berg LH, Bockett NA, de la Concha EG, Dema B, Fehrmann RS, Fernández-Arquero M, Fiatal S, Grandone E, Green PM, Groen HJ, Gwilliam R, Houwen RH, Hunt SE, Kaukinen K, Kelleher D, Korponay-Szabo I, Kurppa K, MacMathuna P, Mäki M, Mazzilli MC, McCann OT, Mearin ML, Mein CA, Mirza MM, Mistry V, Mora B, Morley KI, Mulder CJ, Murray JA, Núñez C, Oosterom E, Ophoff RA, Polanco I, Peltonen L, Platteel M, Rybak A, Salomaa V, Schweizer JJ, Sperandeo MP, Tack GJ, Turner G, Veldink JH, Verbeek WH, Weersma RK, Wolters VM, Urcelay E, Cukrowska B, Greco L, Neuhausen SL, McManus R, Barisani D, Deloukas P, Barrett JC, Saavalainen P, Wijmenga C, van Heel DA. Multiple common variants for coeliac disease influencing immune gene expression. Nat Genet. 2010 Apr;42(4):295-302. Epub 2010 Feb 28.

Dubois PC, van Heel DA. Translational mini-review series on the immunogenetics of gut disease: immunogenetics of coeliac disease. Clinical and Experimental Immunology 2008;153(2):162–73.

Düwel M, Welteke V, Oeckinghaus A, Baens M, Kloo B, Ferch U, Darnay BG, Ruland J, Marynen P, Krappmann D. A20 negatively regulates T cell receptor signaling to NF-kappaB by cleaving Malt1 ubiquitin chains. J Immunol. 2009 Jun 15;182(12):7718-28.

Eagar, N.J. Karandikar, J.A. Bluestone and S.D. Miller, The role of CTLA-4 in induction and maintenance of peripheral T cell tolerance, Eur J Immunol 32 (2002), pp. 972–981

Elder JT. Genes Immun. 2009 Apr;10(3):201-9. Epub 2009 Mar 5. Genome-wide association scan yields new insights into the immunopathogenesis of psoriasis.

El-Mezayen RE, Matsumoto T. In vitro responsiveness to IL-18 in combination with IL-12 or IL-2 by PBMC from patients with bronchial asthma and atopic dermatitis. Clin Immunol. 2004 Apr;111(1):61-8.

EURODIAB ACE Study Group. Variation and trends in incidence of childhood diabetes in Europe. Lancet 2000; 355:873–6.

Eyre S, Hinks A, Bowes J, Flynn E, Martin P, Wilson AG, Morgan AW, Emery P, Steer S, Hocking LJ, Reid DM, Harrison P, Wordsworth P; Yorkshire Early Arthritis Consortium; Biologics in RA Control Consortium, Thomson W, Worthington J, Barton A. Overlapping genetic susceptibility variants between three autoimmune disorders: rheumatoid arthritis, type 1 diabetes and coeliac disease. Arthritis Res Ther. 2010;12(5):R175. Epub 2010 Sep 20.

Fahl SP, Crittenden RB, Allman D, Bender TP. c-Myb is required for pro-B cell differentiation. J Immunol. 2009 Nov 1;183(9):5582-92.

Falchuk ZM, Rogentine GN, Strober W. Predominance of histocompatibility antigen HL-A8 in patients with gluten-sensitive enteropathy. The Journal of Clinical Investigation 1972;51(6):1602–5.

Falzetti F, Carotti A, Perruccio K, Latgé JP, Rodrigues F, Velardi A, Aversa F, Romani L, Carvalho A.

Fasano A, Berti I, Gerarduzzi T, et al. Prevalence of coeliac disease in at-risk and not-at-risk groups in the United States: a large multicenter study. Archives of Internal Medicine 2003;163(3):286–92.

Ferguson A, Arranz E, O'Mahony S. Clinical and pathological spectrum of coeliac disease--active, silent, latent, potential. Gut 1993; 34: 150-1.

Fischer AM, Katayama CD, Pagès G, Pouysségur J, Hedrick SM. The role of erk1

and erk2 in multiple stages of T cell development. Immunity. 2005 Oct;23(4):431-43.

Fitau J, Boulday G, Coulon F, Quillard T, Charreau B. The adaptor molecule Lnk negatively regulates tumor necrosis factor-alpha-dependent VCAM-1 expression in endothelial cells through inhibition of the ERK1 and -2 pathways. J Biol Chem. 2006 Jul 21;281(29):20148-59. Epub 2006 Apr 27.

Fleckenstein B, Molberg O, Qiao SW et al. Gliadin T cell epitope selection by tissue transglutaminase in coeliac disease. Role of enzyme specificity and pH influence on the transamidation versus deamidation process. J Biol Chem 2002; 277:34109–16.

Forton, J., D. Kwiatkowski, et al. (2005). "Accuracy of haplotype reconstruction from haplotype-tagging single-nucleotide polymorphisms." Am J Hum Genet 76: 438-448.

Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, Lees CW, Balschun T, Lee J, Roberts R, Anderson CA, Bis JC, Bumpstead S, Ellinghaus D, Festen EM, Georges M, Green T, Haritunians T, Jostins L, Latiano A, Mathew CG, Montgomery GW, Prescott NJ, Raychaudhuri S, Rotter JI, Schumm P, Sharma Y, Simms LA, Taylor KD, Whiteman D, Wijmenga C, Baldassano RN, Barclay M, Bayless TM, Brand S, Büning C, Cohen A, Colombel JF, Cottone M, Stronati L, Denson T, De Vos M, D'Inca R, Dubinsky M, Edwards C, Florin T, Franchimont D, Gearry R, Glas J, Van Gossum A, Guthery SL, Halfvarson J, Verspaget HW, Hugot JP, Karban A, Laukens D, Lawrance I, Lemann M, Levine A, Libioulle C, Louis E, Mowat C, Newman W, Panés J, Phillips A, Proctor DD, Regueiro M, Russell R, Rutgeerts P, Sanderson J, Sans M, Seibold F, Steinhart AH, Stokkers PC, Torkvist L, Kullak-Ublick G, Wilson D, Walters T, Targan SR, Brant SR, Rioux JD, D'Amato M, Weersma RK, Kugathasan S, Griffiths AM, Mansfield JC, Vermeire S, Duerr RH, Silverberg MS, Satsangi J, Schreiber S, Cho JH, Annese V, Hakonarson H, Daly MJ, Parkes M. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet. 2010 Dec;42(12):1118-25.

Fu G, Vallée S, Rybakin V, McGuire MV, Ampudia J, Brockmeyer C, Salek M, Fallen PR, Hoerter JA, Munshi A, Huang YH, Hu J, Fox HS, Sauer K, Acuto O, Gascoigne NR. Department of Immunology and Microbial Science, The Scripps Research Institute, La Jolla, California, USA. Themis controls thymocyte selection through regulation of T cell antigen receptor-mediated signaling. Nat Immunol. 2009 Aug;10(8):848-56. Epub 2009 Jul 13.

Fung EY, Smyth DJ, Howson JM, Cooper JD, Walker NM, Stevens H, Wicker LS, Todd JA. Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus. Genes Immun. 2009 Mar;10(2):188-91. Epub 2008 Dec 25.

G. De Palma, J. Cinova, R. Stepankova, L. Tuckova, and Y. Sanz. Pivotal Advance: Bifidobacteria and Gram-negative bacteria differentially influence immune responses in the proinflammatory milieu of celiac disease. J Leukoc Biol May 2010 87:765-778; doi:10.1189/jlb.0709471

Gandolfi L, Pratesi R, Cordoba JC, Tauil PL, Gasparin M, Catassi C. Prevalence of celiac disease among blood donors in Brazil. Am J Gastroenterol. 2000 Mar;95(3):689-92.

Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, Ortmann W, Kosoy R, Ferreira RC, Nordmark G, Gunnarsson I, Svenungsson E, Padyukov L, Sturfelt G, Jönsen A, Bengtsson AA, Rantapää-Dahlqvist S, Baechler EC, Brown EE, Alarcón GS, Edberg JC, Ramsey-Goldman R, McGwin G Jr,

Reveille JD, Vilá LM, Kimberly RP, Manzi S, Petri MA, Lee A, Gregersen PK, Seldin MF, Rönnblom L, Criswell LA, Syvänen AC, Behrens TW, Graham RR. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. Nat Genet. 2009 Nov;41(11):1228-33. Epub 2009 Oct 18.

Gee K, Guzzo C, Che Mat NF, Ma W, Kumar A. The IL-12 family of cytokines in infection, inflammation and autoimmune disorders. Inflamm Allergy Drug Targets. 2009 Mar;8(1):40-52.

Gee, SJ (1888). "On the coeliac affection". St Bartholomew's Hospital Report 24: 17–20.

Giedraitis V, He B, Huang WX, Hillert J. Cloning and mutation analysis of the human IL-18 promoter: a possible role of polymorphisms in expression regulation. J Neuroimmunol. 2001 Jan 1;112(1-2):146-52.

Giordano M, Marano C, Mellai M et al: A family-based study does not confirm the association of MYOIXB with coeliac disease in the Italian population. Genes Immun 2006; 7: 606– 608.

Glas J, Stallhofer J, Ripke S, Wetzke M, Pfennig S, Klein W, Epplen JT, Griga T, Schiemann U, Lacher M, Koletzko S, Folwaczny M, Lohse P, Göke B, Ochsenkühn T, Müller-Myhsok B, Brand S. Novel genetic risk markers for ulcerative colitis in the IL2/IL21 region are in epistasis with IL23R and suggest a common genetic background for ulcerative colitis and celiac disease. Am J Gastroenterol. 2009 Jul;104(7):1737-44. Epub 2009 May 19.

Gobbi G: Coeliac disease, epilepsy and cerebral calcifications. Brain Dev 2005; 27: 189–200.

Goddard CJR, Gillet HR. Complications of coeliac disease: are all patients at risk? Postgrad Med J 2006;82:705-712

Goldstein, D.B., and Cavalleri, G.L. Genomics: Understanding human diversity. Nature, 437:1241-1242. 2005.

Gorenne I, Jin L, Yoshida T, Sanders JM, Sarembock IJ, Owens GK, Somlyo AP, Somlyo AV. LPP expression during in vitro smooth muscle differentiation and stent-induced vascular injury. Circ Res. 2006 Feb 17;98(3):378-85. Epub 2006 Jan 5.

Goulding C, McManus R, Murphy A, MacDonald G, Barrett S, Crowe J, Hegarty J, McKiernan S, Kelleher D. The CCR5-delta32 mutation: impact on disease outcome in individuals with hepatitis C infection from a single source. Gut. 2005 Aug;54(8):1157-61. Epub 2005 Apr 29.

Gracie JA, Koyama N, Murdoch J, Field M, McGarry F, Crilly A, Schobel A, Madhok R, Pons-Kühnemann J, McInnes IB, Möller B. Disease association of two distinct interleukin-18 promoter polymorphisms in Caucasian rheumatoid arthritis patients. Genes Immun. 2005 May;6(3):211-6.

Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM, Burtt NP, Guiducci C, Parkin M, Gates C, Plenge RM, Behrens TW, Wither JE, Rioux JD, Fortin PR, Graham DC, Wong AK, Vyse TJ, Daly MJ, Altshuler D, Moser KL, Gaffney PM. Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. at Genet. 2008 Sep;40(9):1059-61.

Greco L, Romino R, Coto I, et al. The first large population based twin study of Coeliac disease. Gut 2002;50(5):624–8.

Greco, L., Babron, M-C., Corazza, G. et al. 2001 Existence of a genetic risk factor on chromosome 5q in Italian Coeliac Disease families. Ann Hum Genet 65, 35–41.

Greco, L., Corazza, G., Babron, M. C. et al. 1998 Genome search in Coeliac Disease. Am J Hum Genet 62, 669–675.

Green PH, Rostami K, Marsh MN. Diagnosis of coeliac disease. Best Pract Res Clin

Gastroenterol. 2005 Jun;19(3):389-400.

Gregersen PK, Amos CI, Lee AT, Lu Y, Remmers EF, Kastner DL, Seldin MF, Criswell LA, Plenge RM, Holers VM, Mikuls TR, Sokka T, Moreland LW, Bridges SL Jr, Xie G, Begovich AB, Siminovitch KA. REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. Nat Genet. 2009 Jul;41(7):820-3. Epub 2009 Jun 7.

Griffin M, Casadio R, Bergamini CM. Transglutaminases: nature's biological glues. The Biochemical Journal 2002;368(Pt 2):377–96.

Gringhuis SI, den Dunnen J, Litjens M, van der Vlist M, Wevers B, Bruijns SC, Geijtenbeek TB. Dectin-1 directs T helper cell differentiation by controlling noncanonical NF-kappaB activation through Raf-1 and Syk. Nat Immunol. 2009 Feb;10(2):203-13. Epub 2009 Jan 4.

Gudbjartsson DF, Arnar DO, Helgadottir A, Gretarsdottir S, Holm H, Sigurdsson A, Jonasdottir A, Baker A, Thorleifsson G, Kristjansson K, Palsson A, Blondal T, Sulem P, Backman VM, Hardarson GA, Palsdottir E, Helgason A, Sigurjonsdottir R, Sverrisson JT, Kostulas K, Ng MC, Baum L, So WY, Wong KS, Chan JC, Furie KL, Greenberg SM, Sale M, Kelly P, MacRae CA, Smith EE, Rosand J, Hillert J, Ma RC, Ellinor PT, Thorgeirsson G, Gulcher JR, Kong A, Thorsteinsdottir U, Stefansson K. Variants conferring risk of atrial fibrillation on chromosome 4q25. Nature. 2007 Jul 19;448(7151):353-7. Epub 2007 Jul 1.

Gudjónsdóttir AH, Nilsson S, Ek J, Kristiansson B, Ascher H. The risk of coeliac disease in 107 families with at least two affected siblings. J Pediatr Gastroenterol Nutr. 2004;38:338–42.

