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.

A Molecular Genetic Investigation of Schizophrenia in an Irish study population

Aiden Corvin

MB BCh MRCPsych

PhD Thesis

Department of Psychiatry

University of Dublin, Trinity College

2005

TRINITY COLLEGE
15 MAY 2006
LIBRARY DUBLIN
THESIS
7862

Declaration

Declaration 1:
This work has not been previously accepted in substance for any degree and is not being concurrently submitted in candidature for any degree at this or any other university.

Signed.....
Date.....19.02.06.....

Declaration 2:
This thesis is the result of my own investigations except where otherwise stated in the text. Other sources are acknowledged in the text. A bibliography is appended.

Signed.....
Date.....19.02.06.....

Declaration 3:
I hereby give consent for this thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organizations.

Signed.....
Date.....19.02.06.....

Summary

Schizophrenia (OMIM 181500) is a complex genetic disorder, which affects ~1% of the population, and typically presents in early adulthood with abnormalities of perception (psychosis), cognition, information processing and social functioning. Genetic epidemiological research suggests that schizophrenia is related to other psychotic disorders (including schizoaffective disorder and bipolar affective disorder). As a group these functional psychoses are chronic in course, and associated with substantial morbidity. Current treatments are partially effective and were discovered surreptitiously: establishing genes that contribute to these disorders may be important in understanding the biology involved and developing new therapies.

I established a large case-control sample for genetic investigation of schizophrenia and other psychotic disorders in the Irish population. The final sample included 396 cases (of Irish ancestry) matched with 645 controls obtained from the Irish Blood Transfusion Service. For cases the clinical assessment included a structured diagnostic interview (SCID-P), clinical symptoms assessment (PANSS), examination for physical dysmorphology (MPA-Lane Scale) and neurological assessment (NSS and AIMS).

The molecular studies presented in this thesis focussed on the schizophrenia phenotype (including DSM-IV schizophrenia, schizoaffective disorder and schizophreniform disorder) (n=299). Linkage studies have identified a number of putative chromosomal susceptibility loci from family studies conducted in Ireland and elsewhere (including chromosome 6p24-22, 8p21-22 and 22q11-12). Based on reports suggesting the gene Neuregulin-1 (*NRG1*) as a candidate gene for schizophrenia at chromosome 8p, I performed a case-control study of markers at this gene using samples collected during the first phase of this sample collection (246 cases and 231 controls). This identified significant evidence for association with a risk haplotype ($HapB_{IRE}$) (OR=1.7, p=0.013). This haplotype shared common alleles with the original risk haplotype reported from an Icelandic study population and was also found at significant excess in an independent Scottish schizophrenia case-control sample.

Data from gene expression studies suggested the *APOL* gene family as potential schizophrenia candidate genes at the putative susceptibility locus at chromosome 22q. I performed candidate gene analysis across the *APOL1-3* genes using a 3-stage DNA pooling strategy, which represented a novel variation of the original method to increase efficiency for studies that employ public database SNPs. No evidence was found to suggest that variation at these genes was associated with schizophrenia. The availability of increasingly detailed information on inter-marker LD from a number of ethnic populations (e.g. with the HapMap project) has potential benefits for selecting markers for a study of this type. An ancillary study of HapMap data suggested that the density of markers selected for this study was likely to be insufficient to capture the majority of genetic variation at the locus. The absence of sufficient comparative marker data from other populations at this time hinders a more efficient marker selection without substantially more genotyping.

I analysed genotype data from candidate gene analyses at the D-amino acid oxidase (*DAAO*), *G72* (D-amino acid oxidase activator (*DAAOA*)), dystrobrevin binding protein 1 (*DTNBP1*) and *NRG1* genes in the full sample of 299 cases and 645 controls. These studies

identified evidence of association with haplotypes at *DTNBP1* and individual markers at *DAAO* (M5) (OR=1.4, p=0.003) and *G72* (M12)(1.34, p=0.005). These studies failed to confirm evidence of association with the *NRG1* haplotype ($HapB_{IRE}$) in the extended dataset.

Understanding the nature of the relationships between susceptibility genes is likely to be important in establishing how they contribute to schizophenia susceptibility. A previous report suggested evidence of epistatic interaction between markers at *G72* and *DAAO* in schizophrenia susceptibility in a French-Canadian sample. The pattern of individual marker associations differed between samples, so I examined for evidence of epistatic interaction between the markers proposed by the original report and between the *G72* and *DAAO* markers individually associated with schizophrenia in this population. The study, using a logistic regression model, identified significant interaction between *G72* (M12) and *DAAO* (M5) in this sample (OR=9.3, (CI 1.4-60.5), p=0.008). Novel exploratory methods of testing for interaction between multiple loci are being developed. Using the multifactorial dimension reduction (MDR) method, I tested for evidence of interaction among six loci in this dataset. The markers (at *G72, DAAO, MRDS-1, DTNBP-1, NRG1* and *erbB4*) were selected *a priori* based on the limited literature available for interaction between loci in schizophrenia susceptibility. A single locus model (*DAAO*) provided the best fit to the data using this approach.

Acknowledgements

First and foremost I would like to thank my supervisor Professor Michael Gill, without whose mentorship and encouragement this work would not have been possible. The generosity of all of the individuals who have contributed to the study as participants is truly extraordinary. I would like to thank them all, for giving of their precious time and affording a warm welcome to the clinical research team. Their ongoing support and interest in the results of this project have been a wonderful motivating force for all of us in the last five years.

On the clinical side there are many people who merit special acknowledgement. Among my supportive and interested colleagues at the participating hospitals, the involvement of Prof. John Waddington and Gabriel Sherry warrants particular mention. At the Irish Blood Transfusion Service I would like to thank Dr Willie Murphy. The warmest thanks are reserved for Siobhan Schwaiger and Dr Jeanne Marie Nangle my wonderful and dedicated co-interviewers.

In the neuropsychiatric genetics laboratory I would particularly like to thank Dr Derek Morris for his support and (above all) patience with this 'lab rookie'. Thanks also to all of the rest of the laboratory team past and present, Kevin McGhee, Dr Ziarah Hawi, Dr Gary Donohoe, Niamh Kenny, Kevin Murphy and Sarah Clarke. I would also like to thank my statistical collaborators Dr Jason Moore at Dartmouth Medical School and Dr Cathal Walsh at Trinity College Dublin.

Finally, I would particularly like to thank my family who have listened, supported and believed in my labours over the last few years, most particularly my wife Michele to whom this thesis is dedicated.

Statement of Work

This work is the product of the Psychosis Genetics Research Group at the Department of Psychiatry, University of Dublin, Trinity College. The author designed the study, ascertained cases, and performed much of the clinical recruitment and collection of DNA samples from participating individuals with the assistance of a research nurse, Siobhan Schwaiger (referred to as SS in the thesis) and more recently Dr Jeanne Marie Nangle (JMN). Staff at the Irish Blood Transfusion Service collected the control samples, and the control collection was co-ordinated by Siobhan Schwaiger.

The author performed DNA extraction and quantification for all of the samples used in the studies described in Chapters 5 and 6 with Dr Derek Morris and Kevin McGhee (PhD student). Research technicians, Kevin Murphy and Niamh Kenny, extracted the additional samples contributing to the analyses described in Chapter 7.

The author performed the genotyping and analysis of the *APOL* gene family in collaboration with Kevin McGhee (PhD student). The author performed the comparative analysis of haplotype structure across this locus as described in Chapter 6. Genotyping of the single nucleotide polymorphisms and microsatellite markers at the *NRG1* locus were performed anonymously, on a non-commercial basis by deCODE genetics. The author conducted the analysis of this data. The author performed some of the genotyping for the investigation of the *DTNBP1* and *RGS4* loci, but other members of the research group performed much of the genotyping for the other candidate genes mentioned in this thesis. Analyses of data at these individual loci was conducted by the author and Dr Derek Morris. The interaction analysis of *G72* and *DAAO* described in Chapter 7 was performed by the author with the support of Dr Cathal Walsh at the Dept of Statistics, TCD. Finally, the multifactor-dimensionality reduction (MDR) models described in Chapter 7 were investigated in collaboration with Professor Jason Moore at Dartmouth College, Hanover, New Hampshire, US.

Publications

McGhee KA, Morris DW, Schwaiger S, Nangle JM, Donohoe G, Clarke S, Meagher D, Quinn J, Scully P, Waddington JL, Gill M, Corvin A. Investigation of the apolipoprotein-L (APOL) gene family and schizophrenia using a novel DNA pooling strategy for public database SNPs. Schizophrenia Research 2005; 76: 231-238.

Murtagh A, McTigue O, Hegarty AM, Stallings RL, Green AJ, Ramsay L, Corvin A. Interstitial deletion of chromosome 21 and schizophrenia. Schizophrenia Research (in press).

Corvin A, Morris DW, McGhee K, Schwaiger S, Scully P, Quinn J, Meagher D, Waddington JL, Gill M. Confirmation and refinement of an 'at risk' haplotype for schizophrenia suggests the EST cluster, *Hs.*97362, as a potential susceptibility gene at the Neuregulin-1 locus. Molecular Psychiatry 2004; 9: 208-213 [PMID: 14966480].

Morris DW, Rodgers A, McGhee K, Schwaiger S, Scully P, Quinn J, Meagher D, Waddington JL, Gill M and Corvin A. Confirming RGS-4 as a susceptibility gene for schizophrenia. American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics 2004; 125B: 50-3 [PMID: 14755443]

Williams NM, Preece A, Morris DW, Spurlock G, Bray NJ, Stephens M, Norton N, Williams H, Waddington JL, Gill M, Corvin A, Zammit S, Kirov G, Owen MJ, O'Donovan MC. Confirmation in two independent samples that dystrobrevin binding protein gene (DTNBP1) is a susceptibility locus for schizophrenia. Archives of General Psychiatry 2004; 61: 336-44 [PMID: 15066891]

Corvin A, Nangle JM, Gill M. Schizophrenia genes. Irish Medical Journal 2004; 97:70-71. [PMID: 15164685]

Corvin A, Gill M. Psychiatric Genetics in the Post Genome Age. British Journal of Psychiatry 2003; 182: 95-96. [PMID: 12562733]

Morris DW, McGhee K, Schwaiger S, Scully P, Quinn J, Meagher D, Waddington JL, Gill M and Corvin A. No evidence for association of the dysbindin gene [DTNBP1] with schizophrenia in an Irish population-based study. Schizophrenia Research 2003; 60: 167-172. [PMID: 12591580].

Abbreviations

5HT2A	serotonin 5HT2A receptor gene
A	adenosine
AIMS	Abnormal Involuntary Movement Scale
APOE	apolipoprotein-E
APOL	apolipoprotein-L
C	cytidine
CEPH	Centre d'Etude du Polymorphisme Humain
cM	centiMorgans
CNE	Condensed Neurological Examination
CNS	central nervous system
DA	dopamine
DAAO	D-amino acid oxidase
DdH_2O	distilled deionised H_2O
DISC-1	Disrupted in Schizophrenia-1
DNA	deoxyribonucleic acid
DRD3	dopamine D3 receptor gene
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition.
DTNBP1	dystrobrevobindin-1
DZ	dizygotic
EDTA	ethylenediaminetetraacetic acid
EST	expressed sequence tag
FP	fluorescence polarization
G	guanosine
GCPII	glutamate carboxypeptidase II
GGF	glial growth factors
h^2	heritability
HDL	high density lipoprotein
HLA	human leukocyte antigen
HRB	Health Research Board

ICD-10	International Classification of Mental and Behavioural Disorders.
LD	linkage disequilibrium
MPA	Minor Physical Anomalies
MR	morbid risk
MRI	magnetic resonance imaging
MRNA	messenger ribonucleic acid
MZ	monozygotic
NES	Neurological Evaluation Scale
NRG1	Neuregulin-1
NMDA	N-methyl-D-aspartate receptor
NRG1	Neuregulin-1
OD	Optical density
OMIM	Online Mendelian Inheritance in Man
PANSS	Positive and Negative Syndrome Scale
PCR	polymerase chain reaction
PPR	putative promoter region
PSE	present state examination
RDC	Research Diagnostic Criteria
RGS4	regulator of G-protein signalling-4
SFI	Science Foundation Ireland
SNP	single nucleotide polymorphism
SZ	schizophrenia
T	thymidine
TDT	transmission disequilibrium test
VCFS	velo-cardio-facial syndrome
WTAR	Weschler test of adult reading

Table of contents

Page

Declaration.....i

Summary.....ii

Acknowledgements.....iv

Statement of Work v

Publications.....vi

Abbreviations..... vii

Index of Tables.....xvi

Index of Figures..... xix

CHAPTER 1 INTRODUCTION

1.1 Clinical background

1.1.1 Definition of schizophrenia..... 1

1.1.2 Historical framework for schizophrenia diagnosis..... 1

1.1.3 Modern diagnostic criteria..... 2

1.1.4 Defining schizophrenia for genetic studies..... 2

1.1.5 Differential diagnosis.....5

1.2 The pathogenesis of schizophrenia

1.2.1 Neuropharmacological hypotheses.....7

1.2.1.1 Dopamine hypothesis.....7

1.2.1.2 Glutamate hypothesis.....7

1.2.2 Neurodevelopmental hypotheses.....8

1.2.2.1 Epidemiological support for.....9

1.2.2.2 Neuropathological findings in schizophrenia.....10

1.2.2.3 Gene expression studies in schizophrenia.....10

1.2.2.4 Neuroimaging findings in schizophrenia.....11

1.3 Schizophrenia as a complex genetic disorder

1.3.1 Family studies.....11

1.3.2 Twin studies.....12

1.3.3 Adoption studies.....13
1.3.4 Mode of inheritance..... 13
1.3.5 Demarcating the schizophrenia phenotype.....14
1.3.6 Molecular studies of schizophrenia.....15
1.3.6.1 Cytogenetic abnormalities and schizophrenia.....16
1.3.6.2 Linkage studies..... 16
1.3.6.3 Association studies.....18

1.4 Specific hypotheses for this project

1.4.1 Establishing a psychosis sample for molecular research.....19
1.4.2 Funding applications.....20
1.4.3 Molecular genetic studies of schizophrenia..... 21

CHAPTER 2 METHODOLOGICAL ISSUES IN COMPLEX GENETIC DISORDER RESEARCH

2.1 Introduction

2.2 Positional cloning methods of gene identification

2.2.1 Linkage analysis..... .26
2.2.2 Genetic association studies..... .29
2.2.2.1 Case-control and family-based association methods.....30

2.3 Linkage disequilibrium (LD) mapping

2.3.1 Definition of linkage disequilibrium.....31
2.3.2 Recombination.....32
2.3.3 Genetic drift.....34
2.3.4 Selection.....34
2.3.5 Other population factors.....35
2.3.6 Admixture and admixture mapping.....35
2.3.7 Measuring LD.....36

2.4 Patterns of LD in human populations

2.4.1 Early studies of LD structure in human populations.....37

2.4.2 LD patterns in human populations: a regular structure?.....38
2.4.3 Linkage disequilibrium in the Irish population.....40
2.4.4 Complex genetic disorders: common or rare risk variants?.....40

CHAPTER 3 MATERIALS AND METHODS

3.1 Introduction

3.2 Collaborators

3.2.1 Clinical collaborators.....43
3.2.2 Laboratory collaborations.....43
3.2.3 Statistical collaborations.....44

3.3 **Clinical methodologies**

3.3.1 Ethics approval.....44
3.3.2 Ascertainment.....44
3.3.3 Recruitment.....45
3.3.4 Dissemination of results to participants.....46
3.3.5 DNA collection and storage.....46
3.3.6 Clinical instruments and procedures.....46
3.3.6.1 The Structured Clinical Interview for DSM-IV (SCID).....47
3.3.6.2 The Positive and Negative Symptom Scale (PANSS).....48
3.3.6.3 The Abnormal Involuntary Movement Scale (AIMS).....48
3.3.6.4 Neurological Soft Signs Assessment.....48
3.3.6.5 Minor Physical Anomalies Scale (MPAs).....49

3.4 Laboratory methods

3.4.1 Buffers and solutions.....50
3.4.2 DNA extraction.....50
3.4.2.1 DNA suspension and storage.....53
3.4.3 DNA quantification.....53
3.4.3.1 Quantified DNA stock preparation.....54
3.4.4 DNA storage.....54
3.4.5 Strategy for genotyping individuals.....54

3.4.6 Genotyping of DNA pooled samples.....55
3.4.6.1 Preparation of samples for DNA pools.....56
3.4.7 Genotyping the APOL gene family.....57
3.4.7.1 Primer design.....57
3.4.7.2 DNA amplification/optimisation by PCR.....57
3.4.7.3 Genotyping methodology.....57
3.4.7.4 SNP genotyping on the ABI 377 sequencing system.....58
3.4.7.5 Simulating heterozygosity data for DNA pooling.....59
3.4.7.6 Confirming reliability for this DNA pooling method.....60

3.5 Statistical methods

3.5.1 Allele frequencies and Hardy-Weinberg equilibrium.....61
3.5.2 Relative risk.....61
3.5.3 Odds ratios.....62
3.5.4 Chi-square test.....62
3.5.5 Power calculations.....62
3.5.6 Graphical overview of LD (GOLD).....62
3.5.7 Haploview.....63
3.5.8 Haplotype estimation.....64
3.5.7.1 Haplotype association analyses.....64

CHAPTER 4 CLINICAL SAMPLE DESCRIPTION

4.1 Introduction

4.2 Description of sample

4.2.1 Ascertainment.....75
4.2.2 Inclusion criteria.....75
4.2.3 Exclusion criteria.....76
4.2.4 Exclusions, withdrawals & refusals.....77
4.2.5 Description of the case sample.....77
4.2.5.1 Clinical assessments.....78
4.2.6 Description of the control sample.....79

CHAPTER 5 INVESTIGATION OF THE NEUREGULIN-1 (NRG1) CANDIDATE GENE AT THE CHR 8p SUSCEPTIBILITY LOCUS

5.1 Introduction

5.1.1 Evidence for a SZ susceptibility locus at chromosome 8p21-22.....82

5.1.2 Identification of a positional candidate gene for SZ at this locus.....83

5.1.3 Identification of the NRG1 gene.....84

5.1.4 NRG1 as a functional candidate gene for SZ.....84

5.1.4.1 Human functional studies of NRG1.....84

5.1.4.2 Animal functional studies of NRG1.....85

5.1.5 Rationale for this study.....86

5.2 Methodology

5.2.1 Sample size and study power.....86

5.2.2 DNA handling and genotyping.....87

5.2.3 Statistical methods.....87

5.3 Results

5.3.1 Tests for association between SZ and the NRG1 core haplotype.....88

5.3.2 Single marker association analyses across the NRG1 locus.....88

5.3.3 Haplotype structure at NRG1 in Irish and Icelandic samples.....88

5.3.4 Haplotype analysis of the NRG1 locus in the Irish sample.....88

5.3.5 Overlap between identified Irish and Icelandic risk haplotypes.....89

5.3.6 Replication in another study population.....89

5.4 Discussion

CHAPTER 6 INVESTIGATION OF THE APOLIPOPROTEIN-L (*APOL*) GENE FAMILY AS CANDIDATE GENES FOR SCHIZOPHRENIA

6.1 Introduction

6.1.1 Identification of the *APOL* gene family.....108

6.1.2 Gene ontology.....109

6.1.3 Positional evidence for involvement of the *APOL* locus in SZ susceptibility.....109
6.1.4 Functional evidence for involvement of the *APOL* genes in SZ.....111
6.1.5 Strategy for candidate genes studies of *APOL1-6*.....112
6.1.6 Investigation of haplotype structure across the APOL locus.....113

6.2 Results

6.2.1 Bioinformatics/primer design.....115
6.2.2 PCR conditions.....115
6.2.3 Confirming SNP heterozygosity.....115
6.2.4 DNA pooling experiments for *APOL1-3* (Stage 1).....116
6.2.5 Screening DNA pooling experiments for *APOL1-3* (Stage 2).....116
6.2.6 Individual genotyping at the *APOL* locus.....116
6.2.7 Haplotype structure across this locus.....117
6.2.7.1 Overlap with the marker set for association studies.....117
6.2.7.2 Comparison of LD-structure at increased marker density...117
6.2.7.3 Stability of the haplotype structure at increased marker density.....118
6.2.7.4 Haplotype blocks tagged by markers in this study.....118

6.3 **Discussion**

CHAPTER 7 INVESTIGATING GENE-GENE INTERACTION IN SCHIZOPHRENIA

7.1 Introduction

7.1.1 Epistatic interaction.....128
7.1.2 Evidence for gene-gene interaction in Schizophrenia.....129
7.1.3 Testing for epistatic interaction.....130

7.2 Investigation of epistatic interaction in this dataset

7.2.1 Candidate gene selection.....133
7.2.1.1 Dysbindin-1 (*DTNBP1*) and Schizophrenia.....133
7.2.1.2 *MRDS-1* and Schizophrenia.....134

7.3 **Methodology**

7.3.1 Case-control study sample.....134

7.3.2 Marker selection and genotyping methods.....135

7.3.3 Statistical methods.....135

7.3.3.1 Logistic regression modelling.....135

7.3.3.2 Multi-dimension reduction analysis (MDR).....136

7.4 Results

7.4.1 Single marker association studies.....137

7.4.1.1 Results of candidate gene studies at *G72* and *DAAO*.....137

7.4.1.2 Results of candidate gene studies at *NRG1* and *erbB4*.....138

7.4.1.3 Results of association studies at *DTNBP1*.....138

7.4.1.4 Results of association studies at *MRDS1*.....138

7.4.2 Interaction between *G72* and *DAAO* 138

7.4.3 Interaction study using MDR.....139

7.5 Discussion

CHAPTER 8 DISCUSSION

REFERENCES.....162

APPENDICES

Appendix I.....187

Appendix II.....189

Appendix III.....191

Appendix IV.....193

Appendix V.....195

INDEX OF TABLES

Page

Table 1.1 Diagnostic criteria for Schizophrenia in DSM-IV.....3

Table 1.2 Diagnostic criteria for Schizophrenia in ICD-10.....4

Table 1.3 Differential diagnosis of Schizophrenia.....6

Table 3.1 Polymerase Chain Reaction (PCR) cycling parameters.....65

Table 3.2 APOL-1, PCR details.....66

Table 3.3 APOL-2, PCR details.....67

Table 3.4 APOL-3, PCR details.....68

Table 4.1 Ascertainment/inclusion of cases for the project.....80

Table 4.2 Demographic details of the Phase 2 sample.....81

Table 5.1 Results (simulated p-values) from association analyses of single markers and 2-, 3-, 4-marker haplotypes constructed from adjacent markers.....94

Table 5.2 Allele frequencies at the locus D81770 (M1).....95

Table 5.3 Allele frequencies at the locus 29H12-1 (M2).....95

Table 5.4 Allele frequencies at the locus 29H12-7320 (M3).....96

Table 5.5 Allele frequencies at the locus D81711 (M4)……..……96

Table 5.6 Allele frequencies at the locus 29H12-121L21 (M5)……..…..97

Table 5.7 Allele frequencies at the locus 450K14-72458(M6)……..…..97

Table 5.8 Allele frequencies at the locus SNP8NRG221533 (M7)……98

Table 5.9 Allele frequencies at the locus 247B14-642 (M8)……..…..98

Table 5.10 Allele frequencies at the locus SNP8NRG241930 (M9)……..…99

Table 5.11 Allele frequencies at the locus SNP8NRG243177 (M10)……..99

Table 5.12 Allele frequencies at the locus 478-2 (M11)……..……99

Table 5.13 Allele frequencies at the locus 478B14-848 (M12)……..…...100

Table 5.14 Allele frequencies at the locus 420M9-1395 (M13)……..……..100

Table 5.15 Allele frequencies at the locus 420M9-1 (M14)……101

Table 5.16 Allele frequencies at the locus D8S1810 (M15)……..……....101

Table 5.17 Allele frequencies at the locus 420M9-116I12 (M16)……..…...102

Table 5.18 Allele frequencies at the locus 473C15-533 (M17)……..……102

Table 5.19 Allele frequencies at the locus 473C15-439 (M18)……..103

Table 5.20 Allele frequencies at locus 72H22-1 (M19)……103

Table 5.21 Allele frequencies at locus 72H22-36 (M20)……..104

Table 5.22 Simulated p-values for assessment of Hardy-Weinberg Equilibrium at M1-M20 in cases and controls.....104

Table 6.1 Results of DNA pooling experiments for *APOL-4,2,1*.....122

Table 6.2 Results of DNA pooling experiments for *APOL-6,5,3*.....123

Table 6.3 Results of individual genotyping experiments.....124

Table 6.4 Differences between haplotype blocks from HapMap 2004.....124

Table 7.1 Results of single marker analyses for *G72* and *DAAO*.....145

Table 7.2 Genotype marker data from *DTNBP1*, *NRG1* and *MRDS1*.....145

Table 7.3 Results of MDR modelling of interaction between schizophrenia loci.....146

Table 7.4 Results of schizophrenia association studies at the *DAAO* locus.....147

Table 7.5 Comparison of results from published genetic association studies of *G72* and Schizophrenia.....148

INDEX OF FIGURES

Page

Figure 1.1 Schizophrenia risk in relatives of affected probands.....22

Figure 1.2 Summary of linkage findings from meta-analysis of SZ genome-scans.....23

Figure 1.3 Genetic and environmental risk factors in SZ.....24

Figure 1.4 The biochemical pathways in Glutamate metabolism.....25

Figure 3.1 Allele specific primer extension genotyping method.....70

Figure 3.2 Multiplexing primer extension method.....71

Figure 3.3 SNaPSHOT primer extension genotyping reaction Protocol.....72

Figure 3.4 Simulating heterozygosity data for DNA pooling.....73

Figure 3.5 Maximum potential difference.....74

Figure 5.1 The 3 'at-risk' haplotypes (Hap A, B and C) and core 'at-risk' haplotype identified by Stefansson et al, (2002).....105

Figure 5.2 Inter-marker LD at *NRG1* measured using *D'* in the Irish schizophrenia case sample.....106

Figure 5.3 Inter-marker LD at *NRG1* measured using *D'* in the Icelandic schizophrenia case sample.....107

Figure 6.1 Structure of the *APOL* gene family.....125

Figure 6.2 Haplotype blocks/tagging SNPs at the *APOL* locus.....126

Figure 6.3 Haplotype structure of the APOL locus from the CEPH families..... ...127

Figure 7.1 Putative interactions between schizophrenia susceptibility genes.....143

Figure 7.2 Multidimensional Reduction (MDR) method.....144

Chapter 1
Introduction

1.1 Clinical background

1.1.1 Definition of schizophrenia

Schizophrenia (OMIM 181500; http://www.ncbi.nlm.gov/Omim/) is a chronic psychiatric disorder that affects ~1% of the world's population and is only partially responsive to current treatments. The disorder has a typical onset in early adulthood and impacts on perception (delusions and hallucinations), information processing (disordered thinking), cognition and social functioning (Andreasen & Carpenter, 1993). In the absence of defining pathophysiology, the clinical definition of schizophrenia represents an amalgamation of different historical perspectives on the 'true' nature of the disorder.

1.1.2 Historical framework for schizophrenia diagnosis

In the late nineteenth century, Kraepelin developed the first comprehensive and widely accepted definition of what he termed *dementia praecox* (Kraepelin, 1919). He identified a syndrome of chronic course and poor outcome that began early in life and produced persistent impairments in many aspects of behaviour and cognitive function. In contrast, his contemporary Bleuler attempted to identify specific symptoms that represented essential aspects of *dementia praecox.* In Bleuler's formulation the most important feature was a loosening of associational networks and a loss over time of the 'unity' of personality. To emphasise the fundamental importance of the splitting of associations, he renamed the disorder 'schizophrenia', or more correctly, the 'group of schizophrenias' (Bleuler, 1911).

Working half a century later, Schneider also emphasized the importance of discriminatory symptoms rather than course of illness in schizophrenia diagnosis (Schneider, 1959). He identified specific psychotic symptoms that he considered of first-rank in making the diagnosis and argued that these were specific to schizophrenia and other toxic psychotic syndromes. The symptoms he described were implausible and bizarre and fit with the concept of schizophrenia as a disorder of experience beyond the comprehension of others. They included experiences of externally

controlled thought, movement or impulse; thought insertion, withdrawal or broadcasting; and voices discussing a patient in the third person or making a running commentary on a patient's behaviour.

1.1.3 Modern diagnostic criteria

There is general agreement as to the clinical features of schizophrenia in contemporary diagnostic classification systems (Diagnostic and Statistical Manual of Mental Disorder (DSM-IV) and the ICD-10 Classification of Mental and Behavioural Disorders (ICD10)) (see Tables 1.1 and 1.2). Each relies heavily on the interview structure and diagnostic algorithm of the Present State Examination (PSE) and incorporates the PSE time frame of one month (Wing et al, 1974). In each, Schneider's symptoms are prioritised among the delusions and hallucinations described as 'positive' symptoms. The abnormalities in the process of thinking described by Bleuler remain, but with less emphasis, in part because of the difficulty in objectively and reliably defining thought disorder. Other symptoms, described by Bleuler as fundamental, form the basis of the 'negative' symptoms of affective flattening, alogia and avolition. The course of the disorder is recognised in both classifications but formalised in DSM-IV. Prodromal or residual symptoms are also recognised in both modern classifications, but in DSM-IV continuous signs of disturbance must persist for at least 6 months, otherwise the episode is termed schizophreniform. The DSM-IV diagnosis of schizophrenia also requires evidence of functional impairment in interpersonal relationships, at work, or in self-care, for a significant portion of time since onset of the disorder.

1.1.4 Defining schizophrenia for genetic studies

'The student of schizophrenia pursues a moving target'

Andreasen & Carpenter (1993).

Accurately defining the schizophrenia phenotype for epidemiological and genetic studies is important. To insure assessment reliability most interviews implement a standardised structured clinical interview (e.g. the Structured Clinical Interview for DSM-IV (SCID)) and perform inter-rater reliability exercises. With such procedures the diagnostic reliability of ICD-10 and DSM-IV diagnosed SZ matches that available for medical conditions such as diabetes or hypertension (ICD-10, 1992).

Table 1.1: Diagnostic criteria for Schizophrenia in DSM-IV

A. Characteristic symptoms: two (or more) of the following, each present for a significant portion of time during a one-month period (or less if successfully treated): 1) Delusions 2) Hallucinations 3) Disorganized speech (e.g. Frequent derailment or incoherence) 4) Grossly disorganized or catatonic behaviour 5) Negative symptoms, i.e., affective flattening, alogia or avolition *Note: only one criterion a symptom is required if delusions are bizarre or hallucinations consist of a voice keeping a running commentary on the persons behaviour or thoughts, or two or more of the voices converse with each other.*
B. Social/occupational dysfunction: For a significant portion of time since the onset of the disturbance, one or more major areas of functioning such as work, interpersonal relations, or self-care are markedly below the level achieved prior to the onset (or when the onset is in childhood or adolescence, failure to achieve expected levels of interpersonal, academic, or occupational achievement).
C. Duration: Continuous signs of the disturbance persist for at least 6 months. This period must include at least 1 month of symptoms (or less if successfully treated) that meet criterion A and may include periods of prodromal or residual symptoms. During these prodromal or residual periods, the signs of the disturbance may be manifest by only negative symptoms or two or more symptoms listed in Criterion A present in an attenuated form (e.g. odd beliefs, unusual perceptual experiences).
D. Schizoaffective and Mood Disorder exclusion: Schizoaffective Disorder and Mood Disorder with Psychotic Features have been ruled out because either (1) no Major Depressive, Manic, or Mixed Episodes have occurred concurrently with the active-phase symptoms; or (2) if mood episodes have occurred during the active-phase symptoms their total duration has been brief relative to the duration of the active and residual periods.
E. Substance/general medical condition exclusion: The disturbance is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.
F. Relationship to a Pervasive Developmental Disorder: If there is a history of Autistic Disorder or another Pervasive Developmental Disorder, the additional diagnosis of Schizophrenia is made only if prominent delusions or hallucinations are also present for at least a month (or less if successfully treated).

Table 1.2: Diagnostic criteria for Schizophrenia in ICD-10

SECTION 1 SYMPTOMS (a) thought echo, thought insertion or withdrawal, and thought broadcasting; (b) delusions of control, influence, or passivity, clearly referred to body or limb movements or specific thoughts, actions, or sensations; delusional perception; (c) hallucinatory voices giving a running commentary on the patient's behaviour, or discussing the patient among themselves, or other types of hallucinatory voices coming from other parts of the body (d) persistent delusions of other kinds that are culturally inappropriate and completely impossible, such as religious or political identity, or superhuman powers and abilities (e.g. Being able to control the weather, or being in communication with aliens from another world).
SECTION 2 SYMPTOMS (a) Persistent hallucinations in any modality, when accompanied either by fleeting or half-formed delusions without clear affective content, or by persisting overvalued ideas, when occurring every day for weeks or months on end; (b) Breaks or interpolations in the train of thought, resulting in incoherence or irrelevant speech, or neologisms; (c) Catatonic behaviour, such as excitement, posturing, or waxy flexibility, negativism, mutism and stupor (d) 'Negative' symptoms such as marked apathy, paucity of speech, and blunting or incongruity in emotional responses, usually resulting in social withdrawal and lowering in social performance; it must be clear that these are not due to depression or to neuroleptic medication (e) a significant and consistent change in the overall quality of some aspects of personal behaviour, manifest as loss of interest, aimlessness, idleness, a self-absorbed attitude and social withdrawal
A minimum of one very clear (and usually two or more if less clear-cut) symptoms from section 1 or two of the symptoms from section 2 are present for most of the time during a period of one-month or more. Exclusion of overt brain disease or a state of drug intoxification or withdrawal. The diagnosis should not be made in the presence of extensive depressive or manic symptoms unless it is clear that the schizophrenic symptoms antedated the affective disturbance. If both schizophrenia and affective symptoms develop together and are evenly balanced the diagnosis of schizoaffective disorder should be made.

Diagnostic reliability does not equate to construct validity and the biological basis of clinical schizophrenia is unknown. Affected individuals can present with non-overlapping symptoms involving different areas of function such as perception, inferential thinking, volition and social behaviour. Symptoms may also vary during the course of illness. Different symptoms in schizophrenia may represent the involvement of diverse neural pathways and represent variants of normal brain function or discrete disease processes. Clinical symptoms may present in other disorders (e.g. head injury, substance abuse or epilepsy) and can mimic the schizophrenia phenotype. Where available, detailed medical history, collateral clinical information, drug screening and neuroimaging should be considered in reaching the diagnosis.

Psychiatric diagnoses are known to be somewhat unstable over time. Clinically, individuals may present at disease onset with symptoms meeting criteria for an affective disorder but later meet criteria for schizophrenia; the reverse is less usually the case (Schwartz et al, 2000). Indeed, the reliability of lifetime schizophrenia diagnosis 30-40 years after initial assessment has been reported as 92.5% (Tsuang et al, 1981). For this reason most genetic studies consider a lifetime ever diagnosis of schizophrenia.

1.1.5 Differential diagnosis

The differential diagnosis of schizophrenia includes both psychiatric and organic disorders. Psychiatric disorders that may present with psychotic symptoms include the other major psychotic disorders (the affective psychotic disorders and schizoaffective disorder) and other psychotic disorders (e.g. the brief psychotic disorders). Certain personality disorders, pervasive developmental disorders and obsessive-compulsive disorder may also present with features similar to schizophrenia. An extensive list of medical conditions can mimic the symptoms of the disorder, but these account for only a small fraction of cases (Lewis 1990; Johnstone et al, 1988). The list of potential differential diagnoses is described in Table 1.3.

Table 1.3: Differential diagnosis of Schizophrenia

Schizophrenia differential diagnosis
Other psychiatric disorders:
Schizoaffective disorder
Affective disorders
Delusional disorders
Brief psychotic disorders
Psychotic disorder not otherwise specified
Personality disorders (e.g. schizotypal PD)
Obsessive-compulsive disorders
Pervasive developmental disorders
Drug induced psychotic disorders:
Substance abuse (e.g. amphetamines, LSD, cocaine)
Alcohol
Prescription medications (e.g. levodopa, steroids)
Organic disorders:
Epilepsy
Head trauma
Space occupying lesions (e.g. tumours)
Inflammatory disorders (e.g. sarcoidosis)
Infective (e.g. HIV, neurosyphilis)
Autoimmune disorders (e.g. diabetes mellitus)
Metabolic (e.g. vitamin deficiencies)
Degenerative (e.g. dementia)
Vascular (e.g. CVA)
Endocrine (e.g. thyroid dysfunction)
Other genetic disorders:
Mendelian disorders (e.g. Huntington's disease)

1.2 The pathogenesis of schizophrenia

1.2.1 Neuropharmacological hypotheses

The discovery in the 1950's that some of the symptoms of schizophrenia improved with drugs such as chlorpromazine and reserpine opened a new era of neuropharmacological hypotheses of schizophrenia pathogenesis. Of many proposed hypotheses the most persistent and supported relate to dopamine and glutamate dysfunction.

1.2.1.1 Dopamine hypothesis

The original dopamine hypothesis evolved from the observation that early antipsychotic agents were dopamine (D2) receptor agonists. The hypothesis proposes that the positive symptoms of schizophrenia (delusions and hallucinations) result from dopaminergic overactivity, due either to an excess of dopamine (DA) or an excess of sensitivity to the effects of DA (Carlsson & Lindqvist, 1963). This hypothesis is supported by the induction of psychotic symptoms by DA enhancing drugs and the correlation between clinical doses of antipsychotic drugs and D2 receptor blockade (Laruelle, 2003). A revised hypothesis of dopaminergic imbalance between cortical and subcortical structures has been proposed to explain other deficits in SZ (reviewed in Davis et al, 1991).

1.2.1.2 Glutamate hypothesis

Glutamate is the primary excitatory neurotransmitter and a model suggesting abnormal glutamate neurotransmission in schizophrenia was first mooted in 1959 (Luby et al, 1959). However, glutamate function is complex, as highlighted by the numerous metabolic processes with which it is involved (see Figure 1.4) and it is perhaps inevitable that glutamate metabolism would be in some way affected by the disorder or its treatment. More recent pharmacological challenge studies have suggested, more specifically, a role for the N-methyl-D-aspartate receptor (NMDA) in schizophrenia susceptibility. Non-competitive antagonists of the NMDA receptor, such as phencyclidine, induce a psychosis closely resembling schizophrenia. To explain this, a hypothesis of reduced glutamate neurotransmission at the NMDA

receptor has been proposed (Javitt & Zukin, 1991). This has been supported by post mortem studies reporting regional specific changes in glutamate receptors in schizophrenia brain samples and provisional evidence of a therapeutic role for NMDA receptor agonists (Coyle & Tsai, 2004). Reconciling data from both glutamate and dopamine hypotheses, Goff & Coyle (2001) noted that the synaptic relationships between forebrain dopaminergic and glutamatergic systems make it likely that dysregulation of one system is likely to have reciprocal effects on the other.

1.2.2 Neurodevelopmental hypotheses

In 1891, Clouston proposed that an early insult to the developing brain of patients with *dementia praecox* resulted in non-progressive changes that became manifest, with brain maturation, in early adulthood (Clouston, 1891). In the last two decades, as understanding of the neurobiology of brain development has improved, early and late neurodevelopmental models have been suggested. In the early model, expression of the disorder is due to subtle abnormalities of early brain development the full consequences of which are only manifest in adolescence or early adulthood. It is now known that brain maturation in adolescence is characterised by a programmed elimination of synapses, which has been termed synaptic pruning. The original 'late' neurodevelopmental model proposed a primary role for abnormality in peri-adolescent synaptic pruning, but did not specify the nature of this abnormality (Feinberg, 1982/83).

Many subsequent authors have proposed models that accommodate early developmental and late brain maturation abnormalities interacting with other risk factors (Weinberger, 1995; Lewis & Murray, 1987; Keshavan 1999). These models have attempted to reconcile evidence for genetic and environmental contributions to risk and broadly conform to the stress-diathesis model where SZ arises from the impact of environment on the genetically vulnerable (Hanson et al, 1976). A number of environmental risk factors for schizophrenia have been identified including *inter alia* perinatal hypoxia, head injury, cannabis use and exposure to psychosocial stressors. Potential interactions between genetic and environmental risk factors are illustrated in Figure 1.3. Such a wide range of risk factors where exposure may occur from early development to adulthood hints that the relationship between

environmental and genetic risk will prove more complex than suggested by the original stress-diathesis model. For example, the extent to which susceptibility may represent sensitivity to (genotype-environment interaction) or selection of (genotype-environment correlation) environment has not been formally tested (van Os & Marcelis, 1998). In the rest of this section, evidence for the hypothesis that schizophrenia is a developmental disorder is reviewed.

1.2.2.1 Epidemiological support

Using different methodologies, developmental impairments in social, behavioural, intellectual and motor function have been consistently identified in children who later develop SZ. Birth cohort data provides evidence for delays in attaining developmental milestones in infancy, deficits in language and cognition, and abnormalities in social functioning in childhood among individuals who go on to develop schizophrenia (Jones 1999). Complementary data from genetic-high risk studies has allowed more detailed information to be collected from individuals. Several studies have now prospectively followed high-risk individuals into adulthood but not yet through the entire period of risk (Fish et al, 1992; Mirsky et al, 1995; Cannon et al, 2002; Cannon and Mednick, 1993; Erlenmeyer-Kimling et al, 2000). From these studies, 25-60% of high-risk children display developmental abnormalities in infancy; motor, attention information processing deficits and social adjustment problems are more common in adolesence. Data from the New York study suggests that early developmental motor and attentional deficits are predictive of later SZ with positive predictive values of 50% (Erlenmeyer-Kimling et al, 2000).

There is consistent evidence for season-of-birth and place of birth as risk factors in schizophrenia. The cause is unknown but prenatal exposure, particularly in the second trimester; to infection, stress or malnutrition may be contributory (for review, see Cannon et al, 2003). Complications of pregnancy (such as bleeding, diabetes and pre-eclampsia), delivery (such as asphyxia) and abnormal foetal growth (as defined by low birthweight or reduced head circumference) are also known to be significantly associated with schizophrenia, although the pooled estimates of effect sizes are small (OR <2). Markers of early developmental deviance, such as minor physical or dermatoglyphic anomalies, are also more common in SZ than in control populations.

1.2.2.2 Neuropathological findings in schizophrenia

Examination of post mortem tissue in schizophrenia has yet to produce clear evidence of gross abnormalities of brain development. Support for the synaptic pruning hypothesis comes from studies showing reduced cortical dendritic volume/spine density or changes in expression of neurodevelopmental genes and proteins in schizophrenia. However, changes in dendritic spine density and gene expression are dynamic processes that vary throughout life. Neuropathological data collected at the endpoint of illness is not ideal for identifying neurodevelopmental deficits that are likely to be subtle. An additional consideration is that current neuropathological data do not support alternative hypotheses of schizophrenia as a disorder of neurodegeneration (Harrison & Lewis, 2003).