Habib-Agahi M, Phan TT, Searle PF. Co-stimulation with 4-1BB ligand allows extended T-cell proliferation, synergizes with CD80/CD86 and can reactivate anergic T cells. Int Immunol. 2007 Dec;19(12):1383-94. Epub 2007 Oct 31.

Haimila K, Smedberg T, Mustalahti K, Mäki M, Partanen J, Holopainen P. Genes Immun. 2004 Mar;5(2):85-92. Genetic association of Coeliac disease susceptibility to polymorphisms in the ICOS gene on chromosome 2q33.

Han JI, Huang NN, Kim DU, Kehrl JH. RGS1 and RGS13 mRNA silencing in a human B lymphoma line enhances responsiveness to chemoattractants and impairs desensitization.J Leukoc Biol. 2006 Jun;79(6):1357-68. Epub 2006 Mar 24.

Harteneck C, Reiter B. TRP channels activated by extracellular hypo-osmoticity in epithelia. Biochem Soc Trans. 2007 Feb;35(Pt 1):91-5.

Hartl D, Krauss-Etschmann S, Koller B, Hordijk PL, Kuijpers TW, Hoffmann F, Hector A, Eber E, Marcos V, Bittmann I, Eickelberg O, Griese M, Roos D. Infiltrated neutrophils acquire novel chemokine receptor expression and chemokine responsiveness in chronic inflammatory lung diseases. J Immunol. 2008 Dec 1;181(11):8053-67.

Heap GA, van Heel DA. Semin Immunol. 2009 Dec;21(6):346-54. Epub 2009 May 13.Genetics and pathogenesis of Celiac disease.

Hill, I., A. Fasano, et al. (2000). The prevalance of celiac disease in at-risk groups of children in the United States. J Pediatr 136: 86.

Hilliard BA, Mason N, Xu L, Sun J, Lamhamedi-Cherradi SE, Liou HC, Hunter C, Chen YH. Critical roles of c-Rel in autoimmune inflammation and helper T cell differentiation. J Clin Invest. 2002 Sep;110(6):843-50.

Hindorff, L. A. et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl Acad. Sci. 106, 9362–9367 (2009).

Hinks A, Martin P, Flynn E, Eyre S, Packham J, Barton A; Childhood Arthritis

Prospective Study (CAPS), UKRAG Consortium, BSPAR study group, Worthington J, Thomson W. Investigation of type 1 diabetes and coeliac disease susceptibility loci for association with juvenile idiopathic arthritis. Ann Rheum Dis. 2010 Dec;69(12):2169-72. Epub 2010 Jul 20.

Hollis-Moffatt JE, Chen-Xu M, Topless R, Dalbeth N, Gow PJ, Harrison AA, Highton J, Jones PB, Nissen M, Smith MD, van Rij A, Jones GT, Stamp LK, Merriman TR. Only one independent genetic association with rheumatoid arthritis within the KIAA1109-TENR-IL2-IL21 locus in Caucasian sample sets: confirmation of association of rs6822844 with rheumatoid arthritis at a genome-wide level of significance. Arthritis Res Ther. 2010;12(3):R116. Epub 2010 Jun 16.

Holopainen P, Arvas M, Sistonen P et al: CD28/CTLA4 gene region on chromosome 2q33 confers genetic susceptibility to coeliac disease. A linkage and family-based association study. Tissue Antigens 1999; 53: 470–475.

Horwitz DA, Zheng SG, Gray JD. The role of the combination of IL-2 and TGF-beta or IL-10 in the generation and function of CD4+ CD25+ and CD8+ regulatory T cell subsets. J Leukoc Biol. 2003 Oct;74(4):471-8.

Hunt KA, McGovern DP, Kumar PJ, Ghosh S, Travis SP, Walters JR, Jewell DP, Playford RJ, van Heel DA. A common CTLA4 haplotype associated with Coeliac disease. Eur J Hum Genet. 2005 Apr;13(4):440-4.

Hunt KA, Monsuur AJ, McArdle WL, Kumar PJ, Travis SP, Walters JR, Jewell DP, Strachan DP, Playford RJ, Wijmenga C, van Heel DA. Lack of association of MYOIXB genetic variants with eliac disease in a British cohort. Gut. 2006 Jul;55(7):969-72. Epub 2006 Jan 19.

Hunter DJ, Kraft P. Drinking from the fire hose--statistical issues in genomewide association studies. N Engl J Med. 2007 Aug 2;357(5):436-9. Epub 2007 Jul 18.

Huppa JB, Davis MM. T-cell-antigen recognition and the immunological synapse. Nat Rev Immunol. 2003 Dec;3(12):973-83.

Illumina Technical Bulletin

Imboden M, Nicod L, Nieters A, Glaus E, Matyas G, Bircher AJ, et al. The common G-allele of interleukin-18 single-nucleotide polymorphism is a genetic risk factor for atopic asthma. The SAPALDIA Cohort Study. Clin Exp Allergy. 2006 Feb;36(2):211-8.

Imhof I, Gasper WJ, Derynck R. Association of tetraspanin CD9 with transmembrane TGF{alpha} confers alterations in cell-surface presentation of TGF{alpha} and cytoskeletal organization. J Cell Sci. 2008 Jul 1;121(Pt 13):2265-74. Epub 2008 Jun 10.

International Multiple Sclerosis Genetics Consortium (IMSGC). IL12A, MPHOSPH9/CDK2AP1 and RGS1 are novel multiple sclerosis susceptibility loci. Genes Immun. 2010 Jul;11(5):397-405. Epub 2010 Jun 17.

Jabri B,de SerreNP,Cellier C et al. Selective expansion of intraepithelial lymphocytes expressing the HLA-E-specific natural killer receptor CD94 in celiac disease. Gastroenterology 2000; 118:867–79.

Jang WC, Park SB, Nam YH, Lee SS, Kim JW, Chang IS, et al. Interleukin-18 gene polymorphisms in Korean patients with Behcet's disease. Clin Exp Rheumatol. 2005 Jul-Aug;23(4 Suppl 38):S59-63.

Jenner RG, Townsend MJ, Jackson I, Sun K, Bouwman RD, Young RA, Glimcher LH, Lord GM. The transcription factors T-bet and GATA-3 control alternative pathways of T-cell differentiation through a shared set of target genes. Proc Natl Acad Sci U S A. 2009 Oct 20;106(42):17876-81. Epub 2009 Oct 5.

Jhang WK, Kang MJ, Jin HS, Yu J, Kim BJ, Kim BS, Lee JK, Seo EJ, Yoo HW, Park IS,

Hong YM, Hong SJ. The CCR5 (-2135C/T) polymorphism may be associated with the development of Kawasaki disease in Korean children. J Clin Immunol. 2009 Jan;29(1):22-8. Epub 2008 Jul 15.

Jin HR, Jin X, Lee JJ. Zinc-finger protein 91 plays a key role in LIGHT-induced activation of non-canonical NF-κB pathway. Biochem Biophys Res Commun. 2010 Oct 1;400(4):581-6. Epub 2010 Sep 6.

Jin L, Kern MJ, Otey CA, Wamhoff BR, Somlyo AV. Angiotensin II, focal adhesion kinase, and PRX1 enhance smooth muscle expression of lipoma preferred partner and its newly identified binding partner palladin to promote cell migration. Circ Res. 2007 Mar 30;100(6):817-25. Epub 2007 Feb 22.

Jinquan T, Jacobi HH, Jing C, Millner A, Sten E, Hviid L, Anting L, Ryder LP, Glue C, Skov PS, Jarman E, Lamberth K, Malling HJ, Poulsen LK. CCR3 expression induced by IL-2 and IL-4 functioning as a death receptor for B cells. J Immunol. 2003 Aug 15;171(4):1722-31.

Jung BH, Beck SE, Cabral J, Chau E, Cabrera BL, Fiorino A, Smith EJ, Bocanegra M, Carethers JM. Activin type 2 receptor restoration in MSI-H colon cancer suppresses growth and enhances migration with activin. Gastroenterology. 2007 Feb;132(2):633-44. Epub 2006 Nov 16.

Juremalm M, Olsson N, Nilsson G. Selective CCL5/RANTES-induced mast cell migration through interactions with chemokine receptors CCR1 and CCR4. Biochem Biophys Res Commun. 2002 Sep 27;297(3):480-5.

Kagnoff MF. Coeliac disease: pathogenesis of a model immunogenetic disease. J Clin Invest 2007; 117: 41–49.

Kankkunen P, Teirilä L, Rintahaka J, Alenius H, Wolff H, Matikainen S. (1,3)-beta-glucans activate both dectin-1 and NLRP3 inflammasome in human macrophages. J Immunol. 2010 Jun 1;184(11):6335-42. Epub 2010 Apr 26.

Kasaian MT, Whitters MJ, Carter LL, Lowe LD, Jussif JM, Deng B, Johnson KA, Witek JS, Senices M, Konz RF, Wurster AL, Donaldson DD, Collins M, Young DA, Grusby MJ. IL-21 limits NK cell responses and promotes antigen-specific T cell activation: a mediator of the transition from innate to adaptive immunity. Immunity. 2002 Apr;16(4):559-69.

Kavanagh B, O'Brien S, Lee D, Hou Y, Weinberg V, Rini B, Allison JP, Small EJ, Fong L. CTLA4 blockade expands FoxP3+ regulatory and activated effector CD4+ T cells in a dose-dependent fashion. Blood. 2008 Aug 15;112(4):1175-83. Epub 2008 Jun 3.

Kawasaki A, Ito S, Furukawa H, Hayashi T, Goto D, Matsumoto I, Kusaoi M, Ohashi J, Graham RR, Matsuta K, Behrens TW, Tohma S, Takasaki Y, Hashimoto H, Sumida T, Tsuchiya N. Association of TNFAIP3 interacting protein 1, TNIP1 with systemic lupus erythematosus in a Japanese population: a case-control association study. Arthritis Res Ther. 2010;12(5):R174. Epub 2010 Sep 17.

Kelly CP, Feighery CF, Gallagher RB et al. (1991) Mucosal and systemic IgA anti-gliadin antibody in coeliac disease. Dig Dis Sci 36, 743–751.

Keuning JJ, Pena AS, van Leeuwen A, van Hooff JP, va Rood JJ. HLA-DW3 associated with coeliac disease. Lancet 1976;1(7958):506–8.

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J. Complement factor H polymorphism in age-related macular degeneration. Science. 2005 Apr 15;308(5720):385-9. Epub 2005 Mar 10.

Koch MA, Tucker-Heard G, Perdue NR, Killebrew JR, Urdahl KB, Campbell DJ. The transcription factor T-bet controls regulatory T cell homeostasis and function during type 1 inflammation. Nat Immunol. 2009 Jun;10(6):595-602. Epub 2009 May 3.

Koskinen LL, Einarsdottir E, Dukes E, Heap GA, Dubois P, Korponay-Szabo IR, Kaukinen K, Kurppa K, Ziberna F, Vatta S, Not T, Ventura A, Sistonen P, Adány R, Pocsai Z, Széles G, Mäki M, Kere J, Wijmenga C, van Heel DA, Saavalainen P. Association study of the IL18RAP locus in three European populations with coeliac disease. Hum Mol Genet. 2009 Mar 15;18(6):1148-55. Epub 2008 Dec 22.

Koutoulaki A, Langley M, Sloan AJ, Aeschlimann D, Wei XQ. TNFalpha and TGF-beta1 influence IL-18-induced IFNgamma production through regulation of IL-18 receptor and T-bet expression. Cytokine. 2010 Feb;49(2):177-84. Epub 2009 Oct 30.

Kuchen S, Robbins R, Sims GP, Sheng C, Phillips TM, Lipsky PE, Ettinger R. Essential role of IL-21 in B cell activation, expansion, and plasma cell generation during CD4+ T cell-B cell collaboration. J Immunol. 2007 Nov 1;179(9):5886-96.

Kuo MT, Wei Y, Yang X, Tatebe S, Liu J, Troncoso P, Sahin A, Ro JY, Hamilton SR, Savaraj N. Association of fragile site-associated (FSA) gene expression with epithelial differentiation and tumor development. Biochem Biophys Res Commun. 2006 Feb 17;340(3):887-93. Epub 2005 Dec 27.

Lametschwandtner G, Biedermann T, Schwärzler C, Günther C, Kund J, Fassl S, Hinteregger S, Carballido-Perrig N, Szabo SJ, Glimcher LH, Carballido JM. Sustained T-bet expression confers polarized human TH2 cells with TH1-like cytokine production and migratory capacities. J Allergy Clin Immunol. 2004 May;113(5):987-94.

Langman MJS. McConnell TH, Spiegelhalter DJ. Changing patterns of coeliac disease frequency: an analysis of coeliac society membership records. 1985 Gut, 1985, 26, 175-178.

Le Y, Honczarenko M, Glodek AM, Ho DK, Silberstein LE. CXC chemokine ligand 12-induced focal adhesion kinase activation and segregation into membrane domains is modulated by regulator of G protein signaling 1 in pro-B cells. J Immunol. 2005 Mar 1;174(5):2582-90.

Lee HM, Yuk JM, Shin DM, Jo EK. Dectin-1 is inducible and plays an essential role for mycobacteria-induced innate immune responses in airway epithelial cells. J Clin Immunol. 2009 Nov;29(6):795-805. Epub 2009 Jul 25.

Lee MM, Wong YH. CCR1-mediated activation of Nuclear Factor-kappaB in THP-1 monocytic cells involves Pertussis Toxin-insensitive Galpha(14) and Galpha(16) signaling cascades. J Leukoc Biol. 2009 Dec;86(6):1319-29. Epub 2009 Aug 17.

Lesourne R, Uehara S, Lee J, Song KD, Li L, Pinkhasov J, Zhang Y, Weng NP, Wildt KF, Wang L, Bosselut R, Love PE. Themis, a T cell-specific protein important for late thymocyte development. Nat Immunol. 2009 Aug;10(8):840-7. Epub 2009 Jul 13.

Lewis CM, Whitwell SC, Forbes A, Sanderson J, Mathew CG, Marteau TM. Estimating risks of common complex diseases across genetic and environmental factors: the example of Crohn disease. J Med Genet 2007; 44:689–94. J Med Genet. 1999 Sep;36(9):687-90. 26)

Li C, Yan YP, Shieh B, Lee CM, Lin RY, Chen YM. Frequency of the CCR5 delta 32 mutant allele in HIV-1-positive patients, female sex workers, and a normal population in Taiwan. J Formos Med Assoc. 1997 Dec;96(12):979-84.

Li P, Jayarama S, Ganesh L, Mordi D, Carr R, Kanteti P, Hay N, Prabhakar BS. Akt-phosphorylated mitogen-activated kinase-activating death domain protein(MADD) inhibits TRAIL-induced apoptosis by blocking Fas-associated death domain (FADD) association with death receptor 4. J Biol Chem. 2010 Jul 16;285(29):22713-22. Epub 2010 May 18.

Ling-ling Jiang, Bing-ling Zhang, and You-shi Liu. Is adult celiac disease really uncommon in Chinese? J Zhejiang Univ Sci B. 2009 March; 10(3): 168–171.

Lo W, Sano K, Lebwohl B, Diamond B, Green PH. Changing presentation of adult coeliac disease. Digestive Diseases and Sciences 2003;48(2):395–8.

Louka AS, Moodie SJ, Karell K et al: A collaborative European search for non-DQA1*05-DQB1*02 coeliac disease loci on HLADR3 haplotypes: analysis of transmission from homozygous parents. Hum Immunol 2003; 64: 350–358.

Louka AS, Torinsson Naluai A, D'Alfonso S et al: The IL12B gene does not confer susceptibility to coeliac disease. Tissue Antigens 2002; 59: 70– 72.

Luetteke NC, Lee DC. Transforming growth factor alpha: expression, regulation and biological action of its integral membrane precursor. Semin Cancer Biol. 1990 Aug;1(4):265-75.

Lygerou Z, Pluk H, van Venrooij WJ, Séraphin B. hPop1: an autoantigenic protein subunit shared by the human RNase P and RNase MRP ribonucleoproteins. EMBO J. 1996 Nov 1;15(21):5936-48.

M J Langman, T H McConnell, D J Spiegelhalter, R B McConnell. Changing patterns of coeliac disease frequency: an analysis of Coeliac Society membership records. Gut 1985;26:175-178 doi:10.1136/gut.26.2.175

Maki M, Collin P. Coeliac disease. Lancet 1997;349(9067):1755–9.

Maki M, Mustalahti K, Kokkonen J, et al. Prevalence of coeliac disease among children in finland. The New England Journal of Medicine 2003; 348(25):2517–24.

Manousou P, Kolios G, Valatas V, Drygiannakis I, Bourikas L, Pyrovolaki K, Koutroubakis I, Papadaki HA, Kouroumalis E. Increased expression of chemokine receptor CCR3 and its ligands in ulcerative colitis: the role of colonic epithelial cells in in vitro studies. Clin Exp Immunol. 2010 Nov;162(2):337-47.

Manzotti CN, Tipping H, Perry LC, Mead KI, Blair PJ, Zheng Y, Sansom DM. Inhibition of human T cell proliferation by CTLA-4 utilizes CD80 and requires CD25+ regulatory T cells. Eur J Immunol. 2002 Oct;32(10):2888-96.

Mao M, Biery MC, Kobayashi SV, Ward T, Schimmack G, Burchard J, Schelter JM, Dai H, He YD, Linsley PS. T lymphocyte activation gene identification by coregulated expression on DNA microarrays. Genomics. 2004 Jun;83(6):989-99.

Margaritte-Jeannin P, Babron MC, Bourgey M, Louka AS, Clot F, et al. (2004) HLA-DQ relative risks for coeliac disease in European populations: a study of the European Genetics Cluster on coeliac disease. Tissue Antigens 63: 562–7

Marinelli A, Bossi D, Pelicci PG, Minucci S. A redundant oncogenic potential of the retinoic receptor (RAR) alpha, beta and gamma isoforms in acute promyelocytic leukemia. Leukemia. 2007 Apr;21(4):647-50. Epub 2007 Jan 25.

Marsh M (1992). "Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('coeliac sprue')". Gastroenterology 102 (1): 330–54.

Martens HA, Kallenberg CG, Bijl M. Role of CCR5 Delta32 bp deletion in RA and SLE. Autoimmunity. 2009 May;42(4):260-2.

McGargill MA, Ch'en IL, Katayama CD, Pagès G, Pouysségur J, Hedrick SM. Cutting edge: Extracellular signal-related kinase is not required for negative selection of developing T cells. J Immunol. 2009 Oct 15;183(8):4838-42.

Meresse B, Curran SA, Ciszewski C et al. Reprogramming of CTLs into natural killer-like cells in celiac disease. J Exp Med 2006;203:1343–55.

Michalski J, McCombs C, Arai T, Elston R, Cao T, McCarthy C, Stevens F (1996). "HLA-DR, DQ genotypes of coeliac disease patients and healthy subjects from the West of Ireland". Tissue Antigens 47 (2): 127–33

Molteni N, Caraceni MP, Bardella MT, Ortolani S, Gandolini GG, Bianchi P: Bone mineral density in adult coeliac patients and the effect of gluten-free diet from childhood. Am J Gastroenterol 1990; 85: 51– 53.

Monsuur AJ, de Bakker PI, Alizadeh BZ, Zhernakova A, Bevova MR, Strengman E, Franke L, van't Slot R, van Belzen MJ, Lavrijsen IC, Diosdado B, Daly MJ, Mulder CJ, Mearin ML, Meijer JW, Meijer GA, van Oort E, Wapenaar MC, Koeleman BP, Wijmenga C. Myosin IXB variant increases the risk of coeliac disease and points toward a primary intestinal barrier defect. Nat Genet. 2005 Dec;37(12):1341-4. Epub 2005 Nov 13.

Nadiminty N, Chun JY, Hu Y, Dutt S, Lin X, Gao AC. LIGHT, a member of the TNF superfamily, activates Stat3 mediated by NIK pathway. Biochem Biophys Res Commun. 2007 Jul 27;359(2):379-84. Epub 2007 May 25.

Nair RP, Duffin KC, Helms C, Ding J, Stuart PE, Goldgar D, Gudjonsson JE, Li Y, Tejasvi T, Feng BJ, Ruether A, Schreiber S, Weichenthal M, Gladman D, Rahman P, Schrodi SJ, Prahalad S, Guthery SL, Fischer J, Liao W, Kwok PY, Menter A, Lathrop GM, Wise CA, Begovich AB, Voorhees JJ, Elder JT, Krueger GG, Bowcock AM, Abecasis GR; Collaborative Association Study of Psoriasis. Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. Nat Genet. 2009 Feb;41(2):199-204. Epub 2009 Jan 25.

Nakata Y, Brignier AC, Jin S, Shen Y, Rudnick SI, Sugita M, Gewirtz AM. c-Myb, Menin, GATA-3, and MLL form a dynamic transcription complex that plays a pivotal role in human T helper type 2 cell development. Blood. 2010 Aug 26;116(8):1280-90. Epub 2010 May 18.

Naluai AT, Ascher H, Nilsson S, Wahlström J. Searching for genes influencing a complex disease: the case of coeliac disease. Eur J Hum Genet. 2008 May;16(5):542-53. Epub 2007 Aug 29.

Naluai AT, Nilsson S, Gudjonsdottir AH et al: Genome-wide linkage analysis of Scandinavian affected sib-pairs supports presence of susceptibility loci for coeliac disease on chromosomes 5 and 11. Eur J Hum Genet 2001; 9: 938–944.

Nischwitz S, Cepok S, Kroner A, Wolf C, Knop M, Müller-Sarnowski F, Pfister H, Roeske D, Rieckmann P, Hemmer B, Ising M, Uhr M, Bettecken T, Holsboer F, Müller-Myhsok B, Weber F. Evidence for VAV2 and ZNF433 as susceptibility genes for multiple sclerosis. J Neuroimmunol. 2010 Oct 8;227(1-2):162-6. Epub 2010 Jul 2.

Nistico L, Fagnani C, Coto I, et al. Concordance, disease progression, and heritability of Coeliac disease in Italian twins. Gut 2006;55(6):803–8.

Nóbrega MA, Zhu Y, Plajzer-Frick I, Afzal V, Rubin EM. Megabase deletions of gene deserts result in viable mice. Nature. 2004 Oct 21;431(7011):988-93.

Núñez C, Márquez A, Varadé J, Martínez A, Polanco I, Maluenda C, Fernández-Arquero M, de la Concha EG, Urcelay E. No evidence of association of the MYOIXB polymorphisms with celiac disease in the Spanish population. Tissue Antigens. 2006 Dec;68(6):489-92.

Nunez C, Rueda B, Martinez A et al: A functional variant in the CD209 promoter is associated with DQ2-negative coeliac disease in the Spanish population. World J Gastroenterol 2006; 12: 4397– 4400.

Oberhuber G, Granditsch G, Vogelsang H. The histopathology of coeliac disease:

time for a standardized report scheme for pathologists. Eur J Gastroenterol Hepatol 1999;11:1185–94.

Oberwinkler J, Phillipp SE. TRPM3. Handb Exp Pharmacol. 2007;(179):253-67.

Olmos, P., R. A'Hern, et al. (1988). "The significance of the concordance rate of type 1 (insulin-dependent) diabetes in identical twins." Diabetologia 31: 747-750.

Oppermann M. Chemokine receptor CCR5: insights into structure, function, and regulation. Cell Signal. 2004 Nov;16(11):1201-10.

Orozco G, Eyre S, Hinks A, Bowes J, Morgan AW, Wilson AG, Wordsworth P, Steer S, Hocking L; UKRAG consortium, Thomson W, Worthington J, Barton A. [Epub ahead of print] Study of the common genetic background for rheumatoid arthritis and systemic lupus erythematosus. Ann Rheum Dis. 2010 Nov 10.

Orsini B, Calabrò A, Milani S, Grappone C, Herbst H, Surrenti C. Localization of epidermal growth factor/transforming growth factor-alpha receptor in the human gastric mucosa. An immunohistochemical and in situ hybridization study. Virchows Arch A Pathol Anat Histopathol. 1993;423(1):57-63.

Ovcharenko I., Nobrega M.A., Loots G.G., and Stubbs L. Nucleic Acids Research. 32, W280-W286 (2004)

Ovcharenko, I., Loots, G.G., Giardine, B.M., Hou, M., Ma, J., Hardison, R.C., Stubbs, L. and Miller, W. Mulan: multiple sequence local alignment and visualization for studying function and evolution. Genome Res, 15, 184-194. (2005)

Pan J, Kunkel EJ, Gosslar U, Lazarus N, Langdon P, Broadwell K, Vierra MA, Genovese MC, Butcher EC, Soler D. A novel chemokine ligand for CCR10 and CCR3 expressed by epithelial cells in mucosal tissues. J Immunol. 2000 Sep 15;165(6):2943-9.

Park KS, Baek JA, Do JE, Bang D, Lee ES. CTLA4 gene polymorphisms and soluble CTLA4 protein in Behcet's disease. Tissue Antigens. 2009 Sep;74(3):222-7. Epub 2009 Jun 25.

Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, et al. In vivo enhancer analysis of human conserved non-coding sequences. Nature. 2006;444:499–502

Percopo S, Babron MC, Whalen M, De Virgiliis S, Coto I, Clerget-Darpoux F, Landolfo F, Greco L. Ann Hum Genet. 2003 May;67(Pt 3):265-8. Saturation of the 5q31-q33 candidate region for coeliac disease.

Ploski R, Ek J, Thorsby E, Sollid LM: On the HLA-DQ (alpha 1*0501, beta 1*0201)-associated susceptibility in coeliac disease: a possible gene dosage effect of DQB1*0201. Tissue Antigens 1993; 41: 173–177.

Post PL, Tyska MJ, O'Connell CB, Johung K, Hayward A, Mooseker MS. Myosin-IXb is a single-headed and processive motor. J Biol Chem. 2002 Apr 5;277(14):11679-83. Epub 2002 Jan 18.

Post, P.L. et al. Ilumina Technical Bulletin. Whole-Genome Genotyping with the Sentrix® HumanHap300 Genotyping BeadChip and the Infinium™ II Assay. www.illumina.com/technology/publications. Myosin-IXb is a single-headed and processive motor. J. Biol. Chem. 277, 11679–11683 (2002).

Prahalad S, Hansen S, Whiting A, Guthery SL, Clifford B, McNally B, Zeft AS, Bohnsack JF, Jorde LB. Variants in TNFAIP3, STAT4, and C12orf30 loci associated with multiple autoimmune diseases are also associated with juvenile idiopathic arthritis. Arthritis Rheum. 2009 Jul;60(7):2124-30.

Price P, Witt C, Allcock R, et al. The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases. Immunological Reviews 1999;167:257–74.

Qiao SW, Bergseng E, Molberg O, Jung G, Fleckenstein B, Sollid LM.Institute of Immunology, University of Oslo, Rikshospitalet University Hospital, Oslo, Norway. S.w.qiao@labmed.uio.no Refining the rules of gliadin T cell epitope binding to the disease-associated DQ2 molecule in coeliac disease: importance of proline spacing and glutamine deamidation. J Immunol. 2005 Jul 1;175(1):254-61.

Ramsay RG, Ciznadija D, Sicurella C, Reyes N, Mitchelhill K, Darcy PK, D'Abaco G, Mantamadiotis T. Colon epithelial cell differentiation is inhibited by constitutive c-myb expression or mutant APC plus activated RAS. DNA Cell Biol. 2005 Jan;24(1):21-9.

Reid DM, Gow NA, Brown GD. Pattern recognition: recent insights from Dectin-1. Curr Opin Immunol. 2009 Feb;21(1):30-7. Epub 2009 Feb 14.