1.2.2.3 Gene expression studies in schizophrenia

Gene expression studies directly test for differences in mRNA levels between post-mortem brain tissue from cases and controls. Because mRNA transfers information from the genome to the proteome, identifying differences in mRNA levels between cases and controls may indicate involvement of a gene in disease. With the development of microarray technology, research has advanced from investigating single genes to assaying gene expression levels of most known genes simultaneously. The vast amount of data processed may give new insights into the genes and pathways involved in the pathogenesis of schizophrenia.

Recently, it has been demonstrated that known oligodendrocyte-related and myelin-related genes, and transcription factors that regulate these genes, show reduced expression in schizophrenia. Significantly, similar expression changes are evident in bipolar disorder suggesting involvement of common pathophysiological mechanisms (Tkachev et al 2003). Microarray studies of schizophrenia have now been performed by a number of different groups with consistent expression differences being demonstrated for a number of specific genes including regulator of G-protein signalling 4 (*RGS4*) and glutamate carboxypeptidase II (*GCPII*) (Mirnics, 2003).

Despite these preliminary findings, microarray technology has technical and practical limitations. Current technology is unable to detect less abundant mRNA transcripts or to distinguish between different splice variations of a gene, and there are potential procedural difficulties in examining human post-mortem brain tissue (Ryan et al, 2004). As with other neuropathological studies, gene expression research uses tissue sampled after many years of illness and resultant gene expression abnormalities may be secondary to medication, co-morbid substance abuse or many other differences not directly related to disease biology.

1.2.2.3 Neuroimaging findings in schizophrenia

Volumes of various brain structures are smaller and cerebrospinal fluid spaces larger when groups of patients with schizophrenia are compared with control groups. These changes may predate illness and relate to genetic risk of SZ as smaller intracranial volumes and larger ventricles have been detected in MRI studies of MZ twins discordant for SZ and normal MZ twins (Baare et al, 2001). However, they provide only partial support for the neurodevelopmental view, as changes in ventricular volumes may be more dynamic than previously realised (Segal, 2001).

1.3 Schizophrenia as a complex genetic disorder

'It seems...very doubtful whether (it) is inherited as a Mendelian unit character, but that (it) is inherited is obvious for it occurs much oftener in families than in the general population, where it is between 1 and 2 percent'

Castle (1916) -referring to the mode of inheritance of 'insanity'

1.3.1 Family studies

That schizophrenia is a familial disorder was well recognised by 1916; in describing *dementia praecox* Kraepelin noted that ~70% of his case series had an affected family member. More than 100 family studies performed during the twentieth century confirmed that having an affected relative is the most significant risk factor for schizophrenia (Cardno & Gottesman, 2000). The lifetime morbid risk (MR) of schizophrenia in the general population is approximately 1% but increases to approximately 10% in siblings or offspring of individuals with schizophrenia. Risk increases exponentially with an increase in the number and degree of relatedness of

affected relatives (see Figure 1.1). For example, the risk increases to 17% if one parent and one sibling are affected but to 46% if two parents are affected. Critics of these data have commented that early studies lacked standardized diagnostic criteria, were not sampled systematically, did not interview blind to affectation status of the index case and in certain cases did not have a matching control group. A re-analysis of family data, which excluded studies failing to meet these criteria, yielded seven studies from which 15 cases of schizophrenia were identified in 3,035 lifetimes at risk in the control families and 116 cases in 2,418 lifetimes at risk in families of an affected proband, confirming a ten-fold increased risk of schizophrenia in relatives of schizophrenia patients (Kendler and Diehl, 1993).

'From the start the child who is destined to become schizophrenic in adult life has, it is said, an unusual and unsuitable emotional environment...the mother of the schizophrenic is often a challenging specimen-cold, detached, and forceful; the father, too may be odd and is often the non-dominant partner'
Hays (1967).

For much of the twentieth century, particularly in the US, psychodynamic approaches dominated family research and the primacy of environmental influence was assumed. This changed with the development of formal genetic epidemiological methods and their application to the natural experimental conditions provided by adoption and twin-births. Applying these paradigms, a new generation of schizophrenia studies investigated relative genetic and environmental effects in illness development.

1.3.2 Twin studies

In 1967, Margit Fischer introduced the twin study strategy to schizophrenia research. By cross-referencing the Danish National Psychiatric and Twin Registers she identified a cohort of 62 SZ twin-pairs with pair-wise concordance rates of 48% for monozygotic (MZ) and 19% for dizygotic (DZ) co-twins (Fischer et al, 1973). Pooled data including 17 subsequent studies indicate rates in MZ twins of 48% and in DZ twins of 12%. From these data it is possible to estimate the proportion of individual differences in risk in a population at a given time that are due to genetic differences between individuals (termed heritability (h^2)). This can be calculated from the

equation $h^2 = 2(r_{MZ}-r_{DZ})$, where r is the correlation in the trait between twin pairs (Kimble 1993). Based upon the five most recent, systematically ascertained samples, the heritability for DSM-IIIR or ICD-10 defined schizophrenia is estimated at 75-85% (Cardno & Gottesman, 2000). These studies (almost exclusively of European samples) confirm a substantial genetic contribution to schizophrenia in these study populations. Variation in the estimation of heritability may be because of differences in the populations being studied due to methodological (e.g. in ascertainment, phenotype measurement or statistical methods), environmental or other differences. The risk of schizophrenia in different classes of relatives is summarized in Figure 1.1

1.3.3 Adoption studies

Adoption offers an alternative approach to examining relative genetic and environmental influences. Adoption studies investigate the effect of being adopted into, or out of, a family with a specific disorder using a variety of study designs. The original adoptee study found that 5 of 47 (16.6%) adopted-away offspring of mothers with schizophrenia themselves had schizophrenia compared to 0 of 50 adopted-away offspring of control mothers (Heston, 1966). This finding met with support from a study based on the Danish National Psychiatric Register, which investigated the siblings and half siblings of adult probands who had been adopted (Kety, 1988). A re-analysis of the Danish study using DSM-III criteria found an increased rate of schizophrenia in biological first-degree relatives of probands (7.9%) but not in adoptive family members (0.9%) who share the same environment as probands (Kendler et al, 1994). These findings have since been replicated in other adoptee studies and by using cross-fostering and adoptee family study designs (Tienari et al, 2000; Kety et al, 1994; Wender et al, 1974; Rosenthal at al, 1971).

1.3.4 Mode of inheritance

Taken together these studies indicate a substantial genetic component to schizophrenia susceptibility. The rate at which MZ twins are discordant for the disorder, despite sharing the same genes, indicates an incomplete relationship between genotype and phenotype. Studies of the recurrence risk in various classes of relatives excludes the possibility that schizophrenia is (in most cases) due to a single gene, or collection of single gene disorders, even allowing for models of incomplete

penetrance (Gottesman & Shields, 1967; McGuffin et al, 1995). Instead, schizophrenia is a complex genetic disorder where genes of moderate or small effect interact with environmental risk to bring about disease expression. The number of genes, the degree of risk conferred by each gene and the relationships between genes are almost completely unknown. Statistical modelling of the contribution of individual genes to a complex disorder uses the expression λ_s (the relative risk to siblings of developing the disorder resulting from possession of the risk allele). The data for recurrence risks in the relatives of probands with schizophrenia are incompatible with the existence of a single locus of $\lambda_s > 3$ (Risch, 1990). This has been borne out by recent genome-wide linkage studies of sufficient statistical power to exclude the presence of individual susceptibility loci contributing this level of risk (Williams et al, 2003; deLisi et al, 2001).

1.3.5 Demarcating the schizophrenia phenotype

The authors of DSM-IV and ICD-10 are unlikely to have defined a clinical entity that precisely maps to the genome. As yet, the nature or number of biological steps between genotype and clinical phenotype are undetermined. In designing molecular studies we are limited by the clinical phenotype in our ability to identify carriers of the 'risk' genotype, yet the boundaries of this phenotype are artificial (Kendler et al, 1994). Patterns of co-aggregation in families indicate a genetic relationship between schizophrenia, other psychotic disorders (including affective disorders), and schizotypal, schizoid and paranoid personality disorders. From this has developed the concept of a spectrum of disorders that are 'to some extent genetically related to schizophrenia' (Kety et al, 1988).

Including schizophrenia spectrum disorders (such as schizoaffective disorder and SZ-related personality disorders) in a broader schizophrenia spectrum phenotype may increase study power by assigning 'affected' status to additional risk genotype carriers. Data from the Finnish Adoptive Family study appears to support this strategy. This study identified a morbid risk for the broader phenotype of 22.5% in adoptees whose mother also had a schizophrenia spectrum disorder; compared with a morbid risk for SZ of 5.34% in adoptees whose mothers had SZ. The rates of schizophrenia spectrum disorder (4.4%) and SZ (1.74%) were much lower in a low-

risk adoptee group (Tienari et al, 2003). To date, most researchers have focussed on the narrow SZ phenotype, principally because of the robust evidence for its heritability but also because of a lack of consensus as to the boundaries of the schizophrenia spectrum and concerns that including schizophrenia spectrum disorders could come at the expense of diagnostic reliability.

Including other psychotic disorders (e.g. bipolar affective disorder) in SZ analysis has also been mooted. However, twin research indicates that the major psychotic disorders do not lie on a single continuum of liability: different psychotic disorders may have common but also unique elements (Cardno et al, 1999; Cannon et al, 1998). Studies of the offspring of affected and unaffected MZ twins indicate that the risk of schizophrenia-like psychosis is similar in the offspring of both groups (Gottesman and Bertelsen, 1989; Kringlen and Cramer 1989). Recognising that genotype carriers may not express a psychiatric phenotype, many groups have investigated proxy-measures of susceptibility, termed endophenotypes, as a method of increasing study power. As yet, the heritability of specific endophenotypes is less well established than for the clinical phenotype militating against this approach.

1.3.6 Molecular studies of schizophrenia

Investigators seeking to identify schizophrenia susceptibility genes have employed a range of complementary research strategies including positional cloning (employing cytogenetic, linkage and linkage disequilibrium (LD) mapping) and candidate gene studies. These methods have been applied, with remarkable success, to identify genes for Mendelian disorders where a direct relationship between manifestation of illness (phenotype) and genetic variation (genotype) is assumed. However, even for Mendelian disorders this assumption is not always justified, as mutations in a gene can be expressed as different Mendelian disorders or can be modified by other gene effects (Krakow et al, 2004; The US-Venezuela Collaborative Research Project 2004). Applying this approach to a complex genetic disorder such as SZ raises additional issues. Particular difficulties in defining the SZ phenotype for genetic research have already been described (Section 1.3.5). In addition, schizophrenia susceptibility may involve different combinations of genes contributing to illness risk in different patient populations (locus or genetic heterogeneity). The impact of a specific genetic

variation may vary at the level of the individual (gene penetrance) and different variants of a gene may have different effects on illness susceptibility (allelic heterogeneity). Differences in post-translational modification of DNA, imprinting and mitochondrial inheritance may also influence the expression of the clinical phenotype.

1.3.6.1 Cytogenetic abnormalities and schizophrenia

Cytogenetic methods contribute to localising susceptibility genes by identifying chromosomal abnormalities such as deletions or translocations, which segregate with a disorder in families or are found more commonly in cases than in control populations. Of the many reported associations between SZ and chromosomal anomaly two have provided consistent evidence of a contribution to SZ risk, albeit in specific populations (reviewed in Murtagh et al, 2005).

In a large Scottish family a balanced reciprocal translocation t(1;11)(q42.1;q14.3) co-segregates with SZ (St. Clair et al, 1990). Interestingly, this translocation also co-segregates with a broader phenotype including affective disorders within this pedigree. The translocation was discovered to disrupt a gene called *DISC1* on chromosome 1 which is located in a chromosomal region of putative SZ susceptibility identified by several linkage studies (described below in Section 1.3.6.2).

Small interstitial deletions of chromosome 22q11 are among the commonest known chromosomal abnormalities. These cause velo-cardio facial syndrome (VCFS), which increases the risk of psychosis by at least 20-fold, and also presents with cranio-facial dysmorphology and congenital heart disease. The contribution of VCFS to SZ in the general population is likely to be small, as the condition affects one in 4000 live births, and cytogenetic studies in schizophrenia rarely identify VCFS cases (Murphy, 2002). However, this region of chromosome 22q11 has also been implicated in SZ linkage studies suggesting a more general role in schizophrenia or psychosis susceptibility.

1.3.6.2 Linkage studies

Since the early 1990's more than 20 genome-wide linkage studies of schizophrenia have been reported from multiply affected pedigrees or larger collections of smaller

families. Four studies have met criteria for significant linkage at a genome-wide level ($p<0.05$), albeit for different loci: chromosome 6p24-22 (OMIM 600511)(Straub et al, 1995); chromosome 13q32-34 (OMIM 603176) (Blouin et al, 1998); chromosome 1q21-22 (OMIM 604906) (Brzustowicz et al, 2000); and chromosome 10q25-26 (Williams et al, 2003). For each region suggestive evidence for linkage has also been reported in other studies (Moises et al, 1995; SLCG 1996, Maziade et al, 1997; Schwab et al, 2000; Gurling et al, 2001; Lindholm et al, 2001; Mowry et al, 2000). At six other loci suggestive evidence of linkage has been reported in more than one study 8p21-22 (OMIM 603013); 6q21-25 (OMIM 603175); 22q11-12 (OMIM 600850); 5q21-33 (OMIM 181510); 10p15-11 (Kendler et al, 2000) and 1q42 (OMIM 606271).

Interpretation of these findings is complicated by differences between studies in the phenotype employed (broad or narrow); the number of mode of inheritance models tested (for parametric linkage analysis); differences in markers selected between studies and between marker locations in earlier genetic maps. No single region has been replicated in a majority of genome scans: this may relate to the differences between samples, false positive findings or individual studies being underpowered. For example, most of these studies lack statistical power to identify individual gene effects of the size predicted by statistical model of family data (single loci of $\lambda_s \leq 3$).

As a partial solution to this problem several meta-analyses of schizophrenia genome scans have been performed. The largest of these included 20 completed genome scans representing 1,200 affected families, but did not include complete data sets for the two largest individual studies (Lewis et al, 2003). This meta-analysis found significant statistical evidence for linkage to 12 chromosomal regions and lesser evidence to implicate a number of other regions. The regions providing significant evidence for linkage were 2p12-q22.1, 5p23.2-q34, 3p25.3-q22.1, 11q22.3-q24.1, 6pter-q22.3, 2q22.1-q23.3, 1p13.3-q23.3, 22pter-q12.3, 8p22-p21.1, 6p22.3-p21.1, 20p12.3-p11, and 14pter-q13.1 (see Figure 1.2). An earlier meta-analysis by Badner and Gershon (2002) taking a different approach examined linkage data from a smaller (but overlapping) set of published genome scans, found significant evidence for schizophrenia susceptibility loci at 8p, 13q and 22q.

Taken together, the linkage studies provide significant evidence for the involvement of the 6p24-22, 13q32-34, 1q21-22 and 10q25-26 chromosomal loci. Many of the other loci implicated across studies and supported by meta-analysis may also harbour schizophrenia susceptibility genes. Confirming these findings will require larger samples for linkage analysis with sufficient power to detect moderate individual gene effects and ultimately proof by identification of the genes themselves. This will be a major undertaking as the implicated chromosomal loci are large and each may contain hundreds of genes.

1.3.6.3 Association studies

At the onset of this project many putative functional candidate genes could be identified from the neuropharmacological literature. Genes involved in regulation of DA, glutamate and serotonin neurotransmission were extensively tested for association. The list of individual positive findings defies succinct summation, but only two genes have received sufficient confirmatory support to indicate that they represent true positive findings. Although more negative than positive reports have accrued across many genes, definitive exclusion of small genetic effects can only be made in a few cases. As with the linkage literature, this is because most reported studies have been of small sample size and the power of individual studies for replication is often based on an overestimate of effect sizes in the initial studies (Göring et al, 2001).

In 1992, association was reported between schizophrenia and homozygosity for a Ser9Gly polymorphism in exon 1 of *DRD3*, the gene encoding the dopamine D3 receptor (Crocq et al, 1992). Many replications and non-replications were subsequently reported. A meta-analysis of these data, including over 5,000 individuals confirmed association, albeit of small effect size (OR=1.23; p=0.0002) (Williams et al, 1998). A more recent meta-analysis of 48 studies excluding the original report also supported the association, but demonstrated a relative risk of ~1.1 (Lohmueller et al, 2003).

In 1996, Williams et al, published evidence for involvement of the *5HT2a* receptor gene in schizophrenia susceptibility. Once again, this finding received mixed support

from subsequent studies. In 1997 a meta-analysis of all available data (including ~3,000 individuals) supported association with the *C* allele (OR=1.2; p=0.0009). A more recent meta-analysis of 28 studies (excluding the original report) provided further support for this association, but indicated a relative risk of only 1.07 (Lohmueller et al, 2003). However, there is no convincing data that this polymorphism or DNA variants in LD with either marker in either gene has functional significance (Bray et al, 2003). Therefore, if *DRD3* and *5HT2A* play a role in schizophrenia susceptibility, their contribution to risk of illness is minor.

Genetic association studies have been heralded as a powerful method of identifying susceptibility genes (Risch & Merikangas, 1996). An important caveat is that study power is dependent on the prior probability that the candidate gene being investigated is involved in disease pathogenesis. For a disorder such as schizophrenia, because our understanding of pathogenesis is limited, the prior probability of success for each gene investigated is extremely low. Restricting studies to genes compatible with current understanding of schizophrenia, for example those involved in synaptic functioning or cell signalling could still plausibly implicate ~10,000 genes. Methods of increasing this prior probability are discussed in the next chapter.

1.4 Specific hypotheses for this project

1.4.1 Establishing a psychosis sample for molecular research

The project was made possible by the award of a Wellcome Trust Mental Health Training Fellowship to the author. The aim was to establish a case-control psychosis sample from the Irish population for genetic research of schizophrenia and other psychotic disorders. A key objective was to gather optimal phenotypic data using the methods outlined in Chapter 3. The contribution of individual genes to psychotic disorder susceptibility is predicted to be modest and power calculations (see Chapter 3) indicate that relatively large samples will be required to detect such effects. The limitations of having a single researcher attempt to collect such a sample and conduct all of the laboratory work within the time frame of a Training Fellowship was self-evident. By necessity, a significant proportion of time was devoted to securing additional funding to recruit staff and in managing a research team.

1.4.2 Funding applications

During the four years of the project the author was responsible for grant applications to maximise the resources of this project. A project grant was awarded from the Health Research Board in Ireland (HRB), which funded a research nurse (SS) who worked with the author on the Phase 1 sample collection described in Chapter 3. In 2001 a project grant from the same agency funded a PhD studentship (KMcG) to collaborate on the studies described in Chapter 5 and to work on other aspects of schizophrenia and psychosis genetics.

More recently, a programme grant awarded to the author (with Professor Michael Gill) by Science Foundation Ireland (SFI) has funded a clinical researcher (JMN) to collect additional samples (the Phase 2 sample described in Chapter 6). This Phase 2 sample provided the basis for the analyses conducted by the author in Chapter 7. The award also funded a postdoctoral researcher (DM) and two laboratory technicians (KM, NK) to complete genotyping of potentially interesting findings in the extended sample and to investigate additional putative susceptibility loci and molecular pathways. The author was responsible for the co-ordination of both the clinical and laboratory sides of the project in addition to completing laboratory work and analysis. This also involved the training and supervision of clinical research staff and co-supervision of laboratory personnel.

1.4.3 Molecular genetic studies of schizophrenia

This study was designed to investigate a number of putative schizophrenia susceptibility loci, which had previously been identified using linkage analysis by other groups. Based on all available linkage data and particularly on linkage findings reported previously from the Irish population, the original focus was on the chromosome 6p, 8p and 10p susceptibility loci. The aim was to investigate specific functional candidate genes at these loci for association with schizophrenia using a case-control study design. A second goal was to establish the feasibility of a linkage disequilibrium (LD) mapping strategy across one of these chromosomal regions.

During the course of this project evidence for association was reported with genes at the 6p (*Dysbindin-1*) and 8p (*Neuregulin-1*) schizophrenia susceptibility loci (Straub

et al, 2002; Stefansson et al, 2002). Both of these genes were investigated as part of this study and the results are reported in Chapter 4 and 6. Also, reproducible gene expression abnormalities in post mortem schizophrenia samples were reported for a novel gene family (Apolipoprotein-L (OMIM 603743)), which map to the putative chromosome 22q susceptibility locus (Mimmack et al, 2002). Consequently, association analyses were performed across these genes applying an LD-mapping strategy. These experiments are discussed in Chapter 6.

Finally, during the course of the project evidence for association between a number of different genes and schizophrenia susceptibility in this study population was identified. These findings are summarised in Chapter 7, as are methods of investigating relationships between genes in contributing to complex disorder susceptibility. The application of two different methods of analysing gene-gene interactions to the study dataset is also described in this chapter. Methodological considerations relevant to each of these studies are discussed in the next chapter.

Figure 1.1 Schizophrenia risk in relatives of affected probands

(adapted from Gottesman, 1991)

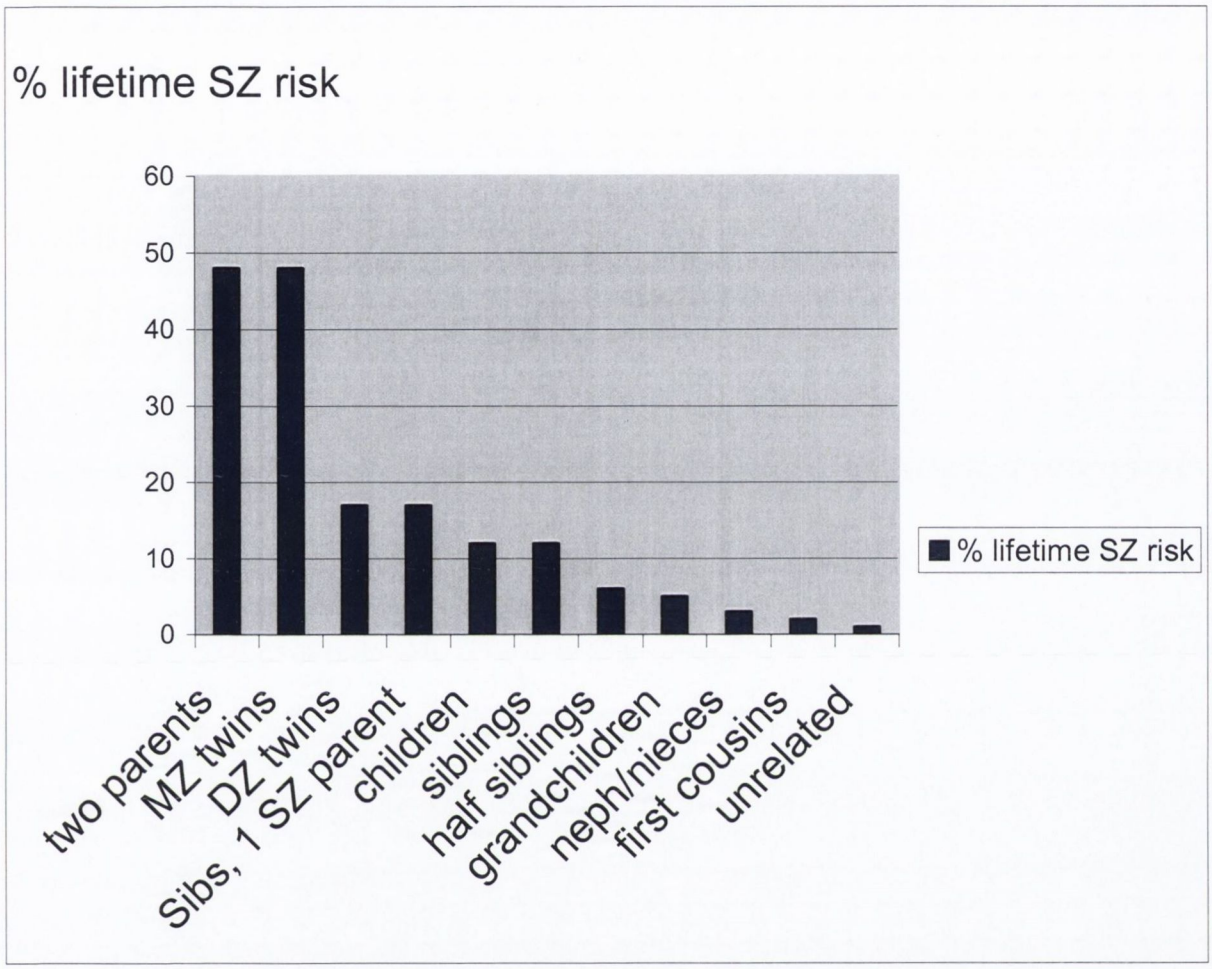


Figure 1.2 **Summary of linkage findings from meta-analysis of SZ genome-scans (Lewis et al, 2003)**

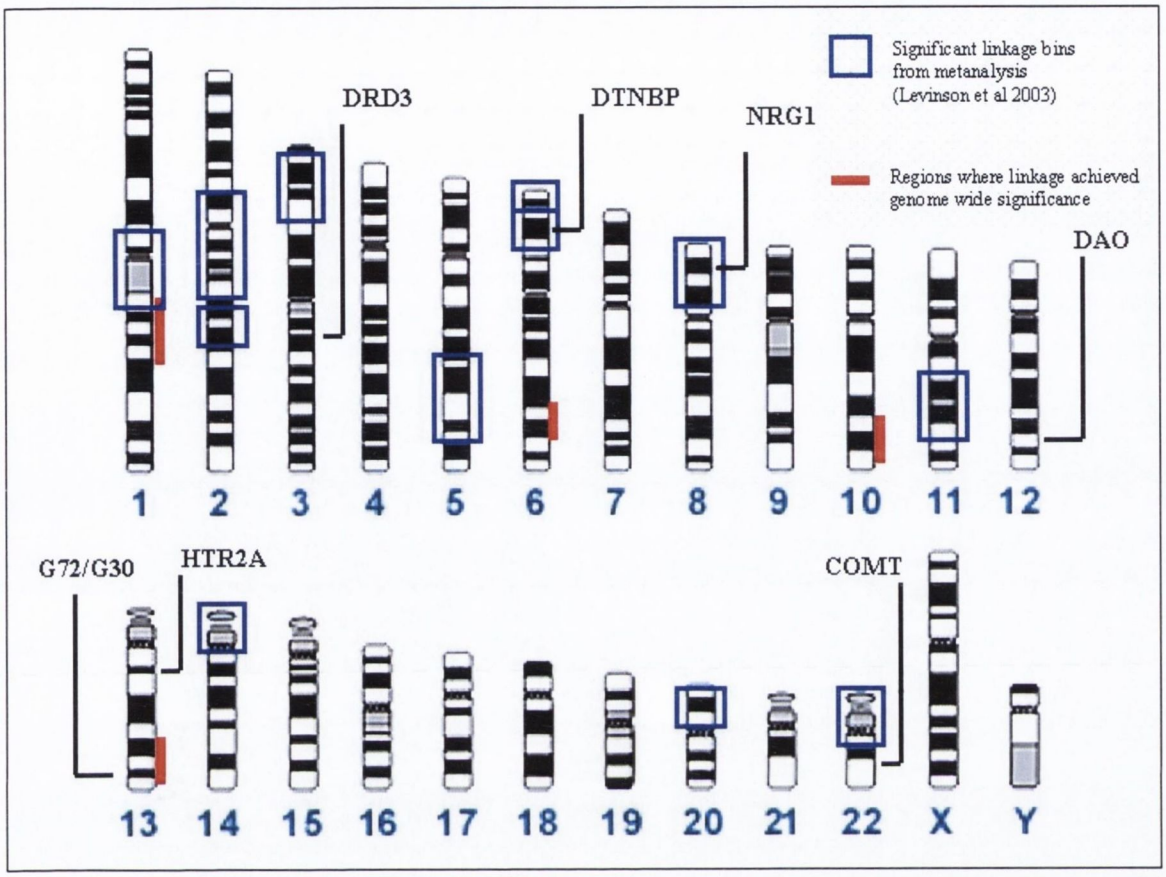


Figure 1.3 Genetic and environmental risk factors in SZ

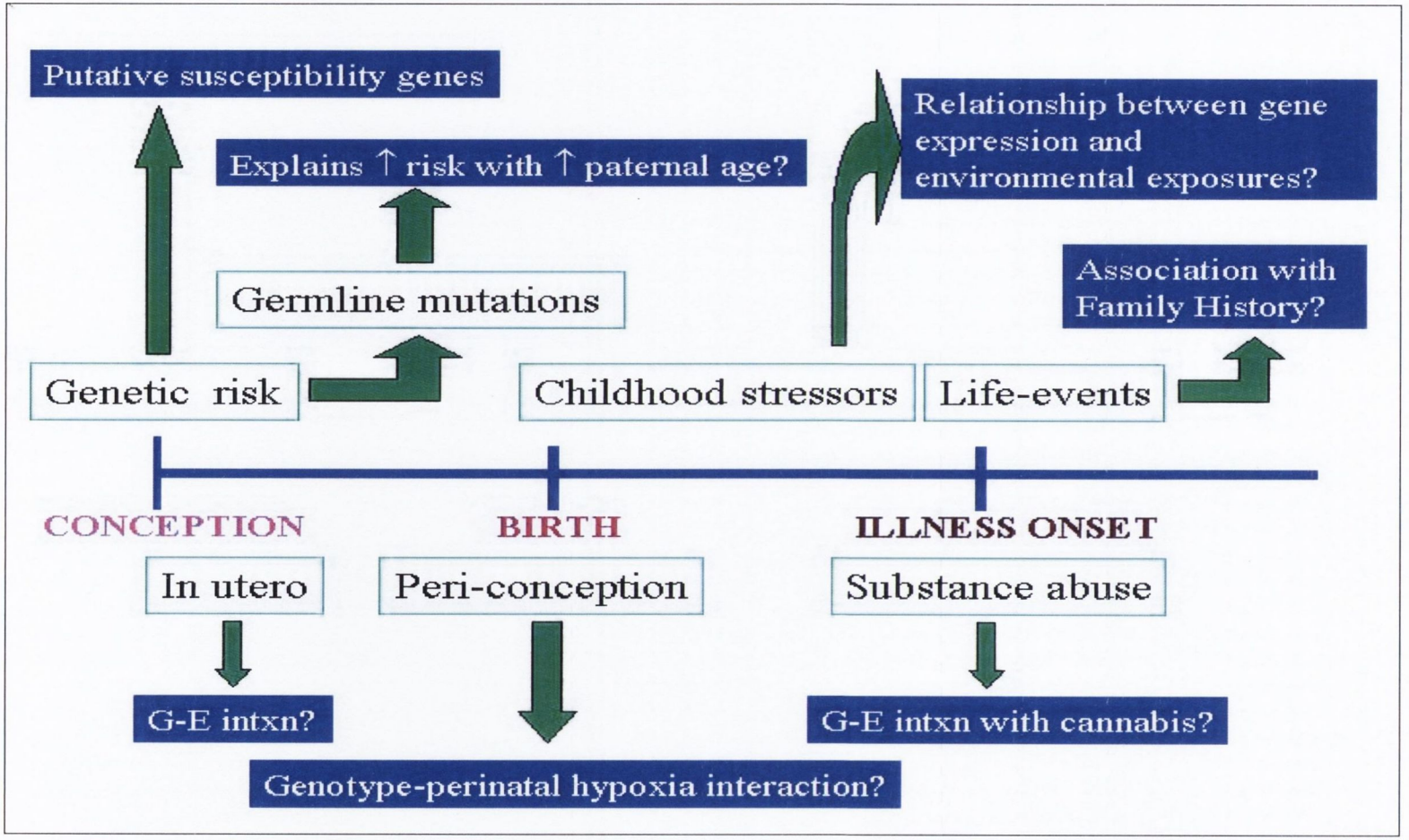
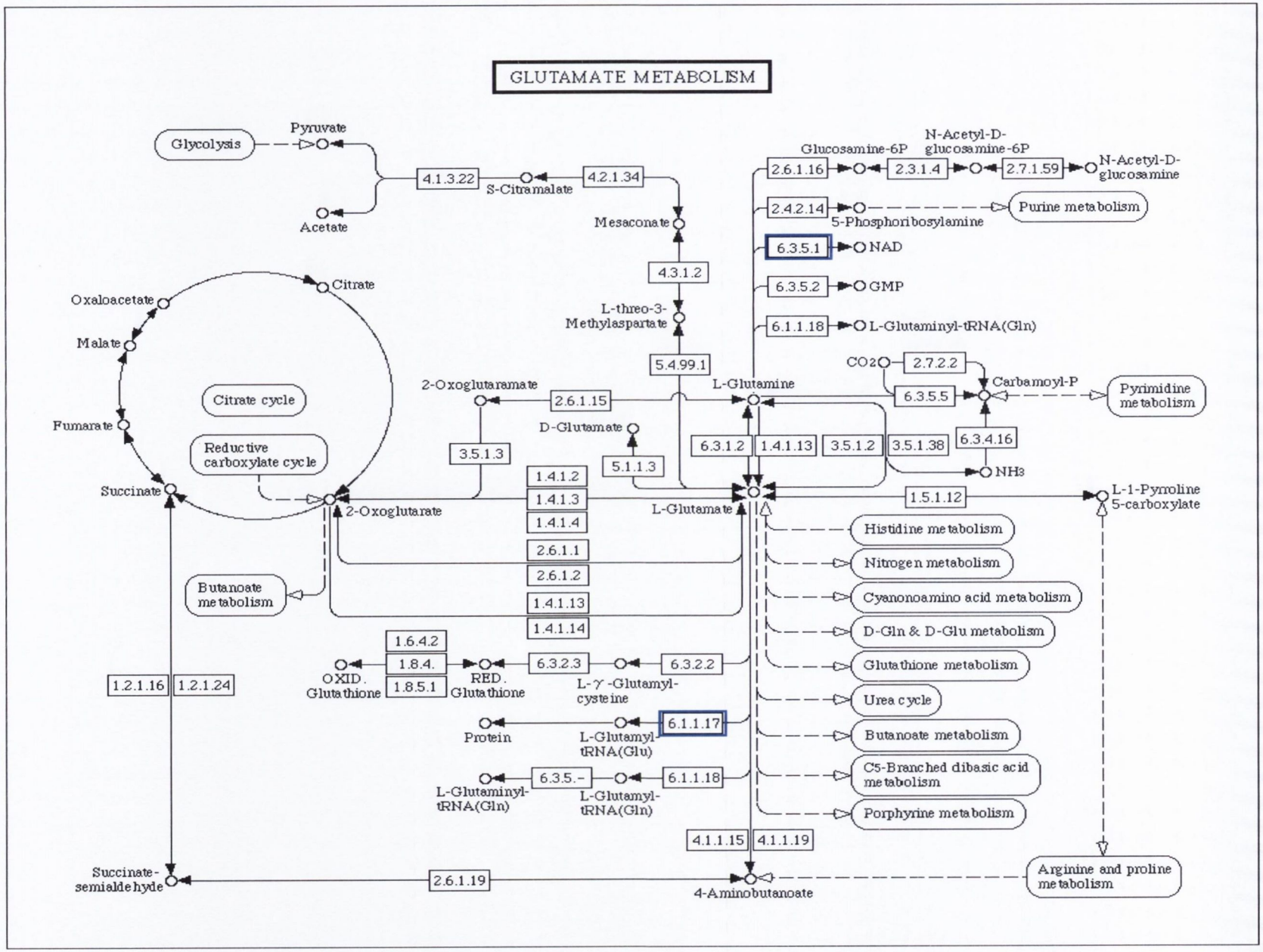


Figure 1.4 The biochemical pathways involved in Glutamate Metabolism



Chapter 2

Methodological issues in complex genetic disorder research

2.1 Introduction

For most, if not all cases of schizophrenia, the effects of alleles at any single susceptibility locus are insufficient to cause disease. Individual gene effects contributing to the phenotype may be modest and expression of the phenotype dependent on interaction between genes, or between genetic and environmental or stochastic factors. This chapter discusses strategies used to identify susceptibility genes under these conditions, with particular reference to the approach employed in this project.

2.2 Positional cloning methods of gene identification

Positional cloning methods aim to localise DNA variation that contributes to disease susceptibility by applying genetic principles. These methods do not require any *a priori* understanding of the function of a gene or variant to identify susceptibility loci. In investigating disorders such as SZ where understanding of disease pathogenesis is poor, this approach is attractive. At present three positional cloning strategies are available:

i) Linkage analyses that investigate meiotic recombination in pedigrees.

ii) Genetic association analyses that may be direct or indirect (based on linkage disequilibrium (LD)).

iii) Admixture disequilibrium mapping.

In practice, these are typically applied in sequence or combination. For example, linkage studies in pedigrees affected with a disease may identify chromosomal regions of susceptibility, which are then investigated by direct or indirect association mapping to identify DNA variants in genes that contribute to disease.

2.2.1 Linkage analysis

Linkage studies investigate families for evidence of co-segregation of genetic markers with a disease phenotype looking for a departure from independent segregation (Mendel's second law). Linkage analysis can estimate the amount of recombination between the marker and the putative disease gene thus indicating the likely distance between the genetic marker and the disease gene. By determining that a marker and disease phenotype co-segregate, it can be inferred that the marker and disease locus are linked and in close physical proximity to each other. This is a factor of meiotic recombination and linkage analysis essentially consists of identifying recombinants and nonrecombinants at marker loci within pedigrees. From this the recombination fraction (θ) is derived by dividing the total number of recombinant gametes by the total number of gametes. Where markers (or a marker and a genetic variation causing a disease phenotype) segregate independently, θ takes the value of 0.5. If θ is less than 0.5, markers are not segregating independently and are linked, meaning that in physical terms they lie close to each other on the same chromosome.

As the recombination fraction is proportional to the physical distance between markers, it is possible to measure genetic distance between markers. One Morgan (1M) corresponds to the length of DNA sequence on which an average of 1 recombination event is expected per meiosis. This is more routinely expressed as centiMorgans (cM), where 1cM corresponds to a genetic distance of $\theta = 0.01$. Genetic distance approximates physical distance, but the relationship is imprecise because the rate of recombination varies across chromosomes and between the sexes. The conventional statistical method of assessing the statistical significance of the value θ is to calculate a LOD (log of the odds) score. This is the common logarithm of the probability that the recombination fraction has some given value, divided by the probability that the value is 0.5. Interpretation of the LOD score requires that Bayesian theory is taken into account whereby the prior odds that two loci are linked is ~ 1:50. The observed odds ratio of a LOD score of 3 is 1000:1 leading to a posterior probability of about 20:1, roughly equivalent to a p-value of 0.05.Where the mode of inheritance is known and the phenotype can be clearly defined, a LOD score of 3.0 (odds in favour of linkage 1000:1) is accepted as statistical evidence of linkage and a score of –2.0 (odds against of 100:1) as evidence of exclusion of linkage.

Linkage analysis has proved most effective in mapping genes for Mendelian diseases, such as cystic fibrosis where there is a strong correlation between the measured phenotype and the inferred genotype. For complex genetic disorders many factors may conspire to weaken the correlation between phenotype and inferred genotype. For example, including individuals with different diseases (phenotypic heterogeneity) will significantly reduce study power. This may occur due to misdiagnosis or the presence of phenocopies (e.g. psychosis due to medical causes in a schizophrenia study). Study power can also be reduced by genetic heterogeneity where mutations at different loci result in the same phenotype (non-allelic heterogeneity). By contrast, where distinct mutations at the same locus contribute to a phenotype (allelic heterogeneity), linkage is maintained. The penetrance at each locus, high or very low frequency susceptibility alleles, and genotyping error can all significantly impact on the power of linkage studies (Thomson & Esposito, 1999; Abecasis et al, 2001).

Linkage studies follow the transmission of DNA markers down generations in pedigrees and investigate whether they segregate with an illness phenotype. Single multiply affected pedigrees, collections of many such pedigrees or even multiple pairs of affected siblings may be studied, and each approach has advantages and disadvantages (further details are provided in Lander & Kruglyak 1995). Early SZ linkage studies investigated small chromosomal regions, but as sufficient genetic markers became available in the 1990's this strategy was expanded taking a genome-wide approach using regularly spaced markers. Despite recent increases in mapping density the resolution of linkage studies remains the number of meioses at which recombination might be observed (Kong et al, 2002). Even for studies of reasonable sample size, very few meioses will be observed in the families between markers that are relatively close together (e.g. <1cM), limiting the resolution that can be achieved with linkage analysis.

Traditional methods of linkage analysis require the assumption of a particular mode of inheritance that explains the transmission pattern of a disorder in pedigrees. Model-free or non-parametric methods of linkage analysis have been developed and applied to affected sibling pairs, for example, in an attempt to combat this problem. These analyses assess the degree of allele sharing at marker loci between related affected individuals and are less affected by misspecification of mode of inheritance. If a

disease mutation exists in a large proportion of families, affected individuals will share alleles at marker loci linked to the mutation more often than expected by chance. Affected sibling pair methods offer three potential advantages over traditional LOD-score methods for investigating complex disorders (reviewed, Holmans & Craddock, 1997). First, a model of disease susceptibility is not required. Second, under certain simulations of oligogenic inheritance affected sib-pair methods are more likely than large, densely affected pedigrees to be informative for linkage (Risch, 1990). Finally, in practical terms it is more efficient to ascertain nuclear families containing pairs of affected siblings rather than large, multiply affected, multigenerational families.

2.2.2 Association studies

Association studies are based on populations instead of pedigrees, and compare frequencies of marker alleles in affected individuals with frequencies in unaffected individuals from the general population or some other control group (Burmeister, 1999). At its simplest a direct association study tests whether a mutant allele (one that disrupts function, structure or expression of a gene) is at a significantly higher frequency in a case group (non-related affected individuals) compared with a control group (unaffected individuals). Where the gene has been chosen because of its likely involvement in disease pathogenesis this can be termed the candidate gene approach. Association-based methods are potentially more powerful than linkage studies for identifying common variants conferring moderate risk of a disorder (Risch & Merikangas 1996). However, study power is dependent on the prior probability that the candidate gene being investigated is involved in disease pathogenesis.

In the positional cloning approach the prior probability is increased by investigating genes at chromosomal susceptibility loci previously identified by linkage or cytogenetic studies. If all genetic variants at a locus are transmitted independently then an association study of a locus could potentially involve testing hundreds of thousands of genetic markers. However, contrary to Mendel's second law, allelic variants in many cases do exist together more often than expected by chance because of the phenomenon of linkage disequilibrium (LD). Because of the existence of LD, much of the DNA variation at a locus may potentially be represented by assaying a

subset of polymorphisms at that locus. From the early 1990's this lead to speculation that LD-mapping could be applied to large chromosomal regions or indeed the whole genome (Risch & Merikangas, 1996).

2.2.2.1 Case-control and family based association methods

The simplest study design for assessing correlation between genotype and phenotype is the case-control method. This method, applied in the studies described in Chapters 5-7, makes the assumption that observed differences in allele frequencies between case and control groups relate to the phenotype being measured. However, because each population has a unique genetic and social history, allele frequencies may vary between and within populations. Allele frequency differences between case and control populations due to background population effects are termed population stratification. For population stratification to affect genetic association studies, two circumstances must be met: variation in allele frequency between groups must be present, but also differences in disease prevalence must exist between case and control groups (Cardon & Palmer, 2003).