Reijmerink NE, Postma DS, Bruinenberg M, Nolte IM, Meyers DA, Bleecker ER, Koppelman GH. Association of IL1RL1, IL18R1, and IL18RAP gene cluster polymorphisms with asthma and atopy. J Allergy Clin Immunol. 2008 Sep;122(3):651-4.e8.

Reiman EM, Webster JA, Myers AJ, et al. GAB2 alleles modify Alzheimer's risk in APOE ε4 carriers. Neuron. 2007;54(5):713–720. [PMC free article] [PubMed]

Reunala TL. Dermatitis herpetiformis. Clin Dermatol 2001; 19: 728-736

Rioux JD, Karinen H, Kocher K et al: Genomewide search and association studies in a Finnish coeliac disease population: identification of a novel locus and replication of the HLA and CTLA4 loci. Am J Med Genet A 2004; 130: 345–350.

Rodriguez-Galán MC, Bream JH, Farr A, Young HA. Synergistic effect of IL-2, IL-12, and IL-18 on thymocyte apoptosis and Th1/Th2 cytokine expression. J Immunol. 2005 Mar 1;174(5):2796-804.

Romanos J, Barisani D, Trynka G, Zhernakova A, Bardella MT, Wijmenga C.J. Six new coeliac disease loci replicated in an Italian population confirm association with coeliac disease. Med Genet. 2009 Jan;46(1):60-3. Epub 2008 Sep 19.

Ryan AW, Thornton JM, Brophy K et al: Chromosome 5q candidate genes in coeliac disease: genetic variation at IL4, IL5, IL9, IL13, IL17B and NR3C1. Tissue Antigens 2005; 65: 150– 155.

Ryan AW, Thornton JM, Brophy K et al: Haplotype variation at the IBD5/SLC22A4 locus (5q31) in coeliac disease in the Irish population. Tissue Antigens 2004; 64: 195– 198.

Sabeti PC, Walsh E, Schaffner SF, Varilly P, Fry B, Hutcheson HB, Cullen M, Mikkelsen TS, Roy J, Patterson N, Cooper R, Reich D, Altshuler D, O'Brien S, Lander ES. The case for selection at CCR5-Delta32. PLoS Biol. 2005 Nov;3(11):e378. Epub 2005 Nov 1.

Sandberg ML, Sutton SE, Pletcher MT, Wiltshire T, Tarantino LM, Hogenesch JB, Cooke MP. c-Myb and p300 regulate hematopoietic stem cell proliferation and differentiation. Dev Cell. 2005 Feb;8(2):153-66.

Scarpa M, Behboo R, Angriman I, Termini B, Barollo M, Ruffolo C, Polese L, D'Incà R, Sturniolo GC, D'Amico DF. The role of costimulatory molecules CD80 and CD86 and IFNgamma in the pathogenesis of ulcerative colitis. Dig Dis Sci. 2004 Nov-Dec;49(11-12):1738-44.

Schumacher JM, Lee K, Edelhoff S, Braun RE. Distribution of Tenr, an RNA-binding protein, in a lattice-like network within the spermatid nucleus in the mouse. Biol Reprod. 1995 Jun;52(6):1274-83.

Scott A Ness C-Myb UCSD-Nature Molecule Pages Published online: 11 Apr 2006 | doi:10.1038/mp.a000671.01

Seegers D, Borm ME, van Belzen MJ et al: IL12B and IRF1 gene polymorphisms and susceptibility to coeliac disease. Eur J Immunogenet 2003; 30: 421–425.

Shahbazi M, Ebadi H, Fathi D, Roshandel D, Mahamadhoseeni M, Rashidbaghan A, Mahammadi N, Mahammadi MR, Zamani M. CCR5-delta 32 allele is associated with the risk of developing multiple sclerosis in the Iranian population. Cell Mol Neurobiol. 2009 Dec;29(8):1205-9.

Shaikh RB, Santee S, Granger SW, Butrovich K, Cheung T, Kronenberg M, Cheroutre H, Ware CF. Constitutive expression of LIGHT on T cells leads to lymphocyte activation, inflammation, and tissue destruction. J Immunol. 2001 Dec 1;167(11):6330-7.

Shan L, Molberg Ø, Parrot I, et al. Structural basis for gluten intolerance in coeliac sprue. Science 2002; 297: 2275–79.

Shevach EM, McHugh RS, Piccirillo CA, Thornton AM. Control of T-cell activation by CD4+ CD25+ suppressor T cells. 2Immunol Rev. 2001 Aug;182:58-67.

Shiina, T., H. Inoko, et al. (2004). "An update of the HLA genomic region, locus information and disease associations:2004." Tissue Antigens 64: 631-649.

Simon DP, Vadakkadath Meethal S, Wilson AC, Gallego MJ, Weinecke SL, Bruce E, Lyons PF, Haasl RJ, Bowen RL, Atwood CS. Activin receptor signaling regulates prostatic epithelial cell adhesion and viability. Neoplasia. 2009 Apr;11(4):365-76.

Singh N, Chandler PR, Seki Y, Baban B, Takezaki M, Kahler DJ, Munn DH, Larsen CP, Mellor AL, Iwashima M. Role of CD28 in fatal autoimmune disorder in scurfy mice. Blood. 2007 Aug 15;110(4):1199-206. Epub 2007 Apr 26.

Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat. Genet. 38, 209–213 (2006).

Smyth DJ, Plagnol V, Walker NM, Cooper JD, Downes K, Yang JH, Howson JM, Stevens H, McManus R, Wijmenga C, Heap GA, Dubois PC, Clayton DG, Hunt KA, van Heel DA, Todd JA. Shared and distinct genetic variants in type 1 diabetes and celiac disease. N Engl J Med. 2008 Dec 25;359(26):2767-77. Epub 2008 Dec 10.

Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, Thorsby E. Evidence for a primary association of coeliac disease to a particular HLA-DQ alpha/beta heterodimer. The Journal of Experimental Medicine 1989;169(1):345–50.

Sollid LM: Molecular basis of coeliac disease. Annu Rev Immunol 2000; 18: 53– 81

Sollid, L. M. (2002). "Coeliac disease: dissecting a complex inflammatory disorder." Nat Rev Immunol 2(9): 647-655.

Spencer SD, Di Marco F, Hooley J, et al. The orphan receptor CRF2-4 is an essential subunit of the interleukin 10 receptor. J Exp Med 1998;187:571-578

Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, Li Y, Kurreeman FA, Zhernakova A, Hinks A, Guiducci C, Chen R, Alfredsson L, Amos CI, Ardlie KG; BIRAC Consortium, Barton A, Bowes J, Brouwer E, Burtt NP, Catanese JJ, Coblyn J, Coenen MJ, Costenbader KH, Criswell LA, Crusius JB, Cui J, de Bakker PI, De Jager PL, Ding B, Emery P, Flynn E, Harrison P, Hocking LJ, Huizinga TW, Kastner DL, Ke X, Lee AT, Liu X, Martin P, Morgan AW, Padyukov L, Posthumus MD, Radstake TR, Reid DM, Seielstad M, Seldin MF, Shadick NA, Steer S, Tak PP, Thomson W, van der Helm-van Mil AH, van der Horst-Bruinsma IE, van der Schoot CE, van Riel PL, Weinblatt ME, Wilson AG, Wolbink GJ, Wordsworth BP; YEAR Consortium, Wijmenga C, Karlson EW, Toes RE, de Vries N, Begovich AB, Worthington J, Siminovitch KA, Gregersen PK, Klareskog L, Plenge RM.

Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet. 2010 Jun;42(6):508-14. Epub 2010 May 9.

Steele R; CRF. Diagnosis and management of coeliac disease in children. Postgrad Med J. 2011 Jan;87(1023):19-25. Epub 2010 Dec 3.

Stenszky V, Kozma L, Ambro I, Karmazsin L: Calculation of disease susceptibility gene frequency in insulin-dependent diabetes mellitus. Haematologia (Budap) 1978; 12: 141– 148.

Swaan PW, Bensman T, Bahadduri PM, Hall MW, Sarkar A, Bao S, Khantwal CM, Ekins S, Knoell DL. Bacterial peptide recognition and immune activation facilitated by human peptide transporter PEPT2. Am J Respir Cell Mol Biol. 2008 Nov;39(5):536-42. Epub 2008 May 12.

Swinson CM, Levi AJ. Is coeliac disease under-diagnosed? Br Med J 1980; 281: 1258-60.

Takahashi T, Kuniyasu Y, Toda M, Sakaguchi N, Itoh M, Iwata M, Shimizu J, Sakaguchi S. Immunologic self-tolerance maintained by CD25+CD4+ naturally anergic and suppressive T cells: induction of autoimmune disease by breaking their anergic/suppressive state. Int Immunol. 1998 Dec;10(12):1969-80.

Tamás P, Solti Z, Buday L. Membrane-targeting is critical for the phosphorylation of Vav2 by activated EGF receptor. Cell Signal. 2001 Jul;13(7):475-81.

Tartare-Deckert S, Monthouel MN, Charvet C, Foucault I, Van Obberghen E, Bernard A, Altman A, Deckert M. Vav2 activates c-fos serum response element and CD69 expression but negatively regulates nuclear factor of activated T cells and interleukin-2 gene activation in T lymphocyte. J Biol Chem. 2001 Jun 15;276(24):20849-57. Epub 2001 Mar 21.

Teresa Fritz1, Lukas Niederreiter,Timon Adolph1, Richard S Blumberg, Arthur Kaser. Crohn's disease: NOD2, autophagy and ER stress converge. Gut doi:10.1136/gut.2009.206466

The International HapMap Consortium. A haplotype map of the human genome. Nature 437, 1299-1320 27 October 2005.

Thomas DM, Nasim MM, Gullick WJ, Alison MR. Immunoreactivity of transforming growth factor alpha in the normal adult gastrointestinal tract. Gut. 1992 May;33(5):628-31.

Thornton AM, Shevach EM. CD4+CD25+ immunoregulatory T cells suppress polyclonal T cell activation in vitro by inhibiting interleukin 2 production. J Exp Med. 1998 Jul 20;188(2):287-96.

Tiret L, Godefroy T, Lubos E, Nicaud V, Tregouet DA, Barbaux S, et al. Genetic analysis of the interleukin-18 system highlights the role of the interleukin-18 gene in cardiovascular disease. Circulation. 2005 Aug 2;112(5):643-50.

Tollefsen S, Arentz-Hansen H, Fleckenstein B et al. HLA-DQ2 and-DQ8 signatures of gluten T cell epitopes in coeliac disease. J Clin Invest 2006; 116:2226–36.

Tong P, Prendergast JG, Lohan AJ, Farrington SM, Cronin S, Friel N, Bradley DG, Hardiman O, Evans A, Wilson JF, Loftus B. Sequencing and analysis of an Irish human genome Genome Biol. 2010;11(9):R91. Epub 2010 Sep 7.

Toscano V, Conti FG, Anastasi E et al: Importance of gluten in the induction of endocrine autoantibodies and organ dysfunction in adolescent coeliac patients. Am J Gastroenterol 2000; 95:1742– 1748

Tosi R, Vismara D, Tanigaki N, et al. Evidence that coeliac disease is primarily associated with a DC locus allelic specificity. Clinical Immunology and Immunopathology 1983;28(3):395–404.

Ulker M, Yazisiz V, Sallakci N, Avci AB, Sanlioglu S, Yegin O, Terzioglu E. CTLA-4

gene polymorphism of exon 1(+49 A/G) in Turkish systemic lupus erythematosus patients. Int J Immunogenet. 2009 Aug;36(4):245-50.
Vader LW, Stepniak DT, Bunnik EM, et al. Characterization of cereal toxicity for coeliac disease patients based on protein homology in grains. Gastroenterology2003;125:1105–13.
Vader W, Stepniak D, Kooy Y, et al. The HLA-DQ2 gene dose effect in coeliac disease is directly related to the magnitude and breadth of gluten-specific T cell responses. Proceedings of the National Academy of Sciences of the United States of America 2003;100(21):12390–5.
van Belzen MJ, Koeleman BP, Crusius JB, et al. Defining the contribution of the HLA region to cis DQ2-positive eliac disease patients. Genes Immun 2004;5:215–20.
Van Belzen MJ, Meijer JW, Sandkuijl LA et al: A major non-HLA locus in coeliac disease maps to chromosome 19. Gastroenterology 2003; 125: 1032–1041.
van de Veerdonk FL, Teirlinck AC, Kleinnijenhuis J, Kullberg BJ, van Crevel R, van der Meer JW, Joosten LA, Netea MG. Mycobacterium tuberculosis induces IL-17A responses through TLR4 and dectin-1 and is critically dependent on endogenous IL-1. J Leukoc Biol. 2010 Aug;88(2):227-32. Epub 2010 Mar 18.
van de Wal Y, Kooy Y, van Veelen P et al. Selective deamidation by tissue transglutaminase strongly enhances gliadin-specific T cell reactivity. J Immunol 1998; 161:1585–8.
van de Wal Y, Kooy YM, Drijfhout JW, Amons R, Papadopoulos GK, Koning F. Unique peptide binding characteristics of the disease-associated DQ(alpha 1*0501, beta 1*0201) vs the non-disease-associated DQ(alpha 1*0201, beta 1*0202) molecule. Immunogenetics 1997;46(6):484–92.
van Heel DA, Hunt K, Greco L, Wijmenga C. Genetics in coeliac disease. Best Practice & Research 2005;19(3):323–39.
van Heel DA, West J. Recent advances in coeliac disease. Gut 2006;55(7): 1037–46.
Van Heel *et al.* A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nat Genet. 2007 July ; 39(7): 827–829.
van Heel, D. A., J. Satsangi, et al. (2000). "Inflammatory bowel disease: progress toward a gene." Can J Gastroenterol 14: 207-218.
Vartdal F, Johansen BH, Friede T et al. The peptide binding motif of the disease associated HLA-DQ (alpha 1*0501, beta 1*0201) molecule. Eur J Immunol 1996; 26:2764–72.
Vasilevko V, Ghochikyan A, Holterman MJ, Agadjanyan MG. CD80 (B7-1) and CD86 (B7-2) are functionally equivalent in the initiation and maintenance of CD4+ T-cell proliferation after activation with suboptimal doses of PHA. DNA Cell Biol. 2002 Mar;21(3):137-49.
Vereecke L, Beyaert R, van Loo G. The ubiquitin-editing enzyme A20 (TNFAIP3) is a central regulator of immunopathology. Trends Immunol. 2009 Aug;30(8):383-91. Epub 2009 Jul 28.
Vivas S, Ruiz de Morales JG, Riestra S, Arias L, Fuentes D, Alvarez N, Calleja S, Hernando M, Herrero B, Casqueiro J, Rodrigo L. Duodenal biopsy may be avoided when high transglutaminase antibody titers are present. World J Gastroenterol. 2009 Oct 14;15(38):4775-80.
Wang J, Anders RA, Wang Y, Turner JR, Abraham C, Pfeffer K, Fu YX. The critical role of LIGHT in promoting intestinal inflammation and Crohn's disease. J Immunol. 2005 Jun 15;174(12):8173-82.

Wang K, Baldassano R, Zhang H, Qu HQ, Imielinski M, Kugathasan S, Annese V, Dubinsky M, Rotter JI, Russell RK, Bradfield JP, Sleiman PM, Glessner JT, Walters T, Hou C, Kim C, Frackelton EC, Garris M, Doran J, Romano C, Catassi C, Van Limbergen J, Guthery SL, Denson L, Piccoli D, Silverberg MS, Stanley CA, Monos D, Wilson DC, Griffiths A, Grant SF, Satsangi J, Polychronakos C, Hakonarson H. Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects. Hum Mol Genet. 2010 May 15;19(10):2059-67. Epub 2010 Feb 22.

Wang Y, Zhu M, Yu P, Fu YX. Promoting immune responses by LIGHT in the face of abundant regulatory T cell inhibition. J Immunol. 2010 Feb 1;184(3):1589-95. Epub 2009 Dec 30.

Wapenaar MC, Monsuur AJ, Poell J et al: The SPINK gene family and coeliac disease susceptibility. Immunogenetics 2007; 59: 349– 357.

Wieser H. Chemistry of gluten proteins. Food Microbiol 2007; 24: 115–19.

Wieser H. The precipitating factor in coeliac disease. Baillieres Clin Gastroenterol 1995; 9: 191–207.

Wildenberg ME, van Helden-Meeuwsen CG, van de Merwe JP, Moreno C, Drexhage HA, Versnel MA. Lack of CCR5 on dendritic cells promotes a proinflammatory environment in submandibular glands of the NOD mouse. J Leukoc Biol. 2008 May;83(5):1194-200. Epub 2008 Feb 19.

Willer, C. J., D. A. Dyment, et al. (2003). "Twin concordance and sibling recurrence rates in multiple sclerosis." Proc Natl Acad Sci USA 100: 12877-12882.

Willment JA, Gordon S, Brown GD. Characterization of the human beta -glucan receptor and its alternatively spliced isoforms. J Biol Chem. 2001 Nov 23;276(47):43818-23. Epub 2001 Sep 20.

Wysocki, C.A. Panoskaltsis-Mortari, A. Blazar, B.R. Serody, J.S. (2005). Leukocyte migration and graft-versus-host disease. Blood 105, 4191–4199.

Yan F, Li W, Jono H, Li Q, Zhang S, Li JD, Shen H. Reactive oxygen species regulate Pseudomonas aeruginosa lipopolysaccharide-induced MUC5AC mucin expression via PKC-NADPH oxidase-ROS-TGF-alpha signaling pathways in human airway epithelial cells. Biochem Biophys Res Commun. 2008 Feb 8;366(2):513-9. Epub 2007 Dec 10.

Yarilin, A. A., and I. M. Belyakov. 2004. Cytokines in the thymus: production and biological effects. Curr. Med. Chem. 11:447.

Yeşilkaya E, Koç A, Bideci A, Camurdan O, Boyraz M, Erkal O, Ergun MA, Cinaz P. CTLA4 gene polymorphisms in children and adolescents with autoimmune thyroid diseases. Genet Test. 2008 Sep;12(3):461-4.

Yuhong Zhou, Takao Kurihara, Rolf-Peter Ryseck, Yi Yang, Carol Ryan, James Loy, Glenn Warr and Rodrigo Bravo1, Impaired Macrophage Function and Enhanced T Cell-Dependent Immune Response in Mice Lacking CCR5, the Mouse Homologue of the Major HIV-1 Coreceptor . The Journal of Immunology, *1998, 160: 4018-4025.*

Zhernakova A, Eerligh P, Barrera P et al: CTLA4 is differentially associated with autoimmune diseases in the Dutch population. Hum Genet 2005; 118: 58–66.

Zhernakova A, Elbers CC, Ferwerda B, Romanos J, Trynka G, Dubois PC, de Kovel CG, Franke L, Oosting M, Barisani D, Bardella MT; Finnish Celiac Disease Study Group, Joosten LA, Saavalainen P, van Heel DA, Catassi C, Netea MG, Wijmenga C. Evolutionary and functional analysis of celiac risk loci reveals SH2B3 as a protective factor against bacterial infection. Am J Hum Genet. 2010 Jun 11;86(6):970-7.

Zhernakova A, Festen EM, Franke L, Trynka G, van Diemen CC, Monsuur AJ,

Bevova M, Nijmeijer RM, van 't Slot R, Heijmans R, Boezen HM, van Heel DA, van Bodegraven AA, Stokkers PC, Wijmenga C, Crusius JB, Weersma RK. Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring CARD9 and IL18RAP. Am J Hum Genet. 2008 May;82(5):1202-10. Epub 2008 Apr 24.

Zlotogora J. High frequencies of human genetic diseases: founder effect with genetic drift or selection? Am J Med Genet. 1994 Jan 1;49(1):10-3.

Zou GM, Hu WY. LIGHT regulates CD86 expression on dendritic cells through NF-kappaB, but not JNK/AP-1 signal transduction pathway. J Cell Physiol. 2005 Dec;205(3):437-43.

# APPENDIX 1: GWAS follow-up population frequency tables

This appendix contains supplementary tables referred to at the end of Chapter 5.

**Table 5.a1 Significant Irish SNPs in other populations.**

| Gene | Chr | SNP | P GWAS | P UK2 | P Dutch |
|---|---|---|---|---|---|
| SLC15A2 | 3 | rs2141615 | 0.007733 | 0.6388 | 0.8445 |
| TGFA | 2 | rs377122 | NA | 0.815 | 0.4234 |
| TGFA | 2 | rs432203 | NA | 0.7626 | 0.4121 |
| HMHB1 | 5 | rs325238 | 0.004223 | 0.5595 | 0.5307 |
| TNFSF14 | 19 | rs2279627 | NA | 0.6294 | 0.6601 |
| MACROD2 | 20 | rs6135590 | 0.003319 | 0.718 | 0.4713 |
| LOC401014 | 2 | rs4662248 | 0.00287 | 0.2335 | 0.4179 |
| ACVR2A | 2 | rs1385363 | 0.003185 | 0.2431 | 0.4906 |
| TNFAIP3 | 6 | rs2327832 | 0.002934 | 0.04865 | 0.1682 |
| VAV2 | 9 | rs2797825 | NA | 0.3657 | 0.002348 |
| loc235965 | 7 | rs11760312 | 0.002846 | 0.8527 | 0.6319 |
| TRIM60 | 4 | rs7684974 | 0.003674 | 0.6288 | 0.3964 |
| DAO | 12 | rs3825251 | 0.003096 | 0.8446 | 0.2132 |
| TMEM132B | 12 | rs10083119 | 0.002675 | 0.6409 | 0.3504 |

**Table 5.a2 Most Significant SNPs overall in each population.**

| Gene | Chr | SNP | P CHM | UK2 | P in Irish | Dutch |
|---|---|---|---|---|---|---|
| IL-18RAP | 2 | rs7559479 | 1.266e-006 | 0.003857 | 0.4862 | 2.746e-006 |
| IL-18RAP | 2 | rs3817465 | 3.304e-006 | 0.002849 | 0.8642 | 2.758e-006 |
| TNFAIP3 | 6 | rs2327832 | 6.607e-006 | 0.04865 | 0.005157 | 0.1682 |
| REL | 2 | rs842639 | 8.005e-006 | 0.1111 | 0.04439 | 0.00164 |
| REL | 2 | rs842647 | 9.256e-006 | 0.09767 | 0.03128 | 0.002142 |
|  | 2 | rs4439920 | 1.184e-005 | 0.01939 | 0.06171 | 0.3011 |
| CD80 | 3 | rs1599796 | 4.07e-005 | 0.09609 | 0.33 | 0.01292 |
| ITGA4 | 2 | rs4324288 | 4.474e-005 | 0.04485 | 0.1902 | 0.04367 |
| MyoIXb | 19 | rs2279001 | 4.599e-005 | 0.1175 | 0.02181 | 0.0002804 |
|  | 2 | rs1160542 | 6.684e-005 | 0.2999 | 0.2377 | 0.003405 |
| TAGAP | 6 | rs212402 | 8.101e-005 | 0.2056 | 0.199 | 0.0214 |
|  | 13 | rs9533935 | 9.132e-005 | 0.1452 | 0.4862 | 0.08694 |
|  | 2 | rs4851274 | 9.762e-005 | 0.5138 | 0.8642 | 0.0003707 |

**Table 5.a3 Allele Frequencies of SNPs significant Irish SNPs in control populations .**

| Gene | Irish Control | UK Control | Dutch Control |
|---|---|---|---|
| SLC15A2 | 0.4791 | 0.4696 | 0.4728 |
| TGFA | 0.4188 | 0.4388 | 0.4611 |
| TGFA | 0.4186 | 0.4391 | 0.4594 |
| HMHB1 | 0.3745 | 0.3769 | 0.3733 |
| TNFSF14 | 0.2565 | 0.2369 | 0.2581 |
| MACROD2 | 0.0849 | 0.07844 | 0.09167 |
| LOC401014 | 0.4161 | 0.4353 | 0.4444 |
| ACVR2A | 0.4166 | 0.4363 | 0.4433 |
| TNFAIP3 | 0.2078 | 0.229 | 0.2069 |
| VAV2 | 0.5104 | 0.4757 | 0.4717 |
| loc235965 | 0.0849 | 0.1045 | 0.1268 |
| TRIM60 | 0.3391 | 0.353 | 0.3661 |
| DAO | 0.1557 | 0.151 | 0.1367 |
| TMEM132B | 0.1104 | 0.118 | 0.1128 |

**Table 5.a4 Allele Frequencies of SNPs significant Irish SNPs in Coeliac populations .**

| Gene | Irish Case | UK Case | Dutch Case |
|---|---|---|---|
| SLC15A2 | 0.4043 | 0.462 | 0.4766 |
| TGFA | 0.4853 | 0.435 | 0.4455 |
| TGFA | 0.483 | 0.4342 | 0.4435 |
| HMHB1 | 0.3142 | 0.3677 | 0.3852 |
| TNFSF14 | 0.3096 | 0.2303 | 0.2656 |
| MACROD2 | 0.0540 | 0.0816 | 0.08366 |
| LOC401014 | 0.360 | 0.4544 | 0.4287 |
| ACVR2A | 0.3609 | 0.4551 | 0.43 |
| TNFAIP3 | 0.2551 | 0.2562 | 0.229 |
| VAV2 | 0.457 | 0.4903 | 0.5311 |
| loc235965 | 0.116 | 0.1064 | 0.1206 |
| TRIM60 | 0.2894 | 0.3605 | 0.3502 |
| DAO | 0.1185 | 0.1533 | 0.1537 |
| TMEM132B | 0.07883 | 0.1229 | 0.1245 |

**Table 5.a5 Allele Frequencies of SNPs significant via CHM Population in Control populations.**

| Gene | UK Control | Irish Control | Dutch Control |
|---|---|---|---|
| IL-18RAP | 0.2071 | 0.2052 | 0.2133 |
| IL-18RAP | 0.2046 | 0.201 | 0.2125 |
| TNFAIP3 | 0.229 | 0.2078 | 0.2069 |
| REL | 0.346 | 0.3733 | 0.3537 |
| REL | 0.3404 | 0.3615 | 0.3522 |
| | 0.4134 | 0.425 | 0.436 |
| CD80 | 0.1806 | 0.1839 | 0.1885 |
| ITGA4 | 0.4896 | 0.4984 | 0.48 |
| MyoIXb | 0.3935 | 0.4028 | 0.3774 |
| | 0.4639 | 0.4525 | 0.4333 |
| TAGAP | 0.3649 | 0.3809 | 0.3343 |
| | 0.4235 | 0.4177 | 0.4038 |
| | 0.323 | 0.3203 | 0.2856 |

# APPENDIX ii: References for Tables 6.1 and 6.2

## Tabled 6.1 and 6.2 References


64) J Immunol. 2007 Nov 1;179(9):5886-96.
Essential role of IL-21 in B cell activation, expansion, and plasma cell generation during CD4+ T cell-B cell collaboration.

Kuchen S, Robbins R, Sims GP, Sheng C, Phillips TM, Lipsky PE, Ettinger R.

65) Eur J Immunol. 1996 Feb;26(2):294-9.
Interleukin-2 receptor common gamma-chain signaling cytokines regulate activated T cell apoptosis in response to growth factor withdrawal: selective induction of anti-apoptotic (bcl-2, bcl-xL) but not pro-apoptotic (bax, bcl-xS) gene expression.

Akbar AN, Borthwick NJ, Wickremasinghe RG, Panayoitidis P, Pilling D, Bofill M, Krajewski S, Reed JC, Salmon M.