Concern that population stratification may lead to spurious reports of false-positive findings has resulted in alternative family-based designs being developed. The most popular of these is the transmission-disequilibrium test (TDT) (Spielman et al, 1993). The classic family based association study design requires an affected proband and their parents. Parental alleles transmitted to the affected proband represent 'case' alleles and the non-transmitted alleles represent 'control' alleles. Because case-control pairs are matched within a family, allele frequency differences in the wider study population are no longer relevant. This study design has limitations and was not applied in this project for a number of reasons. First, ascertaining suitable families is problematic for an adult onset disorder such as schizophrenia. Second, because genotype information from three people is required for each case-control pair, genotyping efficiency is reduced. Third, information is only available where parents are heterozygous for the marker being investigated. For SNP data this will only happen at best, 50% of the time, which significantly reduces study power.

Several methods have been proposed to protect against population stratification in traditional case-control association studies. These start with genotyping additional anonymous genetic markers throughout the genome to indicate the amount of background diversity in cases and controls populations (Pritchard & Rosenberg, 1999). In the first method, genomic control, the standard χ^2-statistic is increased in proportion to the degree of stratification (Devlin & Roeder, 1999). An alternative method, called structure assessment, attempts to identify subgroups in stratified samples, which can then be matched appropriately. The test for disease association is then a statistical combination of results from each subgroup (Pritchard & Donnelly, 2001). Applying either of these approaches is problematic, principally because of uncertainty as to the number of anonymous markers to be tested but also because the study sample size required to detect significant population differences is unknown. In the original report, Pritchard & Rosenberg suggested that ~30 SNPs may be sufficient to indicate the level of background diversity. With subsequent reports the suggested number has inflated at least 3-fold with substantial implications for genotyping cost.

Population stratification can potentially induce important biases, however, few clear examples of spurious results due to undetected population substructure have been reported. The classic cited example comes from a study of the association between a HLA haplotype and diabetes. In this case the confounding factor was admixture of white European and Pima Indian populations in the sample (Knowler et al, 1988). Morton & Collins (1998) have suggested that if the basic principles of good epidemiological design are followed such pitfalls can be avoided. From empirical data, under most conditions, the bias introduced by population stratification is likely to be small unless there are major correlated differences in allele frequency and disease prevalence across ethnic groups and the ethnic groups in a study cannot be distinguished (Wacholder et al, 2000).

2.3 Linkage disequilibrium (LD) mapping

2.3.1 Definition of linkage disequilibrium

Linkage disequilibrium (LD) is the non-random association between alleles at different loci in a population. This association arises because variants share a common population ancestry. When a mutation event occurs the mutant allele is transiently

associated with alleles at each polymorphic locus on the mutant chromosome (the ancestral haplotype). After multiple generations, and many meioses, this association dissipates until only those alleles that are physically close to the mutant allele survive as an intact haplotype from the original chromosome.

For gene mapping purposes, if an allele at one locus predicts which allele will be found at another locus, one of these observations is redundant. If there is significant LD at a genomic level, understanding LD patterns in a population could increase the efficiency of gene mapping strategies as most DNA variation at a chromosomal locus could be represented by the subset of markers that represent the haplotypes across the region. In considering such a strategy, it is important to realise that LD patterns within a population are dependent on the mechanisms for exchange of genetic material at meiosis (recombination and gene conversion); factors influencing individual loci (mutation/natural selection); and factors that affect the whole genome (population history). The rest of this chapter considers these factors individually, before looking at methods of measuring LD and the assumptions underlying LD mapping approaches for complex human disorders. In the last section, experimental data on LD structure in human populations is reviewed.

2.3.2 Recombination

Recombination denotes the reciprocal exchange of chromosomal fragments during meiosis. Current models of meiosis suggest two methods of allelic exchange that arise when homologous chromosomes overlap for strand exchange, crossing over events and gene conversion.

Crossing over involves the exchange of large chromosomal segments and on average, one such event occurs per chromosomal arm at meiosis. Over time the number of crossing over events between two loci is dependent on the recombination fraction between the loci and the number of doubly heterozygous crossing over events that have occurred between the current population and the ancestral mutant chromosome. The recombination fraction varies across the genome, between the sexes and within individuals from one meiosis to the next (Kong et al, 2002). Assuming random mating, the number of crossing over events is related to the number of generations

separating the sampled population from the founder population. Greater LD is generally expected between markers in populations that have had few meioses (recent populations with small numbers of founders). For this reason, recently founded populations may be useful for LD mapping studies because LD may be present over large regions. However, as a consequence of this, mapping resolution will be poor and more ancient populations with LD over shorter distances may provide more accurate resolution for fine mapping.

Gene conversion is a different method of allele exchange. In this case, when homologous chromosomes overlap during meiosis only small stretches of DNA, termed 'conversion tracts', are exchanged. The average length of gene conversion tracts is estimated to be 55-250 base pairs, but this is based on experimental data from a small number of loci (Zangenberg et al, 1995; Jeffreys & May, 2004). Current models of gene conversion estimate that conversion events will rarely affect more than one SNP, because tracts are short or SNPs are sparse (Wiehe et al, 2000). For this reason, the effects of gene conversion and crossing over events on LD will be very different. Consider a gene conversion event that affects the middle SNP in a triplet of SNPs in complete LD. Gene conversion will create the equivalent of two crossing over events between the central and two outer SNPs. The outer SNPs can still be strongly associated with each other but neither is associated with the middle SNP.

The relationship between LD and gene conversion is independent of marker spacing for distances greater than the length of a conversion tract (Andolfatto & Nordborg 1998). Therefore, gene conversion can disrupt LD between linked markers at a faster rate than predicted by the recombination fraction, but has less of a role in disrupting LD between more distant markers. Gene conversion may be frequent in the human genome (Ardlie et al, 2001; Przeworski & Wall, 2001; Frisse et al 2001). The relative contribution of gene conversion to the decay of short range LD across the genome is as yet unknown. A recent study modelling available SNP data of gene conversion rates across chromosome 21 by Padhukasahasram et al (2004), suggests that this contribution may be substantial. If confirmed at a genome-wide level, this would limit the usefulness of an LD map based only on crossing over events.

2.3.3 Genetic drift

Genetic drift occurs where gene or haplotype frequencies change between generations due to random sampling of gametes. For most 'disease' phenotypes, at a given locus 'disease' alleles will be less common than wild-type alleles. Allele frequencies at this locus will fluctuate from one generation to the next because of the chance assortment of chromosomes at meiosis. The extent of this fluctuation will depend on the relative frequencies of the alleles and also on the total population of chromosomes. The resultant variance in allele frequency from one generation to the next will be more dramatic where the population of chromosomes carrying the 'disease' allele is substantially smaller than the number of chromosomes carrying the 'wild-type'. For larger populations, the variance in allele frequency between generations will be smaller, and changes in allele frequency due to genetic drift will occur more slowly. As populations expand the absolute numbers of 'disease' alleles increase, the variance between generations decreases, reducing genetic drift. Where genetic markers are in close proximity to such a disease locus this will generate LD. In the Irish population, the influence of genetic drift has been demonstrated, at different geographical locations, but the effect is not pronounced (Hill et al, 2000).

2.3.4 Selection

Natural selection is the process whereby inherited genetic variation between individuals will affect their ability to survive and reproduce successfully. Depending on the advantage or disadvantage conferred by a specific variant selection can be positive or negative. A negative selection reduces fitness and without new mutations leads to a reduction in frequency of the mutant allele. The opposite case may also occur, the classical example being in autosomal recessive disorders where heterozygosity at the locus improves biological fitness, termed the 'heterozgote advantage'. For example, heterozygotes for the thalassaemia group of disorders or sickle-cell anaemia show some resistance to malaria and led to a heterozygote advantage in African populations.

A number of authors have considered the possible effects of positive selection on LD between a disease locus and linked loci (Terwilliger & Weiss, 1998; Nachman 2001). This may be particularly applicable where there is allelic heterogeneity at a disease

locus. If a specific risk allele is positively selected for it will increase in frequency in the expanding population of disease chromosomes. The frequency of other markers linked to this allele will also increase in frequency in the disease population, generating LD. This is termed the 'hitchhiking effect'. Related to the issue of selection is the mutation rate within a disease gene. With a greater mutation rate comes more allelic heterogeneity and more risk haplotypes that contribute to susceptibility at the locus. Positional cloning methods will be less powerful in identifying susceptibility stemming from many different alleles or haplotypes than from a single risk allele/haplotype.

2.3.5 Other population factors

LD between markers in a population is related to the number of recombination events between markers. Inbreeding and assortative mating can reduce the rate at which LD decays between markers. This is because there may be less doubly heterozygous recombination events than would occur in outbred populations. Even where mating is random, factors related to the evolutionary history of the population may influence LD. Population bottlenecks result from temporary, but significant reductions in population size (for example due to famine). If the population size is reduced sufficiently, the number of haplotypes represented in the population may also be reduced so that not all possible haplotypes are represented in the proportion expected by the frequency of each allele. If the population bottleneck has occurred relatively recently there may be extensive LD in the population because of the reduced opportunity for recombination to decay haplotypes.

2.3.6 Admixture and admixture mapping

Admixture is the interbreeding of two different populations. If allele frequencies at two loci differ between these populations, admixture will produce LD between the loci in the resultant population (Chakraborty and Weiss, 1988). The amount of LD between loci will depend on the differences in allele frequencies in the founder populations and the relative contribution of each population to the new admixed population.

Where disease-causing variants differ in frequency between different populations, admixture mapping has been proposed as an alternative positional cloning strategy. This approach involves studying populations descended from the recent mixing of ethnic groups (such as Americans of European or African descent). For example, hypertension is more common in West Africans, so screening European Americans for regions where the proportion of African ancestry is higher may be an efficient method of identifying susceptibility loci. Several admixture-mapping panels are now available for whole genome scans using admixture-generated LD mapping (Patterson et al, 2004). In addition to having a study population of this type available, the approach also depends on the assumption that disease causing variants are known to differ substantially in frequency across the populations being investigated. Current evidence suggests that this is not the case for schizophrenia where the incidence of disease is similar across different ethnic groups. However, this approach may be powerful if different genes contribute to susceptibility in different populations.

2.3.7 Measuring linkage disequilibrium

Many pair-wise measures of LD have been proposed, but in practise only the measures D' and r^2 are widely used. Both of these are variations of the basic pairwise-disequilibrium coefficient D. The statistical significance of linkage disequilibrium can be measured by a standard contingency table. Given two biallelic markers at two loci, the frequencies of the alleles at these loci are P_A, P_a, P_B and P_b.

	B	B
A	P_{AB}	P_{Ab}
a	P_{aB}	P_{ab}

A Fisher exact test measures the probability of obtaining the observed result under the null hypothesis of independence of the two SNPs. This can be expressed as:

$$D=P_{AB}-P_AP_B$$

This is the difference between the probability that two marker alleles will be observed on the same haplotype and that they will be observed independently in the population. The range of possible values of D is a function of the allele frequencies:

$$D_{max} = \min (P_A P_b - P_B P_a) \text{ if } D>0$$

$$D_{max} = \min (P_A P_B - P_a P_b) \text{ if } D<0$$

Whether D is positive or negative depends on how the alleles are labelled. To avoid this the measure D' is used where $D'=D/D_{max}$. This measure is scaled between 0 and 1; a value of 0 implies independence and 1 implies that all copies of the rare allele occur exclusively with one of the two alleles at the other marker. D' has several disadvantages, for example it may inflate the level of LD in small samples. An alternative measure of LD is the square of the correlation coefficient between the loci:

$$r^2 = D^2/p_A p_a p_B p_b.$$

This measure has several properties that make it more useful. The sample size required to detect statistically significant LD is inversely proportional to r^2. As previously an r^2 value of 0 also implies independence, but $r^2=1$ has a more strict interpretation than the equivalent D' value: $r^2=1$ only when each occurrence of an allele perfectly predicts the allele at the other locus and the marker loci have identical allele frequencies (Zondervan & Cardon, 2004). D' can reach a value of 1, where the allele frequencies vary widely, as this value only reflects the correlation since the most recent mutation occurred.

2.4 LD patterns in human populations

2.4.1 Early studies of LD structure in human populations

Establishing the pattern of LD at a genomic level in human populations has important implications for future association studies. The extent of correlation between markers across the genome and conservation of this pattern across populations will define the feasibility of genome-wide association analysis. Should there be substantial correlation between markers; a large proportion of total genetic variation could be represented in association studies by genotyping a smaller subset of markers. At the

start of this project, the investigation of LD focussed on average LD levels across the genome. The estimates of useful LD range from 3kb to 60kb from experimental human data, suggesting that ~20% of markers separated by 0.25cM (~250kb) show significant LD. Although physical proximity increases the likelihood that markers will be correlated, local variation in LD is substantial (reasons for this have been considered in Section 2.3). Many adjacent markers are independent, whereas markers separated by millions of base pairs may be correlated (Weiss & Clark, 2002). This limits how informative average LD data is in selecting markers to capture local genetic variation. The wide variation in the estimated range of LD related as much to differences in LD between chromosomal regions, due to local rates of mutation and recombination, as to differences between populations in the distance over which LD is detectable (Abecasis et al, 2001; Reich et al, 2001; Wright et al, 1999).

2.4.1 LD patterns in human populations: a regular structure?

In 2001 a hitherto unrecognised pattern to LD structure emerged (Daly et al, 2001). A series of papers, investigating higher marker densities across larger genomic regions, suggested that LD patterns conform to a more regular structure (Daly et al, 2001; Reich et al, 2001; Patil et al, 2001; Gabriel et al, 2002; Dawson et al, 2002; Phillips et al, 2003). Daly et al, (2001) described a high-resolution analysis of the haplotype structure across 500 kb of chromosome 5q31 using 103 common (<5% minor allele frequency) SNPs in a European population. They identified 11 blocks of consistently high LD across the region, which contained 75% of the total sequence but showed no evidence for recombination and very limited haplotype diversity. Regions of higher recombination separated these blocks. Subsequent studies identified similar patterns of successive regions of high and low LD within the HLA region, at other loci and across chromosomes 21 and 22 (Reich et al, 2001; Patil et al, 2001; Gabriel et al, 2002; Dawson et al, 2002; Phillips et al, 2003).

From these studies it was proposed that the genome is highly structured and arranged into discrete 'haplotype blocks' of high LD separated by short segments of very low LD, typically termed recombination 'hotspots' (Daly et al, 2001). Within haplotype blocks, markers are highly correlated and a large percentage of genetic variation at a block could potentially be represented for association analyses using a fraction of the

total available markers. At regions of low LD (hotspots), where the correlation between markers is much lower, most if not all markers would be necessary to define genetic variation. If the boundaries of these regions could be defined predictably, it was proposed that this block structure could inform a more efficient marker selection for LD-mapping across a locus. As an extension of this, Patil et al, (2001) proposed that all common haplotypes in the genome could potentially be tested in a genome-wide association study using 300,000-1,000,000 SNPs.

In support of this hypothesis, data from other organisms and from humans demonstrates that recombination events are more likely to occur in some parts of the genome than in others (reviewed by Kauppi et al, 2004). Recombination 'hotspots' of 1-2Kb in length have been identified in the human genome (Templeton et al, 2000; Jeffreys et al, 2000; Jeffreys et al, 2001). However, most of these studies have investigated chromosomal regions previously identified as having a high rate of recombination. There is insufficient human data available as yet to indicate how representative these 'hotspots' are of recombination across the human genome. Evolutionary modelling of haplotype-block distribution data across chromosome 19 provided a close fit without requiring recombination hotspots to explain the distribution of blocks (Phillips et al, 2003).

As a result of these studies the National Institutes of Health funded the Haplotype Mapping (HapMap) Project, an international genome-wide initiative to identify common haplotype blocks in multiple human populations (http://www.hapmap.org). This project made its first public data release in November 2003 and is due for completion in 2005.

To what extent this project will inform future association studies, and the feasibility of genome-wide association analysis, depends on a number of critical considerations. First, that block boundaries are stable and that haplotype diversity remains stable as the marker density increases in investigated regions. Second, identifying blocks does not indicate their value for association analysis *per se*, as other evolutionary factors, random recombination events or mutations could contribute to this structure. Evidence of conservation of blocks between populations would increase their usefulness for association mapping. In the absence of such conservation,

understanding the events that had contributed to LD-structure at specific loci in the specific populations being studied would be a necessary perquisite to LD-mapping. Finally, the focus of HapMap is on identifying common alleles and haplotypes, there role in disease susceptibility is uncertain and is only likely to be established experimentally.

For the purposes of this project, comparisons of haplotype structure across the *NRG1* locus in three European populations were made. These are described in Chapter 5. In completing the association analyses in Chapter 6, the stability of haplotype structure across the *APOL* locus was tested by comparing available CEPH marker data published by Dawson et al, (2001) with more recent HapMap data at a higher marker density in the same subjects (HapMap, 2004).

2.4.3 Linkage disequilibrium in the Irish population

Certain features of the Irish population appeared to lend themselves to LD mapping studies. Although a good deal remains to be understood, the present population is thought to have arisen relatively recently (~12,000 yrs) from a small number of individuals that migrated from central Europe and expanded before remaining constant for a prolonged period (Hill et al, 2000; McEvoy et al, 2004). Recent analysis of Y-chromosome haplotypes supports this theory. From a sampling perspective the population is relatively genetically homogenous.

Data from autosomal polymorphisms suggests LD in the Irish population may extend over larger distances than in other populations. In a study of markers from chromosomes 5,6 and 8, LD was significant at the 5% level between 96% of all markers within 0.5cM, 67% of all markers from 0.5-1 cM, and 35% of all markers within 1-2 cM (Kendler et al, 1999). However, this study was based on a relatively small marker set. There is insufficient marker data is available to establish haplotype structure in the Irish population. As yet, it is also uncertain how representative the patterns of DNA sequence variation identified by the HapMap project will be to other populations of European ancestry such as the population of Ireland.

2.5 Complex genetic disorders: common or rare risk variants?

Most of the human disease genes thus far identified contribute to Mendelian disorders. For such disorders the frequency and number of disease causing alleles varies significantly, from diseases caused by a large number of different mutations to those caused by a small number of relative common disease alleles. The allelic structure representing common, complex genetic disease is less well understood, but fundamental to the likely success of positional cloning strategies.

A Common Disease/Common Variant (CD/CV) hypothesis, has been proposed which predicts that the genetic risk for common diseases will often be due to susceptibility alleles of relatively high frequency (>0.01-0.1), that have undergone little selection in early populations and are likely to date back at least 100,000 years (Lander, 1996). By this hypothesis, one or a small number of risk alleles will contribute to disease risk at most susceptibility loci. This hypothesis is supported by the prediction that allelic diversity for neutral alleles in a small founder population is likely to be low and disease-risk alleles common in such a population are unlikely to have been diluted by the effects of new alleles in the rapid expansion of the human population from that point to the present time. Supporters of the CD/CV hypothesis cite a number of examples such as the association of *ApoE*-e4 with Alzheimer's disease and FactorV^{Leiden} with deep venous thrombosis (reviewed in Weiss & Clark, 2002).

An alternative, common disease/rare variant (CD/RV) model has also been proposed (Weiss & Terwilliger, 2000; Pritchard, 2001). Supporters of this model have raised a number of objections to the CD/CV hypothesis. First, they argue that these disorders are common because of common environmental influences, not disease alleles, in the population. Second, that new model predictions of complex disease suggest that most genetic variance is due to rare alleles under weak selection rather than common alleles that are close to fixation (Pritchard, 2001). Finally, they argue that data from quantitative trait research in animal models and from late-onset Mendelian disorders show broad allelic diversity at odds with the CD/CV hypothesis (Wright et al, 2003).

The power of specific positional cloning methods, such as association or linkage disequilibrium mapping, to detect candidate genes is dependent on the existence of a relatively simple underlying structure of genetic susceptibility. If a gene contains many rare allelic variants, which contribute to disease susceptibility, it has been

argued that the power of current statistical tests of association will be much reduced (Slager et al, 2000). At a population level this is partly because rare variants (<5% frequency) are more likely to be population specific (Risch, 2000; Ptak & Przeworski, 2002). Such variants may be recent in origin and exclusive to a specific founder population or less recent and found in a specific ethnic group, for example, the haemochromatosis mutation *C282Y* is found only in Caucasians. Critics of the CD/CV hypothesis maintain that LD-mapping resources such as the HapMap project, by including common variants only, are 'designed to find what is most probably not there to be found.' This view has been contested by members of the International HapMap consortium, who argue that rare alleles associated with complex disorders will be identifiable where these have a large effect size and site the successful mapping of genes for cystic fibrosis and diastrophic dysplasia as examples (The International HapMap Consortium, 2003).

Chapter 3

Materials and methods

3.1 Introduction

Clinical, laboratory and statistical methods are presented in this chapter. An important aspect of the study was developing an understanding of the context of this work within a local and international framework. This chapter begins with a brief discussion of the research collaborations that were vital in contributing to the sample collection described in Chapter 4.

3.2 Collaborators

3.2.1 Clinical collaborators

At the onset of this project, collaborations were established with Professor John Waddington at the Royal College of Surgeons in Ireland (RCSI) and Professor Eadbhard O'Callaghan at University College Dublin (UCD). Funded by the Stanley Foundation (US), these researchers had collected detailed clinical/epidemiological information on large urban and rural cohorts of patients with major psychotic disorders. Details of these sample collections have been published elsewhere (Gervin et al, 1998; Waddington et al, 1997). The clinical instruments and procedures involved were complementary to those implemented in this sample collection and detailed below. In a similar manner to the general ascertainment method described in this chapter, all available individuals who had been involved in the Stanley Foundation studies were approached about providing DNA samples and access to their clinical research information for the genetic project.

3.2.2 Laboratory collaborations

Collaboration was established with Dr Hreinn Stefansson, at the biopharmaceutical company DeCODE genetics (Reykjavik, Iceland). This involved the provision by deCODE of anonymous genotyping of single nucleotide polymorphism (SNP) and microsatellite markers at the Neuregulin-1 (*NRG-1*) locus for the Phase 1 sample. Professor David St Clair at the University of Edinburgh provided genotype data at this

locus from his Scottish schizophrenia case-control sample for analysis. The results of these analyses are detailed in Chapter 5.

3.2.3 Statistical collaborations

Collaboration was established with Dr Jason Moore, Associate Professor of Genetics at Dartmouth Medical School (New Hampshire, US). Professor Moore's group specialise in computational genetics and are involved in developing methods of investigating gene-gene interaction for complex genetic disorders. This collaboration, detailed in Chapter 7, involved investigating gene-gene interaction models of six putative schizophrenia susceptibility loci. Dr Cathal Walsh at the Department of Statistics, University of Dublin, Trinity College provided statistical advice for the regression analyses described in Chapter 7.

3.3 Clinical methodologies

3.3.1 Ethics approval

Ethics approval was sought and approved from all hospitals participating in this study. Since 1999 the ethics committees representing the following Dublin hospitals have provided ethics approval:

i. St James's Hospital
ii. St. Patrick's Hospital
iii. St Vincent's Hospital (Fairview)
iv. St. Brendan's Hospital

Approval has also been obtained from the ethics committees of the former North Eastern Health Board (for sampling at St. Davnett's Hospital, Monaghan) and St John of God Hospital (Dublin) for review of clinical research data and blood sampling at their facilities. At the request of the Irish Blood Transfusion Service additional approval has been obtained by the Dublin Federated Hospitals' Ethics Committee for collection of control DNA samples from their attending blood donors.

3.3.2 Ascertainment

Potential cases were ascertained through the treating teams at the facilities listed in the previous section. At each centre an educational session was arranged to describe the rationale and methodology of the project and to answer questions from potential

referrers. An information brochure (included as Appendix I) was developed for potential participants, control participants (from the Irish Blood Transfusion Service) and for staff at these facilities. This was distributed on hospital wards, outpatient departments and day hospitals at the participating centres. All attending patients with a diagnosis of a major psychotic disorder were eligible for recruitment. Before contacting patients, we established with the treating teams the best strategy for making contact (e.g. at clinics or on home visits with community nurses). In addition we confirmed with the clinical teams that their patient's were in a position to give informed consent, rather than in a period of relapse or detained in hospital under Irish Mental Health Legislation.

By necessity, the process of ascertainment was influenced by the treating teams and thus varied between and within hospitals. At St James's Hospital, all patients who presented between 1999 and 2004 with a major psychotic disorder were assessed for study suitability. Those meeting study criteria were approached individually about the study. For this reason, patients attending this service represent the largest subject group in the study. The ascertainment procedure differed at St. Vincent's, St John of God and St. Patrick's Hospitals' where all referrals were from consultants. The method of recruitment at these centres also differed between clinical teams, with some teams providing a list of patients that could be directly contacted and others writing to patients' individually to confirm that we could make contact regarding the study.

Procedural differences in ascertainment were an unavoidable factor of differences in policy between the multiple ethics committee's and the individual preferences of referring consultants and their teams. To address this, we have broadened our ascertainment strategy by publishing details of our work in the general media and through patient support groups (e.g. Schizophrenia Ireland). We are increasingly accepting direct referrals from affected individuals.

3.3.3 Recruitment

Potential participants were provided with a copy of the information sheet for the study, which could be discussed with a research team member. Those interested in

being involved were offered an appointment, at a venue convenient to them, with an interviewer (AC, SS or JMN). Participants were interviewed in hospital, at outpatient suites, at day centre facilities, at hostels, in their own homes or at the Trinity Centre for Health Sciences, St James's Hospital, Dublin. At this first interview, written informed consent was obtained, demographic information collected and the clinical interviews completed (details of all clinical and neurological assessments are provided in section 3.3.6). In most cases, blood sampling was also completed at this assessment. Arrangements were made for a follow-up assessment to complete additional clinical and neurological examination. We also established whether participants wished to be actively informed of progress in the study. At this point recruited subjects were assigned an identifier in the form of a four-digit number where the first digit represented the study centre.

3.3.4 Dissemination of results to participants

We provide regular updates on the activities of our research group on our website (http://www.tcd.ie/Psychiatry/Neuropsychiatry/schizo1.htm). From 2003 the research team have produced a bi-annual newsletter to update participating individuals and institutions about progress and results from this project, and developments in the field of schizophrenia genetics research. Copies of recent newsletters are also available at our website (http://www.tcd.ie/Psychiatry/Neuropsychiatry/ Newsletter1Schiz.pdf).

3.3.5 DNA collection and storage

Blood samples were obtained for DNA by venepuncture. Three 6ml samples of blood, in two EDTA and one serum tube, were obtained from all participating cases using 18 gauge needles in Vacuette® containers (Greiner BioOne, Gloucestershire, Stonehouse, UK). A single serum tube (6ml) was collected from participants from the Irish Blood Transfusion Service as part of their blood donation. Blood samples for DNA were stored in Vacuettes® containing EDTA at -70°C until they were processed. Serum samples were spun at 6,000 rpm at room temperature and the supernatant was stored at -70°C.

3.3.6. Clinical instruments and procedures

The author received training in clinical/neurological assessment as part of a research methods training programme which incorporated videotape and concurrent direct assessment of patients for inter-rater reliability purposes. This programme (co-ordinated by Dr Abbie Lane at St John of God Hospital, Dublin) was conducted by researchers who had specific training in the implementation of each of the research instruments described. The author attained the programme reliability criteria for all measures (kappa >0.75-0.90) and subsequently trained the other clinical investigators (SS, JMN) in implementing these instruments using the same training procedure and reliability guidelines.

3.3.6.1 The Structured Clinical Interview for DSM-IV (SCID)

The SCID is a semi-structured interview, administered by clinicians, from which major DSM-IV Axis I-diagnoses can be made (Spitzer et al, 1990). This study used the Research Edition of SCID-P, the standard SCID designed for use with subjects who are identified as psychiatric patients. The interview takes the form of nine modules including an introductory overview section and individual modules with questions covering the seven major axis I diagnostic classes. Because of this structure the SCID can readily be adapted to study individual diagnostic categories.

Reliability for DSM-IV major psychotic disorder diagnoses using SCID-P are reported to be in the fair to excellent range (kappa=0.64-0.94) (http://www.scid4.org/scidI_summary_table.htm). In interviewing cases for this study, the five sections of the SCID relevant to psychotic disorders were included, namely:

A-Mood syndromes

B-Psychotic and associated symptoms

C-Psychotic disorders (differential diagnosis)

D-Mood disorders

E-Psychoactive substance use disorders

The interviewer also reviewed the case notes and where available accessed diagnostic collateral information. For each individual participant, a case vignette was derived for inter-rater reliability purposes. The three clinical interviewers rated all cases

independently. Agreement between raters for psychotic disorder diagnoses in this study was in the excellent range (kappa >0.8).

3.3.6.2 The Positive and Negative Symptom Scale (PANSS)

The Positive and Negative Syndrome Scale (PANSS) is a 30-item scale that assesses severity of illness on positive, negative and general psychopathology domains. The scale measures positive and negative symptoms, their differential and general severity of illness. These four scales have been shown to be normally distributed in schizophrenia populations and demonstrate good reliability (Kay et al, 1987). The scale is included as Appendix II.

3.3.6.3 Abnormal Involuntary Movement Assessment

Movement disorders in schizophrenia may occur spontaneously or in response to antipsychotics and other medications. These typically present as involuntary movements involving the oro-facial musculature and limbs; both acute (akathesia and dystonia) and chronic (parkinsonism, akathesia, dystonia, dyskinesia) forms are recognised. The assessment performed in this study incorporated the Abnormal Involuntary Movement Scale (AIMS), a global instrument, divided into seven body regions (four of which relate to the face), with each item rated on a five-point ordinal scale (Guy, 1976). A copy of the scale is included as Appendix III.

3.3.6.4 Neurological Soft Signs Assessment

Subjects with schizophrenia are more likely to have neurological abnormalities than individuals with other mental illnesses or controls. Neurological dysfunction in schizophrenia is correlated with earlier age at onset, negative or deficit symptoms and poorer outcome. These abnormalities can be classified as 'hard' signs if they can be localized to a specific brain area or 'soft' signs when this is not the case. Soft signs include impairments in motor sequencing and coordination, confusion in right-left discrimination, and deficits in sensory-perceptual abilities (Flashman et al, 1996). Differences in performance in subjects with schizophrenia do not appear to be attributable to either medication or medication-induced neurological side effects (Browne et al, 2000). The aetiology of these symptoms is uncertain and some authors

have suggested that specific subcortical structures may be involved (Buchanan & Heinrichs, 1989). Increased scores for neurological soft signs in first-degree relatives of schizophrenia subjects and the unaffected twins of pairs discordant for schizophrenia suggests that these deficits are at least partially genetically mediated (reviewed by Niethammer et al, 2000).

A standardized assessment of neurological function, which incorporated two instruments, was performed. The Neurological Evaluation Scale (NES) (Buchanan & Heinrichs, 1989) is a battery of 26 items, 14 of which are tested bilaterally. Each item is scored as normal (0), mildly abnormal/questionable (1) or abnormal (2) except for the snout and suck reflexes which are rated 0 or 2. The Condensed Neurological Examination (CNE) (Rossi et al, 1990) consists of 19 items, 7 of which are bilateral. Items are related as present or absent (0,1) or on ordinal scales having maximum values between 2 and 6. The neurological examination battery is included as Appendix IV

3.3.6.5 Minor Physical Anomalies Scale (MPAs)

Minor physical anomalies are persistent, subtle abnormalities of external physical appearance, which represent developmental deviation from the usual morphological form. Such anomalies are thought to develop during the first or early second trimester of gestation and involve the mouth/palate, skull structure, eyes, ears, hands and feet. However, accurate timing may be precluded by the different anatomical sites involved, which develop at different times. Also the developmental impact an insult depends on its severity as well as its timing (reviewed Murphy & Owen 1996). Higher MPA scores are reported in many neurodevelopmental disorders, including schizophrenia.

As early as the 1890's Kraepelin noted an association between abnormalities of skull and ear structure and dementia praecox (Kraepelin, 1896). More recently, 12 of 13 studies have demonstrated higher rates of MPAs in schizophrenia subjects. More detailed anthropometric or 3-D morphometry measures indicate involvement mainly of mid-facial and lower facial regions (Lane et al, 1997; Hennessy et al, 2004). The occurrence of MPAs in schizophrenia may be partially genetically mediated, but the

aetiological significance of this is uncertain (Ismail et al, 1998). This study employed the rating scale devised by Lane et al, (1997) and specifically measured the variables that had discriminated between cases and controls in the Irish sample described by these authors. The assessment protocol is included as Appendix V.

3.4 Laboratory methods

3.4.1 Buffers and solutions

(1) TE Buffer:
2M Tris-Hcl (pH 7.5)
0.5M EDTA (pH 8.0)

(2) Extraction Buffers/ Solutions:
Lysis Buffer (x10)
50mM Tris-HCl, pH 7.5
25mM $MgCl_2$
0.6M sucrose
5% (v/v) Triton x100
Distilled H_20

Suspension Buffer (x10)
2M Tris-Hcl (pH 7.5)
0.5M EDTA (pH 5.0)
5M NaCl

(3) Electrophoresis Buffers:
10x TAE:
0.4M Trisma Base
0.5M EDTA
The pH of the buffer was brought to 8.2 with glacial acetic acid.

TBE (x10) (Invitrogen, Paisley, UK.):
0.9M Boric Acid
0.01M EDTA

(4) PCR Buffer:
10X PCR Buffer:
100mM Tris HCl pH 8.9
500 mM KCl
15mM $MgCl_2$
1% Triton x100 (Sigma, Dublin, Ireland)

3.42 DNA extraction procedure

For the laboratory studies presented in this thesis the author and other laboratory personnel performed DNA extractions using a Phenol/Chloroform method (Sambrook

et al, 1982). This method is labour intensive, so the additional case and control samples contributing to the Phase 2 sample described in Chapter 7, were extracted using the QIAamp DNA Blood Maxi Kit (catalogue number 51194), (Qiagen, Crawley, UK). Instructions for using this kit come with the accompanying manual.

The Phenol/Chloroform method involves a two-day protocol:

On day one, 6ml blood samples were removed from the freezer and allowed to thaw slowly on ice for a two- four hour period. Next, the following steps were performed:

(1) Blood samples were transferred to labelled 50ml Falcon tubes.

(2) The blood tubes were rinsed with sterile distilled water, which was added to each Falcon tube to make up a volume of 12.5ml.

(3) The tubes were topped up with Lysis buffer (x1) to a volume of 25 ml, inverted and placed on ice on a rotary shaker for 30 minutes.
(*Lysis buffer is a salt solution that is hypermolar to the contents of red blood cells causing them to absorb water from the solution by osmosis, swell and rupture. The white cells remain intact.*)

(4) The tubes were centrifuged at 3, 500 rpm for 15 minutes.

(5) The supernatant was poured off to approximately 4 ml, without displacing the white cell layer.

(6) Lysis buffer was again added to 25mls. The tubes were placed on ice for 10 minutes on a rotary shaker to dislodge the pellet into solution.

(7) The tubes were again centrifuged at 3, 500rpm for 15 minutes.

(8) The supernatant was poured off until the white cell layer on the surface of the pellet began to move.

(9) The pellets were suspended in 2 ml of suspension buffer (x1) and the contents were transferred to a labelled 15ml sterile Falcon Tube.

(10) 150µl of 10% Sodium Dodecyl Sulphate (SDS) and 60µl of Proteinase K was added to each tube.

(SDS is negatively charged with a high affinity for protein, thus encouraging proteolysis. Proteinase K is a powerful proteolytic enzyme that ensures the degradation of nucleoproteins).

(11)The tubes were placed on a rotary shaker overnight at room temperature.

Day 2 of the extraction process involves the use of Phenol and Chloroform/ Isoamylalcohol (24/1, v/v). As Phenol is neurotoxic, the Phenol/ Chloroform steps were conducted in a fume hood.

(1) The tubes were removed from the rotary shaker and 2ml of Phenol was added to each and thoroughly mixed.

(2) The tubes were placed on a rotary shaker for 10 minutes and then centrifuged at 3,800rpm for 10 minutes.

(3) The aqueous white layer was removed into a clean, labelled 15ml tube and 1ml of Phenol and 1ml of Chloroform/IAA was added. The tubes were again placed on the rocker for 10 minutes and centrifuged at 3, 800rpm for 10 minutes.

(4) The top layer was removed into a clean 15ml tube and 2ml of Choloroform/ IAA was added to each tube. The tubes were rocked for 10 minutes and centrifuged at 3,800 rpm and the supernatant was removed into a clean 15ml tube.

(5) 50µl of 3M Sodium Acetate was added followed by Absolute Alcohol at a temperature of -20°C to a volume of 6mls.

(6) The tubes were gently inverted until the formation of a white fluffy DNA pellet was observed.

(7) The pellet was removed from the tube using a sterile 1ml pipette tip and transferred to a labelled 1.5ml sterile Eppendorf.

(8) 1 ml of 70% ethanol (-20°C) was added to remove excess salt and the samples were microfuged at 13, 000rpm for 5 minutes. The excess alcohol was carefully removed with a pipette tip and this step was repeated.

(9) After removal of the ethanol the tubes were again microfuged at 13, 000 rpm for 5 minutes.

(10) The DNA was dried overnight at room temperature or for 15 minutes in a vacuum dessicator.

3.4.2.1 DNA suspension and storage

The dried DNA pellet was suspended in 250µl of TE Buffer and stored in the fridge in 1ml Eppendorf tubes at 4°C for four to five days until the pellet had resuspended. The resultant stock DNA samples were stored at -20°C pending quantification.

3.4.3 DNA quantification

Two-stage DNA quantification was carried out: in the first stage using spectrometry to obtain an indication on the quality and amount of DNA extracted. In stage 2, an accurate measure of DNA concentration was made using fluorimetry.

i) Spectrometry: DNA absorbs light at 260nm whereas RNA absorbs light at 280nm. Comparing a ratio of these two values gives an indication of the quality of the DNA. A ratio of <1.8 it indicated that the extracted sample contained proteins and aromatic hydrocarbons. In these cases, DNA was re-extracted from a backup blood sample. A ratio of >2 it indicated that the sample contained a substantial quantity of RNA. In preparing the samples for testing, 5µl of stock DNA was made up to 1000µl with 1x TAE. The DNA concentration was calculated using the following formula:

$$\textbf{DNA concentration} = \frac{\textbf{Optical Density (OD)}_{260nm} \textbf{ x 50}}{\textbf{5}} \ \mu g/\mu l$$

For example, in a sample where the OD_{260nm} was 0.67µg/µl (670ng/µl), because the DNA is dissolved in 250 µl of 1x TAE, this equates to 250 x 670 = 167500 ng of DNA.

ii) Fluorimetry: fortunately, our group were provided with access to a fluorimeter by the Department of Microbiology, TCD. For this we are grateful. A standard curve was obtained for each run of samples by concurrently running a set of standard DNA concentrations (ng) (750, 500, 250, 100 and 50). The DNA sample concentrations were calculated by entering the fluorometer readings into the equation:

y=mx

Samples were loaded onto the fluorimeter using 100µl of 1x PicoGreen [dsDNA Quantification Kit, Molecular Probes, Eugene, Oregon, US] and 100µl of a 1:50 dilution of stock solution.

3.4.3.1 Quantified DNA stock preparation

Once DNA quantification was completed the diluted samples of 80ng and 8ng stocks were prepared from the raw stocks by diluting appropriate volumes in 1x TE buffer. Final volumes were 350µl for 80ng/µl stocks and 1000µl of 8ng/µl stocks.

3.4.4 DNA storage

DNA stock solutions, DNA pools and DNA working solutions were stored at -20°C. Approximately 600µl of DNA at 8ng/µl were stored in 96 deep well boxes for individual genotyping. When individual genotyping was performed these boxes were stored at 4°C to reduce the effects of freeze-thaw damage. DNA pooled samples were constantly stored at -20°C. Continuous use of the pools in PCR amplification reactions revealed no perceivable degradation.

3.4.5 Strategy for genotyping individuals

The use of genetic association studies to identify susceptibility genes for complex disorders such as SZ requires the investigation of large numbers of genetic markers in

large samples. The task of identification of polymorphisms (as markers) in candidate genes has been greatly aided by the recent explosion in the numbers of SNPs catalogued on public databases on the web. Additional advances in genotyping methodologies have allowed cheaper and faster assaying of these SNPs using a number of different biochemical reaction principles (for review see Syvanen, 2001).

For the experiments conducted in Chapters 6 and 7, an allele specific primer extension method was used; this is summarised in Figure 3.1. The method uses primers that abut the SNP site. During extension, specific di-deoxyribonucleotides are incorporated which bring about a sequence-specific termination of the reaction resulting in allele-specific extension products. These products can be detected using different methods, for these studies, the SnaPshot method of fluorescent tagging and electrophoresis was employed (Applied Biosystems, Warrington, UK).

Markers at the NRG1 locus were genotyped by deCODE Genetics using a method based on a template-directed primer extension and detection by fluorescence polarization (FP) (Chen et al, 1999; Gretarsdottir et al, 2002). Other members of the research team did the additional SNP genotyping for the studies discussed in Chapter 7. For these studies the total size of the sample had increased substantially, making it cheaper and more efficient to employ a different genotyping method. These studies were performed using Custom Taqman® SNP Genotyping Assays on an ABI Prism 7900HT Sequence Detection System (Applied Biosystems).

3.4.6 Genotyping of pooled DNA samples

In an attempt to increase genotyping efficiency, a number of methods of measuring allele frequencies of SNPs and microsatellite markers in pooled DNA samples have been developed (Carmi et al, 1995; Barcellos et al, 1997; Plomin et al, 2002; Hoogendoorn et al, 2000; Norton et al, 2002). Rather than genotyping individual DNA samples, pools are made up of DNA from many individuals. An approximation of allele frequencies at a marker can be measured from the resultant pools (e.g. of 1,000 cases in one pool and 1,000 controls in another), without genotyping all of the individuals directly (reducing the number of genotyping reactions from 2,000 to 2 in this example). Two of these reports, from a UK group, demonstrate that this DNA

pooling approach can provide accurate and reliable SNP allele frequency measurement using the SNaPshot™ genotyping platform (Hoogendoorn et al, 2000; Norton et al, 2002).

There are two main difficulties associated with DNA pooling using a primer extension genotyping method (Sham et al, 2002). First, for a number of reasons the observed signal for an allele may not represent the true frequency in the pooled sample. For example, SNPs may show differential amplification during PCR, causing the signal for the more efficiently amplified allele to be inflated above its true frequency in the pooled sample. The same outcome may occur because nucleotides may not be incorporated into the extension reaction with the same efficiency or because alleles may not be detected with the same efficiency. To counteract this potential bias a correction factor provided by the known allele frequencies of an identified heterozygous individual is applied (Norton et al, 2002). Secondly, self- annealing may occur at the 3'-end of primers allowing extension to occur independently of the target template. Including a negative control in the reaction can readily identify this problem.

3.4.6.1 Preparation of samples for DNA pools

Additional quantification and sample preparation procedures were required for the DNA pooling method detailed below. Dr Derek Morris prepared the DNA pools for the experiments conducted in Chapter 6 using the principles described by Norton et al (2002). Dilute stock samples were employed to ensure homogeneous DNA concentration in individual samples and all were checked by PCR to confirm that they yielded a reliable product.

The 8ng/µl dilutions from all case and control samples were re-quantified using the fluorimeter. This confirmed that the sample concentrations varied from 3-12ng/µl. Of the original individual cases (n=219) and controls (n=231), for 3 cases and 1 control sample the DNA concentrations were insufficient for their inclusion in the pools (<3ng/µl). The case and control pools were formed by aliquoting an equal amount of DNA (75ng) from each individual sample into case and control eppendorfs respectively. Because of the differences in individual numbers making up each pool

this resulted in a final case pool DNA concentration of 4.8ng/μl and control pool DNA concentration of 5.9ng/μl. At this point the control pool DNA was diluted with TE buffer to match the case pool DNA concentration.

3.4.7 Genotyping the APOL gene family

3.4.7.1 Primer design

Standard primer extension reactions were conducted for the amplification of genetic markers at the APOL locus. Primers for all fragments were designed using Primer3 (http://wwwgenome.wi.mit.edu/cgibin/primer/primer3_www.cgi). Primer pairs were obtained from Invitrogen (Paisley, Scotland).

3.4.7.2 DNA amplification/optimisation by polymerase chain reaction (PCR)

Standard PCR reactions in volumes of 10μl were conducted for each marker on PTC-225 Peltier Thermocyclers (MJ Research, Waltham, MA, US) in 96-well PCR Plates. The PCR reaction volume was 10 μl: 3μl of DNA (at 8ng/μl); 1μl of PCR Buffer; dNTPs (50mmol); F primer (5pmol/μl); R primer (5pmol/μl); Taq polymerase (0.25/μl (6u/μl)) and 2.75 ml of ddH_20. The parameters for PCR reactions are included in Table 3.1. A sample of PCR products for individual markers were electrophoresed on 1.5% agarose gels stained with ethidium bromide and examined under a UV transilluminator to examine the quality of the PCR product. Comprehensive details on PCR conditions and primers are included in Tables 3.2-3.4.

3.4.7.3 Genotyping methodology

The author performed the genotyping described in Chapter 6 using the SNaPshot™ method of single base extension (Applied Biosystems, Warrington, UK). First, the region of genomic DNA surrounding the SNP of interest is amplified using PCR. An unlabelled oligonucleotide primer is designed to anneal to the PCR product with its 3' base adjacent to the SNP of interest. Next the primer is extended in the presence of fluorescently labelled ddNTPs, each of the 4 ddNTPs being tagged by a different fluorescent dye. When the primer extension products are run on an ABI DNA Sequencer or ABI Genetic Analyser (in this case an ABI PRISM 377 DNA Sequencer, Applied Biosystems, Warrington, UK), the specific allele products can be differentiated from each other on the basis of which fluorescent dye they carry.

Designing extension primers of different lengths increased genotyping efficiency, as 4 different fluorescent dyes are available making it possible to mix extension products and run them in the same gel lane (multiplexing).

The SNaPshot™ genotyping system involves an initial PCR reaction; a thermal cycling reaction (to anneal unlabelled primers and extend primers with target ddNTPs); a clean up step; gel electrophoresis and analysis. This process is summarized in Figure 3.3 and detailed below:

Step 1:PCR reaction

This involved the standard reaction described in section 3.4.7.2 to amplify the genomic region containing the SNP of interest.

Step 2: Removing dNTPs and primers

For each reaction Shrimp Alkaline Phosphatase (SAP) (0.5µl), SAP dilution buffer (0.5 µl) and Exonuclease 1 (*ExoI*) (0.05µl) were added sequentially to ddH_20 (1µl) and PCR product (5 µl). The samples were then run on a PTC-225 Peltier Thermocycler (MJ Research) at 37^OC for 30 minutes (to activate the enzymes), 80^OC for 15 minutes (to deactivate the enzymes) and 4^OC for ≥2 minutes (for storage).

Step 3: Primer extension reaction step

An aliquot of the product from step 2 (2µl) was combined with Primer extension primer (0.5µl at 5p mol/µl), SNaPshot multiplex reaction mix (1.5µl) and SNaPshot buffer (1µl). This was made up to a volume of 10µl with ddH_20 and run on a PTC-225 Peltier Thermocycler using a standard protocol (see Table 3.1).

Step4: Final cleanup step

The SNaPshot product (5µl) from step 3 was added to SAP (0.5 µl) and SAP buffer (0.5µl) and the thermal cycling reaction in Step 2 was repeated.

3.4.7.4 SNP genotyping on the ABI 377 DNA sequencing system

The ABI 377 DNA sequencing system employs electrophoresis across a polyacrylamide gel for separation of DNA fragments. The author used 10%

polyacrylamide gels. These contained 11g of urea, 13ml of ddH_2O, 6ml of 50% Long Ranger gel solution (BioWhittaker Molecular Application, Rockland, ME, USA), 3ml of 10X TBE (0.89M Tris Borate pH 8.3+ 20mM Na_2EDTA), 150 µl of 10% Ammonium Per-Sulphate ($(NH_4)_2S_2O_8$) and 21 µl of NNN'N'-Tetramethyethylenediamine (TEMED).

Prior to loading the gel, 2 µl of sample (extension product from individual or pooled DNA samples) was mixed with 0.5 µl of size standard and 2 µl of loading buffer (blue dextran, 50mg/ml; EDTA, 25mM; diluted 1 in 5 with deionised formamide). For economy, the size standard, required to size DNA fragments, was made from 3 ROX labelled oligonucleotides (Genset Oligos, Paris, France) of 14, 26 and 30bp. The samples were then denatured for 2 min at $95^{O}C$ and placed on ice for loading.

Once loaded, the samples (a maximum of 64 lanes) were electrophoresed through the gel. One negative control per SNP (ddH_2O, which had been through the same PCR and SNaPshot protocol) was used to test for contamination. As DNA fragments pass a laser, the attached dye labels are excited causing them to fluoresce. The emitted light is separated by wavelength (each dye has a separate spectral profile/colour) and processed by 377 data collection software. The raw data were analysed using Genescan Analysis 3.1.2 (Applied Biosystems), with all DNA fragments being sized based on their migration through the gel relative to the internal size standard. All samples were genotyped using Genotyper 2.5 (Applied Biosystems). Typical outputs for the individual and pooled DNA samples are represented in Figure 3.1 and 3.2.

3.4.7.5 Simulating heterozygosity data for DNA pooling

In common with other methods, the DNA pooling technique of Norton et al, (2002) requires the identification of a heterozygous individual to provide accurate SNP allele frequencies to determine the correction factor applied in testing for evidence of association in pooled samples of cases and controls. Because the investigated SNPs at the APOL locus were identified from public databases, this introduced the additional time-consuming step of genotyping a panel of individual DNA samples to identify a heterozygous individual at each SNP. To circumvent this task our group have developed a novel variation of this DNA approach (McGhee et al, 2005). We have

applied a range of simulated heterozygote data to the calculation of allele frequencies of each SNP in the DNA pools. This modified DNA pooling strategy is more time and cost efficient than existing strategies for analysis of public database SNPs.

The Norton et al, (2002) method used the SNaPshot genotyping assay (Applied Biosystems) to determine peak heights for a pair of SNP alleles in a pool of case DNA samples and in a pool of control DNA samples, and to determine the peak heights for one individual known to be heterozygous for this SNP. The ratio of the peak heights in the heterozygous sample (k) is used to correct for the unequal representation of alleles in a DNA pool. In practice, the frequency f in the pool of allele A is calculated as $f(a)=A/(A+kB)$ where A and B are the peak heights of the SNaPshot products representing alleles A and B.

Instead, we applied simulated heterozygote ratios (range of *0.1* to *10*) to the analysis of each SNP. Each simulated k results in an estimate of the allele frequencies in the case and control pools (Figure 3.4). Computing allele frequencies with a range of these k values allows the identification of the maximum *potential* difference in allele frequencies between the case and control pools (Figure 3.5). Where this maximum potential difference in SNP allele frequency reaches nominal significance levels, further investigation is warranted – this involves identification of a heterozygous individual to determine the true k value for the SNP and hence accurate allele frequency estimation in the case and control pools. However, where this maximum potential difference in SNP allele frequency does not reach nominal significance levels, it can be categorized as showing no evidence of association in the case-control sample and thus excluded from further analysis. It is possible that the maximum potential difference in SNP allele frequency will not be identified within the range of simulated k values stated (*0.1* to *10*). In such cases the maximum difference will have been found at $k = 0.1$ or $k = 10$. In these cases the range of k values should be extended (e.g. *0.1* to *0.01*, or *10* to *100*) until the maximum potential difference has been found.

3.4.7.6 Confirming reliability for this DNA pooling method

The case and control pools were genotyped for four SNPs that had previously been genotyped individually in our phase 1 case-control sample. Making the assumption that the individual genotyping was 100% accurate, the difference in allele frequencies between the cases and controls were compared for both the individual genotyping and the DNA pooled genotyping experiments. The mean pooling error was 0.004, which compares favourably with results published previously for this method. In conducting the APOL pooling studies, 3 SNPs were individually genotyped, giving a mean error in estimating the difference between cases and controls of only 2 alleles per thousand. These results show that the DNA pools are able to give accurate estimations of absolute allele frequencies as well as accurate estimations of the difference in allele frequency between different pooled DNA samples

3.5 Statistical methods

3.5.1 Allele frequencies and Hardy-Weinberg equilibrium

Allele frequencies for all markers were calculated as the proportion of occurrences of each allele in the sample. The Hardy Weinberg law states that in the presence of random mating and in the absence of disturbances including mutation, natural selection, migration, inbreeding or random genetic drift, allele frequencies from generation to generation will conform to the following equation:

$$\mathbf{p^2 + 2pq + q^2 = 1}$$

Where p^2 = the expected fraction of the population homozygous for p
q^2 = the expected fraction homozygous for q
2pq = the expected fraction of heterozygotes

Each marker was assessed for deviation from Hardy-Weinberg equilibrium based on observed and expected allele frequencies and implementing the above formula using the programme FSTAT (http://www.unil.ch/izea/software/fsta.html).

3.5.2 Relative risk

The relative risk is a measure of the raised risk in one group compared to another. In genetic association studies relative risk gives a measure of the strength of association between a genetic marker or haplotype and disease. This can be expressed as:

$$RR = \frac{a / (a + c)}{b / (b + d)}$$

Where a = frequency of the associated allele in case sample
b = frequency of the associated allele in the control sample
c = frequency of the non-associated allele in the case sample(s)
d = frequency of the non-associated allele in the control sample

3.5.3 Odds ratio

In the context of this project, the odds compare the probability that a disease is present with the probability that a disease is absent. The odds ratio (OR) is the ratio of odds in the case group to the odds in the control group. This is measured, using the same terms as in the previous formula by:

$$OR = \frac{a / b}{c / d}$$

3.5.4 Chi-square test

SNPs were tested for association with the SZ phenotype by using a 2 X 2 contingency table to calculate a $\chi 2$-statistic. This was implemented using the website (http://www.unc.edu/~preacher/chisq/chisq.htm). Microsatellite markers were tested for association using the T2 statistic of CLUMP (Sham & Curtis, 1995) and significance was assessed using Monte Carlo simulations.

3.5.5 Power calculations

Power calculations for the case-control association studies were performed using the formula derived in Sham et al, (2000) and implemented using the programme available at their website (http://statgen.kcl.ac.uk/gpc/).

3.5.5 Graphical overview of linkage disequilibrium (GOLD)

GOLD (http://www.well.ox.ac.uk/asthma/GOLD/index.html) calculates inter-marker linkage disequilibrium (D' and delta^2) and represents this graphically (Abecasis & Cookson, 2000). The program runs in MS-DOS mode and utilizes an input file with population genotype data for SNP and/or microsatellite markers. This was used to calculate D' across the *APOL* and *NRG1* candidate loci in the experiments detailed in Chapters 5-7.

3.5.6 Haploview

Haploview is a software package in development at the Whitehead Institute (Massachusetts Institute of Technology, US) http://www-genome.wi.mit.edu/personal/jcbarret/haplo/. It is designed to simplify and expedite the process of haplotype analysis by providing a common interface to several tasks relating to such analyses (Barrett et al, 2005). Haploview calculates D' and r^2 measures of linkage disequilibrium, allows users to examine block structures, generate haplotypes in these blocks, run association tests, and save the data in a number of formats. It also tags haplotypes with single nucleotide polymorphisms (SNPs) that are most informative about a given haplotype. The program takes input SNP data in standard linkage format, as phased haplotypes or as HapMap Project data using the GBrowse interface. It also takes a second file with marker position information. The three alternative default algorithms for haplotype-block definition from Haploview were implemented in the experiments described in Chapter 6.

i) Blocks defined by a contiguous set of markers where D' is above a threshold.

This applies the method suggested by Gabriel et al, (2002) where 95% confidence intervals are placed around D' and a block is defined where 95% of informative comparisons (excluding markers with minor allele frequency of <0.05) are in 'strong' LD. In applying this method I used the default setting where the minimum CI for 'strong' LD is 0.98-0.7.

ii) Blocks defined by a minimum number of SNPs that account for common haplotypes

The application in Haploview is a variant of the algorithm described by Wang et al, 2002. For each marker pair, the population frequencies of the 4 possible two-marker haplotypes are computed. If all 4 are observed with at least a frequency of 0.01, a recombination is deemed to have occurred. Blocks are then defined by consecutive markers where only 3 gametes are observed.

iii) Blocks defined by a solid spine of LD

This approach searches for a "spine" of strong LD running from one marker to another. With this method, developed by the group responsible for Haploview, the first and last markers in a block are in strong LD with all intermediate markers but the intermediate markers are not necessarily in LD with one another.

3.5.7 Haplotype estimation

Investigating haplotypes in case-control samples of unrelated individuals is complicated by the absence of genetic phase information. A number of different statistical approaches have been applied to inferring haplotypes from such genotype data. Examples of these include the parsimony method (Clark, 1990), maximum likelihood methods, implemented via the expectation-maximization (EM) algorithm (Excoffier & Slatkin, 1995), and Bayesian approaches (Stephens et al, 2001; Niu et al, 2002).

3.5.7.1 Haplotype association analyses

The haplotype association analyses described in Chapters 5-7 were performed using FASTEHPLUS, a form of the EM algorithm (Zhao & Sham, 2002). FASTEHPLUS performs a global test of association for each combination of markers chosen. P-values were calculated by simulation. A separate program, GENECOUNTING was used to output individual haplotype frequencies from cases and controls for testing specific haplotypes (Zhao at al, 2002). The author also used the Bayesian approach to reconstructing haplotypes described by Stephens et al, (2003) and implemented in PHASE (v2.0.2). This version introduces a model that allows for recombination and decay of LD with distance in haplotype estimates and includes a facility to test for haplotype frequency differences between cases and controls. Using the latter function the haplotype association analyses at the loci investigated were repeated. This confirmed that maximum likelihood (FASTEHPLUS) and Bayesian (PHASE v2.0.2) methods of analysis yielded comparable results.

Table 3.1: Polymerase Chain Reaction (PCR) cycling parameters

PCR Steps	Programme 1 (Q)	Programme 2 (R)	Programme 3 (Touchdown)	SNaPSHOT
Initial Denaturation	95°C - 2 min	95°C - 2 min	95°C - 3 min	95°C - 2 min
Denaturation	95°C - 7 sec	95°C – 20 sec	95°C – 15 sec	95°C - 5 sec
Annealing	n°C – 7 sec	n°C – 20 sec	n to n-4°C – 15 sec (1 cycle each)	43°C- 5 sec
Extension	72°C- 30 sec	72°C- 30 sec	72°C-30 sec	60°C-5 sec
No. of cycles	41	35	29 (at n-4°C)	49
Final extension	72°C- 15 min	72°C- 15 min	72°C- 10 min	-

Table 3.2: APOL-1, PCR details

Fragment	Primer sequence	Fragment Size (bp)	PCR protocol	Marker	SNP ID	SNP	Primer extension details
APOL1-1	F: tctgtttttggttttgtttgaa R: cctgtgtgtcttccttaggc	235	57Q	APOL-1	dbSNP:713791	(A/G)	F pex(17) aggggattcctaaatcc
APOL1-3,4	F: attgaaaacatggggttgaa R: tagagatggggaggaaggtt	262	59Q	APOL1-3 APOL1-4	dbSNP:136147 dbSNP:136148	(G/T) (C/T)	Fpex (17) tgtgtttgaatgggaaa Fpex (20) tgtgtgatcatcgaatgaga
APOL1-5	F: gcatcctgacctctccttag R: gtcagcttggagcagagc	223	59Q	APOL1-5	dbSNP:136149	(G/A)	Fpex (20) cttggggaagtggaatcagc
APOL1-6	F: cctcccttctcctgtaccat R:ggtaagcaggtgcttgtgat	231	60R	APOL1-6	dbSNP:136150	(T/A)	Fpex (17) cctccctaagggctggt
APOL1-7,8	F:agtgcaaaacagaagtgcag R:tcacatgacagacacagaagg	417	54Q	APOL1-7 APOL1-8	dbSNP:2413395 dbSNP:136151	(A/G) (A/G)	Rpex (17) cccaaactccactgggc Fpex (17) caggatcctgtgggccc
APOL1-18,19	F: gggcacacaattccctctat R: cctcattatcagtcagcagga	245		APOL1-18 APOL1-19	dbSNP:713929 dbSNP:136163	(A/G) (G/T)	Fpex (17) tgcaatggctgttatgc Fpex (20) tcctccaaccttatcctttc
APOL1-20	F: ggaaaaagtgagcacaga R:caagaaggtcatcctcaagc	218	54Q	APOL1-20	dbSNP:136164	(T/C)	Fpex (20) gtaagctccatggggttacc
APOL1-21	F: gggcttatgtcagagactgg R: gccttcctgcttgtcctac	218	57Q	APOL1-21	dbSNP:713753	(C/T)	Fpex (17) caaaggtggggaaccaa
APOL1-29,30,31	F: tggagctcatttcacagaca R: agctttacggagctcatctg	474	57Q	APOL1-29 APOL1-30 APOL1-31	dbSNP:136171 dbSNP:136172 dbSNP:136173	(A/G) (A/G) (G/A)	Fpex(17) agacagggtttcaccgt Rpex(17) atgcagttttggctcgg Fpex(20) caggtgtgagccaccacacc
APOL1-32,33	F:aaaggccagcagtacagaaa R:gcttgtgtccaccacttctt	320	57Q	APOL1-32 APOL1-33	dbSNP:2239785 dbSNP:136174	(A/G) (A/C)	Fpex (20) agtttcctcggttgaaaagt Fpex (17) gagttgggaatcacagc
APOL1-34,35	cacgacctggtcatcaaaa gactcctctgctcatttcca	279	59Q	APOL1-34 APOL1-35	dbSNP:136176 dbSNP:136177	(G/A) (G/A)	Fpex (20) tgacaaattgaaggaggtga Rpex (20) aggatgctgggttcattaac

Table 3.3: APOL-2, PCR conditions

Fragment PCR	Primer sequence	Fragment Size (bp)	PCR protocol	Marker	SNP ID	SNP	Primer extension details
APOL2-1,2	F: gggtctaattaattggcttgc R: atgcgtatcccaggatgaa	400	57Q	APOL2-1 APOL2-2	dbSNP 1315 dbSNP:1317	(A/C) (G/T)	Fpex (17) : aattagtcaagcaggtc Fpex (20) : ttgtctgctagatgtattag
APOL2-3,4	F: ccagccttgtagtgtcctct R: ctccagctctttggagtttg	527	54Q	APOL2-3 APOL2-4	dbSNP:2016708 dbSNP:763086	(C/T) (A/G)	Fpex (17): agggggattgaagaat Fpex (17): gagtttcta gacagcca
APOL2-5	F: tgactcacagatggaaacca R: gctctccccatcttgtgtt	304	57Q	APOL2-5	dbSNP:132757	(T/C)	Fpex (20) : cagtgagcga gcagaagggg
APOL2-6,7	F: gcacccttagtaacctgga R: tttggcatttgtgtagctga	484	57Q	APOL2-6 APOL2-7	dbSNP:2017689 dbSNP:129607	(A/G) (T/C)	Fpex (20): atcaaac ctgggggggggtc Fpex (20): ttccccacac tctccagtcc
APOL2-8	F: tcagctacacaaatgccaaa R: ccaggacaaaatgcagactt	415	57Q	APOL2-8	dbSNP:132759	(T/C)	Fpex (23) gccgaggtgggca gatcacgagg
APOL2-9	F: tgactcatatgcaaggctga R: ggtgatgaaggagtttgtgg	307	57Q	APOL2-9	dbSNP:132760	(T/C)	Fpex (17): tgagatt cgcccaatga
APOL-10	F: atctgatcatgctcctcacc R: gctgtgttctcctgtgtgaa	340	57Q	APOL2-10	dbSNP:2005998	(C/T)	Fpex (20): gcttccctggttggcaacac
APOL2-11,12,13,14	F:ctaacacagatgaggggactg R: tcctggctagttcctactcg	815	58TQ	APOL2-11 APOL2-12 APOL2-13 APOL2-14	dbsnp:2413392 dbsnp:2010467 dbsnp:2003813 dbsnp:2003814	(A/G) (C/T) (A/G) (C/T)	fpex (17): gtaccctta taatattt fpex (20): tcttc ccccactccc ctctc fpex (20): acctctcctcgggg tgacac fpex (17): ggcca atgagtctgc cc
APOL2-15	F: gaaagaggaagcgagccta R: tccctcccatctctattcct	420	57Q	APOL2-15	dbSNP:2010499	(A/T)	Fpex (17): t cgttccagct tcctct
APOL2-16, 17, 18	F: tccaagaacctaggaagtgtgt R: cttggcattgctgtggtt	301	57Q	APOL2-16 APOL2-17 APOL2-18	dbSNP:1001293 dbSNP:1001294 dbSNP:2157249	(C/T) (C/T) (T/C)	Fpex (17): ta aagactgggttccca Rpex (23): taaattttcattggcagcagagg Fpex (20): gg gcaacaggag atcagaga
APOL2-19,20,21	F: ccaggagggagaagagga R: agctacctggaaggctgag	531	61Q	APOL2-19 APOL2-20 APOL2-21	dbSNP:1557534 dbSNP:2157251 dbSNP:1557535	(A/G) (G/T) (A/C)	Fpex (17): ctggat actgcccctca Fpex (20): ctccttgggcagggaaagag Fpex (20): ga aatgatttcc tgagaaat
APOL2-22,23	tacctttcttctggggtcag agtgtgatctacccccatgt	401	61Q	APOL2-22 APOL2-23	dbSNP:136143 dbSNP:136144	(G/A) (A/G)	Fpex (20): cctctctcta cagtttaccc Fpex (17): ccttata gagctattgc

Table 3.4: APOL-3 PCR details

Fragment	Primer sequence	Fragment Size (bp)	PCR protocol	Marker	SNP ID	SNP	Primer extension details
APOL3-1,2	F: aaaatctcggctgacttgc R: catgggtcagaatggaaact	892	54Q	APOL3-1 APOL3-2	dbSNP:2051609 dbSNP:2413381	(G/T) (C/T)	Fpex (17) ttcatcatgttggccac Fpex (20) gcgccacctcctggctctgc
APOL3-3,4	F: atggccttgagagtgaactg R: tgggggttttgtaaagtcag	722	54Q	APOL3-3 APOL3-4	dbSNP:23085 dbSNP:105161	(T/A) (G/A)	Fpex (17) gaggaat catcatttgg Fpex (20) ctcaagcact tcctgcccac
APOL3-5	F: cctctcaaggacggttcc R: aaactagccaccagacatgg	601	57Q	APOL3-5	dbSNP:132615	(C/G)	Fpex (17) tcagaac aacggtagaa
APOL3-6	F: agaaggtggagatgctgtgt R: gcttgggttagtgactcttctt	401	59Q	APOL3-6	dbSNP:132617	(C/T)	Fpex (20) gtgtgtgcagtagcaag
APOL3-8,9	F: cctcagcagatgcagacttt R: cccttgcaaatggtattgaa	606	54Q	APOL3-8 APOL3-9	dbSNP:132618 dbSNP:132619	(A/T) (T/G)	Fpex (17) gttctct ctgcttgacc Fpex (20) acttccttaaataccttcaa
APOL3-10,11,12	F: ccacatgccctttgatctt R: tgagtccataagaggttccat	260	54Q	APOL3-10 APOL3-11 APOL3-12	dbSNP:80575 dbSNP:80576 dbSNP:80577	(G/T) (A/G) (G/A)	Fpex (20) gaactctcaagccacttgac Fpex (17) tctcactgtcacatgtc Fpex (23) gtgacagcttcacctgtgtgcat
APOL3-13,14,15,16,17	F: tcctcagaggtggaagaaaa R: acatggtaggtgctccgtta	517	59Q	APOL3-13 APOL3-14 APOL3-15 APOL3-16 APOL3-17	dbSNP:132624 dbSNP:132625 dbSNP:132626 dbSNP:132627 dbSNP:1807498	(C/T) (A/T) (G/A) (G/A) (A/G)	Fpex (17) aaaaacacttggtcggg Rpex (15) caccgcatccggcta Fpex (17) cgggaggcggaggttgc Fpex (20) gggaggcggaggttgcggtg Fpex23) ctgtctcaaaaaaaaaaaaaaa
APOL3-18,19,20	F: ctctcccggaagtatttgg R: caaaaattccaaggtgaagg	599	57Q	APOL3-18 APOL3-19 APOL3-20	dbSNP: 1807672 dbSNP:132628 dbSNP: 132629	(G/T) (C/T) (T/A)	Fpex (17) agaataaggtggttcga Rpex (23) ccaaatatagttactggggttta Fpex (20) tctctctctctcacacac
APOL3-21,22	F: ggaacgttctgacaatgacc R:acatacaggcagaagcaagg	525	57Q	APOL3-21 APOL3-22	dbSNP:2032518 dbSNP:132644	(C/T) (T/C)	Fpex (17) caggagatg ggtgcaac Fpex (20) gactcatgcatacatgctta

Table 3.4: APOL3-marker details (continued)

Fragment	Primer sequence	Fragment Size (bp)	PCR protocol	Marker	SNP ID	SNP	Primer extension details
APOL3-23,24	F: gtggcaccaccataactca R: cagggagcagtgaggaaa	464	59Q	APOL3-23 APOL3-24	dbSNP 2105915 dbSNP 132648	(C/T) (C/T)	Fpex (17) ttttgcagagatggggt Fpex (20) ggatcctcctgcattagcct
APOL3-25	F: gattgtcaagctgggatcat R: gatcgcaccagtcactcc	392	59Q	APOL3-25	dbSNP:80584	(C/T)	Fpex (17) cctgctcaacttcctga
APOL3-26, 27	F: ggagatggggaaggtcag R: aacccccaaccctaagtatg	540	59Q	APOL3-26 APOL3-27	dbSNP:132653 dbSNP:916334	(G/T) (C/T)	Fpex (17) ccagacacgttctccag Rpex (17) gatcagcaggagggtgg
APOL3-28,29,30	F: tgattttgaaggggtgaaat R: gtcttgcaagcctgctgt	797	54Q	APOL3-28 APOL3-29 APOL3-30	dbSNP: 132655 dbSNP:132656 dbSNP: 916335	(G/T) (C/T) (A/G)	Fpex (17) tgtgtggcttctggaca Rpex (23)ctttgtcttttactcccacatcc Fpex (20) caggtaattgtcagtcagcc
APOL3-31,32	F: cagattttcagggaggaaca R: ccgactgtgacttgtgtctct	310	57Q	APOL3-31 APOL3-32	dbSNP:132662 dbSNP:132663	(G/A) (A/T)	Fpex (17) ctgagtttacaggtgag Fpex (20)gccaagatttgtaaaccaag

Figure 3.1: Allele specific primer extension genotyping method

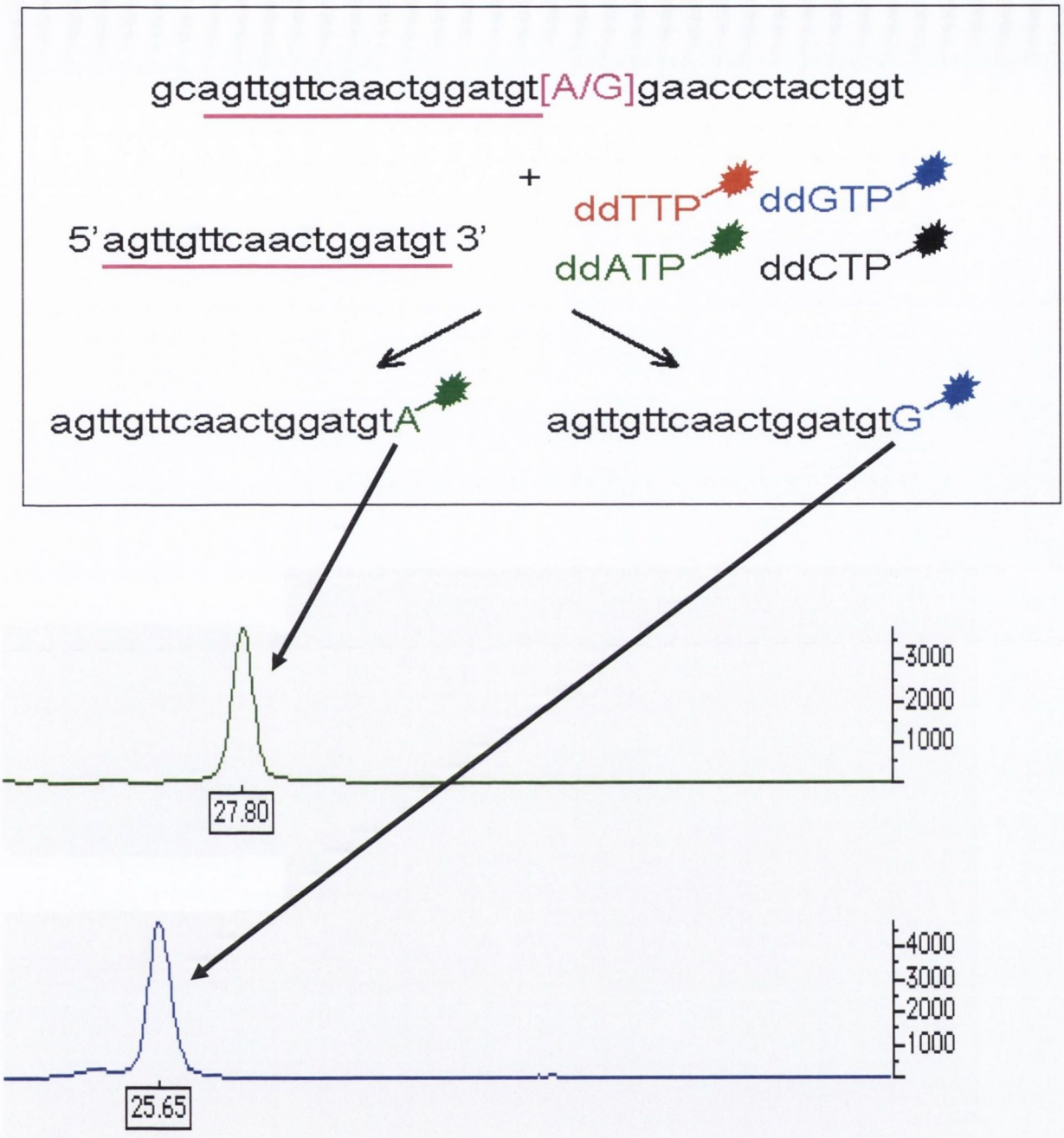


Figure 3.2: Multiplexing primer extension products

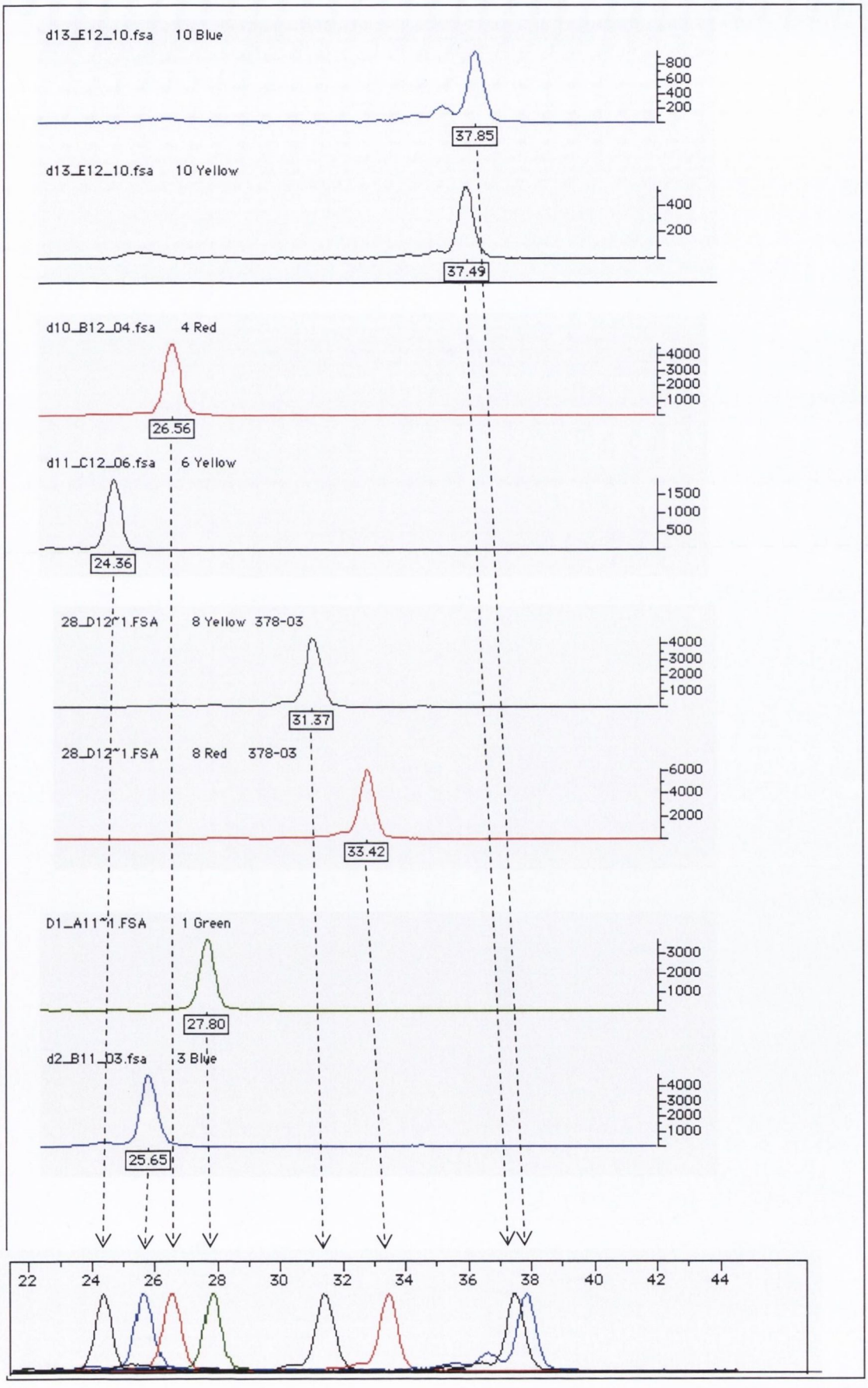


Figure 3.3: SNaPSHOT primer extension genotyping reaction protocol

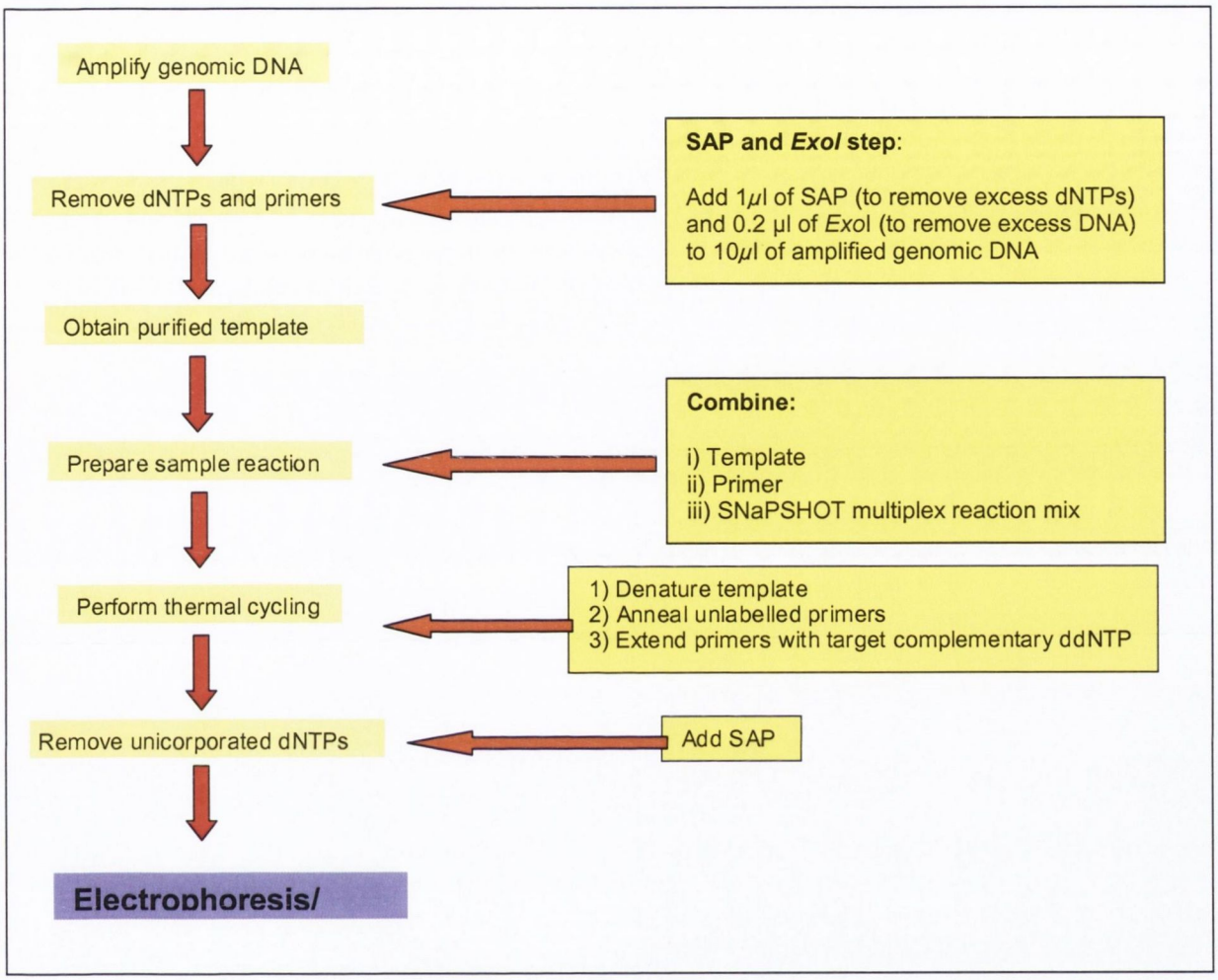


Figure 3.4: Simulating heterozygosity data for DNA pooling

Example of sample data for a SNP genotyped in the case DNA pool and control DNA pool using the SNaPshot reaction. The raw data for each pool is fed into the formula $f(a) = A/(A + kB)$ where k is in the range *0.1* to *10*. The difference between the results for each pool is listed for each simulated value of k.

Case Pool	
Allele	Peak Height
A	1798
B	1135

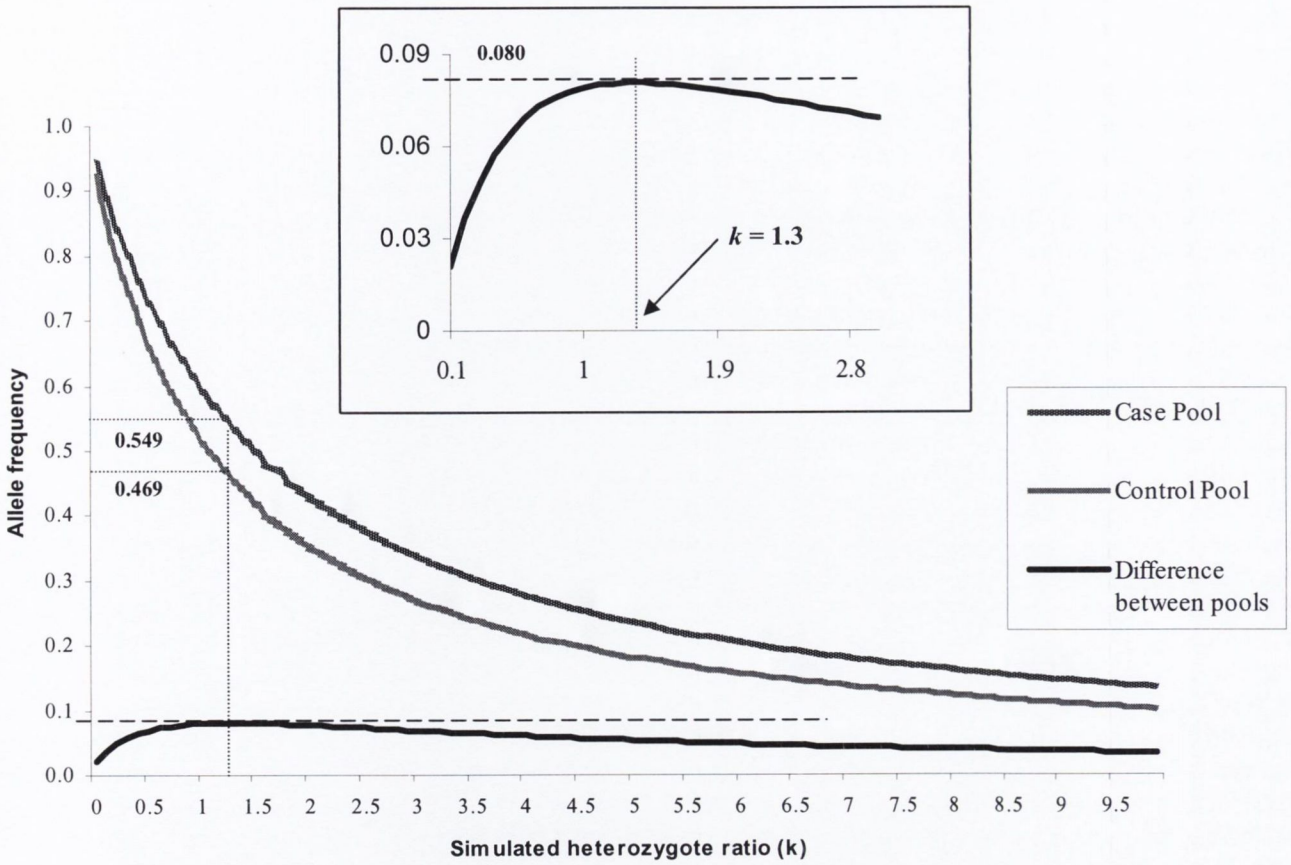
Control Pool	
Allele	Peak Height
A	1500
B	1309

$$f(a) = A/(A + kB)$$

k	**Case Pool Data**	**Control Pool Data**	**Differen**
0.1	1798/(1798 + *0.1**1135) = **0.941**	1500/(1500 + *0.1**1309) = **0.920**	**0.021**
0.2	1798/(1798 + *0.2**1135) = **0.888**	1500/(1500 + *0.2**1309) = **0.851**	**0.037**
:	:	:	:
1	1798/(1798 + *1**1135) = **0.613**	1500/(1500 + *1**1309) = **0.534**	**0.079**
:	:	:	:
3.6	1798/(1798 + *3.6**1135) = **0.306**	1500/(1500 + *3.6**1309) = **0.242**	**0.064**
:	:	:	:
10	1798/(1798 + *10**1135) = **0.137**	1500/(1500 + *10**1309) = **0.103**	**0.034**

Figure 3.5: Maximum potential difference

The results for each pool, and the respective difference between them, are plotted against the range of simulated k values. The line plotting the difference between the pools peaks at a value of 0.080 on the y-axis. This maximum potential difference between the pools occurs at $k = 1.3$ on the x-axis. This can be seen more clearly in the magnified section of the graph displayed in the box above the graph itself. At $k = 1.3$, the frequency of allele A is estimated at 0.549 in the case pool and 0.469 in the control pool. These estimates can be computed to allele counts and tested for evidence of association using a 2 x 2 contingency table to a calculate a χ^2 statistic.