66) Trends Immunol. 2006 Mar;27(3):109-11. Epub 2006 Feb 3.
The dichotomous role of IL-2: tolerance versus immunity.

Fehérvari Z, Yamaguchi T, Sakaguchi S.

67) J Clin Immunol. 1987 Jul;7(4):277-87.
Interleukin-2 effects on human B cells activated in vivo.

Miyawaki T, Suzuki T, Butler JL, Cooper MD.

68) J Immunol. 2005 Mar 1;174(5):2796-804.
Synergistic effect of IL-2, IL-12, and IL-18 on thymocyte apoptosis and Th1/Th2 cytokine expression.

Rodriguez-Galán MC, Bream JH, Farr A, Young HA.

69) Gastroenterology. 1999 Nov;117(5):1078-88.
Anti-interleukin 12 treatment regulates apoptosis of Th1 T cells in experimental colitis in mice.
Fuss IJ, Marth T, Neurath MF, Pearlstein GR, Jain A, Strober W.

70) J Immunol. 1994 Mar 15;152(6):2729-35.
IL-12 influences intrathymic T cell development.

Godfrey DI, Kennedy J, Gately MK, Hakimi J, Hubbard BR, Zlotnik A.

71) Inflamm Allergy Drug Targets. 2009 Mar;8(1):40-52.
The IL-12 family of cytokines in infection, inflammation and autoimmune disorders.

Gee K, Guzzo C, Che Mat NF, Ma W, Kumar A.

72) J Immunol. 2009 Oct 1;183(7):4537-44. Epub 2009 Sep 4.
SOCS-1 binding to tyrosine 441 of IFN-gamma receptor subunit 1 contributes to the attenuation of IFN-gamma 224eliac224ing in vivo. Starr R, Fuchsberger M, Lau LS, Uldrich AP, Goradia A, Willson TA, Verhagen AM, Alexander WS, Smyth MJ.

73) J Immunol. 2005 Mar 1;174(5):2796-804.
Synergistic effect of IL-2, IL-12, and IL-18 on thymocyte apoptosis and Th1/Th2 cytokine expression. Rodriguez-Galán MC, Bream JH, Farr A, Young HA

74) Clin Immunol. 2006 Oct;121(1):13-8. Epub 2006 Jun 21. Regulation of T:B cell interactions by the inducible costimulator molecule: does ICOS "induce" disease? Shilling RA, Bandukwala HS, Sperling AI.

75) J Immunol. 2009 Mar 15;182(6):3349-56. The role of ICOS in directing T cell responses: ICOS-dependent induction of T cell anergy by tolerogenic dendritic cells. Tuettenberg A, Huter E, Hubo M, Horn J, Knop J, Grimbacher B, Kroczek RA, Stoll S, Jonuleit H.

75) Starr, T. K., Jameson, S. C. and Hogquist, K. A., Positive and negative selection of T cells. Annu. Rev. Immunol. 2003. 21: 139–176.

76) EMBO J. 2002 Mar 15;21(6):1379-88.
Promoter-specific functions of CIITA and the MHC class II enhanceosome in transcriptional activation. Masternak K, Reith W.

77) J Biol Chem. 2006 Jul 21;281(29):20148-59. Epub 2006 Apr 27.
The adaptor molecule Lnk negatively regulates tumor necrosis factor-alpha-dependent VCAM-1 expression in endothelial cells through inhibition of the ERK1 and -2 pathways.
Fitau J, Boulday G, Coulon F, Quillard T, Charreau B.

78) Liston, A., Lesage, S., Wilson, J., Peltonen, L. and Goodnow,C. C., Aire regulates negative selection of organ-specific T cells. Nat. Immunol. 2003. 4: 350–354.

79) Rudd, C. E. (2001) Lnk adaptor: novel negative regulator of B cell lymphopoiesis Sci. STKE **2001**,PE1[Medline]

Anderson, M. S., Venanzi, E. S., Klein, L., Chen, Z., Berzins, S. P., Turley, S. J., von Boehmer, H., Bronson, R., Dierich, A., Benoist, C. and Mathis, D., Projection of an immunological self shadow within the thymus by the AIRE protein. Science 2002. 298: 1395–1401.

80) Am J Physiol Cell Physiol. 2010 Mar;298(3):C656-64. Epub 2009 Dec 23.
Ca2+ influx via TRPC channels induces NF-kappaB-dependent A20 expression to prevent thrombin-induced apoptosis in endothelial cells. Thippegowda PB, Singh V, Sundivakkam PC, Xue J, Malik AB, Tiruppathi C.

81) Gut. 2009 Aug;58(8):1078-83. Epub 2009 Feb 24.
Coeliac disease-associated risk variants in TNFAIP3 and REL implicate altered NF-kappaB signalling. Trynka G, Zhernakova A, Romanos J, Franke L, Hunt KA, Turner G, Bruinenberg M, Heap GA, Platteel M, Ryan AW, de Kovel C, Holmes GK, Howdle PD, Walters JR, Sanders DS, Mulder CJ, Mearin ML, Verbeek WH, Trimble V, Stevens FM, Kelleher D, Barisani D, Bardella MT, McManus R, van Heel DA, Wijmenga C.

82) The CD160, BTLA, LIGHT/HVEM pathway: a bidirectional switch regulating T-cell activation. Review articles Immunological Reviews. 229(1):244-258, May 2009.
Cai, Guifang 1; Freeman, Gordon J. 1

83) J Immunol. 2005 Mar 1;174(5):2627-36. Overexpression of the Runx3 transcription factor increases the proportion of mature thymocytes of the CD8 single-positive lineage. Kohu K, Sato T, Ohno S, Hayashi K, Uchino R, Abe N, Nakazato M, Yoshida N, Kikuchi T, Iwakura Y, Inoue Y, Watanabe T, Habu S, Satake M.

84) Leukemia. 2008 Dec;22(12):2201-7. Epub 2008 Sep 4. Regulation of IL-2 expression by transcription factor BACH2 in umbilical cord blood CD4+ T cells. Lesniewski ML, Haviernik P, Weitzel RP, Kadereit S, Kozik MM, Fanning LR, Yang YC, Hegerfeldt Y, Finney MR, Ratajczak MZ, Greco N, Paul P, Maciejewski J, Laughlin MJ.

85) Genomics. 2004 Jun;83(6):989-99. T lymphocyte activation gene identification by coregulated expression on DNA microarrays. Mao M, Biery MC, Kobayashi SV, Ward T, Schimmack G, Burchard J, Schelter JM, Dai H, He YD, Linsley PS.

86) DNA Cell Biol. 2002 Mar;21(3):137-49. CD80 (B7-1) and CD86 (B7-2) are functionally equivalent in the initiation and maintenance of CD4+ T-cell proliferation after activation with suboptimal doses of PHA. Vasilevko V, Ghochikyan A, Holterman MJ, Agadjanyan MG.

87) Int Immunol. 2008 Dec;20(12):1507-15. Epub 2008 Oct 1.
CD28 stimulation triggers NF-kappaB activation through the CARMA1-PKCtheta-Grb2/Gads axis. Takeda K, Harada Y, Watanabe R, Inutake Y, Ogawa S, Onuki K, Kagaya S, Tanabe K, Kishimoto H, Abe R.

88) J Immunol. 2008 Aug 15;181(4):2855-68.
CD28 costimulation is essential for human T regulatory expansion and function.
Golovina TN, Mikheeva T, Suhoski MM, Aqui NA, Tai VC, Shan X, Liu R, Balcarcel RR, Fisher N, Levine BL, Carroll RG, Warner N, Blazar BR, June CH, Riley JL.

89) J Immunol. 2002 Mar 15;168(6):2729-36.
CD28 costimulation mediates down-regulation of p27kip1 and cell cycle progression by activation of the PI3K/PKB 226eliac226ing pathway in primary human T cells.
Appleman LJ, van Puijenbroek AA, Shu KM, Nadler LM, Boussiotis VA.

90) Blood. 2008 Aug 15;112(4):1175-83. Epub 2008 Jun 3.
CTLA4 blockade expands FoxP3+ regulatory and activated effector CD4+ T cells in a dose-dependent fashion. Kavanagh B, O'Brien S, Lee D, Hou Y, Weinberg V, Rini B, Allison JP, Small EJ, Fong L.

91) EMBO J. 2000 Dec 1;19(23):6351-60.
The anti-apoptotic activities of Rel and RelA required during B-cell maturation involve the regulation of Bcl-2 expression. Grossmann M, O'Reilly LA, Gugasyan R, Strasser A, Adams JM, Gerondakis S.

92) Int Immunol. 1999 Mar;11(3):361-71.
c-Rel is crucial for lymphocyte proliferation but dispensable for T cell effector function.

Liou HC, Jin Z, Tumang J, Andjelic S, Smith KA, Liou ML.

93) FEBS Lett. 2002 Dec 4;532(1-2):61-6.
A nuclear protein tyrosine 226eliac226ing226 activates p53 and induces caspase-1-dependent apoptosis. Gupta S, Radha V, Sudhakar Ch, Swarup G.

94) J Immunol. 2003 Jul 15;171(2):726-32. Gr-1+ myeloid cells lacking T cell protein tyrosine 226eliac226ing226 inhibit lymphocyte proliferation by an IFN-gamma- and nitric oxide-dependent mechanism. Dupuis M, De Jesús Ibarra-Sánchez M, Tremblay ML, Duplay P.

95) Nat Immunol. 2009 Aug;10(8):848-56. Epub 2009 Jul 13.
Themis controls thymocyte selection through regulation of T cell antigen receptor-mediated signaling.

Fu G, Vallée S, Rybakin V, McGuire MV, Ampudia J, Brockmeyer C, Salek M, Fallen PR, Hoerter JA, Munshi A, Huang YH, Hu J, Fox HS, Sauer K, Acuto O, Gascoigne NR.

96) Nat Immunol. 2009 Aug;10(8):840-7. Epub 2009 Jul 13.
Themis, a T cell-specific protein important for late thymocyte development.

Lesourne R, Uehara S, Lee J, Song KD, Li L, Pinkhasov J, Zhang Y, Weng NP, Wildt KF, Wang L, Bosselut R, Love PE.

97) Ets-1 Regulates Plasma Cell Differentiation by Interfering with the Activity of the Transcription Factor Blimp-1.

Received for publication, June 27, 2007, and in revised form, October 9, 2007 Published, JBC Papers in Press, October 31, 2007, DOI 10.1074/jbc.M705262200
Shinu A. John‡, James L. Clements§, Lisa M. Russell‡, and Lee Ann Garrett-Sinha‡1

98) Oncogene. 1997 Jun 12;14(23):2845-55.
ETS1, NfkappaB and AP1 synergistically transactivate the human GM-CSF promoter.
Thomas RS, Tymms MJ, McKinlay LH, Shannon MF, Seth A, Kola I.

99) J Leukoc Biol. 2006 Jun;79(6):1357-68. Epub 2006 Mar 24.
RGS1 and RGS13 mRNA silencing in a human B lymphoma line enhances responsiveness to chemoattractants and impairs desensitization. Han JI, Huang NN, Kim DU, Kehrl JH.

110) Biochem Biophys Res Commun. 2009 Dec 18;390(3):489-93. Epub 2009 Oct 1.
Protein-tyrosine 227eliac227ing227-kappa regulates CD4+ T cell development through ERK1/2-mediated 227eliac227ing. Erdenebayar N, Maekawa Y, Nishida J, Kitamura A, Yasutomo K.

112) Nature. 2001 Jan 4;409(6816):97-101.
ICOS co-stimulatory receptor is essential for T-cell activation and function.

Dong C, Juedes AE, Temann UA, Shresta S, Allison JP, Ruddle NH, Flavell RA.

102) Arthritis Rheum. 2004 Oct;50(10):3211-20.
Involvement of inducible costimulator in the exaggerated memory B cell and plasma cell generation in systemic lupus erythematosus.

Hutloff A, Büchner K, Reiter K, Baelde HJ, Odendahl M, Jacobi A, Dörner T, Kroczek RA.

103) Cell Signal. 2009 Dec;21(12):1748-57. Epub 2009 Jul 28.
Inhibition of dynamin prevents CCL2-mediated endocytosis of CCR2 and activation of ERK1/2.

García Lopez MA, Aguado Martínez A, Lamaze C, Martínez-A C, Fischer T.

104) J Leukoc Biol. 2009 Dec;86(6):1319-29. Epub 2009 Aug 17.
CCR1-mediated activation of Nuclear Factor-kappaB in THP-1 monocytic cells involves Pertussis Toxin-insensitive Galpha(14) and Galpha(16) signaling cascades.
Lee MM, Wong YH.

105) J Immunol. 2003 Aug 15;171(4):1722-31.
CCR3 expression induced by IL-2 and IL-4 functioning as a death receptor for B cells.

Jinquan T, Jacobi HH, Jing C, Millner A, Sten E, Hviid L, Anting L, Ryder LP, Glue C, Skov PS, Jarman E, Lamberth K, Malling HJ, Poulsen LK.

106) Arthritis Rheum. 2010 Feb;62(2):323-9.
Confirmation of an association between rs6822844 at the Il2-Il21 region and multiple autoimmune diseases: evidence of a general susceptibility locus.
Maiti AK, Kim-Howard X, Viswanathan P, Guillén L, Rojas-Villarraga A, Deshmukh H, Direskeneli H, Saruhan-Direskeneli G, Cañas C, Tobön GJ, Sawalha AH, Cherñavsky AC, Anaya JM, Nath SK.

107) Expert Rev Clin Immunol. 2008 Sep;4(5):565-71.
Genome-wide association scan in psoriasis: new insights into chronic inflammatory disease.
Schrodi SJ.

108) Hum Mol Genet. 2009 Nov 1;18(21):4195-203. Epub 2009 Jul 31.
Common and different genetic background for rheumatoid arthritis and 227eliac disease.