Chapter 4

4.1 Introduction

This chapter outlines the composition and clinical characteristics of the case sample. Included are details on the subgroup of cases with the narrower schizophrenia/schizoaffective disorder phenotype investigated in the molecular genetic studies described in Chapters 5-7.

4.2 Description of sample

4.2.1 Ascertainment

A total of 743 individuals were identified as suitable for the study (having had a diagnosis of a functional psychotic disorder) by the referring agents and referred for assessment. These included 551 individuals identified at the referring Dublin centres and 192 individuals from the Monaghan psychosis study. All who could be contacted (n=680) were approached with information about the study and offered the opportunity to participate by a member of the Dublin clinical research team (AC, SS, JMN) or by their local Community Psychiatric Nurse in Monaghan. At the Monaghan site only 129 of the potential 192 cases could be contacted. This probably reflected the delay between the initial research contact and this current study (n=1-6 years).

4.2.2 Inclusion criteria

Individuals were included in the psychosis sample if:

i) They met criteria for a major psychotic disorder as defined by DSM-IV (see Table 1.1 for these criteria).

ii) Individuals were of self-reported Irish ancestry, being born of two parents from Ireland and having Irish grandparents.

iii) Were over age 16 and gave informed consent.

The studies reported in Chapters 5-7 were an investigation of the Schizophrenia/Schizoaffective phenotype. All included individuals met DSM-IV criteria for Schizophrenia, Schizoaffective disorder or Schizophreniform disorder.

4.2.3 Exclusion criteria

Individuals were excluded where:

i) There was evidence of learning disability, documented in their medical records or, where possible, confirmed by a measure of Premorbid IQ (Weschler Test of Adult Reading (WTAR), Psychological Corporation, Harcourt Assessment Company, San Antonio, US).

ii) Individuals with evidence of a Psychotic Disorder due to a general medical condition, delirium or a dementia were excluded. Diagnosis in this case was made at assessment and from chart review/collateral information from history, laboratory/radiological investigation, and physical examination.

iii) Diagnosis of a substance induced psychotic disorder was made from SCID assessment, chart review and available collateral information. Following DSM-IV guidelines, for example in relation to a diagnosis of schizophrenia, these cases were distinguished by evidence that a substance (e.g. drug abuse or a medication) was likely to be aetiologically related to the symptoms meeting criterion A for that diagnosis.

iv) Individuals were excluded where the age of onset (first evidence of psychotic symptomatology) was over 65 years.

v) Individuals who met criteria for a Pervasive Developmental Disorder were excluded, even if such individuals also met criteria for Schizophrenia. This exclusion criterion was selected because there may be additional genetic and aetiological factors contributing to the development of psychosis in this population.

4.2.4 Exclusion, withdrawals and refusals

Across centres, 124 (18%) of those approached refused to participate. An additional 131 individuals were excluded from the study, principally on failing to meet ethnicity criteria (n=75), on diagnostic grounds (n=47), or because insufficient diagnostic information was available to reach a consensus diagnosis (n=9). Of those excluded on diagnostic grounds, 15 individuals met DSM-IV criteria for a Substance Induced Psychotic Disorder or a Psychotic Disorder secondary to a general medical condition. Eighteen were excluded because of evidence of a co-morbid learning disability. A small number failed to meet the study criteria for DSM-IV major functional psychotic disorder: six cases were excluded because of a diagnosis of a bipolar II disorder; four cases met criteria for a schizoid or schizotypal personality disorder; three cases met criteria for obsessive-compulsive disorder and one individual had a diagnosis of a pervasive developmental disorder.

A small number of potential participants were excluded on health grounds (n=28) and one Dublin case later withdrew from the study. Of the excluded patients, the majority were chronic, elderly patients at the Monaghan site who were unable to give informed consent. At the Dublin centres, six cases were too unwell to participate from a mental health perspective and had not improved sufficiently when re-assessed to be included in the study. These data are summarized in Table 4.1.

4.2.5 Description of the case sample

From the total number referred, 396 individuals met the inclusion criteria, agreed to participate and completed the assessment. From this referred group 299 cases met criteria for inclusion in the studies of schizophrenia/schizoaffective disorder. Laboratory analysis of this sample began before the full sample was collected; therefore 219 cases were available for the investigation of the *APOL* gene family described in Chapter 6 (this is designated the Phase 1 sample). The analyses described in Chapter 7 include the full case sample (n=299); this is described as the Phase 2 sample. The *NRG1* study, described in Chapter 5, included an intermediate number of cases (n=243).

Of the total cases included in the schizophrenia/schizoaffective disorder phenotype (n=299) the majority (n=197 (66%)) were male patients. The average age of case participants was 44.8 years (range 17-76). The average age at onset of illness was 24.2 years (range 12-64). Where it could be unambiguously established from interview, collateral or case note review (n=257), the majority of cases had no family history of a psychotic disorder in a first or second degree relative (n=170 (64%)).

Half of the participants had a course of illness defined by DSM-IV as 'episodic with interepisode residual symptoms' (n=155 (52%)); a further quarter had a 'continuous' course of illness (n=75 (25%)); the remainder had either a course defined as 'episodic with no interepisode residual symptoms' (n=63 (21%)) or had a single episode of illness (n=6 (2%)). Participants also had significant levels of symptomatology (as measured by worst-ever PANSS scores); summarized in Table 4.2. There were no significant differences between the 219 cases defined as the Phase 1 study sample and the additional cases in age at onset, sex, duration of illness at interview, course of illness or severity of worst-ever PANSS scores. Other demographic details are detailed in Table 4.2

4.2.5.1 Clinical assessments

The full clinical assessment included a diagnostic interview (SCID-P), clinical symptoms assessment (PANSS), a physical assessment of dysmorphology (MPA measurement) and neurological symptoms (AIMS, NSS). These measures are described more fully in Chapter 3. In scoring the PANSS, current symptomatology was assessed and from the available information on previous episodes a score for worst-ever symptomatology was also derived. These measures were all applied in the Dublin assessments but in the Monaghan dataset only current PANSS measures were recorded and dysmorphology was not assessed.

In piloting these assessments with the early cases, the total assessment was taking 3-4 hours. At the request of the participants, we divided the assessment into two interviews performed over two periods of ~2 hours duration. The first meeting involved collection of clinical, diagnostic and demographic information. The second meeting was for completion of the measures of dysmorphology and neurological

assessment. Of the 396 individuals who have been included in the sample collection to date, 307 have completed the full assessment.

4.2.5. Description of the control sample

The controls were obtained from the Irish Blood Transfusion Service from attendees at their clinic in St. James's Hospital. These were ethnically matched with the case population by the same Inclusion criterion employed for cases (see Section 4.2.2). As a condition of the Ethics Committee approval, no additional information was available on individual control participants. Therefore, it was not possible to screen donors for psychiatric disorder, although in Ireland individuals who are on regular medications are excluded from donation and there is no financial recompense provided to donors. For these reasons, it is unlikely that affected individuals are over-represented among the controls sampled for this study. This would obviously be more of an issue in a case-control study of a disorder of greater population prevalence. The full control sample included 645 individuals. For future studies the availability of a more representative sample of the Irish population with demographic information and a screening interview would be beneficial and is indeed planned as part of the Irish GeneBank project (http://www.dmmc.ie/BIORESOURCES.htm).

Table 4.1: Ascertainment/inclusion of cases for the project

Cases	N
Potential participants identified	743
Approached re: participating	680
Included	396
Schizophrenia	230
Schizoaffective disorder	65
Bipolar I disorder	65
Schizophreniform disorder	4
Psychotic disorder NOS	6
Delusional disorder	8
Depressive disorder (with psychosis)	18
Refused to participate	124
Excluded	131
Ethnicity	75
Diagnosis	47
Co-morbid learning disability	18
Medical/substance induced	15
Bipolar II disorder	6
Schizotypal/schizoid PD	4
Obsessive-compulsive disorder	3
Pervasive developmental disorder	1
Insufficient diagnostic information	9

Table 4.2 Demographic details of the Phase 2 sample

		N
Sex	Male	197 (66%)
	Female	102 (33%)
Age at onset		24 (range 12-64 years)
Age at interview		44.8 (range 17-76 years)
Family history	Positive	170 (64%)
	Negative	87 (46%)
PANSS scores	Positive symptoms scale	26.6 (range 8-41)
	Negative symptom scale	19.8 (range 7-36)
	General psychopathology Scale	41.2 (range 14-73)
Marital status	Married	23 (7.7%)
	In a relationship	21 (7.0%)
	Single	26 (8.7%)
	Separated	226 (75.6%)
	Widowed	3 (1.0%)
Work	Full-time employment	41 (13.7%)
	Part-time employment	38 (12.7%)
	Retired	24 (8.0%)
	Home duties	17 (5.6%)
	Unemployed/disability allowance	179 (60.0%)
Accommodation	Independent	150 (50.1%)
	Independent hostel	15 (5.0%)
	Living with carer	63 (21.0%)
	Hostel	57 (19.1%)
	Hospital	14 (4.6%)

Chapter 5

Investigation of the Neuregulin-1 (NRG1) candidate gene at the chromosome 8p susceptibility locus

5.1 Introduction

5.1.1 Evidence for a SZ susceptibility locus at chromosome 8p21-22

The chromosome 8p21-22 locus has been implicated in schizophrenia susceptibility by a number of studies. The original published report, by Pulver et al, (1995), suggested evidence for a SZ susceptibility locus at 8p22-23 in 57 US multiplex pedigrees using affective sibling pair (minimum $p=4x10^{-5}$ at marker D8S258) and LOD score analyses (maximum LOD score 2.35 at marker D8S136). This finding was supported by linkage data from the Irish Study of High-Density Schizophrenia Families (ISDHF) of 265 multiplex families containing 1,408 individuals ascertained throughout the island of Ireland (Kendler et al, 1996). In this study four different phenotype models were investigated; the narrowest included cases with schizophrenia and poor outcome schizoaffective disorder; intermediate and broad models included, in addition, non-affective psychotic disorders and spectrum personality disorders; the very broad model also included cases with psychiatric disorders such as non-psychotic affective disorders, anxiety disorders and alcohol dependence syndrome. The narrowest model was equivalent to the single model investigated by Pulver et al, (1995).

In the Irish study the strongest evidence for linkage, using two-point heterogeneity LOD scores, was found for markers D8S1731 (maximum LOD=2.00), D8S1715 (maximum LOD=2.52) and D8S133 (maximum LOD=2.08) using the broadest phenotype. The Kendler study also employed multipoint nonparametric linkage analysis and identified two distinct peaks at this locus using the intermediate diagnostic model-at marker D8S1731 (maximum z=2.43, p=0.008) and D8S298 (maximum z= 2.51, p=0.006). Markers providing suggestive evidence of linkage overlapped between the studies and both found evidence for linkage over a broad chromosomal region of 20-25 cM defined by markers D8S552 and D8S136. Because

of the differences in diagnostic models the range of phenotypes influenced by this putative susceptibility locus was less clear.

Further evidence for linkage has been suggested by an extension of the original report (Blouin et al, 1998) and in independent family samples of Canadian, Icelandic and African-American pedigrees (Brzustowicz et al. 1999; Gurling et al 2001; Kaufmann et al, 1998). As has been the case with other complex genetic disorders, this linkage finding has not been universally replicated, and multiple negative studies have also been reported. The recent meta-analysis of 20 completed schizophrenia genome scans has provided significant statistical support for chromosome 8p as a schizophrenia susceptibility locus (Lewis et al, 2003).

5.1.2. Identification of a positional candidate gene for SZ at this locus

In 2002, the deCODE group reported suggestive evidence of linkage to the 8p12-21 locus in a study of 33 Icelandic families with 105 individuals affected with schizophrenia or schizoaffective disorder (Stefansson et al, 2002). Of note, this family sample may overlap with the sample previously reported by Gurling et al, (2001). Using a model independent, affecteds-only allele sharing method (Gudbjartsson et al, 2000), the group reported a maximum multipoint LOD score of 3.06 at marker D8S352 in a 30 cM linkage peak. This fell in the region of susceptibility identified by the meta-analysis, but was 10-15 cM centromeric of the region reported by most other studies. Following up on this finding, the authors investigated the region using their full sample, which included patients with no family history of schizophrenia. The final study sample consisted of 476 cases (440 with schizophrenia, 32 with schizoaffective disorder and 4 with unspecified functional psychosis) diagnosed according to Research Diagnostic Criteria (RDC) (Spitzer et al, 1978).

After genotyping markers at high density across the linked region the authors identified several risk haplotypes for schizophrenia. The haplotypes mapped to a region containing the neuregulin-1 gene (*NRG1*). Stefansson and colleagues identified three 'at risk' haplotypes (designated Hap A, B and C), which spanned the *NRG1* locus and shared a common core haplotype ~191kb in length (simulated-p=0.67-$8.7x10^{-5}$ OR=2.1) (see Figure 5.1).

5.1.3 Identification of the *NRG1* gene

Neuregulin-1 (*NRG1*; OMIM 142445) was identified as a 44-kD glycoprotein that interacts with the *NEU/ERBB2* receptor kinase to increase its phosphorylation on tyrosine residues. The *NRG1* gene produces a significant number of isoforms by alternative splicing, these include the heregulins (*HRGs*) and glial growth factors (*GGFs*). The gene maps to chromosome 8p21-p12 and extends from 31554907 to 32683142 (NCBI build 34, August 2003) on the direct strand, a distance of 1128.24 kb. A family of neuregulin genes (*NRG1-3*) have now been identified, these are growth and differentiation factors that are related to epidermal growth factor (*EGF*; OMIM 131530). The receptors for neuregulins are the *ERBB* family of tyrosine kinase transmembrane receptors, which includes the EGF receptor (*EGFR*; OMIM 131550), *ERBB2* (OMIM 164870), *ERBB3* (OMIM 190151), and *ERBB4* (OMIM 600543). Through interaction with *ERBB* receptors, neuregulins induce the growth and differentiation of many different cell types including neuronal and glial cells (Burden & Yarden, 1997).

5.1.4 *NRG1* as a functional candidate gene for schizophrenia

As the aetiology of schizophrenia is obscure, *a priori* evidence for *NRG1* as a functional candidate gene is limited. By necessity, current hypotheses of schizophrenia pathobiology are broad and the prevailing view is that schizophrenia is a neurodevelopmental disorder which impacts on CNS maturation. This may be mediated through abnormal cortical wiring or dysregulation of synaptic function with secondary downstream effects on neurotransmitter systems (Weinberger, 1995). *NRG1* isoforms may play a role at each stage in this process through influencing gliogenesis and neuronal migration during CNS development or by inhibiting long-term potentiation, a form of synaptic plasticity (Falls, 2003). *NRG1* also regulates the development and function of cortical GABAerbic interneurones that have been implicated in schizophrenia (Yau et al, 2003). However, it is important to point out that many other CNS expressed genes may have similar effects and be equally plausible candidates for involvement in schizophrenia pathogenesis.

5.1.4.1 Human functional studies of *NRG1*

NRG1-ERBB3 signalling is known to be involved in modulating gliogenesis. Two recent microarray studies suggest that aberrant functioning at this receptor may be involved in schizophrenia. Both studies identified altered expression of the neuregulin receptor gene *ERBB3* (OMIM 190151) in post mortem samples of prefrontal cortex from schizophrenia subjects. Hakak et al, (2001) compared gene expression profiles in the prefrontal cortex of 12 chronic schizophrenia cases on antipsychotic medication, 4 similar cases that had been off medication for at least 6 weeks and 12 control patients from nursing homes. The authors detected a significant reduction in the level of *ERBB3* expression with levels in the cases being 58% of control levels. A comparison of expression levels between cases that had been on or off medication at the time of death did not indicate a significant difference. More recently, Tkachev et al, (2003) replicated this finding in an independent study of 11 schizophrenia cases and controls. With quantitative RT-PCR they confirmed a 1.97-fold ($p=5x10^{-4}$) down-regulation of *ERBB3* in schizophrenia cases. These results have been interpreted as evidence that alteration in NRG1-ERBB3 signalling either leads to, or is a result of, oligodendrocyte deficits in schizophrenia. Other possibilities, discussed more fully in Chapter 1, must also be considered as the pathological samples available for such studies represent gene expression changes decades after the onset of illness and may be secondary to many other factors.

5.1.4.2 Animal functional studies of *NRG1*

Stefansson et al, (2002) also presented functional evidence of involvement of *NRG1* in SZ pathobiology from animal models of SZ. They first confirmed a previous report that *NRG1* hypomorphic mice show hyperactivity under a number of behavioural conditions including in the novel open field test. Second, by administering a non-sedating dose of the atypical antipsychotic drug clozapine, they demonstrated that this hyperactivity responded to treatment at a dose that had no behavioural effects on wild-type mice.

Stefansson et al, (2002), reviewed the body of literature indicating prepulse inhibition (PPI) deficits in schizophrenia subjects. As a further experiment they demonstrated that similar deficits were evident in the *NRG1* hypomorphic mice and again were reversed with clozapine treatment. Finally, the *NRG1* mutant mice also had

approximately 16% fewer functional NMDA receptors than wild-type mice. The phenotypes modelled in the animal studies reported by the deCODE group, approximate non-specific features of schizophrenia and the biological validity of these animal models is yet to be established. However, the deficiency of NMDA receptors in *NRG1* mutant mice is consistent with evidence that *NRG1* has a role in NMDA receptor function and subsequent glutamate transmission in schizophrenia (Ozaki et al, 1997).

5.1.5 Rationale for this study

A family-based study in the Irish population provided suggestive evidence of linkage to the chromosome 8p susceptibility locus that contains the *NRG1* gene. By a combination of linkage and association approaches, Stefansson et al, (2002) identified a core haplotype ~191kb in length, mapping to the *NRG1* locus which was associated with schizophrenia in the Icelandic population. Although not conclusive, independent gene expression and animal studies provide suggestive evidence for *NRG1* as a functional candidate gene for schizophrenia.

There are historical connections between the Irish and Icelandic populations: the founder population of Iceland being of Norse and Gaelic ancestry (Helgason et al, 2001). Thus an independent case-control sample from the larger, ethnically related and more out-bred Irish population might have the capacity both to test the finding of association in this region, and refine the location of the putative susceptibility variants.

5.2 Methodology

5.2.1 Sample size and study power

This investigation was performed using an extension of the Phase 1 schizophrenia/schizoaffective disorder sample (this is described in detail in Chapter 4). The study included patients diagnosed with DSM-IV defined schizophrenia (n = 196) or schizoaffective disorder (n = 47). The control sample, drawn from Irish blood donors, was not specifically screened for psychiatric illness but individuals were not taking regular prescribed medications. A power calculation assuming the same haplotype frequencies in our sample as in the original report by Stefansson et al,

(2002) indicated that our study had 65% power to detect an association of similar effect size at a significance level of 0.05 (Genetic Power Calculator; Purcell et al, 2003).

5.2.2 DNA handling and Genotyping

In total, 17 microsatellite markers and 3 SNPs were genotyped in our sample. These markers spanned a region of ~911 KB, which covers the putative susceptibility locus identified by Stefansson et al, (2002). The names and relative positions of all 20 markers are detailed in Table 5.1. The SNPs were genotyped using a method based on template-directed primer extension and detection by fluorescence polarization (FP) (Chen et al, 1999). All samples were genotyped for the above markers by deCODE Genetics using protocols previously outlined (Stefansson et al, 2002; Gretarsdottir et al, 2002). A strict sample handling protocol was employed and deCODE remained blind to the affection status of all samples until all genotypic data was returned to Dublin and statistical analysis was complete.

5.2.3 Statistical methods

SNPs were tested for association with the phenotype by using a 2 x 2 contingency table to calculate a χ^2 statistic. Microsatellite markers were tested for association using the T2 χ^2 statistic of CLUMP and significance was assessed using Monte Carlo simulations (Sham and Curtis, 1995). Inter-marker linkage disequilibrium (LD) was measured using D' (Lewontin, 1964); the D' values were calculated using GOLD (http://www.well.ox.ac.uk/asthma/GOLD/index.html). Haplotype association analysis was performed using FASTEHPLUS and p-values were calculated by simulation (Zhao and Sham, 2002). GENECOUNTING, was used to output individual haplotype frequencies from cases and from controls for the purpose of testing specific haplotypes (Zhao et al, 2002). All markers were tested for Hardy-Weinberg Equilibrium (HWE) in both case and control samples as described in Chapter 3.

5.3 Results

All markers were in HWE (simulated p>0.05) and the details are available at our research group website (http://www.tcd.ie/Psychiatry/Neuropsychiatry). The individual allele frequencies at each marker locus are detailed in Tables 5.2-5.22.

5.3.1 Tests for association between SZ and the core NRG1 haplotype

The author first tested the core 'at risk' haplotype as defined by Stefansson et al, (2002) for evidence of association with SZ in our sample. This haplotype is represented by alleles 1, 0 and 0 from markers SNP8NRG221533 (M7), 478B14-848 (M12) and 420M9-1395 (M13) respectively. In our study, the frequency of the *1 0 0* haplotype was 9.6% in cases and 8.4% in controls (Odds Ratio (OR) = 1.18 (95% Confidence Interval (95%CI) 0.67 – 2.09), p = 0.534) (see Figure 5.1)

5.3.2 Single marker association analyses across the NRG1 locus

All 20 markers genotyped across this region were then systematically tested for evidence of association. The results of these single marker association analyses are detailed in Table 5.1. Two markers show evidence of association, 450K14-72458 (M6; p = 0.039) and 473C15-439 (M18; p = 0.036).

5.3.3 Haplotype structure at NRG1 in Irish and Icelandic samples

D' values for all marker combinations were calculated for cases and controls. Both case and control data sets produce a similar pattern of LD across the region. Two 'blocks' of markers (450K14-72458 (M6) to SNP8NRG243177 (M10); 487-2 (M11) to 420M9-116I12 (M16)) exhibit strong inter-marker LD, separated by a segment of very low LD. This is graphically represented for our case sample in Figure 5.2. The Icelandic core haplotype extends across these two blocks. A similar structure was identified when using the r^2 definition of LD. Haplotype structure may differ between populations; hence we investigated whether a different pattern of association was present in the Irish sample.

5.3.4 Haplotype analyses of the NRG1 locus in the Irish sample

Association analyses were performed using haplotypes constructed from adjacent markers across the region in blocks of 2, 3 or 4 markers using the 'sliding-window'

method. The results (see Table 5.1) suggest evidence of association in two sections of the region under investigation. The strongest evidence came from 2-, 3-, and 4 marker haplotypes in the region of markers 420M9-1395 (M13) to 473C15-439 (M18). These markers lie within the second haplotype 'block' and overlap with the 3' end of the Icelandic core haplotype.

5.3.5 Overlap between identified Irish and Icelandic risk haplotypes

The next investigation was whether haplotype associations in our sample shared common allelic structure with the Icelandic 'risk' haplotypes (HapA, B and C; see Figure 5.1). In such an instance, further refinement of 'risk' haplotypes could be possible. To test this hypothesis combinations of markers that had been genotyped in both samples and shared common alleles at individual loci were specifically examined. This identified a two-marker haplotype (*0 18*) constructed from microsatellite markers 420M9-1395 (M13) and D8S1810 (M15), which shares common alleles with the HapB haplotype in the Icelandic sample (see Figure 5.1). The *0 18* haplotype from markers M13 and M15 is in significant excess when patients are compared with controls (19.4% v 12.3%; OR 1.71 (95%CI 1.09 - 2.68), p = 0.013). Henceforth this refined 'at risk' haplotype is termed $HapB_{IRE}$.

5.3.6 Replication of evidence for this association in other populations

None of the single marker or haplotype findings reported above were corrected for multiple testing. Standard statistical measures, such as the Bonferroni correction assume that each test performed is independent. The tests performed in this study at each marker locus are not independent as many of the markers investigated are highly correlated. Employing the Bonferroni correction would overcorrect for potential Type I error and reduce study power. At the time of this study, no readily accessible method of correcting for multiple testing for SNP or microsatellite markers in LD was available. More recently, methods for correcting data from multiple markers in LD have emerged, but these can only applied to SNP data. For example, the method developed by Nyholt (2004) was applied to correct for multiple testing of SNPs in the analyses reported in Chapter 7.

As an alternative, the author sought to replicate this finding by investigating whether this haplotype contributes to schizophrenia susceptibility in two other European populations. For this analysis, Dr Stefansson (deCODE) kindly agreed to provide the relevant marker data from the Icelandic SZ sample and confirmed that another group, based in Scotland, had provided DNA to deCODE for anonymous genotyping of a case-control SZ sample. The principle investigator of this group, Professor David St Clair (Dept of Mental Health, University of Aberdeen Medical School, Aberdeen, Scotland) also kindly agreed to provide genotypic data from this Scottish sample for the purpose of further investigating this association.

The Scottish sample consisted of 609 unrelated patients, diagnosed with DSM-IIIR defined schizophrenia or schizoaffective disorder. This sample, ascertained in Scotland, was ethnically matched with 618 controls drawn from the same population. Diagnosis was based on psychiatric case note inspection and additionally through the use of the Schizophrenia and Affective Disorders Schedule (Spitzer & Endicott, 1977). Diagnosis was confirmed by consensus of two senior psychiatrists. Informed consent was obtained from all patients and control individuals (further details are provided in Stefansson et al, 2003).

The same LD pattern was evident in Scottish cases and controls. This represented the same structure of two haplotype blocks separated by a discrete region of low LD, typical of what is termed in the literature a 'recombination hotspot' (Figure 5.3). The author next investigated the $HapB_{IRE}$ haplotype in the Scottish sample. This demonstrated an excess of this two-marker haplotype in Scottish SZ cases compared to controls (17.0% v 13.5%; OR 1.31 (95%CI 1.01 – 1.70), p = 0.036). However, there was no evidence of an excess of $HapB_{IRE}$ in the Icelandic cases (20.3%) compared to controls (20.1%).

Finally, to delineate this haplotype the extent of the association in the region of $HapB_{IRE}$ was tested with the additional marker data. This identified a three-marker haplotype (HapD) that extended 3' of $HapB_{IRE}$ and included 420M9-116I12 (M16) (see Figure 5.2). This haplotype was in significant excess in our sample (7.3% in cases v 3.5% in controls; OR 2.26 (95%CI 1.04 – 4.98), p = 0.024), but not in the Scottish sample (5.2% in cases v 5.8% in controls). From these data the author

concluded that the most likely position of the causative variant(s) is in the region of overlapping allelic association ($HapB_{IRE}$) across these three study populations.

5.4 Discussion

This investigation provides additional support for a susceptibility locus for schizophrenia on chromosome 8p21-22 by replicating the finding of association previously reported with *NRG1* in an Icelandic sample (Stefansson et al, 2002). The study indicates that a risk haplotype ($HapB_{IRE}$), which shares common markers with one of the risk haplotypes in the original report (HapB), is associated with schizophrenia in an Irish schizophrenia case-control population. In addition, this novel risk haplotype is in significant excess in Scottish but not Icelandic SZ samples. In the Irish sample the original core risk haplotype, identified as associated with SZ in Icelandic and subsequently Scottish samples, was not significantly in excess in cases.

$HapB_{IRE}$ is derived from 2 markers separated by a distance of ~99 kb. $HapB_{IRE}$ is within the second of two haplotype 'blocks' that were identified from Irish inter-marker LD data (Figure 5.2) and which extends from 487-2 (M11) to 420M9-116I12 (M16). However, it is of interest to note that at this marker resolution a similar pattern of LD is present in Irish, Scottish and Icelandic samples (Figure 5.3). The Icelandic core haplotype extends across the low LD 'gap' between (SNP8NRG243177 (M10) and 420M9-116I12 (M11) and includes all of the first haplotype 'block' between markers 450K14-72458 (M6) to SNP8NRG243177 (M10). The same pattern is evident in the Scottish but not in the Irish sample. This challenges the view that LD-mapping can confirm association within discrete haplotype blocks and is in keeping with evidence from other studies that neighbouring haplotype blocks can be highly correlated or overlap substantially (Daly et al, 2001; Dawson et al, 2002).

The differences in patterns of association between these samples may be due to a number of factors. This study may have lacked power to detect association with the Icelandic/Scottish core haplotype as a result of the smaller sample size. Alternatively, there may be differences in haplotype structure at this locus between these populations, which would complicate direct comparison of haplotype data between studies. In a population isolate such as Iceland there have been relatively few meioses

since the original founding population, which likely represented a small subset of the genetic diversity of the general European population. It is reasonable to expect that a particular haplotype or haplotypes containing a susceptibility variant might bridge a region of low LD and be represented by an extended haplotype in this sample. However, this does not explain the association detected with this haplotype in the more out-bred Scottish population. One possibility relates to differences in ascertainment, as the Scottish sample originated from the eastern seaboard and may share more ancestry with the Icelandic population than an Irish population. The complexity of haplotype structure is well recognised, as a result definitive interpretation of the three data sets is difficult and further investigation at greater marker resolution will be required. The possibility that different variations within the region of *NRG1* contribute to SZ susceptibility must also be considered. But in such a case it may be more plausible that the associated risk alleles would fall on different marker haplotypes.

From these data, the likely susceptibility allele (or alleles) are within the second haplotype 'block' currently delineated by markers 487-2 (M11) and 420M9-116I12 (M16). Further analysis of the position of the susceptibility variant or variants would then be dependent on the extent of LD between it and surrounding markers. Were this region to display extensive LD, more exact positioning of a causative variant may require investigation in older populations where LD may have dissipated to a greater degree.

The refinement of the at-risk haplotype to the region of $HapB_{IRE}$ has potentially important implications for the identity of the putative susceptibility gene (see figure 1). These markers are located between Exon 1 of Glial Growth Factor 2 (*GGF2*), an alternatively spliced isoform of *NRG1* and the putative first exon of an EST cluster (*Hs.*97362; http://www.ncbi.nlm.nih.gov/UniGene/). Exon 1 of *GGF2* is not known to be shared by any other isoforms of the *NRG1* gene, and is located ~70 kb upstream of 487-2 (M11) and ~2kb downstream of SNP8NRG243177 (M10). This probably positions this exon within the first haplotype 'block'. Although *GGF2* is widely expressed in the brain during development, *in vivo* function of this isoform has yet to be demonstrated (Falls, 2003). The next *NRG1* exon is located ~633 kb downstream of 420M9-116I12 (M16). Positioned between 420M9-116I12 (M16) and the second

NRG1 exon is the EST cluster (*Hs.*97362). In addition to the 3 ESTs identified in UniGene, we have identified 2 further ESTs (BU567364 and BG202570) in the NCBI GenBank database that share homology with this cluster and map to this locus. Together these 5 ESTs identify 9 putative exons, all positioned from ~30kb to ~340 kb downstream of 420M9-116I12 (M16). As yet this EST cluster is of unknown function. However, given its close physical proximity to $HapB_{IRE}$, it warrants consideration in the search for the putative susceptibility variant or variants at this locus.

Table 5.1: Results (simulated p values) from association analyses of single markers and of 2-, 3- and 4-marker haplotypes constructed from adjacent markers

	Marker Name	Inter-Marker Distances (Kb)	Single Markers[a]	2-Marker Haplotypes	3-Marker Haplotypes	4-Marker Haplotypes
M1	D8S1770	-	0.369	0.250[d]	0.043[e]	0.201[f]
M2	29H12-1	128.9[c]	0.349	0.059	0.029	0.067
M3	29H12-7320	5.4	0.116	0.368	0.606	0.914
M4	D8S1711	49.8	0.419	0.638	0.472	0.498
M5	29H12-121L21	65.1	0.278	0.163	0.572	0.884
M6	450K14-72458	79.0	0.039	0.552	0.570	0.233
M7	**SNP8NRG221533[b]**	5.7	0.724	0.419	0.432	0.387
M8	478B14-642	18.8	0.177	0.225	0.217	0.179
M9	**SNP8NRG241930[b]**	1.4	0.546	0.109	0.821	0.791
M10	**SNP8NRG243177[b]**	1.2	0.244	0.699	0.652	0.643
M11	487-2	74.9	0.753	0.696	0.643	0.505
M12	**478B14-848**	18.4	0.749	0.307	0.157	0.166
M13	**420M9-1395**	76.6	0.106	0.183	0.037	0.006
M14	420M9-1	22.4	0.274	0.174	0.082	0.220
M15	D8S1810	3.9	0.133	0.045	0.118	0.004
M16	420M9-116I12	33.8	0.521	0.710	0.169	0.236
M17	473C15-533	108.6	0.380	0.126	0.205	0.573
M18	473C15-439	66.2	0.036	0.341	0.370	-
M19	72H22-1	36.2	0.058	0.347	-	-
M20	72H22-36	114.2	0.919	-	-	-

[a] Individual allele frequencies for all markers are not displayed. These are available on request.
[b] SNP markers; p values for single marker analysis of SNPs are not simulated.
[c] 128.9 kb represents the distance between M1 and M2. Data is presented in the same format for each adjacent pair of markers down this column.
[d] Simulated p value for 2-marker haplotype analysis of markers M1 and M2. Data is presented in the same format for each adjacent pair of markers down this column.
[e] Simulated p value for 3-marker haplotype analysis of markers M1, M2 and M3. Data is presented in the same format for each set of 3 adjacent markers down this column.
[f] Simulated p value for 4-marker haplotype analysis of markers M1, M2, M3 and M4. Data is presented in the same format for each set of 4 adjacent markers down this column.
Markers in bold are those that define the Icelandic core haplotype.

Table 5.2: Allele frequencies at locus D8S1770 (M1)

Allele	Cases	Controls
0	0.353	0.321
6	0.015	0.009
8	0.053	0.034
10	0.560	0.615
12	0.017	0.018
14	0.002	0.002

Table 5.3: Allele frequencies at locus 29H12-1 (M2)

Allele	Cases	Controls
-10	0.274	0.240
-8	0.016	0.037
-6	0.002	0.007
-2	0.011	0.005
0	0.354	0.346
1	0.000	0.007
2	0.005	0.012
4	0.290	0.309
6	0.043	0.039
8	0.002	0.000
10	0.002	0.000

Table 5.4: Allele frequencies at locus 29H12-7320 (M3)

Allele	Cases	Controls
-8	0.017	0.002
-4	0.332	0.279
-2	0.013	0.025
0	0.212	0.194
2	0.004	0.005
4	0.033	0.069
6	0.063	0.081
8	0.105	0.130
9	0.002	0.000
10	0.131	0.135
12	0.035	0.039
14	0.046	0.032
16	0.000	0.000
18	0.007	0.010

Table 5.5: Allele frequencies at locus D8S1711 (M4)

Allele	Cases	Controls
-6	0.007	0.002
-4	0.000	0.002
-2	0.335	0.295
0	0.651	0.686
2	0.007	0.015
4	0.000	0.000

Table 5.6: Allele frequencies at locus 29H12-121L21 (M5)

Allele	Cases	Controls
0	0.709	0.715
4	0.182	0.150
8	0.106	0.136
12	0.002	0.000

Table 5.7: Allele frequencies at locus 450K14-72458 (M6)

Allele	Cases	Controls
-13	0.003	0.003
-11	0.098	0.054
-9	0.000	0.003
-7	0.031	0.030
-6	0.000	0.003
-5	0.152	0.116
-4	0.008	0.000
-3	0.052	0.043
-2	0.003	0.003
-1	0.015	0.027
0	0.273	0.323
1	0.003	0.003
2	0.119	0.169
3	0.010	0.003
4	0.204	0.196
5	0.000	0.003
6	0.021	0.013
8	0.008	0.011
9	0.003	0.000

Table 5.8: Allele frequencies at locus SNP8NRG221533 (M7)

Allele	Cases	Controls
1	0.371	0.383
3	0.629	0.617

Table 5.9: Allele frequencies at locus 478B14-642 (M8)

Allele	Cases	Controls
-2	0.006	0.002
0	0.002	0.000
2	0.356	0.373
4	0.258	0.217
6	0.025	0.012
8	0.038	0.064
10	0.025	0.012
12	0.169	0.160
14	0.083	0.099
16	0.023	0.031
18	0.002	0.002
20	0.011	0.024
22	0.000	0.002
24	0.000	0.002

Table 5.10: Allele frequencies at locus SNP8NRG241930 (M9)

Allele	Cases	Controls
2	0.686	0.665
3	0.314	0.335

Table 5.11: Allele frequencies at locus SNP8NRG243177 (M10)

Allele	Cases	Controls
1	0.615	0.577
3	0.385	0.423

Table 5.12: Allele frequencies at locus 487-2 (M11)

Allele	Cases	Controls
-4	0.005	0.010
0	0.175	0.200
4	0.081	0.094
8	0.304	0.280
12	0.016	0.002
16	0.175	0.181
20	0.235	0.225
24	0.009	0.007

Table 5.13: Allele frequencies at locus 478B14-848 (M12)

Allele	Cases	Controls
-2	0.109	0.129
0	0.367	0.325
2	0.154	0.157
4	0.201	0.210
6	0.154	0.173
8	0.002	0.000
10	0.012	0.007

Table 5.14: Allele frequencies at locus 420M9-1395 (M13)

Allele	Cases	Controls
-4	0.058	0.031
-2	0.187	0.212
0	0.408	0.411
2	0.232	0.259
4	0.016	0.026
6	0.005	0.008
8	0.092	0.052
10	0.003	0.000
12	0.000	0.000

Table 5.15: Allele frequencies at locus 420M9-1 (M14)

Allele	Cases	Controls
-12	0.090	0.056
-8	0.030	0.036
-4	0.118	0.101
0	0.219	0.189
4	0.258	0.291
8	0.215	0.248
12	0.062	0.068
16	0.009	0.011

Table 5.16: Allele frequencies at locus D8S1810 (M15)

Allele	Cases	Controls
0	0.244	0.295
8	0.000	0.005
10	0.012	0.005
12	0.007	0.011
14	0.060	0.077
16	0.320	0.332
18	0.272	0.191
20	0.016	0.024
22	0.069	0.053
24	0.000	0.003
26	0.000	0.003

Table 5.17: Allele frequencies at locus 420M9-116I12 (M16)

Allele	Cases	Controls
-18	0.002	0.000
-16	0.019	0.009
-14	0.002	0.000
-12	0.106	0.096
-10	0.087	0.082
-8	0.111	0.108
-6	0.231	0.209
-4	0.101	0.073
-2	0.038	0.047
0	0.269	0.336
2	0.031	0.033
4	0.002	0.007

Table 5.18: Allele frequencies at locus 473C15-533 (M17)

Allele	Cases	Controls
0	0.799	0.836
8	0.004	0.005
14	0.002	0.000
16	0.004	0.007
18	0.011	0.000
19	0.004	0.000
20	0.091	0.072
22	0.085	0.079
24	0.000	0.002

Table 5.19: Allele frequencies at locus 473C15-439 (M18)

Allele	Cases	Controls
-12	0.004	0.008
-6	0.009	0.005
-4	0.009	0.003
-2	0.057	0.018
0	0.189	0.217
2	0.506	0.497
4	0.087	0.089
6	0.081	0.086
8	0.045	0.071
10	0.013	0.005

Table 5.20: Allele frequencies at locus 72H22-1 (M19)

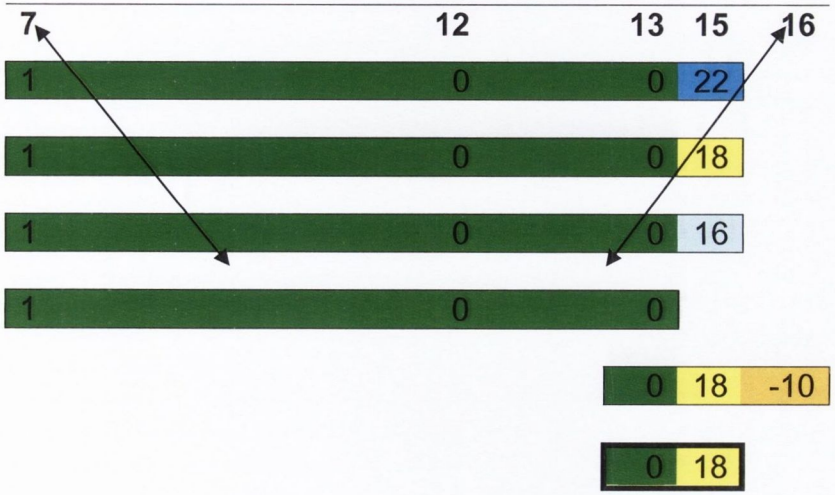
Allele	Cases	Controls
-6	0.002	0.002
-4	0.013	0.016
-2	0.040	0.078
0	0.695	0.610
2	0.162	0.176
4	0.027	0.034
6	0.009	0.005
8	0.053	0.078
12	0.000	0.002

Table 5.21: Allele frequencies at locus 72H22-36 (M20)

Allele	Cases	Controls
-4	0.005	0.000
-2	0.353	0.367
0	0.583	0.569
1	0.002	0.002
2	0.057	0.062

Table 5.22: Simulated p values for assessment of Hardy-Weinberg equilibrium at M1 – M20 in cases and controls

Marker	Cases	Controls
M1	0.513	0.603
M2	0.228	0.233
M3	0.295	0.168
M4	0.643	0.540
M5	0.753	0.190
M6	0.800	0.155
M7	0.905	0.068
M8	0.995	0.015
M9	0.240	0.458
M10	0.908	0.038
M11	0.920	0.783
M12	0.870	0.313
M13	0.718	0.370
M14	0.963	0.985
M15	0.845	0.283
M16	0.498	0.248
M17	0.055	0.695
M18	0.073	0.165
M19	0.105	0.328
M20	0.495	0.630



Data from Irish schizophrenia case-control sample				
	Cases	**Controls**	**OR (95%CI)**	***p* value**
HapA	4.2%	3.0%	1.43 (0.55 – 3.74)	0.427
HapB	4.5%	3.1%	1.55 (0.61 – 4.00)	0.317
HapC	0.7%	1.0%	0.70 (0.08 – 5.19)	0.698
core	9.6%	8.4%	1.18 (0.67 – 2.09)	0.534
HapD	7.3%	3.5%	2.26 (1.04 - 4.98)	0.024
$HapB_{IRE}$	19.4%	12.3%	1.71 (1.09 – 2.68)	0.013

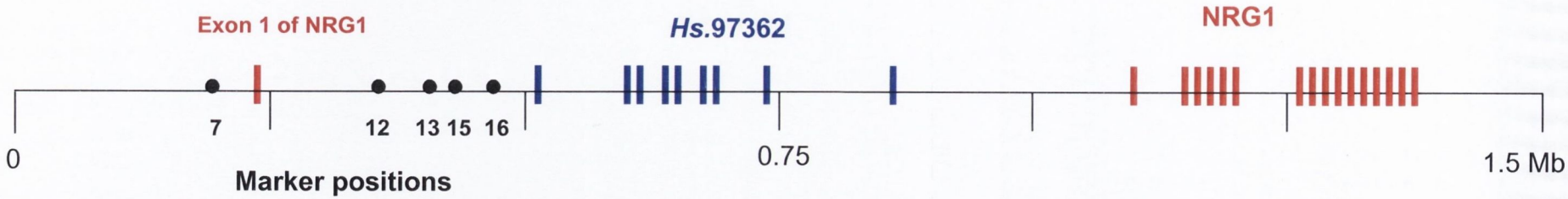
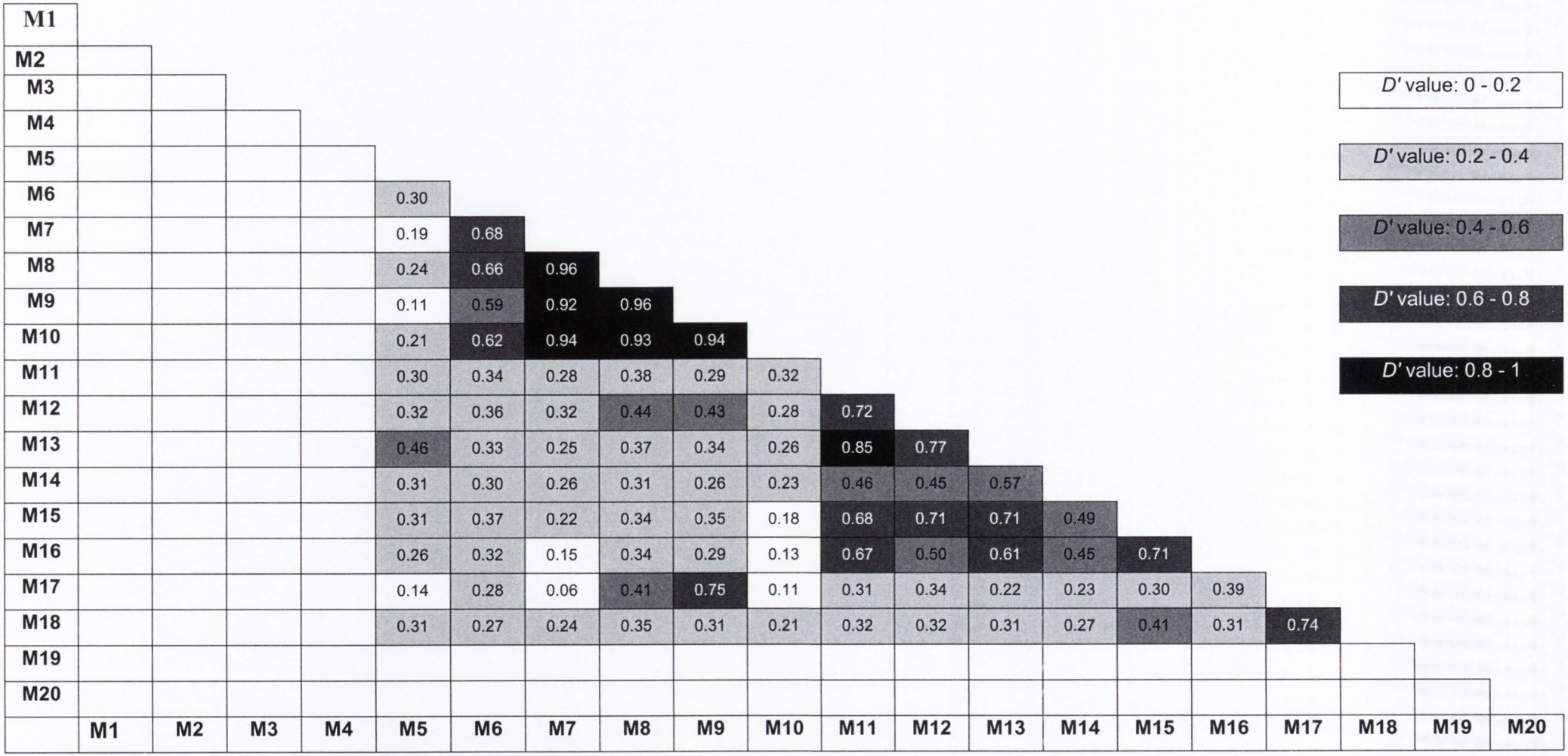


Figure 5.1: The 3 'at-risk' haplotypes (HapA, B and C) and core 'at-risk' haplotype identified by Stefansson et al. (2002).
The physical positioning of haplotypes are indicated by black arrows relative to *NRG1* (exons displayed in red) and the EST cluster *Hs.*97362 (exons displayed in blue). The frequencies of these haplotypes in the case and control samples from the Irish population are detailed. In addition, we present the haplotype frequencies of $HapB_{IRE}$ and HapD that were found to be associated in the Irish sample.

Figure 5.2: Inter-marker LD at *NRG1* measured using *D'* in the Irish schizophrenia case sample



	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12	m13	m14	m15	m16	m17	m18	m19	m20
m1																				
m2																				
m3																				
m4																				
m5																				
m6					0.30															
m7					0.19	0.68														
m8					0.24	0.66	0.96													
m9					0.11	0.59	0.92	0.96												
m10					0.21	0.62	0.94	0.93	0.94											
m11					0.30	0.34	0.28	0.38	1.00	0.32										
m12					0.32	0.36	0.32	0.44	0.43	0.28	0.72									
m13					0.46	0.33	0.25	0.37	0.34	0.26	0.85	0.77								
m14					0.31	0.30	0.26	0.31	0.26	0.23	0.46	0.45	0.57							
m15					0.31	0.37	0.22	0.34	0.35	0.18	0.68	0.71	0.71	0.49						
m16					0.26	0.32	0.15	0.34	0.29	0.13	0.67	0.50	0.61	0.45	0.71					
m17					0.14	0.28	0.06	0.41	0.75	0.11	0.31	0.34	0.22	0.23	0.30	0.39				
m18					0.31	0.27	0.24	0.35	0.31	0.21	0.32	0.32	0.31	0.27	0.41	0.31	0.74			
m19																				
m20																				

Figure 5.3: Inter-marker LD at NRG1 using D' values in the Icelandic case sample

Chapter 6

Investigation of the Apolipoprotein-L (*APOL*) gene family as candidate genes for schizophrenia

6.1 Introduction

This chapter details a SZ genetic association analysis of the Apolipoprotein-L gene family (OMIM603743). In the first instance the structure, position and known function of these genes are reviewed. Second, the functional and positional evidence suggesting that the *APOL* genes may be candidate genes in SZ susceptibility is assessed. Third, the results of the genetic association studies performed at these genes in the Phase 1 schizophrenia sample (detailed in Chapter 4) are described. Finally, the haplotype structure at this locus was explored by using public database information to establish whether haplotype structure can be inferred from comparisons of populations of European and non-European origin. Establishing haplotype structure would allow assessment of how informative the markers selected were in investigating this locus.

6.1.1 Identification of the Apolipoprotein-L (*APOL*) gene family

This gene family encodes a group of abundant plasma proteins, which are the major apoproteins of High Density Lipoprotein (HDL). Investigating peptide sequences of a novel high-density lipoprotein, Duchateau and colleagues (1997), cloned cDNAs encoding *APOL*. By genomic sequence analysis the authors mapped *APOL1* to chromosome 22q12.1-q13.1. By sequencing a number of *APOL1* clones the authors identified a minor splice variant that includes exon 2. This variant predicts a protein that contains a 43-residue signal peptide, the more common transcript (without exon 2), predicts a protein with a 27-signal residue peptide. Page et al, (2001) identified *APOL1* as being in a cluster of six genes that share significant identity and span 619kb of chromosome 22q. Semiquantitative RT-PCR revealed expression of *APOL1* in a wide variety of tissues, including the brain, where it is more strongly expressed in adult than in foetal samples. The genomic organization of the *APOL* gene family is illustrated in Figure 6.1. Noting homology within intronic sequences, Monajemi et al,

(2002) concluded that the *APOL1, APOL2, APOL3* and *APOL4* gene cluster is a result of tandem gene duplication, whereas *APOL5* and *APOL6* are more distantly related. The *APOL* gene family are thought to have emerged recently in evolutionary history as *APOL1* is only found in humans and African green monkeys (Monajemi et al, 2002).

6.1.2 Gene ontology

Experimental data indicates that the *APOL* gene family are involved in the formation of most cholesteryl esters in plasma and in lipid transport. Inference from electronic annotation suggests a role in lipoprotein metabolism and lipid binding. *APOL2* and *APOL3* are found in the cytoplasm where they may affect movement of lipids or allow the binding of lipids to organelles. Expression of *APOL3* is up regulated by tumour necrosis factor-alpha (TNFα) in endothelial cells (Horrevoets et al, 1999). TNFα is a pro-inflammatory cytokine that is implicated in neuronally-mediated response to disease or injury. A large gene expression study (150,000 cDNAs) also identified *APOL3* as one of 83 genes that activate the NK-κB pathway (Matsuda A et al, 2003). This pathway has a general role in modulation of immune and inflammatory responses (Bonizzi & Karin 2004). Together these studies suggest that the *APOL* family have two important functions:

i) A role in the development and maintenance of cell membrane structure and function (Page et al, 2001; Sutcliffe 2003). This in turn may have a profound impact on modulation of gene expression and signal transduction during neurodevelopment and in adult life.

ii) As a component of the immune and inflammatory response.

6.1.3 Positional evidence for involvement of the *APOL* locus in SZ susceptibility

APOL1-6 map to chromosome 22q12.3 a region, which has been implicated in SZ susceptibility by a number of independent studies. In 1994, two studies of independent US multiplex family samples provided suggestive evidence for linkage (LOD <3) to markers on chromosome 22q12-q13 (Pulver et al, 1994; Coon et al, 1994). Subsequently, two larger US studies identified suggestive linkage evidence to

this region in independent samples of 383 sibling pairs and 54 multiplex pedigrees (Blouin et al, 1998; deLisi et al, 2002). In addition, a family-based association study by Takahashi et al, (2003) reported evidence of association between the microsatellite marker D22S683, located at 22q12.3, and SZ in two small samples of European American and Chinese origin respectively. This microsatellite marker is situated 27.4 Kb from the 5' end of the Apolipoprotein-3 gene (*APOL3*) and 33.85Kb from the 3' end of the Apolipoprotein-5 gene (*APOL5*). However, as has been the case with other putative schizophrenia susceptibility loci, many other studies have failed to identify evidence for linkage or association in this region. This may be because most studies that have been performed in the field to date have lacked statistical power to detect the small to moderate gene effects likely to be contributing to schizophrenia susceptibility (other potential reasons have been discussed in Chapter 2). In the absence of genetic heterogeneity, meta-analysis is one method of overcoming this difficulty.

Three combined analyses support this region as contributing to schizophrenia susceptibility. Analysis of genotypic data from the marker D22S278 in multiply affected schizophrenia families derived from 11 independent research groups showed 252 alleles shared compared to 188 not shared where parental genotype data was known (λ^2=9.31, 1df, p=0.001) (Gill et al, 1996). More recently, two meta-analyses using different methods, have found significant evidence for a susceptibility locus at chromosome 22q (Badner & Gershon, 2002; Lewis et al, 2003). The first meta-analysis, by Badner & Gershon (2002), was limited to linkage data from published genome scans and found strongest evidence for schizophrenia susceptibility loci on 8p, 13q and 22q. The second, larger study included data from 20 completed schizophrenia genome scans. This provided statistically significant evidence that at least 12 chromosomal regions, including 22pter-q12.3, are likely to contain schizophrenia susceptibility genes (Lewis et al, 2003). These combined analyses cannot be viewed as independent studies as the datasets analysed overlap substantially, however they are consistent in supporting involvement of a chromosome 22q locus in SZ susceptibility.

The *APOL* gene family lie approximately 15 Mb from the chromosomal region implicated in velo-cardio-facial syndrome (VCFS) (OMIM192430). In almost 85% of cases, VCFS, also known as DiGeorge syndrome, is caused by microdeletions in

chromosome 22q11. This is the most frequent known interstitial deletion identified in humans and arises in about 1/4,000 live births. The resultant phenotype is variable and complex with congenital abnormalities affecting multiple tissues, many of which are derived from neural crest cells. Common clinical features include characteristic dysmorphology, cleft palate, learning disability and congenital heart defects. Interestingly, high rates of psychiatric disorder have been reported in VCFS patients (reviewed in Murphy, 2002). In particular, individuals with chromosomal deletions of chromosome 22q11 have a 30-fold increased risk of SZ and related psychotic disorders (Murphy, 2002). The contribution of VCFS to schizophrenia in the general population is likely to be small, as the condition is rare and cytogenetic studies of schizophrenia populations seldom identify VCFS cases. Few cases of schizophrenia may be explained directly by VCFS, but genetic variation in this region may contribute to schizophrenia susceptibility.

6.1.4 Functional evidence of involvement of the *APOL* genes in schizophrenia

The known functions of the *APOL* genes suggest several mechanisms by which these genes could be involved in SZ pathobiology. Additionally, the genes are expressed in brain regions that have been implicated in SZ. For example, *APOL1* and *APOL2* show a strong neuronal distribution of expression in prefrontal cortex. This brain region has been implicated in SZ pathobiology by neuropsychological, neuroimaging and neuropathological studies (Harrison & Weinberger, 2004). SZ is increasingly considered a neurodevelopmental disorder where expression of the disorder may relate to abnormalities in the formation and maintenance of cell membranes integral to normal synapse function (Tkachev et al, 2003). The *APOL* gene family, through their role in HDL transport are attractive candidate genes for modulating the development and maintenance of cell membrane structure and function (Page et al, 2001; Sutcliffe 2003). Cell membrane integrity is critical in allowing normal signal transduction and normal neurotransmitter function. The *APOL* gene family could contribute to the development, maturation and maintenance of the neurotransmitter systems thought to be aberrant in schizophrenia.

APOL3 is up-regulated by the pro-inflammatory cytokine, tumour necrosis factor-alpha (TNFα) and both genes have a role in activates the NK-κB pathway. The *NK-κB* gene family are known to regulate many of the genes involved in immune and inflammatory responses. Prenatal infection and immune system dysfunction have both been mooted as risk factors in schizophrenia pathogenesis (Cannon et al, 2003; Wright et al, 1999).

The *APOL* gene cluster has arisen recently in evolutionary terms, that is, it is shared only with other primates, which is compatible with the theory of SZ as a disorder of evolution (Crow 2003, Monajemi et al, 2002). Intriguingly, gene expression studies in two independent SZ samples data indicate that *APOL1, APOL2* and *APOL4* are up regulated in SZ. Initially, 10 SZ and 10 control post mortem brains from New Zealand and Japan were screened with a custom made cDNA array of more than 300 genes. The most significant finding was a 2.6-fold up-regulation of *APOL1*, which was confirmed by real-time PCR and replicated in an independent sample from the Stanley Foundation (which included 15 SZ samples). Further investigation of the *APOL* gene family using real-time PCR demonstrated >2.4-fold up-regulation of *APOL2* in both samples and also up-regulation of *APOL4* in the Stanley Foundation sample only (Mimmack et al, 2002). An analysis of covariance showed no relationship between gene expression and lifetime drug exposure, number of years of antipsychotic and type of antipsychotic (typical v atypical) at time of death. This lead the authors to conclude that these changes were more likely to be secondary to the disease process than an artefact of antipsychotic therapy. Although, as discussed in Chapter 1, there are many other potential confounding factors which could also explain this finding.

6.1.5 Strategy for candidate gene studies of *APOL1-6*

The primary study objective was to perform LD-mapping across the *APOL* gene family in the Phase 1 schizophrenia sample of 219 cases and 231 controls described in Chapter 4. When commencing these studies the extent and variability of LD across the genome in different populations was a cause of much debate, for example, estimates of average LD length at different chromosomal loci in different populations varied from 5-100kb. Based on a review of the literature, discussed in Chapter 2, the author took a conservative view to ensure reasonable marker coverage. This involved

a planned investigation of a marker set of SNPs at ~ 5kb inter-marker intervals across the chromosome 22q12.3 locus containing these genes. As this approach necessitated a significant amount of genotyping the author applied for, and was rewarded, a Health Research Board Project Grant [HRB1020 (2000) with Prof Michael Gill] to complete these studies. This grant funded a postdoctoral student (Kevin McGhee) who conducted the studies on the *APOL4-6* genes, applying the same methodology as applied by the author to investigate *APOL1-3*.

As the APOL gene family span 620 kb of chromosome 22, alternative genotyping methods for screening this region were also investigated. To increase genotyping efficiency, and reduce cost, a DNA pooling strategy incorporating a novel variation developed by our group was performed. The method is detailed in Chapter 3. In summary, a three-stage strategy was employed:

i) Initial screening of identified SNPs in case and control pooled DNA samples using simulated heterozygosity values for each SNP. A low threshold for significance ($p \leq 0.10$) was selected to minimise Type II error.

ii) Investigation of associations that met these criteria using true heterozygosity values. Where this confirmed a significant difference at $p \leq 0.10$) markers were investigated in stage iii).

iii) Individual genotyping of SNPs that meet stage ii) criteria. Allele frequency differences at a level of $p \leq 0.05$ were considered as evidence of association for further individual genotyping for LD-mapping at higher marker density.

6.1.6 Investigation of Haplotype structure across *APOL* locus

In selecting markers to be genotyped for this study the average inter-marker distance was 4.3kb and median inter-marker distance 2.9kb. SNPs in the vicinity of the putative promoter region (PPR), 5' UTR and in exonic regions were prioritised. A disadvantage of the DNA pooling method is that in the absence of individual genotypes it is not possible to estimate phase to generate information on haplotype structure. Had such information been available at the start of this study, this haplotype structure information would have been important in selecting markers to maximise the

extent of coverage of DNA variation across the *APOL* locus achieved by this study. The recent availability of publicly available LD-map data has made comparison possible with identified LD-structure across this region in other populations of European ancestry. From this information a *post-hoc* investigation of the marker coverage achieved by this study was performed.

In 2002, Dawson et al, published a draft LD map based on analysis of 1,500 markers across the complete sequence of chromosome 22. This was constructed using genotype data from the Centre d'Etude du Polymorphisme Humain (CEPH) reference families (a US population of North and West European ancestry consisting of 30 trios of two parents and an adult child). The authors identified a highly variable pattern of LD along the chromosome, where regions of almost complete LD extending for 800kb were interspersed with regions of little or no LD. This pattern proved to be remarkably similar in a panel of unrelated UK Caucasians, suggesting that, at least to some extent, common LD patterns may apply in different populations from the same ethnic group.

Making the assumption that LD structure at the *APOL* locus in this study sample could be inferred from the CEPH dataset, the LD structure at this locus was investigated using the dataset described in Dawson et al, (2002) and subsequent marker data available through the HapMap project (http://www.hapmap.org/). With the development of the HapMap the marker density of the LD map across Chromosome 22 has increased ten-fold (freeze, 22-11-04). In addition to the CEPH families, Nigerian, Japanese and Chinese populations are also being sampled.

The purpose of this comparative haplotype analysis was to:

i) Infer possible haplotype structure in this region

ii) Establish the consistency of this structure across populations

iii) Identify the degree of overlap with the marker set in this study

iv) Establish whether haplotype structure in the CEPH dataset remained stable despite at the increased marker density in the 2004 dataset

v) Identify haplotype tagging SNPs from the CEPH dataset and estimate, if possible, the coverage achieved by the markers genotyped in the study.

6.2 Results

6.2.1 Bioinformatics/ primer design

The *APOL1-6* gene family map to chromosome 22, at 22q13.1: *APOL1* covers 14.46kb from 34920931 to 34935390 (NCBI build 34, August 2003) on the direct strand (http://www.ncbi.nih.gov/IEB/Research/Acembly/). At least three transcript variants encoding two different isoforms exist for this gene (http://www.ncbi.nlm.nih.gov/LocusLink). *APOL2* covers 20.43 kb from 34907813 to 34887389, on the reverse strand. Two transcript variants encoding the same protein have been identified. *APOL3* covers 26.08kb from 34834038 to 34807961 also on the reverse strand. Six transcript variants encoding three different isoforms have been found for this gene. The structure of the gene family is represented schematically in Figure 6.1.

A total of 201 SNPs spanning *APOL1-3* were identified from the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/) (from NCBI build 34, August 2003). SNPs in the vicinity of the putative promoter region (PPR), 5' UTR and in exonic regions were prioritised for investigation. With these criteria, a total of 143 SNPs were selected for examination of *APOL1-6*. From an initial pool of 73 SNPs spanning the *APOL1-3* loci, a marker set of 62 SNPs were selected *a priori*, as likely to provide the marker density required for this study. The included markers are detailed in Tables 3.2-3.4.

6.2.2 PCR conditions

Full optimisation details for the *APOL1-APOL3* markers are included in Tables 3.2-3.4. One fragment [*APOL2-11, 12,13,14*] failed to provide a reliable PCR product and was dropped from further analysis.

6.2.3 Confirming SNP heterozygosity in the study population

The 58 SNPs were screened for heterozygosity in a pooled DNA sample of 100 individuals using SNaPshot™ (Applied Biosystems). Of 62 initial potential markers, 30 (46.9%) were confirmed as polymorphic in the control population sample. This provided a mean inter-marker distance of 3.7kb and a median inter-marker distance of 4.5kb across the *APOL1-3* genes. From the original set of 143 markers across *APOL1-6*, 51 markers (36%) were identified as polymorphic in the Irish control population.

6.2.4 DNA pooling experiments for *APOL1-3* (Stage 1)

The identified heterozygous SNPs (n=30) were then investigated in case and control pooled DNA samples. Each SNP was assayed 5 times for each pool. Simulated *k* values were applied to each replicate and allele frequency estimates were an average of the 5 assay results. At the point of maximum potential difference in allele frequencies between the pools, allele frequencies were computed to allele counts and SNPs were tested for association using a 2 x 2 contingency table to calculate χ^2 statistics and p values. A conservative cut off value of $p< 0.1$ was set for transfer of any SNP to Stage 2 to allow for experimental error. The results of all Stage 1 experiments are summarized in Tables 6.1 and 6.2. Of 51 SNPs tested across the *APOL1-6* genes, only 2 markers *APOL2-6* (dbSNP:2017689) and *APOL4-46* (dbSNP:132734) met the screening criteria for inclusion in Stage 2. This was less than the number of markers (n=5) predicted to reach Stage 2 by chance.

6.2.5 Screening DNA pooling experiments for *APOL1-3* (Stage 2)

Each SNP that reached Stage 2 was genotyped in a panel of individuals to identify a heterozygous individual. This provided the true allele peak height ratio (*k*) and was used to determine the estimated allele frequencies in the case and control pools. These frequencies were computed to allele counts and SNPs were tested for association using a 2 x 2 contingency table to calculate χ^2 statistics and p values. Marker *APOL4-46* (χ^2=2.96, p=0.085) but not marker *APOL2-6* (χ^2=2.57, p=0.11) met screening criteria for individual genotyping.

6.2.6 Individual genotyping (Stage 3)

Individual genotyping using the same method in our total study sample of 219 cases and 231 controls (performed by Kevin McGhee) failed to detect statistically significant evidence of association between *APOL4-46* and SZ (χ^2=2.65, p=0.128) (Table 6.3).

6.2.7 Haplotype structure across the *APOL* locus

From publicly available linkage and marker files from the Dawson et al, (2002) and HapMap (22-11-04) datasets, the author investigated the sequence containing the *APOL* gene family (34353913 to 34988986 from NCBI build 34, November 2004).

6.2.7.1 Overlap with the marker set from the APOL association studies

The Dawson et al, (2002) dataset identified 15 SNPs that had been genotyped across the region, but there was no overlap with the SNPs selected for these association analyses. The HapMap 2004 dataset identified 291 SNPs across this region, 16 (5.49%) of these had been genotyped in the association analyses.

6.2.7.2 Comparison of LD-structure between ethnic populations

In addition to the CEPH families, Nigerian, Japanese and Chinese populations are also being sampled as part of the HapMap Project. To date, more than 15,000 markers have been genotyped across chromosome 22q in the Caucasian population (from the November 2004 freeze of HapMap). However, the total number of completed genotypes for chromosome 22 in the other population pools was much lower (ranging from 6 markers across the whole chromosome in Japanese and Chinese populations to 1,242 markers in the Nigerian population). This marker density was insufficient to perform comparative analyses of LD-structure between populations.

6.2.7.3 Stability of haplotype structure at increased marker density

Haplotype structure at the *APOL* locus was investigated using Haploview (detailed in Chapter 3) with the three test algorithms implemented in this program for both the 2002 and 2004 datasets. The graphical outputs of identified block-structure and tagging SNPs for the 2002 data are represented in Figure 6.2. From the 2002 dataset, each algorithm identified 3 haplotype blocks. There was significant overlap between the defined blocks. The solid-spine of LD method proved the most conservative identifying 2, 6 and 6-marker blocks; the 4-gamete rule identified 3, 4 and 4 marker blocks and the confidence interval method identified 3, 2 and 2 marker blocks.

The haplotype structure and tagging SNPs generated from the 2004 dataset using the confidence-interval method are represented in Figure 6.3. The number of haplotype

blocks defined by the three algorithms varied from 28 (for the solid-spine of LD method) to 37 (using the 4-gamete rule) and the amount of total sequence identified as being represented by blocks varied from 76.3-93.8% depending on the algorithm selected. This data is summarized in Table 6.4.

Similarly, the number of tagging SNPs varied according to method of block definition (n=87-122); there was substantial if incomplete correspondence between the tagging SNPs identified across the three algorithms. Across the three algorithms a total of 150 tagging SNPs were identified. The three algorithms identified 59 common SNPs (representing 39.3% of the total number of tagging SNPs); 53 SNPs were selected by two different methods (35.3%) and 38 SNPs were unique to a single algorithm (25.3%).

6.2.7.4 Haplotype block tagging by markers in this study

Given the large number of SNPs required to 'tag' the haplotype blocks identified in the 2004 dataset (n=87-122) it was unsurprising that none of the SNPs genotyped in the *APOL* association studies completely tagged any of the identified haplotype blocks.

6.3 Discussion

The *APOL* genes are both positional and functional candidate genes for SZ. This gene family maps to chromosome 22q12.3, a region implicated by SZ linkage studies as likely to contain one or more SZ susceptibility genes. A recent gene expression study demonstrated up-regulation of *APOL1*, *APOL2* and *APOL4* in post-mortem brain samples from SZ patients compared to controls in two independent samples.

In this chapter, an extensive genetic association study of the six known *APOL* genes and SZ was performed using a modified method of DNA pooling. To test for genetic association, 143 SNPs (identified from dbSNP) were investigated. Of these, 51 (40.6%) were polymorphic in this sample and were genotyped using a three-stage DNA pooling approach. The author found no evidence to support the hypothesis that genetic variation at these genes contributes to SZ susceptibility in this sample.

In completing this study a three-stage DNA pooling strategy was developed to increase efficiency for studies that employ public database SNPs in large-scale association studies (McGhee et al, 2005). This strategy represents a modification of the DNA pooling method proposed by Norton et al, (2002). That method required the identification of an individual heterozygous for the SNP under analysis in order to accurately test for association at that SNP. By simulating the information likely to be provided by the heterozygous individual, the strategy circumvents this requirement, resulting in a more economical but no less accurate method of association analysis. In this study 51 assayed SNPs could be categorized as not showing evidence for association without the need to identify a heterozygous sample in each case.

There are several possible interpretations of these findings. First, it may well be that up regulation of *APOL* genes is of relevance to the pathobiology of SZ, but that this is influenced by genetic variation at other loci as has been described for other disorders (Kleinjan & van Heyningen, 2005). Second, *APOL* expression abnormalities in SZ in post-mortem samples may not be directly related to the genetic aetiology of SZ. For example, expression differences may represent secondary effects caused by differences in factors such as drug treatment, substance abuse or smoking behaviour between case and control groups. Third, genetic variation in the *APOL* gene family is of aetiological importance in SZ but because the effect size is unknown, it is possible that our study lacked statistical power to detect association.

After completion of the laboratory phase of this work, a provisional negative report was published in abstract form from a Japanese case-control schizophrenia study at the *APOL* locus (Ohtsuki et al, 2003). No other studies for association between the *APOL* gene family and SZ have been reported to date. During the course of this study association was reported between markers at *APOL1/APOL2* and Bipolar Affective Disorder (BPAD) in a US sample (McInnes et al, 2002). This suggested that the locus may contribute to susceptibility to both disorders and indeed linkage data has suggested that this chromosomal region confers susceptibility to both disorders (Park et al, 2004). Following this initial report of association between markers at *APOL1/APOL2* and Bipolar Affective Disorder (BPAD) in the Johns Hopkins BPAD sample, the region has been investigated at a similar marker density (1 SNP/6.9kb) in

a larger US sample (Potash et al, 2004). The follow-up study, which applied case-control and family-based association methods to the Johns Hopkins BP sample, NIHM Genetics Initiative BPAD waves 1 and 2, and the Clinical Neurogenetics BP samples, failed to support the provisional finding (Potash et al, 2004).

Power calculations indicate that for the markers tested, our sample has 80% power to detect an effect of odds ratio (OR) >2 with 95% confidence. Estimates for the individual effect sizes of genes implicated in schizophrenia susceptibility may in fact be more modest. Power calculations of this type also assume that at the marker density employed for this study (1 SNP/3.7kb) a majority of the genetic variation present at the locus was tested for association. Investigation of haplotype structure in this region in another population of European ancestry indicates that this is unlikely to have been the case.

The initial published haplotype information from the CEPH families provided genotype information on 15 markers across the *APOL* locus (Dawson et al, 2002). More recent data from the International HapMap project identified 291 SNPs across this locus in the same family sample. Insufficient data was available to allow comparisons of haplotype structure between the populations investigated in the HapMap project: the availability of such data will be important in establishing how useful HapMap will be in informing mapping studies in different study populations.

A range of different definitions of haplotype blocks has been applied to genotype data (e.g. Daly et al, 2001; Gabriel et al, 2002; Zhang et al, 2002; Wang et al, 2003; Barrett et al, 2005). These can be broadly grouped as methods that define blocks as regions of limited haplotype diversity (Daly et al, 2001; Zhang et al, 2002; Wang et al, 2003) and those that use measures of LD and define blocks where all pair-wise coefficients exceed a pre-defined threshold (Gabriel et al, 2002; Barrett et al, 2005). The method described in Gabriel et al, (2002) places confidence intervals on the measure of pair-wise LD, which allows some account for inaccuracy in the estimation of LD. By either approach, the threshold of LD or haplotype number used to define haplotype blocks is arbitrary and the relationship between defined blocks and patterns of recombination across the genome is uncertain. In this study, methods representative of

each approach (and implemented in Haploview) were applied to the Dawson et al, (2002) and more recent HapMap datasets.

The twenty-fold increase in SNP density across the *APOL* locus between the 2002 and 2004 datasets had a marked effect on the haplotype block structure defined by each approach. The three algorithms identified three reasonably consistent haplotype blocks at the lower marker density. At greater density the algorithms identified between 28 and 37 blocks, with less consistency between the blocks defined, although by each method a significant proportion of the sequence at this locus (76.3-93.8%) was defined as being included in haplotype blocks. There was partial agreement between methods in selecting tagging SNPs at this locus, but 25% of all tagging SNPs were identified by one algorithm but not by the others.

Exploration of the HapMap data suggests that this association study tested only a limited amount of the likely genetic variation present at the *APOL* locus. The inconsistency in defining block structure between methods and instability of the apparent 'blocks' at greater marker density suggests that identifying SNPs to 'tag' haplotypes at this locus may provide only a modest increase in genotyping efficiency. However, it could also be argued that the majority of the markers identified were included in haplotype blocks even by the most conservative block definition. A disadvantage of the study was that LD was measured using D', rather than r^2, which for reasons discussed in Chapter 2 may have advantages for LD-mapping studies. The issues and application of HapMap data to future LD-mapping studies is considered further in Chapter 8.

Taken together, these studies fail to support the hypothesis that the *APOL* gene family contribute susceptibility to SZ. However, it is likely that studies at much greater marker density, with more available information on haplotype structure and larger sample sizes will be required to definitively exclude smaller genetic effects at this locus.

Table 6.1 : Results of DNA pooling experiments for *APOL 4, 2, 1*

Gene	dbSNP ID	ID	Chromsomal Position[a]	SNP	Allele Ratio in Case Pools[c,d]	Frequency in Case Pools (simulated k)	Allele Ratio in Control(Pools)[c,d]	Frequency in Control Pools (simulated k)	Simulated P-value
	dbSNP:2227169	APOL4-8	34859765	C/T	0.713	*0.554*	0.689	0.526	0.267
	dbSNP:2097466	APOL4-9	34859925	G/A	0.592	*0.592*	0.573	0.573	0.543
APOL 4	dbSNP:132704	APOL4-17	34863087	T/C	0.912	*0.674*	0.898	0.638	0.268
	dbSNP:2007468	APOL4-18	34863193	G/A	0.457	*0.457*	0.405	0.405	0.115
	dbSNP:132712	APOL4-27	34867057	G/A	0.133	*0.434*	0.146	0.461	0.409
	dbSNP:132719	APOL4-33	34868829	G/T	0.340	*0.507*	0.314	0.478	0.391
	dbSNP:132720	APOL4-34	34868911	G/A	0.298	*0.563*	0.265	0.522	0.226
	dbSNP:132721	APOL4-35	34868954	G/A	0.805	*0.508*	0.819	0.531	0.480
	dbSNP:132734	**APOL4-46**	**34869690**	**G/A**	**0.720**	***0.462***	**0.792**	**0.559**	**0.004**
	dbSNP:1807673	APOL4-55	34870580	A/G	0.202	*0.502*	0.175	0.459	0.191
	dbSNP:1315	APOL2-1	34887948	A/C	0.788	*0.482*	0.798	0.497	0.628
	dbSNP:1317	APOL2-2	34887976	G/T	0.705	*0.544*	0.692	0.529	0.645
APOL 2	dbSNP:763086	APOL2-4	34890870	A/G	0.794	*0.491*	0.806	0.510	0.548
	dbSNP:132757	APOL2-5	34893522	T/C	0.655	*0.486*	0.659	0.491	0.879
	dbSNP:2017689	**APOL2-6**	**34894533**	**A/G**	**0.764**	***0.447***	**0.802**	**0.503**	**0.099**
	dbSNP:129607	APOL2-7	34894719	T/C	0.661	*0.494*	0.696	0.534	0.236
	dbSNP:132759	APOL2-8	34894961	T/C	0.909	*0.666*	0.922	0.703	0.232
	dbSNP:2005998	APOL2-10	34897317	C/T	0.351	*0.520*	0.391	0.562	0.226
	dbSNP:2010499	APOL2-15	34901279	A/T	0.346	*0.514*	0.391	0.562	0.157
	dbSNP:2157249	APOL2-18	34902852	T/C	0.858	*0.548*	0.830	0.494	**0.101**
	dbSNP:2157251	APOL2-20	34905240	G/T	0.758	*0.440*	0.794	0.491	0.126
	dbSNP:1557535	APOL2-21	34905265	A/C	0.304	*0.570*	0.261	0.517	0.122
	dbSNP:136144	APOL2-23	34907257	A/G	0.779	*0.540*	0.770	0.527	0.698
	dbSNP:713791	APOL1-1	34914183	A/G	0.575	*0.575*	0.562	0.562	0.653
	dbSNP:136147	APOL1-3	34924702	G/T	0.418	*0.589*	0.396	0.568	0.503
	dbSNP:136148	APOL1-4	34924744	C/T	0.907	*0.661*	0.906	0.657	0.898
APOL1	dbSNP:136151	APOL1-8	34926825	A/G	0.715	*0.455*	0.716	0.456	0.998
	dbSNP:713929	APOL1-18	34929409	A/G	0.806	*0.509*	0.823	0.538	0.404
	dbSNP:713753	APOL1-21	34930347	C/T	0.711	*0.450*	0.741	0.488	0.234
	dbSNP:136171	APOL1-29	34932874	A/G	0.750	*0.501*	0.770	0.528	0.397
	dbSNP:136174	APOL1-33	34933349	A/C	0.807	*0.491*	0.828	0.456	0.303

Table 6.2: Results of DNA pooling experiments for *APOL-6,5,3*

Gene	dbSNP ID	ID	Chromsomal Position[a]	SNP	Allele Ratio in Case Pools[c,d]	Frequency in Case Pools (simulated k)	Allele Ratio in Control(Pools)[c,d]	Frequency in Control Pools (simulated k)	Simulated P-value
	dbSNP:926754	Apol6-3	34301481	G/C	0.843	*0.519*	0.832	0.497	0.538
	dbSNP:2103768	Apol6-4	34304784	G/C	0.613	*0.442*	0.600	0.492	0.429
APOL 6	dbSNP:762910	Apol6-5	34304919	T/C	0.585	*0.413*	0.607	0.436	0.464
	dbSNP:1883986	Apol6-6	34305307	T/C	0.810	*0.516*	0.797	0.495	0.541
	dbSNP:1894469	Apol6-7	34309889	A/G	0.686	*0.522*	0.683	0.519	0.920
	dbSNP:2413359	Apol6-9	34310099	C/T	0.682	*0.517*	0.677	0.512	0.876
	dbSNP:2010168	Apol6-12	34322402	A/C	0.907	*0.661*	0.901	0.645	0.590
	dbSNP:2413367	Apol5-2	34373674	A/G	0.549	*0.549*	0.548	0.548	0.992
	dbSNP:1997883	Apol5-5	34375522	T/C	0.586	*0.586*	0.573	0.573	0.690
APOL 5	dbSNP:1970777	Apol5-9	34382149	G/A	0.729	*0.473*	0.730	0.474	0.965
	dbSNP:1540297	Apol5-10	34382278	T/C	0.570	*0.570*	0.588	0.588	0.582
	dbSNP:2413369	Apol5-11	34385614	T/C	0.395	*0.566*	0.388	0.559	0.813
	dbSNP:2009168	Apol5-13	34389667	G/A	0.689	*0.526*	0.698	0.536	0.774
	dbSNP:2009169	Apol5-14	34389705	A/G	0.527	*0.527*	0.527	0.537	0.965
	dbSNP:2073198	Apol5-17	34395126	A/G	0.573	*0.573*	0.617	0.617	0.179
	dbSNP:2413381	APOL3-2	34792045	C/T	0.393	*0.393*	0.406	0.406	0.712
APOL 3	dbSNP:80576	APOL3-11	34811617	A/G	0.749	*0.499*	0.744	0.493	0.899
	dbSNP:1807498	APOL3-17	34812887	A/G	0.382	*0.553*	0.401	0.573	0.566
	dbSNP:2105915	APOL3-23	34824152	C/T	0.554	*0.554*	0.542	0.542	0.730
	dbSNP:132653	APOL3-26	34828636	G/T	0.145	*0.459*	0.139	0.446	0.694

Table 6.3: Results of individual genotyping experiments

dbSNP ID	Marker Name	*k* value	HZ value (true)	ratio *k*	Frequency in Cases (Pools)[a,b]	Frequency in Controls (Pools)[a]	*p* value
rs132734	APOL4-49	3	0.24		0.720	0.792	0.128

[a] Ratio for alleles a and b in pooled sample. Formula =A/(A+B), where A and B represent the peak heights of alleles a and b.