Coenen MJ, Trynka G, Heskamp S, Franke B, van Diemen CC, Smolonska J, van Leeuwen M, Brouwer E, Boezen MH, Postma DS, Platteel M, Zanen P, Lammers JW, Groen HJ, Mali WP, Mulder CJ, Tack GJ, Verbeek WH, Wolters VM, Houwen RH, Mearin ML, van Heel DA, Radstake TR, van Riel PL, Wijmenga C, Barrera P, Zhernakova A.

109) Gut. 2009 Jun;58(6):799-804. Epub 2009 Feb 6.
Genetic variants in the region harbouring IL2/IL21 associated with ulcerative colitis.
Festen EA, Goyette P, Scott R, Annese V, Zhernakova A, Lian J, Lefèbvre C, Brant SR, Cho JH, Silverberg MS, Taylor KD, de Jong DJ, Stokkers PC, Mcgovern D, Palmieri O, Achkar JP, Xavier RJ, Daly MJ, Duerr RH, Wijmenga C, Weersma RK, Rioux JD.

110) Eur J Hum Genet. 2006 Feb;14(2):227-35.
Family based association analysis of the IL2 and IL15 genes in allergic disorders.
Christensen U, Haagerup A, Binderup HG, Vestbo J, Kruse TA, Børglum AD.

111) J Rheumatol. 2007 May;34(5):997-1004. Epub 2007 Apr 15.
Association of polymorphisms in the IL1B and IL2 genes with susceptibility and severity of systemic sclerosis. Mattuzzi S, Barbi S, Carletto A, Ravagnani V, Moore PS, Bambara LM, Scarpa A.

112) Eur J Hum Genet. 2010 Feb 24. [Epub ahead of print]
Polymorphisms in the IL2, IL2RA and IL2RB genes in multiple sclerosis risk.
Cavanillas ML, Alcina A, Núñez C, de Las Heras V, Fernández-Arquero M, Bartolomé M, de la Concha EG, Fernández O, Arroyo R, Matesanz F, Urcelay E.

113) Hum Mol Genet. 2009 Nov 1;18(21):4195-203. Epub 2009 Jul 31.
Common and different genetic background for rheumatoid arthritis and 228eliac disease.
Coenen MJ, Trynka G, Heskamp S, Franke B, van Diemen CC, Smolonska J, van Leeuwen M, Brouwer E, Boezen MH, Postma DS, Platteel M, Zanen P, Lammers JW, Groen HJ, Mali WP, Mulder CJ, Tack GJ, Verbeek WH, Wolters VM, Houwen RH, Mearin ML, van Heel DA, Radstake TR, van Riel PL, Wijmenga C, Barrera P, Zhernakova A.

114) Gut. 2009 Jun;58(6):799-804. Epub 2009 Feb 6.
Genetic variants in the region harbouring IL2/IL21 associated with ulcerative colitis.
Festen EA, Goyette P, Scott R, Annese V, Zhernakova A, Lian J, Lefèbvre C, Brant SR, Cho JH, Silverberg MS, Taylor KD, de Jong DJ, Stokkers PC, Mcgovern D, Palmieri O, Achkar JP, Xavier RJ, Daly MJ, Duerr RH, Wijmenga C, Weersma RK, Rioux JD.

115) Int Arch Allergy Immunol. 2009;148(2):137-46. Epub 2008 Sep 19.
A common exonic variant of interleukin21 confers susceptibility to atopic asthma.
Chatterjee R, Batra J, Ghosh B.

116) Ann Rheum Dis. 2008 Apr;67(4):458-61. Epub 2007 Aug 24.
Genetic association of interleukin-21 polymorphisms with systemic lupus erythematosus.
Sawalha AH, Kaufman KM, Kelly JA, Adler AJ, Aberle T, Kilpatrick J, Wakeland EK, Li QZ, Wandstrat AE, Karp DR, James JA, Merrill JT, Lipsky P, Harley JB.

117) J Clin Immunol. 2009 Nov;29(6):747-51. Epub 2009 Jun 20.
IL-2, IFN-gamma, and IL-12 gene polymorphisms and susceptibility to multiple sclerosis.
Shokrgozar MA, Sarial S, Amirzargar A, Shokri F, Rezaei N, Arjang Z, Radfar J, Yousefi-Behzadi M, Ali Sahraian M, Lotfi J.

118) Eur J Neurol. 2008 Oct;15(10):1100-5. Epub 2008 Aug 20.
The role of IL-12 and TNF-alpha in AIDP and AMAN.

Deng H, Yang X, Jin T, Wu J, Hu LS, Chang M, Sun XJ, Adem A, Winblad B, Zhu J.

119) Acta Neurol Scand. 2009 May;119(5):332-7. Epub 2008 Oct 23.
Circulating interleukin-10 and interleukin-12 in Parkinson's disease. Rentzos M, Nikolaou C, Andreadou E, Paraskevas GP, Rombos A, Zoga M, Tsoutsou A, Boufidou F, Kapaki E, Vassilopoulos D.

120) Arthritis Rheum. 2007 Dec;56(12):3977-88.
Rates of infiltration by macrophages and dendritic cells and expression of interleukin-18 and interleukin-12 in the chronic inflammatory lesions of Sjögren's syndrome: correlation with certain features of immune hyperactivity and factors associated with high risk of lymphoma development. Manoussakis MN, Boiu S, Korkolopoulou P, Kapsogeorgou EK, Kavantzas N, Ziakas P, Patsouris E, Moutsopoulos HM.

121) J Allergy Clin Immunol. 2004 Sep;114(3):561-7.
A TLR4 polymorphism is associated with asthma and reduced lipopolysaccharide-induced interleukin-12(p70) responses in Swedish children. Fagerås Böttcher M, Hmani-Aifa M, Lindström A, Jenmalm MC, Mai XM, Nilsson L, Zdolsek HA, Björkstén B, Söderkvist P, Vaarala O.

122) Mod Rheumatol. 2007;17(3):220-3. Epub 2007 Jun 20.
Dysregulated expression of interleukin-23 and interleukin-12 subunits in systemic lupus erythematosus patients. Huang X, Hua J, Shen N, Chen S.

123) World J Gastroenterol. 2006 Sep 21;12(35):5606-10.
Interleukin-12 and Th1 immune response in Crohn's disease: pathogenetic relevance and therapeutic implication. Peluso I, Pallone F, Monteleone G.

124) Am J Hum Genet. 2008 May;82(5):1202-10. Epub 2008 Apr 24.
Genetic analysis of innate immunity in Crohn's disease and ulcerative colitis identifies two susceptibility loci harboring CARD9 and IL18RAP.

Zhernakova A, Festen EM, Franke L, Trynka G, van Diemen CC, Monsuur AJ, Bevova M, Nijmeijer RM, van 't Slot R, Heijmans R, Boezen HM, van Heel DA, van Bodegraven AA, Stokkers PC, Wijmenga C, Crusius JB, Weersma RK.

125) J Allergy Clin Immunol. 2008 Sep;122(3):651-4.e8.
Association of IL1RL1, IL18R1, and IL18RAP gene cluster polymorphisms with asthma and atopy.

Reijmerink NE, Postma DS, Bruinenberg M, Nolte IM, Meyers DA, Bleecker ER, Koppelman GH.

126) Arthritis Res Ther. 2006;8(3):R62. Epub 2006 Mar 22.
Expression and function of inducible co-stimulator in patients with systemic lupus erythematosus: possible involvement in excessive interferon-gamma and anti-double-stranded DNA antibody production. Kawamoto M, Harigai M, Hara M, Kawaguchi Y, Tezuka K, Tanaka M, Sugiura T, Katsumata Y, Fukasawa C, Ichida H, Higami S, Kamatani N.

127) J Neuroimmunol. 2007 May;186(1-2):193-8. Epub 2007 May 3.
ICOS gene haplotypes correlate with IL10 secretion and multiple sclerosis evolution.

Castelli L, Comi C, Chiocchetti A, Nicola S, Mesturini R, Giordano M, D'Alfonso S, Cerutti E, Galimberti D, Fenoglio C, Tesser F, Yagi J, Rojo JM, Perla F, Leone M, Scarpini E, Monaco F, Dianzani U.

128) Arthritis Rheum. 2009 Apr;60(4):931-7.
CTLA4/ICOS gene variants and haplotypes are associated with rheumatoid arthritis and primary biliary cirrhosis in the Canadian population. Walker EJ, Hirschfield GM, Xu C, Lu Y, Liu X, Lu Y, Coltescu C, Wang K, Newman WG, Bykerk V, Keystone EC, Mosher D, Amos CI, Heathcote EJ, Siminovitch KA.

129) Gastroenterology. 2004 Mar;126(3):829-39.
Hyperexpression of inducible costimulator and its contribution on lamina propria T cells in inflammatory bowel disease. Sato T, Kanai T, Watanabe M, Sakuraba A, Okamoto S, Nakai T, Okazawa A, Inoue N, Totsuka T, Yamazaki M, Kroczek RA, Fukushima T, Ishii H, Hibi T.

130) Clin Rheumatol. 2001;20(6):420-2.

Polymorphsims of CTLA-4 exon 1 +49, CTLA-4 promoter -318 and Fas promoter -670 in spondyloarthropathies. Lee YH, Ji JD, Sohn J, Song GG.

131) Neurology. 2007 Jan 23;68(4):277-82.
Expression of CD28-related costimulatory molecule and its ligand in inflammatory neuropathies. Hu W, Janke A, Ortler S, Hartung HP, Leder C, Kieseier BC, Wiendl H.

132) J Clin Endocrinol Metab. 2008 Sep;93(9):3310-7. Epub 2008 Jul 1.
Polymorphisms in CLEC16A and CIITA at 16p13 are associated with primary adrenal insufficiency. Skinningsrud B, Husebye ES, Pearce SH, McDonald DO, Brandal K, Wolff AB, Løvås K, Egeland T, Undlien DE.

133) Hum Mol Genet. 2010 Jun 1;19(11):2331-40. Epub 2010 Mar 8.
CIITA variation in the presence of HLA-DRB1*1501 increases risk for multiple sclerosis. Bronson PG, Caillier S, Ramsay PP, McCauley JL, Zuvich RL, De Jager PL, Rioux JD, Ivinson AJ, Compston A, Hafler DA, Sawcer SJ, Pericak-Vance MA, Haines JL; International Multiple Sclerosis Genetics Consortium, Hauser SL, Oksenberg JR, Barcellos LF.

134) N Engl J Med. 2008 Dec 25;359(26):2767-77. Epub 2008 Dec 10.
Shared and distinct genetic variants in type 1 diabetes and celiac disease. Smyth DJ, Plagnol V, Walker NM, Cooper JD, Downes K, Yang JH, Howson JM, Stevens H, McManus R, Wijmenga C, Heap GA, Dubois PC, Clayton DG, Hunt KA, van Heel DA, Todd JA.

135) Nat Genet. 2009 Nov;41(11):1228-33. Epub 2009 Oct 18.
A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, Ortmann W, Kosoy R, Ferreira RC, Nordmark G, Gunnarsson I, Svenungsson E, Padyukov L, Sturfelt G, Jönsen A, Bengtsson AA, Rantapää-Dahlqvist S, Baechler EC, Brown EE, Alarcón GS, Edberg JC, Ramsey-Goldman R, McGwin G Jr, Reveille JD, Vilá LM, Kimberly RP, Manzi S, Petri MA, Lee A, Gregersen PK, Seldin MF, Rönnblom L, Criswell LA, Syvänen AC, Behrens TW, Graham RR.

136) Ann Rheum Dis. 2010 Jun;69(6):1077-85. Epub 2010 Mar 16.
Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk. Karlson EW, Chibnik LB, Kraft P, Cui J, Keenan BT, Ding B, Raychaudhuri S, Klareskog L, Alfredsson L, Plenge RM.

137) at Genet. 2008 Sep;40(9):1059-61.
Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM, Burtt NP, Guiducci C, Parkin M, Gates C, Plenge RM, Behrens TW, Wither JE, Rioux JD, Fortin PR, Graham DC, Wong AK, Vyse TJ, Daly MJ, Altshuler D, Moser KL, Gaffney PM.

138) Lancet Neurol. 2009 Dec;8(12):1111-9. Epub 2009 Oct 29.
Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score.De Jager PL, Chibnik LB, Cui J, Reischl J, Lehr S, Simon KC, Aubin C, Bauer D, Heubach JF, Sandbrink R, Tyblova M, Lelkova P; Steering committee of the BENEFIT study; Steering committee of the BEYOND study; Steering committee of the LTF study; Steering committee of the CCR1 study, Havrdova E, Pohl C, Horakova D, Ascherio A, Hafler DA, Karlson EW.

139) Arthritis Rheum. 2009 Jul;60(7):2124-30.
Variants in TNFAIP3, STAT4, and C12orf30 loci associated with multiple autoimmune diseases are also associated with juvenile idiopathic arthritis. Prahalad S, Hansen S, Whiting A, Guthery SL, Clifford B, McNally B, Zeft AS, Bohnsack JF, Jorde LB.

140) Nat Genet. 2009 Feb;41(2):199-204. Epub 2009 Jan 25.
Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways.

Nair RP, Duffin KC, Helms C, Ding J, Stuart PE, Goldgar D, Gudjonsson JE, Li Y, Tejasvi T, Feng BJ, Ruether A, Schreiber S, Weichenthal M, Gladman D, Rahman P, Schrodi SJ, Prahalad S, Guthery SL, Fischer J, Liao W, Kwok PY, Menter A, Lathrop GM, Wise CA,

Begovich AB, Voorhees JJ, Elder JT, Krueger GG, Bowcock AM, Abecasis GR; Collaborative Association Study of Psoriasis.

141) Hum Mol Genet. 2010 May 15;19(10):2059-67. Epub 2010 Feb 22.
Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects.