[b] Number represents frequency for first allele listed

Table 6.4: Differences between haplotype blocks from HapMap 2004
(Defined by different methods implemented in Haploview)

	Number of haplotype blocks	% of total sequence defined by blocks	Number of markers in blocks (% of total markers)	Number of tSNPs identified	Total number of markers to define sequence
CI method	29	76.3	78.9	87	143
4-gamete rule	37	86.7	93.6	122	139
Spine of LD	28	93.8	98.1	111	116

Figure 6.1: Structure of the APOL gene family

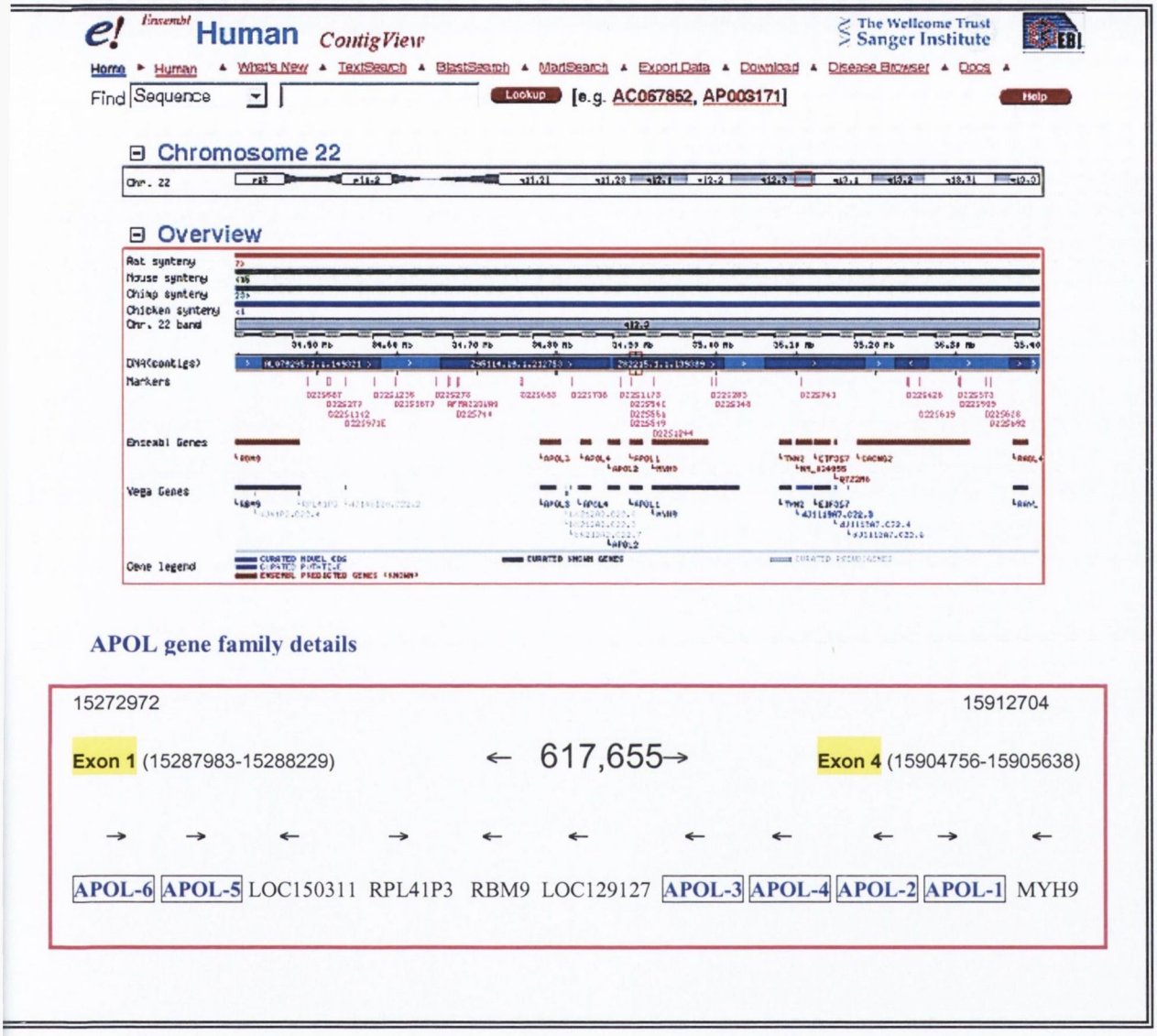
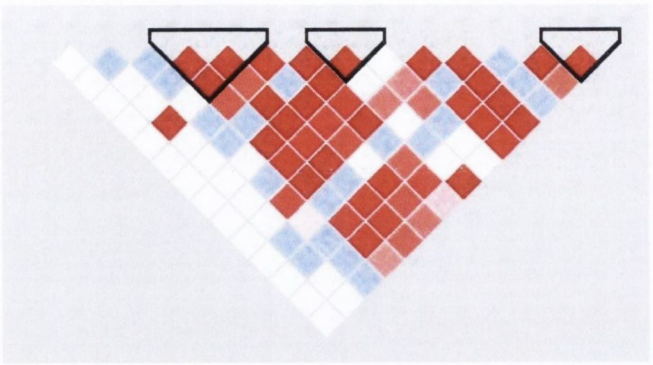
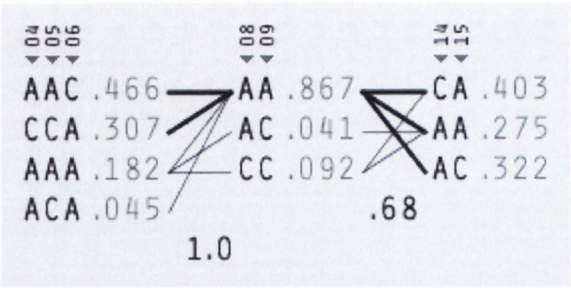


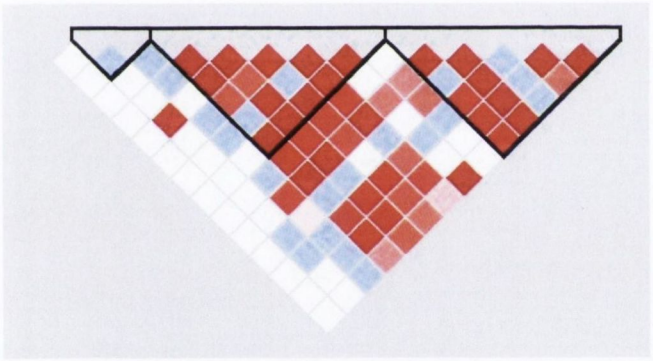
Figure 6.2 : Haplotype blocks/tagging SNPs at the *APOL* locus (from Dawson et al, 2002 dataset)

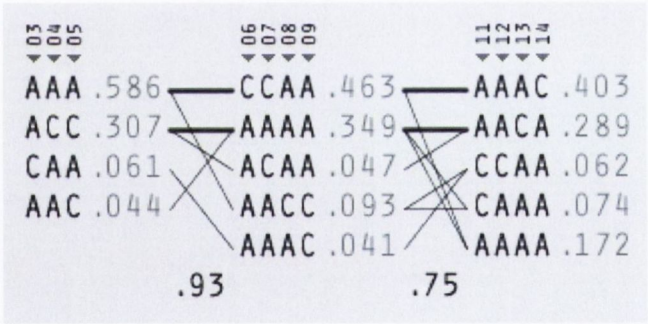
i) Haplotype structure and tagging SNPs defined by the Gabriel et al, 2002 method



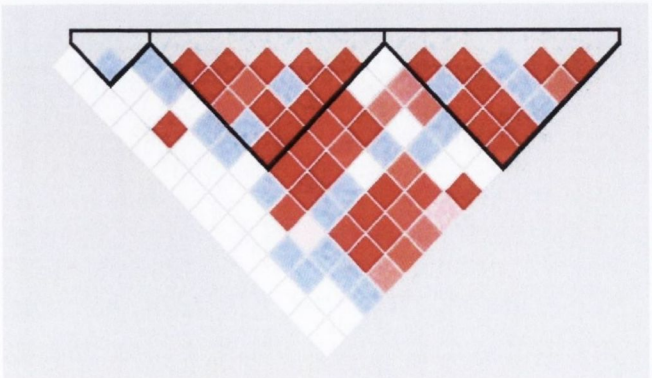


ii) Haplotype structure and tagging SNPs defined by the 4-gamete rule method





iii) Haplotype structure and tagging SNPs defined by the solid spine of LD method



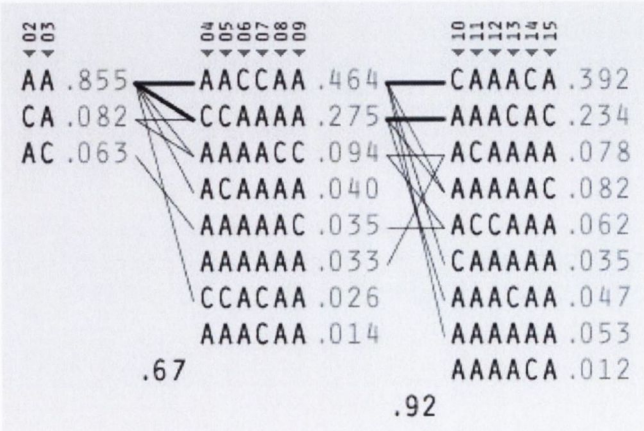
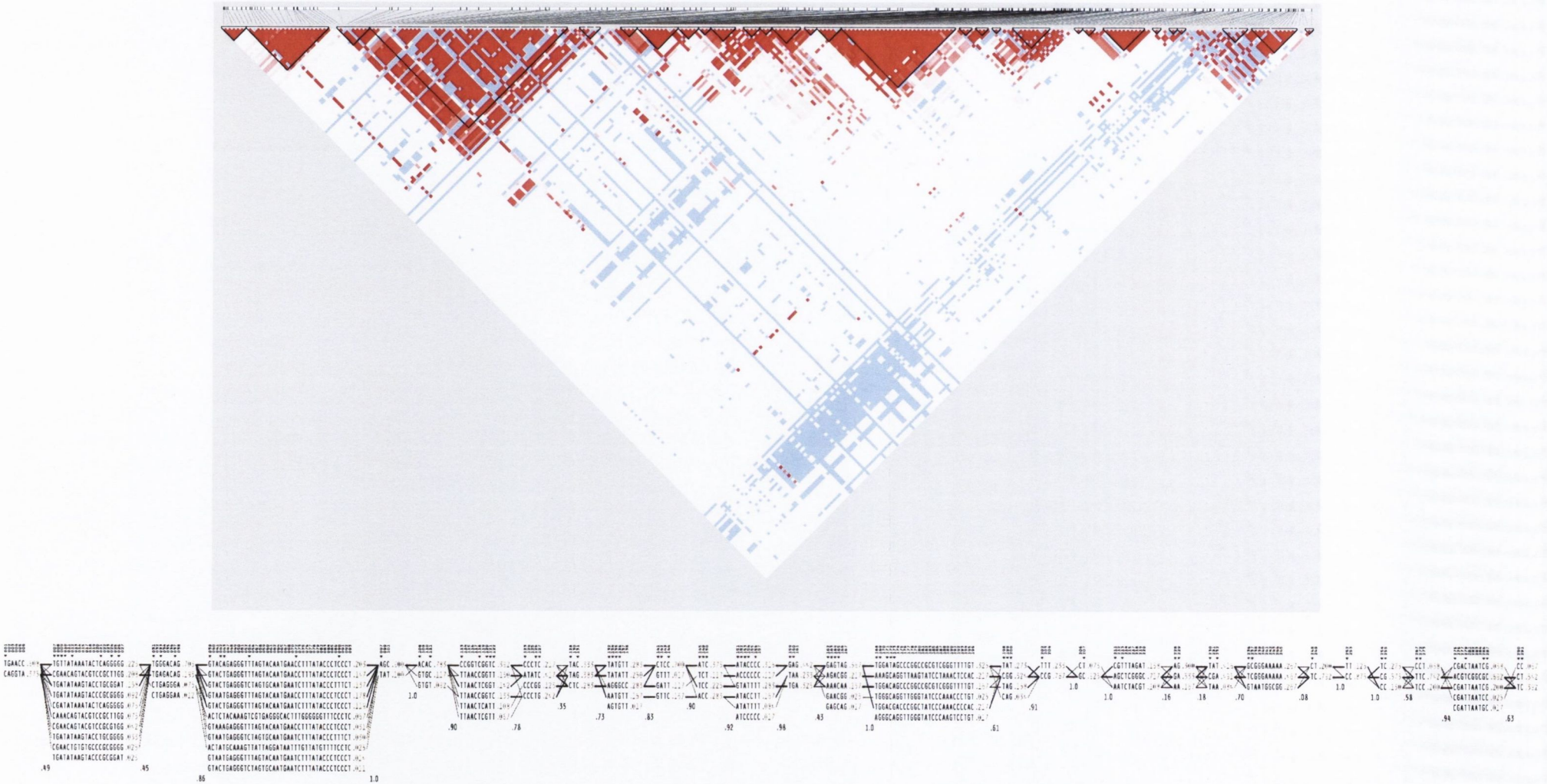


Figure 6.3 : Haplotype structure of the APOL locus from the CEPH families (HapMap, November 2004)



Chapter 7

Investigating other potential schizophrenia candidate genes and testing for gene-gene interactions

7.1 Introduction

Since 2002, there is a growing body of statistical evidence to support the involvement of a number of genes in schizophrenia susceptibility: these include, Dysbindin *(DTNBP1)* (Straub et al, 2002; Schwab et al, 2003; Williams et al, 2004); Neuregulin 1 *(NRG1)* (Stefansson et al, 2002, 2003; Williams et al, 2003; Corvin et al, 2004); *G72* (Chumakov et al, 2002; Schumacher et al, 2004); D-amino-acid-oxidase *(DAAO)* (Chumakov et al, 2002; Schumacher et al, 2004) and Regulator of G-protein signalling-4 (*RGS4*) (Chowdari et al, 2002; Morris et al, 2004; Chen et al, 2004). In most cases understanding of the function of these genes is limited, but driven by their association with schizophrenia, each is being intensively investigated. Understanding the nature of the relationships between genes is likely to be important in establishing how they contribute to a single, or multiple, disease aetiologies. Hypotheses as to how these genes relate to each other are being developed (for an example see Figure 7.1). However, current hypotheses represent broad conceptual frameworks and require formal testing using biostatistical and scientific methods.

7.1.1 Epistatic interactions

In biological terms interaction between proteins is vital to the metabolic and biochemical pathways that maintain normal body homeostasis. In the context of genetics, Bateson coined the term 'epistasis', meaning to 'rest upon', in 1909 to describe a situation where DNA variation at a locus prevented the effect of variation at another locus from being manifest. A later, broader definition, by Fisher, (1918), describes epistatic interaction in statistical terms as existing where the contribution of alleles to the expression of a quantitative phenotype is non-additive.

Identifying non-additive interactions between genes is a potentially powerful method of improving our understanding of complex genetic disorders (Moore & Williams, 2002). In the first instance, the effects of many individual genes may be partially or

fully dependent on interactions with other genes (Templeton, 2000). If this is the case, methods that investigate single genes will have less power than statistical approaches that can examine multiple genes. Methods to investigate epistatic interactions have been applied successfully to other complex disorder phenotypes including breast cancer, hypertension and myocardial infarction (Ritchie et al, 2001; Williams et al, 2004; Coffey et al, 2004). Secondly, by identifying statistical evidence of relationships between genes, it may be possible to design assays to test their biological relationships and establish the systems or pathways that contribute to complex phenotypes (Risch, 1990). This may be of particular pertinence for a condition such as schizophrenia where current understanding of aetiology is limited.

7.1.2 Evidence for gene-gene interaction in schizophrenia

To date, most investigations of gene-gene interaction in schizophrenia have been prompted by existing biological evidence of a relationship between the genes being investigated. Using a case-control linkage disequilibrium mapping strategy, Chumakov et al, (2002) identified the novel gene *G72/G30* (OMIM 607408) as a schizophrenia candidate gene at chromosome 13q32-33 in French Canadian and Russian patient samples. By yeast two-hybrid experiments with the *G72* protein, the authors identified the enzyme D-amino acid oxidase (*DAAO*) as an interacting partner. The *DAAO* gene is expressed in the brain and oxidises D-serine, which is an activator of N-methyl-D-aspartate (NMDA) glutamate receptors. Further *in vitro* experimentation confirmed that *G72* is an activator of *DAAO* (the gene is now also referred to as *D-amino acid oxidase activator*). These experiments suggested a mechanism by which both genes could contribute to schizophrenia susceptibility by modulating NMDA receptor function with downstream effects on glutamatergic transmission.

Subsequently the authors found evidence of association between markers at the *DAAO* gene and schizophrenia in a French Canadian case-control sample and evidence of an epistatic interaction between markers at *G72* and *DAAO* (OR=5.02); that was greater than the single locus effects (marker M-22 at *G72*, OR=1.04 and marker MDAOO6 at *DAAO*, OR=1.89). Association evidence for single locus involvement has been reported for both *G72* and *DAAO* in a number of follow up studies by other groups,

but data from two large studies has failed to support the hypothesis of interaction (Williams et al, 2004; Schumacher et al, 2004).

Investigation of a gene-gene relationship between neuregulin-1 (*NRG1*) and *erbB4* was prompted by evidence that *NRG1* is involved in schizophrenia susceptibility (discussed in chapter 4) and the role of *NRG1* isoforms as ligands for the *erbB* family of tyrosine kinase receptors (Norton et al, 2004). *NRG1* proteins only bind effectively to *erbB3* and *erbB4*; and in putative mouse models for schizophrenia *erbB4* hypomorphs (but not *erbB2* and *erbB3* mice) behave similarly to *NRG1* hypomorphs. For this reason, *erbB4* (OMIM 600543) was identified as a candidate gene with a potential biological interaction with *NRG1* isoforms.

In a large UK schizophrenia case-control sample, Norton et al, (2004) identified significant interaction of *erbB4* IVS12-15C>T genotypes with the *NRG1* risk haplotype identified by Stefansson et al, (2002) ($p=0.019$) in the absence of significant single-gene effects at either locus. Logistic regression analysis showed that individuals heterozygous for the *erbB4* IVS12-15C>T genotype exhibited significantly higher risk of disease as the *NRG1* disease haplotype scores increased ($p = 0.021$, 2-tail, OR= 5.07 (CI 1.28, 20.1)). In collaboration, the authors identified a trend for the same interaction in investigating our Phase 2 sample (290 cases and 634 controls) ($p=0.11$) and stronger evidence for interaction in the combined datasets ($p=0.01$).

7.1.3 Testing for epistatic interactions

Implicit in the last section is the fact that there are relatively few published studies of epistatic interaction in the schizophrenia genetic association literature. There have been no published studies investigating interaction between multiple genetic variants. This is, at least in part, because attempting to detect interactions between multiple variables (e.g. genetic loci) presents analytical difficulties (Freitas, 2001). Traditional parametric statistical methods have been applied with mixed success to investigate multiple-marker loci in genetic association studies of complex disorders (Cruickshanks et al, 1992; Chumakov et al, 2002). Most positive reports have involved interaction between two loci, as statistical models that investigate

interactions between multiple loci generate many contingency-table cells that contain no observations. In turn, this may lead to very large coefficient estimates and standard errors (increasing type I error).

Procedures for fitting regression models are also problematic. Forward selection methods only test variables that have statistically significant independent main effects. However, epistasis may contribute to complex phenotypes in the absence of such single locus effects (as may be the case with *erbB4* in interacting with *NRG1*). Backwards elimination methods may require too many degrees of freedom. For example, a schizophrenia model that aims to include all probable main effects could include dozens of putative susceptibility genes and multiple individual markers or haplotypes at many of these genes. Stepwise procedures, although more flexible, present the same difficulty. Any of these approaches may increase type II errors and decrease study power. As a result, a number of alternative methods for investigating relationships between genes have been proposed, these include variations on standard statistical methods, exploratory data analysis methods, pattern recognition and neural network strategies.

Variations on standard statistical methods typically involve a combination of regression modelling and resampling methods (reviewed in Hoh & Ott, 2003). In the original approach described by Hoh et al, (2000), individual markers are tested for association using the chi-square statistic; the combined effect of all markers is than the sum of all the association statistics. The markers with the smallest test-statistics are then sequentially removed until a pre-selected significance level for the sum of the statistics is achieved (e.g. $p<0.05$). Further resampling is than used to test whether the same markers are consistently selected across a set of replicates. Applying this approach to a case-control dataset of coronary artery restenosis identified statistically significant interactions among SNPs from 9 different genes (Zee et al, 2002).

Exploratory data analysis methods, which are nonparametric and model-free, such as the combinational partitioning method (CPM) were developed to investigate epistatic effects at quantitative trait loci (Nelson et al, 2001). The CPM method pools genotypes from multiple loci into a smaller number of classes (thus reducing the number of tests) to identify the combination of genotypes most strongly associated

with variation in the quantitative trait. An advantage of this method is that it is not dependent on the presence of single gene effects. Applying this method, Nelson et al, (2001) identified non-additive epistatic interactions, in the absence of independent main effects, between loci at putative cardiovascular disease susceptibility genes and variability in plasma triglycerides. A disadvantage of this method is the computational burden posed by investigating interactions between more than two loci. Several variations of this approach, which are considerably more computer efficient, have since been proposed (Ritchie et al, 2001; Culverhouse et al, 2004).

The multifactorial dimensional reduction (MDR) method was developed specifically to improve the power to detect gene-gene interactions in relatively small samples (Ritchie et al, 2001). As with the CPM method on which it was based, the MDR method can detect significant interactions in the absence of main effects. With this method, genotype data from multiple loci is pooled into high-risk or low-risk groups. This reduces the genotype predictors from multiple dimensions to one dimension only. This variable is then tested for its ability to classify and predict disease status through cross-validation and permutation testing. The procedure involved is outlined in Figure 7.2 and in Section 7.3.3.2. The software for implementing MDR is available at http://www.epistasis.org/mdr.html.

Pattern recognition methods using machine learning have only recently been applied to human genetics. These methods are computer-intensive and require further validation and application to real data: the potential application of this methodology to complex disorder genetics has been reviewed elsewhere (Toivonen et al, 2000; Czika et al, 2001).

7.2 Investigation of epistatic interaction in this dataset

This study had two principle aims: first, to test for evidence of marker/haplotype association at schizophrenia candidate genes and second to test for evidence of epistatic interaction between loci. The latter involved testing the hypothesis that genetic variants at the *G72* and *DAAO* loci interact in contributing to schizophrenia susceptibility using standard parametric analysis. In addition, as an exploratory analysis, the MDR method was applied to investigate for evidence of epistatic

interaction between markers at six putative schizophrenia susceptibility genes. In selecting the six genes for investigation, genes were selected for inclusion based on available *a priori* evidence for interaction, albeit between pairs of genes rather than for higher order interactions.

7.2.1 Candidate gene selection

Based on existing schizophrenia studies of gene-gene interaction, variants at the *NRG1, erbB4, G72* and *DAAO* loci were selected. Two additional genes were also included in the model because both have demonstrated evidence for association with schizophrenia in the Irish population. The genes, *DTNBP1* and *MRDS1*, map to the chromosome 6p susceptibility locus identified in the Irish Study of High Density Schizophrenia Family sample (Straub et al, 2002). DNA variation at these genes may interact in contributing to the linkage signal at this locus, although they may also contribute independently to schizophrenia risk (Straub et al, 2002).

7.2.1.1 DTNBP1 and schizophrenia

Investigating the linkage peak on chromosome 6p22 in the Irish High Density Family Study of Schizophrenia sample, Straub et al, (2002) identified significant association between schizophrenia and several individual markers and haplotypes across the gene encoding dystrobrevin-binding protein (*DTNBP1*). A follow up study of *DTNBP1* was performed based primarily on a sample of German proband-parent trios (Schwab et al, 2003). This confirmed evidence for association with individual markers and haplotypes, however the two studies differed as to whether the common or rare alleles at each marker were being preferentially transmitted.

The first published case-control study, in our Phase 1 sample of 219 Irish cases and 231 Irish controls failed to find evidence for association (Morris et al, 2004). Subsequent family-based and case-control studies in Chinese, Bulgarian and Swedish schizophrenia samples identified evidence for association with *DTNBP1*, but negative studies in German and Polish samples have also been reported (reviewed in Kendler, 2004). Identification of association with a 3-marker haplotype in a large UK case-control sample was subsequently confirmed by genotyping an additional marker in our Irish case-control sample (Williams et al, 2004). The same individual risk and

protective haplotypes were present in both UK and Irish samples. Taken together these studies provide evidence that variation at the *DTNBP1* locus is associated with schizophrenia, this association may involve both risk and protective variants, but the risk or protective alleles involved may differ between populations. The function of *DTNBP1* and its role in schizophrenia biology is uncertain, although *DTNBP1* may be involved in presynaptic glutamate regulation (Talbot et al, 2004). Indeed two recent expression studies identified reduced levels of expression of *DTNBP1* mRNA and protein in post-mortem brain samples from patients with schizophrenia (Numakawa et al, 2004; Weickert et al, 2004).

7.2.1.2 *MRDS1* (*OFCC1: orofacial cleft 1 candidate 1*) and schizophrenia

Investigating the chromosome 6p susceptibility locus in the Irish High Density Schizophrenia families, Straub et al, (2003) identified that 10% of the families in this study contributed to the linkage signal at chromosome 6p22 (the site of *DTNBP1*). Additionally, a different 18% of families showed a signal at chromosome 6p24 only. Analysing these families separately, the authors identified evidence of association between two overlapping haplotypes and schizophrenia at this locus. The same group then identified evidence of a gene (which they named *MRDS1*) at this locus (Matsumoto et al, 2002). Expression analysis indicated *MRDS1* transcripts in frontal cortex, hippocampus and substantia nigra, but not other brain regions. However, the function of this gene is unknown. The authors reported additional evidence of association between individual markers and haplotypes at the *MRDS1* locus and schizophrenia in adult-onset (the NIMH CBDB Sibling Pair Study sample) and childhood-onset (the NIMH Child Psychiatry Branch Childhood-Onset sample) US populations (Straub et al, 2003). There was however, a lack of consistency between studies as to the associated individual markers and haplotypes. The authors also failed to indicate the number of markers investigated in the original and follow-up studies.

7.3 Methodology

7.3.1 Case-control study sample

The ascertainment and assessment methods for the extra cases and controls included in these analyses and further details of the sample are described in Chapter 4. The

extended study sample included 299 cases with DSM-IV schizophrenia (n=234) or schizoaffective disorder (n=65) and 645 controls.

7.3.2 Marker selection and genotyping methodology.

At the *G72, DAAO* and *MRDS1* loci, individual SNPs were selected for investigation based on a review of the schizophrenia literature and data presented at the World Congress on Psychiatric Genetics (2002-2004). Markers contributing to the previously described risk haplotypes at the *DTNBP1* and *NRG1* loci, and the *erbB4* SNP described by Norton et al (2004), were also genotyped in the extended sample. The procedures for primer design and PCR optimisation/amplification are described in Chapter 3. Two laboratory technicians (KM and NK) performed the additional SNP and microsatellite genotyping. The SNP genotyping was performed using Custom Taqman® SNP Genotyping Assays on an ABI Prism 7900HT Sequence Detection System (Applied Biosystems). The microsatellite markers were amplified using fluorescently tagged primers and the amplicons were run on an ABI PRISM 3100 Genetic Analyzer along with a fluorescent size standard. Marker alleles were called based on their relative size using ABI Genemapper software (Applied Biosystems).

7.3.3 Statistical methods

The statistical methods employed to analyse single marker and haplotype data are described in Chapter 3. Applying traditional Bonferroni correction to multiple SNPs tested at a locus is conservative as individual SNPs may not be independent because of LD. In this study a correction for multiple testing of SNPs was applied in examining the *G72, DAAO* and *MRDS1* loci, based on spectral decomposition of matrices of pair-wise LD between SNPs described by Nyholt, (2004). This is implemented in the programme SNPSpD and allows for the potential non-independence of SNPs in estimating the total number of tests carried out when multiple markers are genotyped at a locus (http://genepi.qimr.edu.au/general/daleN/SNPSpD/).

7.3.3.1 Logistic regression model

The logistic regression analysis method described by Cordell, (2002) was applied in investigating markers at the *G72* and *DAAO* loci for evidence of statistical interaction.

Two models were tested, the first including the individual markers at these loci which showed the greatest evidence for association with schizophrenia in this study population. The second model tested for the interaction between the markers M-22 and MDAAO-6 which had been reported by Chumakov et al, (2002). These data were fitted using the model described by Cordell, (2002) and implemented in SPSS (version 11.0).

7.3.3.2 MDR methodology

Applying the MDR method to this dataset presented two specific difficulties. First, the methodology was developed for balanced data where the number of cases and controls in the dataset are equal. Second, the method was designed to test for interaction between SNPs or other binary variables, but single-loci effects at a number of loci in this model (*NRG1, DTNBP1*) are best represented by haplotype rather than individual marker data. For this reason, collaboration was sought with Dr Jason Moore (one of developers of the MRD method) in designing, running and interpreting the model described.

In the tested models, cases were oversampled by selecting additional cases at random with replacement. This balanced the number of cases and controls at 645. Two of the loci (*NRG1* and *DTNBP1*) were represented by haplotypic data. Assignment of haplotype carrier status to individual cases and controls was performed using the Bayesian approach to reconstructing haplotypes implemented in PHASE (v2.0.2) (Stephens and Donnelly, 2003). Individuals were only included in the study where the identity of haplotypes at these two loci could be determined at $p>0.95$. The haplotype data was represented in the model as a categorical variable (indicating the presence or absence of the risk haplotype at the locus).

The MDR procedure involves dividing the data into a training set (9/10 of the data) and an independent testing set (1/10 of the data). If testing 2-loci models, there are three possible genotypes and nine possible combinations across any two loci. In the next step, at each of these possible combinations the ratio of cases to controls is determined. If the ratio of cases to controls exceeds a threshold (≥ 1) this combination is labelled as high risk (combinations where the ratio is below this threshold are

considered low risk). The collection of these 'high risk' and 'low risk' combinations represents the MDR model for these two-loci. This model is than used to predict case and control status in the independent testing set. The entire procedure is than repeated 10 times using different random number seeds to estimate the predictive error of the test model. This step is than repeated ten times and the prediction errors are averaged.

The procedure is followed to test each possible model size (i.e. from 1-6 loci). For each model size, as well as testing prediction error, the cross-validation consistency is examined. This is a measure of the number of times that a particular set of loci are identified across the independent testing sets: the maximum value being 10 where the same set of loci are identified across each independent testing set. The 'best' model is that most likely to have the lowest prediction error and have cross-validation consistency. An empirical p-value is also calculated by permutation testing which tests whether the prediction error of the chosen model would be expected under the null hypothesis.

7.4 Results

7.4.1 Single gene association studies

7.4.1.1. Results of candidate gene studies at *G72* and *DAAO*

Five SNPs were genotyped in the Phase 2 case-control sample: rs3916965 (G72-M12); rs2391191 (G72-M15); rs778293 (M22); rs3918342 (G72-M23) and rs1421292 (G72-M24) (see Table 7.1). Application of the SNPSpD programme indicated that the effective number of independent SNPs at this locus was 3.81 and a significance threshold of p= 0.013 was required to keep Type I error at $\alpha = 0.05$. Evidence for single marker association was identified at two SNPs: G72-M12 (χ^2= 7.87, p=0.005, OR 1.34 (95% CI=1.09 – 1.65) and G72-M15 (χ^2= 6.40, p=0.011, OR 1.31 (95% CI= 1.06-1.61)). Based on the r^2 value between M12 and M15 (r^2= 0.99), including both markers in haplotype analysis would provide no additional information, therefore only M12 was included. Haplotype analysis identified three significant two-marker haplotypes; each of these haplotypic associations resulted from excess haplotypes in the case sample containing the *T* allele at M12.

Four markers were genotyped at the *DAAO* locus: M4 (rs2111902), M5 (rs3918346), M6 (rs3741775) and M9 (rs888531). Application of SNPSpD indicated the effective number of independent SNPs at this locus was 3.5 and a significance threshold of p= 0.014 was required to keep Type I error at $\alpha = 0.05$. Significant evidence for single marker association was identified at DAAO-M5 (χ^2= 8.55, p value = 0.0034, OR = 1.43 (1.12 – 1.84) and marginal evidence for association at DAAO-M4 (χ^2= 5.64, p value = 0.018, OR = 1.29 (1.04-1.61). The results of individual *DAAO* marker analyses are included in Table 7.1. As was the case with the *G72* analysis, investigation of haplotypes did not provide additional information.

7.4.1.2 Results of candidate gene studies at *NRG1* and *erbB4*

Neither the 'core' risk haplotype identified in the Icelandic sample nor the overlapping risk haplotype ($HapB_{IRE}$) identified in Chapter 4 were associated with schizophrenia in the extended sample. As in the original report by Norton et al, (2004), there was no evidence for single-locus association of the *ErbB4* marker with schizophrenia in our sample. These results are summarized in Table 7.2.

7.4.1.3 Results of association studies at *DTNBP1*

The three SNPs that define the previously risk haplotype (rs2619539, rs3213207 and rs2619538) were genotyped in the extended sample. This confirmed that the individual markers were not significantly associated, but there was an association between the 3-marker haplotype and schizophrenia (global χ^2=14.94, p=0.006). The individual haplotypes involved are detailed in Table 7.2.

7.4.1.4 Results of association studies at *MRDS1*

Five markers were genotyped across the *MRDS1* locus. There were no significant single marker or haplotype associations with schizophrenia in our sample. The results of the single marker analyses are also presented in Table 7.2.

7.4.2 Interaction between the *G72* and *DAAO* loci

Two logistic regression interaction models were tested. The first used the individual markers with the most significant evidence of association at the *G72* (M12) and *DAAO* (M5) loci. This demonstrated evidence for interaction (OR=9.3 (CI 1.4-60.5),

p=0.008) and the results of the analysis are summarized in Table 7.3. Inspection of the outputs from this model indicated that the presence of at least one *C* allele at M5 and at least one *T* allele at M12 made the most statistically significant contribution to this effect. There was no evidence of interaction between the markers reported by Chumakov et al, (2002) to interact in schizophrenia susceptibility (*G72* (M-22), *DAAO* (M6); (p=0.70)).

7.4.3 Investigation for epistatic interaction between loci

The results of the MDR analysis are presented in Table 7.3. Two models provided the highest cross validation consistency (CVC), these were a single locus model (*DAAO*) and a two-locus model *(DYS-MRDS1)*. In each case, the model was selected in 9 of 10 testing sets. The single locus model had a lower prediction error (or higher testing accuracy in the table). This indicated that the *DAAO* single locus model provided the best fit to this dataset. From permutation testing, the testing accuracy expected under the null hypothesis using traditional criteria for significance (p=0.05), was 0.7 indicating that the ability of the model to predict case or control status fell in the suggestive range of significance (p<0.1).

7.5 Discussion

Investigation of individual loci confirmed evidence of association between the genes *DTNBP1*, *G72* and *DAAO* and schizophrenia susceptibility in the extended case-control sample. The results at the *DTNBP1* locus confirmed the finding reported from the smaller Phase 1 dataset (of 219 cases and 231 controls), which is discussed in Section 7.2.1.1 (Williams et al, 2004).

The results of the individual marker analyses at *G72* and *DAAO* warrant further comment. The association at the M12 *G72* locus represents the first direct replication of the association reported from a French-Canadian case-control population (Chumakov et al, 2002). Evidence for association with this marker has been reported by case-control studies of Chinese and German schizophrenia samples (Wang et al, 2004; Schumacher et al, 2004) and also from a Chinese family-based sample (Zhou et al, 2005): in each case the association was in the opposite direction to that described in the original report. These patterns of association may be more explicable if there

was consistency among ethnic groups, but this was not the case as the association in the German sample differed from that reported in the other samples of European ancestry resembling the results from the Asian samples. In interpreting these results it must also be noted that there have also been a number of negative reports (these results are summarized in Table 7.5).

Conversely, the most significant finding at *DAAO* (of association with the *C* allele at marker M5) was in the opposite direction to the association reported in the original French-Canadian case-control schizophrenia sample, but in the same direction, with similar allele frequencies to the association reported from the German sample (Chumakov et al, 2002; Schumacher et al, 2004). Again there have also been a number of negative reports from analyses of markers at *DAAO* and these are summarized in Table 7.6. These results highlight the difficulty in comparing results from different datasets and may relate to differences in ascertainment/sampling or in allele and haplotype frequencies between populations. Factors that contribute to such differences are discussed more fully in Chapter 3. Alternatively, the results may relate to false-positives or to the effects of multiple risk or protective variants at these loci. These issues are considered further in Chapter 8.

The haplotypic association identified at the *NRG1* locus in the Phase 1 sample (discussed in Chapter 4) was not confirmed in the extended Phase 2 dataset. The frequency of the $HapB_{IRE}$ haplotype in the extended sample remained similar to in the original report but the estimated frequency of this haplotype in the larger control sample increased (from 12.3% to 17.3%). Partial re-genotyping in the extended sample using a different genotyping method (see Section 7.1.4) indicates that this is unlikely to be due to genotyping error. The additional controls were sampled from the same geographical region and source as the original cases. Of note, investigation of this haplotype for association with schizophrenia in the Irish Schizophrenia High Density Families found no evidence of association and reported a similar frequency for this haplotype in these familial cases.

Logistic regression analysis identified an interaction between the M12 marker at *G72* and the M5 marker at *DAAO* in contributing to schizophrenia susceptibility, where the presence of at least one *C* allele at M5 and at least one *T* allele at M12 increased risk

substantially (OR=9.3 (CI 1.4-60.5)). This data adds to earlier statistical and biological evidence of interaction between these genes, however the interacting markers differed between the two studies. This may relate to the differences in the pattern of association between the two studies. The markers investigated for interaction in the original report from a case-control French-Canadian sample were also individually associated with schizophrenia; neither of these markers was associated with schizophrenia in the Irish dataset. The original report does not indicate if additional markers were tested for interaction at the *G72* and *DAAO* loci. Further investigation of interaction between these loci will be required in independent datasets. Possible reasons for differences in patterns of association when investigating complex genetic disorders in different populations are considered more fully in Chapters 2 and 8. Of note, none of the associated SNPs reported in this study or by Chumakov et al, (2002) were functional.

Testing for epistatic interaction between all six loci indicated that the best 'fit' to this data was a single locus model with marker M5 at the *DAAO* locus. Given that the individual association data indicated that this association had a modest effect size (OR=1.43), the predictive power of this model was modest and did not meet the threshold for statistical significance ($p>0.05$). The MDR method also indicated suggestive evidence for interaction for a two-locus model but with two different genes involved (*DTNBP1* and *MRDS1*).

In interpreting these results it is important to note that this was an exploratory analysis. The results provided suggestive evidence for two different models but lacked consistency as to the genes contributing to these models. This highlights some of the difficulties in investigating higher order interaction. First, investigating multiple genes for interaction in the absence of an *a priori* hypothesis may make it difficult to interpret the results of such an analysis particularly in the absence of replication by other statistical approaches. Failure to identify evidence of epistatic interaction may have been as a result of the loci selected for inclusion in the model. These were selected opportunistically and investigation of genes in known molecular pathways may be more scientifically plausible.

Second, evidence of interaction should be confirmed in independent samples but this may prove even more difficult than for single loci because of differences in allele/haplotype frequencies, population structure, or genotyping error. Third, the markers selected for analysis may be in LD with the true risk variants, which would reduce study power. Fourth, for this study the sample size was relatively small and it is difficult to establish the power of the study to identify a set level of interaction. Finally, available methods of investigating interaction are best applied to dichotomous variables (such as SNP genotypes) but this may be over simplistic and doesn't readily allow for the inclusion of microsatellite or haplotype data.

Despite these caveats these results add to a growing literature that epistatic interaction cannot be ignored in attempting to elucidate the genetic architecture of schizophrenia susceptibility. Traditional regression models test for interaction between two loci, but interactions between multiple genes in systems or pathways are equally, if not more, biologically plausible. Simulations by Marchini et al, (2005) suggest that even with conservative correction for multiple testing under a range of different models of locus interaction, methods of investigating for statistical interaction between multiple markers may be more powerful than traditional single marker analyses. Such tests may best be applied to markers for gene families or systems and results confirmed independently: the author is currently applying this method to investigate *DTNBP1* and the other component genes of the biogenesis of lysosome-related organelles complex-1 (*BLOC-1*).

Figure 7.1: Putative interactions between schizophrenia susceptibility genes

Potential points of interaction where putative susceptibility genes could influence glutatamate transmission in schizophrenia

Adapted from Owen MJ et al. J Clin Investigation (2004)

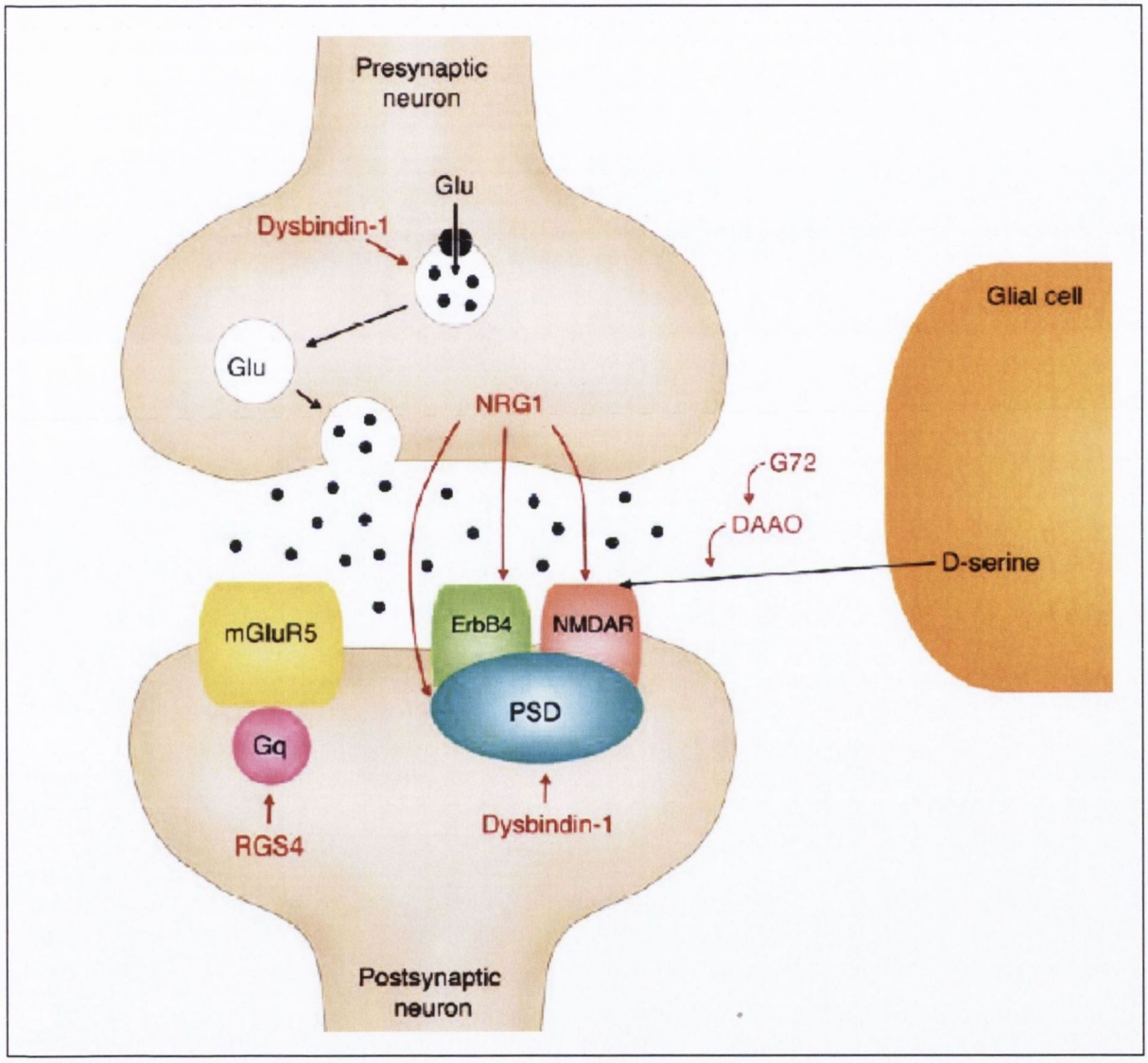
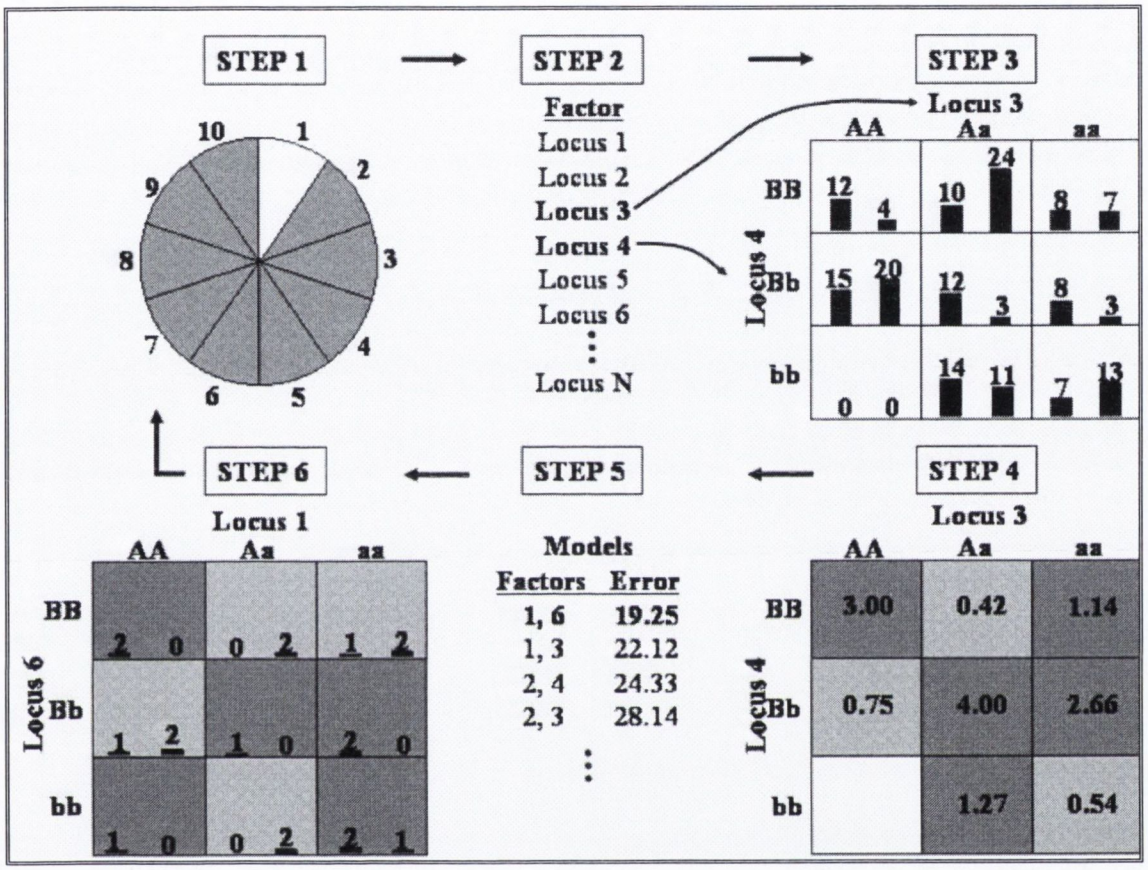


Figure 7.2: Multidimensional Reduction (MDR) method
From Moore, 2003



Summary of the steps involved in implementing the MDR method (in this case for investigating 2-locus models). The sample is divided into a training set (9/10) and test set (1/10). In principle a set of *n* genetic variables is selected; the *n* factors and their multifactor classes are cells are represented in a table (in this example 3x3 representing the possible haplotypes). Each cell is then labeled as 'high-risk' or 'low risk' depending on a predetermined cut off of risk (in this case that frequency of cases>controls in the training set). The prediction error is then estimated in the test set. This is than repeated 10-times to give an average prediction error. This procedure it than repeated ten times using different training sets of data. This provides an average prediction error. The model is retested using different random seeds to select training and test sets. This allows a measure of how reliably a set of loci is identified across independent testing sets (cross-validation consistency).