Wang K, Baldassano R, Zhang H, Qu HQ, Imielinski M, Kugathasan S, Annese V, Dubinsky M, Rotter JI, Russell RK, Bradfield JP, Sleiman PM, Glessner JT, Walters T, Hou C, Kim C, Frackelton EC, Garris M, Doran J, Romano C, Catassi C, Van Limbergen J, Guthery SL, Denson L, Piccoli D, Silverberg MS, Stanley CA, Monos D, Wilson DC, Griffiths A, Grant SF, Satsangi J, Polychronakos C, Hakonarson H.

142) Arthritis Rheum. 2010 Mar;62(3):705-10.
Evidence of epistasis between TNFRSF14 and TNFRSF6B polymorphisms in patients with rheumatoid arthritis.

Perdigones N, Vigo AG, Lamas JR, Martínez A, Balsa A, Pascual-Salcedo D, de la Concha EG, Fernández-Gutiérrez B, Urcelay E.

143) Clin Exp Rheumatol. 2009 Mar-Apr;27(2):260-6.
Changes in the gene expression of peripheral blood mononuclear cells during the menstrual cycle of females is associated with a gender bias in the incidence of systemic lupus erythematosus. Kawasaki M, Sekigawa I, Nozawa K, Kaneko H, Takasaki Y, Takamori K, Ogawa H.

144) Exp Mol Med. 2003 Dec 31;35(6):501-8.
High levels of soluble herpes virus entry mediator in sera of patients with allergic and autoimmune diseases.

Jung HW, La SJ, Kim JY, Heo SK, Kim JY, Wang S, Kim KK, Lee KM, Cho HR, Lee HW, Kwon B, Kim BS, Kwon BS.

145) Genes Immun. 2009 Mar;10(2):188-91. Epub 2008 Dec 25.
Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus.

Fung EY, Smyth DJ, Howson JM, Cooper JD, Walker NM, Stevens H, Wicker LS, Todd JA.

146) Nat Genet. 2009 Dec;41(12):1313-8. Epub 2009 Nov 8.
Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk. Raychaudhuri S, Thomson BP, Remmers EF, Eyre S, Hinks A, Guiducci C, Catanese JJ, Xie G, Stahl EA, Chen R, Alfredsson L, Amos CI, Ardlie KG; BIRAC Consortium, Barton A, Bowes J, Burtt NP, Chang M, Coblyn J, Costenbader KH, Criswell LA, Crusius JB, Cui J, De Jager PL, Ding B, Emery P, Flynn E, Harrison P, Hocking LJ, Huizinga TW, Kastner DL, Ke X, Kurreeman FA, Lee AT, Liu X, Li Y, Martin P, Morgan AW, Padyukov L, Reid DM, Seielstad M, Seldin MF, Shadick NA, Steer S, Tak PP, Thomson W, van der Helm-van Mil AH, van der Horst-Bruinsma IE, Weinblatt ME, Wilson AG, Wolbink GJ, Wordsworth P; YEAR Consortium, Altshuler D, Karlson EW, Toes RE, de Vries N, Begovich AB, Siminovitch KA, Worthington J, Klareskog L, Gregersen PK, Daly MJ, Plenge RM.

147) Dig Dis Sci. 2004 Nov-Dec;49(11-12):1738-44.
The role of costimulatory molecules CD80 and CD86 and IFNgamma in the pathogenesis of ulcerative colitis. Scarpa M, Behboo R, Angriman I, Termini B, Barollo M, Ruffolo C, Polese L, D'Incà R, Sturniolo GC, D'Amico DF.

148) Lupus. 2008 Jul;17(7):630-7.
CD8+CD28-, suppressive T cells in systemic lupus erythematosus. Tulunay A, Yavuz S, Direskeneli H, Eksioglu-Demiralp E.

149) J Neuroimmunol. 2004 Mar;148(1-2):218-30.

Association of common T cell activation gene polymorphisms with multiple sclerosis in Australian patients.

Teutsch SM, Booth DR, Bennetts BH, Heard RN, Stewart GJ.

150) Int J Immunogenet. 2007 Feb;34(1):45-9.
Analysis of CD28 and CTLA-4 gene polymorphisms in Turkish patients with Behcet's disease. Gunesacar R, Erken E, Bozkurt B, Ozer HT, Dinkci S, Erken EG, Ozbalkan Z.

151) Hum Immunol. 2009 Jul;70(7):536-9. Epub 2009 Apr 17.
Association of CTLA4 but not ICOS polymorphisms with type 1 diabetes in two populations with different disease rates. Douroudis K, Laine AP, Heinonen M, Hermann R, Lipponen K, Veijola R, Simell O, Knip M, Uibo R, Ilonen J, Kisand K.

152) Genet Test. 2008 Sep;12(3):461-4.
CTLA4 gene polymorphisms in children and adolescents with autoimmune thyroid diseases.

Yeşilkaya E, Koç A, Bideci A, Camurdan O, Boyraz M, Erkal O, Ergun MA, Cinaz P.

153) Int J Immunogenet. 2009 Aug;36(4):245-50.
CTLA-4 gene polymorphism of exon 1(+49 A/G) in Turkish systemic lupus erythematosus patients. Ulker M, Yazisiz V, Sallakci N, Avci AB, Sanlioglu S, Yegin O, Terzioglu E.

154) Arthritis Rheum. 2009 May;60(5):1255-60.
Confirmation of STAT4, IL2/IL21, and CTLA4 polymorphisms in rheumatoid arthritis. Daha NA, Kurreeman FA, Marques RB, Stoeken-Rijsbergen G, Verduijn W, Huizinga TW, Toes RE.

155) Tissue Antigens. 2009 Sep;74(3):222-7. Epub 2009 Jun 25.
CTLA4 gene polymorphisms and soluble CTLA4 protein in Behcet's disease. Park KS, Baek JA, Do JE, Bang D, Lee ES.

156) Ann Rheum Dis. 2010 Jun 4. [Epub ahead of print]
Confirmation of association of the REL locus with rheumatoid arthritis susceptibility in the UK population. Eyre S, Hinks A, Flynn E, Martin P, Wilson AG, Maxwell JR, Morgan AW, Emery P, Steer S, Hocking LJ, Reid DM, Harrison P, Wordsworth P, Thomson W, Worthington J, Barton A.

157) Exp Dermatol. 2005 Jan;14(1):17-25.
Deficient translocation of c-Rel is associated with impaired Th1 cytokine production in T cells from atopic dermatitis patients. Dieckhoff K, Graf P, Beinhauer B, Schwaerzler C, Carballido JM, Neumann C, Zachmann K, Jung T.

158) Nat Genet. 2008 Jun;40(6):713-5. Epub 2008 Apr 27.Replication of signals from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis.Franke A, Balschun T, Karlsen TH, Hedderich J, May S, Lu T, Schuldt D, Nikolaus S, Rosenstiel P, Krawczak M, Schreiber S.

159) PLoS Genet. 2010 Feb 12;6(2):e1000841.
Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. Yang W, Shen N, Ye DQ, Liu Q, Zhang Y, Qian XX, Hirankarn N, Ying D, Pan HF, Mok CC, Chan TM, Wong RW, Lee KW, Mok MY, Wong SN, Leung AM, Li XP, Avihingsanon Y, Wong CM, Lee TL, Ho MH, Lee PP, Chang YK, Li PH, Li RJ, Zhang L, Wong WH, Ng IO, Lau CS, Sham PC, Lau YL; Asian Lupus Genetics Consortium.

160) Genes Immun. 2009 Jan;10(1):11-4. Epub 2008 Nov 6.
The expanding genetic overlap between multiple sclerosis and type I diabetes. International Multiple Sclerosis Genetics Consortium (IMSGC).

161) Eur J Hum Genet. 2009 Oct;17(10):1304-8. Epub 2009 Apr 1.

Specific association of a CLEC16A/KIAA0350 polymorphism with NOD2/CARD15(-) Crohn's disease patients. Márquez A, Varadé J, Robledo G, Martínez A, Mendoza JL, Taxonera C, Fernández-Arquero M, Díaz-Rubio M, Gómez-García M, López-Nevot MA, de la Concha EG, Martín J, Urcelay E.

162) Lupus. 2009 May;18(6):501-7.
Increased expression of soluble inducible costimulator ligand (ICOSL) in patients with systemic lupus erythematosus. Her M, Kim D, Oh M, Jeong H, Choi I.

163) Nat Genet. 2008 Aug;40(8):955-62. Epub 2008 Jun 29.
Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM, Bitton A, Dassopoulos T, Datta LW, Green T, Griffiths AM, Kistner EO, Murtha MT, Regueiro MD, Rotter JI, Schumm LP, Steinhart AH, Targan SR, Xavier RJ; NIDDK IBD Genetics Consortium, Libioulle C, Sandor C, Lathrop M, Belaiche J, Dewit O, Gut I, Heath S, Laukens D, Mni M, Rutgeerts P, Van Gossum A, Zelenika D, Franchimont D, Hugot JP, de Vos M, Vermeire S, Louis E; Belgian-French IBD Consortium; Wellcome Trust Case Control Consortium, Cardon LR, Anderson CA, Drummond H, Nimmo E, Ahmad T, Prescott NJ, Onnie CM, Fisher SA, Marchini J, Ghori J, Bumpstead S, Gwilliam R, Tremelling M, Deloukas P, Mansfield J, Jewell D, Satsangi J, Mathew CG, Parkes M, Georges M, Daly MJ.

164) Neurology. 2007 Jan 23;68(4):277-82.
Expression of CD28-related costimulatory molecule and its ligand in inflammatory neuropathies. Hu W, Janke A, Ortler S, Hartung HP, Leder C, Kieseier BC, Wiendl H.

165) J Immunol. 2008 Dec 1;181(11):8053-67. Infiltrated neutrophils acquire novel chemokine receptor expression and chemokine responsiveness in chronic inflammatory lung diseases. Hartl D, Krauss-Etschmann S, Koller B, Hordijk PL, Kuijpers TW, Hoffmann F, Hector A, Eber E, Marcos V, Bittmann I, Eickelberg O, Griese M, Roos D.
166) Saudi Med J. 2008 Dec;29(12):1711-8.
Association and preferential transmission of the CCR2V64I polymorphism with absence of asthma in high-risk families. Al-Abdulhadi SA, Al-Rabia MW.

167) Eur J Neurol. 2004 Jul;11(7):445-9.
Chemokine CCL2 and chemokine receptor CCR2 in early active multiple sclerosis.

Sørensen TL, Ransohoff RM, Strieter RM, Sellebjerg F

168) Exp Mol Med. 2007 Aug 31;39(4):499-507.
Expression of CCR2A, an isoform of MCP-1 receptor, is increased by MCP-1, CD40 ligand and TGF-beta in fibroblast like synoviocytes of patients with RA.

Cho ML, Yoon BY, Ju JH, Jung YO, Jhun JY, Park MK, Park SH, Cho CS, Kim HY.

169) Biochem Biophys Res Commun. 2006 Jun 9;344(3):780-5. Epub 2006 Apr 17.
Increased expression levels of monocyte CCR2 and monocyte chemoattractant protein-1 in patients with diabetes mellitus.

Mine S, Okada Y, Tanikawa T, Kawahara C, Tabata T, Tanaka Y.

170) Eur Respir J. 2001 Jan;17(1):59-63.
Genetic polymorphisms of CC chemokine receptor 3 in Japanese and British asthmatics. Fukunaga K, Asano K, Mao XQ, Gao PS, Roberts MH, Oguma T, Shiomi T, Kanazawa M, Adra CN, Shirakawa T, Hopkin JM, Yamaguchi K.

171) Autoimmunity. 2009 May;42(4):260-2.
Role of CCR5 Delta32 bp deletion in RA and SLE.

Martens HA, Kallenberg CG, Bijl M.

172) Scand J Rheumatol. 2008 Jan-Feb;37(1):13-7.
Differential CCR5Delta32 allelic frequencies in juvenile idiopathic arthritis subtypes: evidence for different regulatory roles of CCR5 in rheumatological diseases.

Scheibel I, Veit T, Neves AG, Souza L, Prezzi S, Machado S, Kohem C, Icarelli M, Xavier R, Brenol JC, Chies JA.

173) J Neuroimmunol. 2007 Oct;190(1-2):157-64. Epub 2007 Sep 19.
CCL5 and CCR5 genotypes modify clinical, radiological and pathological features of multiple sclerosis.

van Veen T, Nielsen J, Berkhof J, Barkhof F, Kamphorst W, Bö L, Ravid R, Verweij CL, Huitinga I, Polman CH, Uitdehaag BM.

174) J Neuroimmunol. 2003 Dec;145(1-2):27-39.
Chemokine-induced cell death in CCR5-expressing neuroblastoma cells.

Cartier L, Dubois-Dauphin M, Hartley O, Irminger-Finger I, Krause KH

175) Clin Exp Rheumatol. 2006 Sep-Oct;24(5 Suppl 42):S91-4.
Association of chemokine receptor 5 (CCR5) delta32 mutation with Behçet's disease is dependent on gender in Iranian patients. Mojtahedi Z, Ahmadi SB, Razmkhah M, Azad TK, Rajaee A, Ghaderi A.

176) Clin Exp Immunol. 2002 Mar;127(3):479-85.
Graves' disease is associated with an altered CXCR3 and CCR5 expression in thyroid-derived compared to peripheral blood lymphocytes.

Aust G, Sittig D, Steinert M, Lamesch P, Lohmann T.

177) J Immunol. 2005 Oct 15;175(8):4914-26.
V alpha 24-invariant NKT cells from patients with allergic asthma express CCR9 at high frequency and induce Th2 bias of CD3+ T cells upon CD226 engagement.

Sen Y, Yongyi B, Yuling H, Luokun X, Li H, Jie X, Tao D, Gang Z, Junyan L, Chunsong H, Zhang X, Youxin J, Feili G, Boquan J, Jinquan T.

178) Arthritis Rheum. 2004 Jun;50(6):1806-14.
Up-regulated expression and activation of the orphan chemokine receptor, CCRL2, in rheumatoid arthritis.

Galligan CL, Matsuyama W, Matsukawa A, Mizuta H, Hodge DR, Howard OM, Yoshimura T.