Table 7.1: Results of single marker analyses for *G72* and *DAAO*

SNP marker[a]	SNP rs #	Polymorphism[b]	Distance[c] (kb)	Allele Frequency		p value	Odds Ratio (95% CI)
				Cases[d]	Controls		
G72-M12	rs3916965	C/T	0.000	0.689	0.622	**0.005**	1.34 (1.09, 1.65)
G72-M15	rs2391191	A/G	16.086	0.317	0.378	**0.011**	1.31 (1.06, 1.61)
G72-M22	rs778293	A/G	65.839	0.592	0.625	0.137	1.15 (0.96, 1.38)
G72-M23	rs3918342	C/T	82.389	0.490	0.457	0.186	1.14 (0.94, 1.39)
G72-M24	rs1421292	A/T	94.875	0.473	0.506	0.181	1.14 (0.94, 1.39)
DAAO-M4	rs2111902	G/T		0.266	0.320	**0.018**	1.30 (1.05, 1.61)
DAAO-M5	rs3918346	C/T		0.812	0.751	**0.003**	1.43 (1.12, 1.82)
DAAO-M6	rs3741775	T/G		0.517	0.554	0.136	1.16 (0.95, 1.41)
DAAO-M9	rs888531	T/C		0.708	0.738	0.185	1.16 (0.93, 1.45)

[a] Name of SNP marker used in the Chumakov et al paper (2002)
[b] Alleles called with respect to Taqman assay (Alleles in brackets relate to Chumakov et al 2002 paper)
[c] Distance of SNPs in relation to each other using the format of Korostishevsky et al (2004)
[d] Frequency shown relates to the first allele listed in the polymorphism column

Table 7.2: Genotype marker data from *DTNBP1*, *NRG1* and *MRDS1*

Gene	Marker/haplotype	Variant	Freq in cases	Freq in controls	p-value	OR (CI)
NRG1	Core haplotype		0.096	0.101	0.67	
	HapBIRE		0.189	0.173	0.39	
erbB4	IVS12-15	C/T	0.50	0.53		
DTNBP1	P1655-P1635-SNP A	C-A-T	0.22	0.17	0.013	
		G-G-T	0	0.02	<0.001	
MRDS1	rs855362	C/T	f(T) = 0.1107	f(T) = 0.1342	ns	
	rs4716055	A/C	f(C) = 0.1119	f(C) = 0.1362	ns	
	hCV11320114	A/G	f(A) = 0.1484	f(A) = 0.1372	ns	
	rs1340640	T/C	f(C) = 0.1096	f(C) = 0.1361	ns	
	rs2096022	G/T	f(T) = 0.1475	f(T) = 0.1349	ns	

Table 7.3 Results of MDR modelling of interaction between schizophrenia loci

Number of Loci	Model	CV Consistency	Avg. Testing Accuracy	P-Value
1(*)	*DAAO*	9/10	0.6960	<0.100
2	*DTNBP1* *MRDS1*	9/10	0.6890	>0.100
3	*DTNBP1* *DAAO* *G72*	3/10	0.6624	>0.100
4	*DTNBP1* *NRG1* *ErbB4* *G72*	6/10	0.6445	>0.100

* Best Model – has maximum testing accuracy and a CVC of 10/10

0.7102 is the critical value of the testing accuracy for α = 0.001
0.7044 is the critical value of the testing accuracy for α = 0.010
0.7009 is the critical value of the testing accuracy for α = 0.050
0.6992 is the critical value of the testing accuracy for α = 0.100

Table 7.4 Results of schizophrenia association studies at the DAAO locus

Markers	Alleles	Irish Study			Chumakov et al (Canadian)			Yui et al 2004			Schumacher et al (Germany)		
		Allele Freq Cases	Allele Freq Controls	p value	Allele Freq Cases	Allele Freq Controls	p value	Allele Freq Cases	Allele Freq Controls	p value	Allele Freq Cases	Allele Freq Controls	p value
DAAO-M4	G/T	0.266	0.320	**0.018**	0.36	0.28	**0.017**	0.5074	0.4758	0.319117	0.28	0.34	**0.026**
DAAO-M5	C/T	0.812	0.751	**0.003**	0.72	0.79	**0.007**	0.5054	0.5093	0.934585	0.72	0.78	**0.019**
DAAO-M6	T/G	0.517	0.554	0.136	0.59	0.49	**0.001**	0.7569	0.6623	**0.000001**	0.61	0.53	**0.021**
DAAO-M9	T/C	0.708	0.738	0.185	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Table 7.5 : Comparison of results from published genetic association studies of *G72* and Schizophrenia

Markers	Allele	Irish Study			Chumakov et al (Canadian)			Chumakov et al (Russian)			Schumacher et al (Germany)			Korostishevesky et al (Ashkenazi Jew)			Wang et al (Chinese)			Addington et al (USA)		Zou et al (Chinese)	
		Cases	Controls	p value	Cases	Controls	p value	Cases	Controls	p value	Cases	Controls	p value	Cases	Controls	p value	Cases	Controls	p value	TDT	p value	TDT	p value
																						Overtransmitted	
M12	C/T	0.689	0.622	**0.005**	0.64	0.55	**0.007**	ns	ns	ns	0.59	0.64	**0.048**	ns	ns	ns	0.38	0.43	**0.019**			A allele (A/G)	**0.0018**
M15	G/A	0.683	0.622	**0.011**	0.65	0.58	**0.032**	na	na	na	0.59	0.64	**0.037**	ns	ns	ns	0.38	0.45	**0.001**	TDT	0.031	A allele (A/G)	**0.0108**
M22	A/G	0.592	0.625	ns	0.69	0.60	**0.003**	ns	ns	ns	na	na	na	ns	ns	ns	ns	ns	ns				
M23	T/C	0.510	0.543	ns	0.57	0.49	**0.019**	0.60	0.51	**0.017**	0.47	0.54	**0.033**	62.3	43.3	**0.001**	0.49	0.46	0.16				
M24	T/A	0.527	0.494	ns	0.55	0.47	**0.019**	ns	ns	ns	0.56	0.5	**0.036**	na	na	na	na	na	na				

[1]*Data in the Case and Controls columns represents allele frequencies*

Chapter 8

Discussion

The research outlined in this thesis had two specific aims. The first was to establish and manage a well-characterised sample of individuals with a major psychotic disorder and to obtain controls from the Irish population. The second was to perform laboratory and statistical analyses to investigate genes for evidence of association with schizophrenia.

The procedures involved in establishing the sample are detailed in Chapter 3 and the sample itself is described in Chapter 4. In planning this work it was self-evident that collecting sufficient cases to provide statistical power for genetic association analysis would not be possible for a single investigator within the time frame of a Wellcome Trust Research Training Fellowship. Therefore, during the data collection phase, grant funding was obtained to employ additional clinical research staff (SS and JMN). This greatly accelerated recruitment at the St. James's Hospital site and prompted expansion to other Dublin sites (detailed in Chapter 3). In assessing cases, standard research measures were employed and are also detailed in Chapter 3. The author received research training and in turn, trained the other raters, and conducted inter-rater reliability exercises for these measures through the recruitment phase of the project. Inter-rater reliability for diagnoses and for the other measures was in the excellent range for all measures studied (kappa>0.8) with the exception of the Lane scale for measuring MPAs for which inter-rater reliability was moderate (kappa>0.6).

Most referrals to the project came from clinical teams at the participating institutions, but a small number of self-referrals were prompted by media coverage of the project. A total of 743 potential cases were identified. Of these, 680 were approached about the study. A significant hindrance to ascertainment was that the initial assessments conducted by our collaborators in Monaghan predated this study by 1-5 years and we were unable to make contact with 63 of the 192 individuals in Monaghan who had been identified locally as potential participants. With additional funds to employ a full-time researcher in Monaghan it may have been possible to contact more of this

group. Of all patients approached 124 (18%) refused to participate. An additional 131 individuals were excluded from the study principally on failing to meet ethnicity criteria, but also because they failed to meet diagnostic grounds for inclusion (a small number of patients were identified at assessment to have a co-morbid learning disability, a schizoid or schizotypal personality disorder, obsessive-compulsive disorder or in one case a pervasive developmental disorder). Twenty-three potential participants were deemed to be too unwell to give informed consent to participate both when initially approached and later when this was reassessed. Sadly, 15 of the listed potential participants had passed away before being assessed: this mostly related to the delay in approaching elderly individuals from the Monaghan cohort. In the Dublin group, one participant withdrew consent and was removed from the study and two others agreed to participate but were lost to follow up.

Of the 396 individuals meeting criteria for a DSM-IV functional psychotic disorder the most common diagnosis was schizophrenia/schizophreniform disorder (n=234), followed by schizoaffective disorder (n=65), bipolar affective disorder (n=65), depressive disorder with psychosis (n=18), delusional disorder (n=8) and psychosis not otherwise specified (n=6). As discussed in Chapter 3 individuals in the control sample (n=645) met the same criteria for Irish ethnicity as cases and were recruited from the Irish Blood Transfusion Service at St James's Hospital, Dublin. The controls were sampled anonymously and informed consent was obtained by nursing staff at the Irish Blood Transfusion Service with the support of our clinical team (primarily SS). Due to the stipulations of the local ethics committee no demographic information was available on these individuals and the individual samples were anonymous on receipt by the research team. It would have been methodologically more rigorous to have more demographic information available from the control population and the facility to perform screening interviews. On discussion with the ethics committee this approach is planned for our new independent psychosis sample collection (Resource for Psychoses Genomics, Ireland (RPGI)), which is described in more detail below.

The initial aim of the laboratory phase of this study was to investigate novel candidate genes selected on the basis of positional and functional data in the literature. These studies used a case-control genetic association paradigm. By sampling cases and controls from the same self-reported ethnic background the assumption was made that

that there were no significant differences between the samples in terms of their population of origin. Methods of investigating and dealing with differences in sample structure (such as Genomic Control) were not employed because of the limited resources available, debate as to the necessity of such methods (see Chapter 2) and uncertainty as to the amount of genotyping that would be involved. Recent data from a relatively homogeneous genetic isolate (Iceland) highlights this continuing debate and the potential for substructure in the Irish population may be a consideration in future sampling strategies and analyses (Helgason et al, 2005).

In conducting the genetic studies, particular emphasis was given to identifying potential candidate genes for association analyses at susceptibility loci previously identified from linkage data from Irish and other schizophrenia family studies. Another goal of this project was to establish the feasibility of a more intensive LD-mapping approach across one of these chromosomal loci. As is evident by the completed thesis, the emphasis shifted substantially during the course of this work in response to rapid technological progress, greater availability of human genome data and advances in the field of schizophrenia genetics.

Over the last five years, the debate about schizophrenia genetics has moved through an important transition. While it was generally accepted that genes contributed to schizophrenia susceptibility, by 2000 there were growing doubts that any of these genes could be detected in studies of feasible sample size with the phenotypes being measured. Indeed, it was suggested that funding of schizophrenia genetics research was a prime example of *'a case of there being too much focus on diseases that we want to understand rather than on phenotypes that we are likely to be able to understand'* (Terwilliger & Göring, 2000). Notwithstanding the fact that identifying genes for complex disorders has in general proved difficult, there is now sufficient data to substantially rebut this viewpoint. Variants at a number of candidate genes (including *NRG1, DTNBP1, G72, DAAO, RGS4, DISC-1, 5HT2a* and *DRD3*) have met statistical criteria for association with schizophrenia, and been replicated in independent populations. The genetic association analyses described in this thesis and conducted by the author's research group support the hypothesis that *DTNBP1, RGS4, G72* and *DAAO* contribute to schizophrenia susceptibility in this study population (Williams et al, 2004; McGhee et al, 2004; Morris et al, 2004).

During the initial recruitment phase of this project, evidence for association with genes at two of the target chromosomal loci was reported from Irish and Icelandic schizophrenia studies (Stefansson et al, 2002; Straub et al, 2002). These genes (*NRG1* and *DTNBP1*) were thus prioritised for association analyses.

Investigation of the neuregulin-1 (*NRG1*) candidate gene is described in Chapter 5. In analysing data from 20 markers genotyped across the *NRG1* locus, the author failed to confirm association with the 'risk' haplotype initially identified in the Icelandic population and subsequently replicated in a Scottish case-control sample (Stefansson et al, 2002; Stefansson et al, 2003). However, association was detected with an overlapping haplotype, which shared common markers with the 'risk' haplotype. A significant excess of this novel 'risk' haplotype was also detected in the Scottish cases but not in the Icelandic schizophrenia sample. The associated variants in this study occurred close to an expressed sequence tag (EST) cluster within the first intron of the *NRG1* gene, suggesting involvement of this cluster in susceptibility at the *NRG1* locus (Corvin et al, 2004).

Following up this finding, the *NRG1* locus was investigated in a larger extended case-control sample (described in Chapter 7). This failed to support the initial finding, principally because of an increase in the frequency of the identified risk haplotype in the extended control sample. Since the original report there have been additional replications in other Caucasian populations, evidence for involvement of other markers at this locus in schizophrenia susceptibility in the Chinese samples, and a number of negative reports. In addition there is also recent data suggesting that the Icelandic core risk haplotype may be associated with Bipolar Affective Disorder (Green et al, 2005). Although no functional variants have been identified, meta-analysis of existing datasets may be helpful in confirming the role of *NRG1* in susceptibility to schizophrenia and other functional psychoses.

Additional research funding (an SFI research programme grant with Professor Michael Gill) made it possible to investigate other putative schizophrenia candidate genes (including the *DTNBP1* gene at chromosome 6p). The author contributed to genotyping markers at the *DTNBP1* and *RGS4* loci, using the methodology described in Chapter 3. The author, with Dr Derek Morris, performed the analyses of these data

and also on genotype data from the *G72* and *DAAO* loci. The results support the hypothesis that each of these four genes is associated with schizophrenia susceptibility in the study population, although the effect size at each locus is predictably modest (OR $\leq$ 1.4). The results of the *DTNBP1, G72* and *DAAO* analyses are detailed in Chapter 7 and the *RGS4* data has been published elsewhere (McGhee et al, 2004; Morris et al, 2003).

The second aim of the laboratory work was to investigate novel candidate genes for evidence of association and establish the feasibility of an LD-mapping strategy across a broad locus. Based on positional and functional studies, the apolipoprotein-L (*APOL*) gene family were selected for the study described in Chapter 6; additional research funding was obtained to fund a post-graduate student to share the workload involved. A three-stage DNA pooling strategy was applied to investigate this locus. The strategy was based on that described by Norton et al, (2003), but incorporated a novel variation to increase efficiency (McGhee et al, 2005). These DNA pooling studies identified two markers, which met screening criteria for individual genotyping, but neither marker met standard statistical criteria for evidence of association (p=0.05) when individually genotyped. Given the absence of significant single marker association, even prior to any correction for the total number of markers investigated, these data do not support involvement of the *APOL* gene family in schizophrenia susceptibility.

The identification of schizophrenia susceptibility genes is a major advance, but it is important to caution that the evidence for involvement of genes in schizophrenia susceptibility is principally statistical. Drawing biological inference from these data can be problematic, as neither causality nor the direction of an effect (variants may increase or decrease disease risk) can be assumed. Comparing statistics and interpreting results across studies may be complicated by phenotypic, genotypic or environmental differences between the populations being investigated (as discussed in Chapters 1 and 2). A salient example is the reported association between *DTNBP1* and schizophrenia where different associated markers and haplotypes have been reported even among samples of common European ancestry (this example is discussed in depth by Neale & Sham, 2004). Such results may reflect differences in allele or haplotype frequencies between populations; the effect of multiple risk

variants at this locus; or in some cases false positive findings (van den Oord & Neale, 2003; Owen et al, 2004).

In addition to the genes listed above, other unreplicated studies have reported nominal positive association between genetic variants and schizophrenia. Interpreting these studies is difficult because of the low prior probability of association and the large number of alleles or haplotypes tested without correction. This is an important issue for the wider field of complex genetic disorders. Encouragingly, methods of statistical correction are in development that allow for LD between markers when investigating individual markers (as described in chapter 7); for the multiple tests performed when conducting haplotype analyses (Becker & Knapp, 2004; Seaman & Muller-Myhsok, 2005); for gene-based approaches to association analysis of candidate genes (Neale & Sham, 2004); and for genome-wide association analyses (Lin, 2005). Guidelines in implementing these correction methods would be useful in reconciling results from different datasets and reaching a more complete picture of schizophrenia susceptibility.

Conversely, of the many candidate genes investigated, few can be reliably excluded from the list of potential schizophrenia susceptibility genes. Indeed, false negative reports may be an even larger problem as most studies may have lacked sufficient power to detect the modest individual genetic effects predicted by genetic epidemiological and linkage studies (Lohmueller et al, 2003). Collaboration between groups is an economical method of increasing sample size. For example, the author and colleagues have contributed to a consortium investigating the *RGS4* locus for association with schizophrenia in an analysis involving more than 12,000 individuals (Talkowski et al, 2004). A major disadvantage of this approach is the inclusion of many different populations making interpretation of the results difficult for the reasons outlined previously. Sampling from defined populations may be advantageous; the author and colleagues have recently been awarded funding by the Wellcome Trust to collect a large Irish case-control psychoses sample (Resource for Psychoses Genomics, Ireland). This sample, which will include ~1,000 individuals with schizophrenia and ~3,000 controls from the Irish population, will be independent of existing collections and available as a resource to the international research community.

Study power is not the only issue in attempting to identify susceptibility variants for complex genetic disorders. Present technology dictates that only a subset of total genomic variation can be investigated in a given study. The prospect of genome-wide association studies is enticing (for the reasons discussed in Chapter 1), but as currently envisaged, this approach has limitations. The first generation of genome-wide association studies are likely to concentrate on common SNPs, such as those investigated by the HapMap project (Thomas et al, 2005). This approach, while pragmatic, is likely to omit >94% of SNPs in the human genome (Botstein and Risch, 2003). Until further advances in genotyping technology allow all variation to be directly tested, it will be the role of these studies to establish the relative contribution of common risk alleles to complex disorder susceptibility.

The experiments conducted in Chapter 6 demonstrate that study power and marker selection are not the only issues when planning an LD-mapping strategy. Across the genome, regions of low haplotype diversity are interspersed with regions of low LD, which would necessitate more intensive genotyping efforts. Defining the boundaries between regions is problematic. First, methods of defining LD blocks differ because of the different statistical assumptions applied. Second, as demonstrated at the *APOL* locus, the complexity of the haplotype structure defined may also be a feature of marker density. Maniatis et al, (2002) have developed an alternative approach that avoids defining block structure. This method applies metric maps in LD units to describe patterns of LD and may provide a scale on which to distribute SNPs to minimise marker redundancy when investigating a candidate gene or locus. The success of genome-wide association or LD mapping approaches is also predicated on an ability to infer LD-structure from reference population data to reduce genotyping redundancy. Recently, this has been tested across a number of European populations, suggesting that the assumption may be valid for some genomic regions but not for others (Mueller et al, 2005).

Despite all of these methodological, statistical and technological difficulties, and in the absence of biological markers for the disorder, susceptibility genes for schizophrenia have been identified. Allelic or haplotypic evidence of association with schizophrenia has been replicated at the *NRG1, DTNBP1, G72, DAAO, RGS4, DISC-*

1, 5HT2a and *DRD3* genes. Although the discovery of genes is a step in a larger process, the list of genes identified thus far raises many new questions, including:

1. Where do these findings fit in the broader genetic architecture of the disorder?
The presence of individual susceptibility alleles is not strongly predictive of illness as these alleles are also common in the general population and their contribution to illness risk is modest. For example, the *DTNBP1* risk haplotype has only a small impact on illness risk (OR=1.34), as the prevalence of the disorder in the general population is relatively low (~1%). Another way of considering the impact that a risk allele (or haplotype) may have at a population level is by the Population Attributable Fraction (PAF). This is the fraction of cases which would not have occurred had this risk factor been removed or reduced to population baseline levels (Levin, 1953). The PAF can be calculated using the following formula:

$$PAF= P(E)(RR-1)/1+P(E)(RR-1)$$

Where RR = P(Disease in exposed)/P(Disease in unexposed)
And P(E) =Prevalence of exposure

In the example above, the PAF for the *DTNBP1* risk haplotype is 5.7%, suggesting a small contribution to risk at the population level. By contrast, the estimated PAF for the associated *DAAO* marker (*DAAOM5*) in this sample is more substantial (24%), because the allele is more common in the general population. The PAF may be a useful measure of the impact of a genetic variant, or gene-environment interaction at a population level but is dependent on a number of assumptions that may confound interpretation of results. For instance, the PAF is sensitive to the timing of environmental exposures, is designed primarily for categorical outcomes and exposures, and assumes that the exposure is causal and thus will not be sensitive to the effects of gene-gene interaction. In simple terms, this means that at a population level summed PAF values across susceptibility genes for a complex genetic disorder may exceed 100%.

The genes identified to date are likely to represent a small component of genetic susceptibility to schizophrenia. However, their discovery confirms the predictions

made based on epidemiological data, that schizophrenia is an oligogenic or polygenic disorder where individual genetic effects are modest. Estimating how much of the genetic variance in schizophrenia susceptibility is accounted for by these genes will require an understanding of the relationships between genes. In Chapter 7 the author tested for interaction between risk variants at the *G72* and *DAAO* loci using a logistic regression model. This study provided supportive evidence that there may be epistatic interaction between markers at these loci, but the markers involved differed from those suggested by the original report (Chumakov et al, 2002).

Testing for interaction between multiple genes may be more pertinent in establishing the genetic architecture of the disorder and identifying molecular systems or pathways that contribute to the phenotype. Methods of investigating epistatic interaction between multiple loci are reviewed in Chapter 7; in general these methods are exploratory and any hypotheses generated would require confirmation independently or with more traditional analyses. The author applied one of these methods, multifactorial dimensional reduction (MDR), to test for interaction between six schizophrenia candidate genes (*NRG1, erbB4, MRDS-1, DTNBP-1, DAAO, G72*) and identified a single locus model (*DAAO*) as providing the best fit to the dataset. Based on current data it is unlikely that there are simple relationships between risk and protective variants identified at susceptibility genes, making predictive testing for the disorder in this population unlikely. A more useful application of this methodology may be to investigating gene systems or families: the author is currently applying this method to study the family of genes related to *DTNBP1*, which are know as the Biogenesis of Lysosome-related Organelles Complex-1 (*BLOC-1*) genes.

2. What is the genetic mechanism involved?

Many of the susceptibility genes identified are novel or would not have been suspected as candidates based on their known functions. Intensive sequencing efforts have failed to identify obvious 'pathogenic' variants of known functional significance in most cases. This may be because at these genes the variants associated with schizophrenia are in LD with coding variants that alter protein structure, but other explanations are also possible. Differences in gene expression may be under the influence of *cis*-acting variation within genes (as has been reported for *DTNBP1* (Bray et al, 2005)) or other epigenetic modifiers. Genomic variants may have subtle

effects on transcription or translation, affecting gene expression, as has been described with other disorders (Pagani et al, 2002). Such regulatory variants may profoundly impact on disease phenotype, for example, by inducing exon skipping, activating cryptic splice sites or altering the balance of spliced isoforms (Pagani & Baralle, 2004). At a genomic level, a recent experiment using array technology highlighted that large-scale copy-number variations (LCVs), which are relatively under investigated in most positional cloning strategies, may be more frequent than previously supposed and exert a substantial impact on gene expression (Iafrate et al, 2004).

Protein-coding sequences are the most well defined functional components of the genome, yet only a third of the genomic sequence conserved from the rodent genome is predicted to code proteins (International Human Genome Sequencing Consortium, 2001). Recent data indicates that non-coding genomic regions are extensively transcribed into stable RNAs some of which are transported into the cytosol during development (Cheng et al, 2005). Current understanding of these sequences is summed up by their nomenclature, transcripts of unknown function (TUFs), and it is likely that only a minority of these have been annotated by current experimental and *in silico* methods. Little is known as to the impact these variants have on the relationship between genotype and phenotype.

The long-term aim of the Encyclopaedia of DNA Elements (ENCODE) is to detect all functional elements in the human genome, but the pilot phase is focussed on a specific 30 megabases of the genome across 44 target regions representing approximately 1% of the genome (www.ensembl.ucsc.edu/ENCODE/). This study may offer new insights into the mechanism of gene function in general, although predictably, none of the schizophrenia susceptibility genes discussed above map to these regions. Even complete information on functional genomic elements may not elucidate how DNA variation translates to phenotype. In keeping with the hypothesis that schizophrenia is a neurodevelopmental disorder of late onset, many elements may only be functionally active in certain cell types at particular developmental stages, but impact on phenotype only in adulthood. A molecular mechanism of this type has been demonstrated in an animal model of another adult neuropsychiatric phenotype, anxiety-behaviour. In this example, tissue specific expression of the *5HT1A* receptor

gene in the postnatal period regulates expression of adult anxiety behaviours (Gross et al, 2002).

3. How do these genes contribute to schizophrenia pathobiology?

Alternatively, how does this information fit in to the framework of our understanding of schizophrenia? Many of these genes appear to have a role in neurodevelopment. For example, *DTNBP1* may have a role in synaptogenesis; *RGS4* in neuronal differentiation; and *NRG1* isoforms at almost any neurodevelopmental point between neuronal differentiation/migration and the maintenance of mature functioning neuronal and glial cells. Too little is known about *G72* and *DAAO* function to fit them within this neurodevelopmental framework. As yet there is insufficient data available to establish how these genes relate temporally, whether these effects involve specific neural circuits, and how genetic variation interacts with environmental or stochastic aspects of neurodevelopment.

Neuropathological changes in schizophrenia are subtle and may represent the effects of altered connectivity because of synaptic or signalling dysfunction. Each of the susceptibility genes identified to date may have a role in the development or maintenance of synaptic function or in signal transduction. For example, genetic variation at *G72*, *DAAO* and *RGS4* may modulate receptor function at synapses; *RGS4* and *NRG1* appear to influence signal transduction; and at least three of the identified susceptibility genes (*NRG1*, *DTNBP1*, *DISC1*) are known to be involved in regulating synaptic plasticity.

Neuropharmacological models of are increasingly moving towards an integrated view of schizophrenia as a disorder of the regulation and interaction of dopaminergic and glutamatergic systems (see Section 1.2.1.2 and the more detailed review by Moghaddam & Krystal, 2003). Many of these genes impact on glutamate neurotransmission, particularly through NMDA-receptors, suggesting that genetic susceptibility may partly involve dysregulation of glutamatergic systems. Most directly, a role could be envisaged for the *DAAO* and *G72* susceptibility genes. The NMDA receptor has a secondary site for binding of glycine or D-serine; this site must be occupied for glutamate to gate the NMDA ion channel. Genetic variation at *DAAO* and *G72* may influence the efficiency of glutamate gating of the NMDA ion channel

as the *DAAO* enzyme regulates levels of D-serine and is itself activated by *G72*. Serum levels of D-serine, which are highly correlated with CNS levels, have been found to be decreased in schizophrenia by one recent study (Hashimoto et al, 2003). Reduced levels of D-serine, causing a secondary impairment in NMDA receptor function, may be explained by enhanced metabolism by *DAAO*. Support for this hypothesis comes from a gene expression study of the Stanley Consortium brains, which demonstrated alteration in the ratio of D-serine/*DAAO* expression in schizophrenia (Toro et al, 2004).

Other susceptibility genes may also have a regulatory role on NMDA receptors. An *in vitro* study demonstrated that knockdown of presynaptic dysbindin reduces expression of pre-synaptic proteins and glutamate release (Numakawa et al, 2004). Postsynaptic dysbindin is concentrated at the dystrophin-associated protein complex (DPC) and may be involved in tethering receptors including NMDA. Variation in isoforms of *NRG1* may also be involved in regulating the expression and plasticity of NMDA receptors (Stefansson et al, 2002).

Neuropathological, transgenic and in-vitro research will likely be required to confirm and investigate the biological role of these genes both in schizophrenia and normal brain development. That these genes 'fit' in a hypothetical framework may well be because current aetiological hypotheses are broad and reflect the limits of our understanding of schizophrenia. However, the growth of knowledge drives scientific enquiry and will make possible the testing of novel, more specific hypotheses.

4. What is the relationship between genotype and phenotype?

By any account, there is a significant conceptual leap from genomic variants that contribute to risk of the categorical phenotype 'schizophrenia' and the complex behaviours and perceptual experiences the phenotype represents. A 'bottom up' approach of investigating function and interaction of genes must be met by 'top down' clinical and basic scientific studies of the phenotype. Although the current orthodoxy is that schizophrenia is a discrete entity, Bleuler's formulation that there are a group of schizophrenias cannot be excluded as different molecular and neural pathways could contribute to the clinical phenotype. Also plausible is the hypothesis that genetic susceptibility contributes, by a singular or multiple pathways, to a wider

phenotype not confined by existing diagnostic boundaries. As yet there is no evidence for genetic subtypes of schizophrenia, but molecular studies endorse the earlier view of Andreasen & Carpenter (1993), that the schizophrenia phenotype is a 'moving target'. For example, three of the identified schizophrenia susceptibility genes (*G72, NRG1* and *DISC1*) have also demonstrated association with bipolar disorder, adding to evidence from earlier epidemiological studies that there may be psychosis susceptibility genes (reviewed in Craddock et al, 2005). Whether such genes contribute to other psychiatric disorders (e.g. association has been reported between *G72* and panic disorder (Schumacher et al, 2005)), to personality disorders, or to abnormal perceptual experiences in the general population is yet to be elucidated.

The author's group (and others) are exploring the impact of susceptibility alleles/haplotypes on indices of neurocognitive, neurophysiological and functional neuroimaging performance within the schizophrenia population and non-patient groups. Provisional data from this study suggests that the *DTNBP1* risk haplotype is associated with impaired spatial working memory in schizophrenia. In a comparable US study of *DTNBP1* risk variants, working memory deficits and more global impairment in neurocognitive performance were detected in both schizophrenia and unaffected carrier populations. Similarly, several research groups have identified a functional variant at the putative schizophrenia susceptibility gene *COMT* as associated with impaired performance on indices of frontal lobe function in schizophrenia and control populations (Malhotra et al, 2002; Egan et al., 2001). That these alleles are relatively common in the general population and contribute to variance in neuropsychological performance suggests that certain schizophrenia susceptibility genes may have a more general role in regulation of discrete domains of cognitive processing.

From earlier epidemiological studies, the impact of schizophrenia risk genes appears to be altered by exposure to environmental factors, such as a dysfunctional home environment (Tienari et al, 2004). As susceptibility genes are identified it becomes possible to directly investigate for interaction and correlation between genetic and environmental risk. This is not a simple undertaking, as many of the identified environmental risk factors are conceptually broad (e.g. 'urban living') or difficult to measure reliably 'dysfunctional home environment'). Where environmental effects

may be specific to critical developmental stages (e.g. prenatal infection or hypoxia at birth), the quality of information available may fall foul to the typical problems posed by retrospective assessment.

However work by Caspi et al, (2005) indicates the potential promise of this approach: the authors describe a functional variant at the *COMT* gene that moderates the risk posed by adolescent cannabis use in developing schizophrenia. If confirmed, this indicates that it may be possible to identify specific risk groups where it may be possible to modify environmental risk. With the collection of larger clinical and population-based samples, different study designs, and the application of analytical methods to test for epigenetic and gene-environment interactions, the list of genes implicated in schizophrenia susceptibility and our understanding of what they contribute to can only grow.

In completing this thesis, the author has had the opportunity to establish this sample collection at a key point in the evolution of schizophrenia genetics. For this I am grateful to all who have been involved as participants, colleagues, mentors and funding agencies. The experience has opened up new horizons but has also been humbling in how little we know about schizophrenia. I look forward to continuing my participation in the collaborative effort to better understand this devastating disorder.

REFERENCES

Abecasis GR, Cherny SS, Cardon LR. The impact of genotyping error on family-based analysis of quantitative traits. European Journal of Human Genetics 2001; 9(2): 130-4.

Abecasis GR, Cookson WO. GOLD-graphical overview of linkage disequilibrium. Bioinformatics 2000; 16: 182-3.

Andolfatto P, Nordborg M. The effect of gene conversion on intralocus associations. Genetics 1998; 148(3): 1397-9.

Andreasen NC, Carpenter WT. Diagnosis and classification of schizophrenia. Schizophrenia Bulletin 1993; 19(2): 199-214.

Ardlie K, Liu-Cordero SN, Eberle MA, Daly M, Barrett J, Winchester E, Lander ES, Kruglyak L. Lower than expected linkage disequilibrium between tightly linked markers in humans suggests a role for gene conversion. American Journal of Human Genetics 2001; 69: 582-9.

Baare WF, van Oel CJ, Hulshoff Pol HE, Schnack HG, Durston S, Sitskoorn MM, Kahn RS. Volumes of brain structures in twins discordant for schizophrenia. Archives of General Psychiatry 2000; 58: 33-40.

Badner JA, Gershon ES. Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. Molecular Psychiatry 2002; 7(4): 405-11.

Barcellos LF, Klitz W, Field LL, Tobias R, Bowcock AM, Wilson R, Nelson MP, Nagatomi J, Thomson G. Association mapping of disease loci, by use of a pooled DNA genomic screen. American Journal of Human Genetics 1998; 61: 737-747.

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005; 21(2): 263-5.

Bateson W. Mendel's Principles of Heredity. Cambridge University Press, UK, 1909.
Becker T, Knapp M. A powerful strategy to account for multiple testing in the context of haplotype analysis. American Journal of Human Genetics 2004; 75: 561-570.

Becker T, Knapp M. A powerful strategy to account for multiple testing in the context of haplotype analysis. American Journal of Human Genetics 2004; 75: 21-32.

Bleuler E. Dementia Praecox or the group of Schizophrenias. Translated by J Zinkin. New York. International Universities Press, 1911.

Blouin JL, Dombroski BA, Nath SK, Lasseter VK, Wolyniec PS, Nestadt G, Thornquist M, Ullrich G, McGrath J, Kasch L, Lamacz M, Thomas MG, Gehrig C, Radhakrishna U, Snyder SE, Balk KG, Neufeld K, Swartz KL, DeMarchi N, Papadimitriou, GN, Dikeos DG, Stefanis CN, Chakravarti A, Childs B, Housman DE, Kazazian HH, Antonarakis S, Pulver AE. Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. Nature Genetics 1998; 20(1): 70-73.

Bonizzi G, Karin M. The two NF-kappaB activation pathways and their role in innate and adaptive immunity. Trends in Immunology 2004; 25(6):280-8.

Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. Nature Genetics 2003; 228-237.

Bray NJ, Buckland PR, Williams NM, Williams HJ, Norton N, Owen MJ, O'Donovan MC. A haplotype implicated in schizophrenia susceptibility is associated with reduced COMT expression in human brain. American Journal of Human Genetics 2003; 73: 152-161.

Bray NJ, Preece A, Williams NM, Moskvina V, Buckland PR, Owen MJ, O'Donovan MC. Haplotypes at the dystrobrevin binding protein 1 (DTNBP1) gene locus mediate risk for schizophrneia through reduced DTNBP1 expression. Human Molecular Genetics 2005; 14: 1947-54.

Browne S, Clarke M, Gervin M, Lane A, Waddington JL, Larkin C, O'Callaghan E. Determinants of neurological dysfunction in first episode schizophrenia. Psychological Medicine 2000; 30(6): 1433-41.

Brzustowicz LM, Hodgkinson KA, Chow EW, Honer WG, Bassett AS. Location of a major susceptibility locus for familial schizophrenia on chromosome 1q21-q22. Science 2000; 288: 678-682.

Brzustowicz LM, Honer WG, Chow EW, Little D, Hogan J, Hodgkinson K, Bassett AS. Linkage of familial schizophrenia to chromosome 13q32. American Journal of Human Genetics 1999; 65(4): 1096-1103.

Buchanan RW, Heinrichs DW. The Neurological Evaluation Scale (NES): a structured instrument for the assessment of neurological signs in schizophrenia. Psychiatry Research 1989; 27(3): 335-50.

Burden S, Yarden Y. Neuregulins and their receptors: a versatile signalling module in organogenesis and oncogenesis. Neuron 1997; 18: 847-855.

Burmeister M. Basic concepts in the study of diseases with complex genetics. Biological Psychiatry 1999; 70: 418-430.

Cannon M, Caspi A, Moffit TE, Harrington H, Taylor A, Murray RM, Poulton R. Evidence for early, specific, pan-developmental impairment in schizophreniform disorder: results from a longitudinal birth cohort. Archives of General Psychiatry 2002; 59: 449-457.

Cannon M, Kendell R, Susser E, Jones P. Prenatal and perinatal risk factors for schizophrenia. In: The Epidemiology of Schizophrenia. Murray RM, Jones PB, Susser E, van Os J, Cannon M, eds. University Press, Cambridge, UK, 2003.

Cannon TD, Caprio J, Lonnqvist J, Huttenen M, Koskenvuo M. The genetic epidemiology of schizophrenia in a Finnish twin cohort. A population-based modeling study. Archives of General Psychiatry 1998; 55: 67-74.

Cannon TD, Mednick SA. The schizophrenia High Risk Study of Copenhagen: three decades of progress. Acta Psychiatrica Scandinavica 1993; 87: 33-47.

Cardno AG, Gottesman II. Twin studies of schizophrenia: from bow-and-arrow concordances to Star Wars Mx and functional genomics. American Journal of Medical Genetics 2000; 97:12-17.

Cardno AG, Marshall EJ, Coid B, Macdonald AM, Ribchester TR, Davies NJ, Venturi P, Jones LA, Lewis SW, Sham PC, Gottesman II, Farmer AE, McGuffin P, Reveley AM, Murray RM. Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. Archives of General Psychiatry 1999; 56:162-168.

Cardon LR, Palmer LJ. Population stratification and spurious allelic association. Lancet 2003; 361: 598-604.

Carlsson A, Lindqvist M. Effect of chlorpromazine or haloperidol on formation of 3-methyoxytyramine and normetanephrine in mouse brain. Acta Pharmacologica Toxicologica 1963; 20, 140-144.

Carmi R, Rokhlina T, Kwitek-Black AE, Elbedour K, Nishmura D, Stone EM, Sheffield VC. Use of a DNA pooling strategy to identify a human obesity syndrome locus on chromosome 15. Human Molecular Genetics 1995; 4, 9-13.

Caspi A, Moffitt TE, Cannon M, McClay J, Murray R, Harrington H, Taylor A, Arsenault L, Williams B, Braithwaite A, Poulton R, Craig IW. Moderation of the effect of adolescent-onset cannabis use on adult psychosis by a functional polymorphism in the catechol-O-methyltransferase gene: longitudinal evidence of a gene X environment interaction. Biological Psychiatry 2005; 57: 1117-27.

Castle WE. Genetics and Eugenics: A Textbook for Students of Biology and a Reference Book for Animals and Plant Breeders. Cambridge, MA: Harvard University Press, 1916.

Chakraborty R, Weiss KM. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proceedings of the National Academy of Sciences U S A 1988; 85(23): 9119-23.

Chen X, Dunham C, Kendler S, Wang X, O'Neill FA, Walsh D, Kendler KS. Regulator of G-protein signaling 4 (RGS4) gene is associated with schizophrenia in Irish high density families. American Journal of Medical Genetics. Part B Neuropsychiatric Genetics 2004; 129(1):23-6.

Chen X, Levine L, Kwok, PY. Fluorescence polarization in homogeneous nucleic acid analysis. Genome Research 1999: 9: 492-498.

Cheng 2005

Chowdari KV, Mirnics K, Semwal P, Wood J, Lawrence E, Bhatia T, Deshpande SN, B K T, Ferrell RE, Middleton FA, Devlin B, Levitt P, Lewis DA, Nimgaonkar VL. Association and linkage analyses of RGS4 polymorphisms in schizophrenia. Human Molecular Genetics 2002; 11(12):1373-80.

Chumakov I, Blumenfeld M, Guerassimenko O, Caverec L, Palicio M, Abderrahim H, Bougueleret L, Barry C, Tanaka H, La Rosa P, Puech A, Tahri N, Cohen-Akenine A, Delabrosse S, Lissarrague S, Picard FP, Maurice K, Essioux L, Millasseau P, Grel P, Debailleul V, Simon AM, Caterina D, Dufaure I, Malekzadeh K, Belova M, Luan JJ, Bouillot M, Sambucy JL, Primas G, Saumier M, Boubkiri N, Martin-Suamier S, Nasroune M, Peixoto H, Delaye A, Pinchot V, Bastucci M, Guillou S, Chevillon M, Saintz-Fuertes R, Meguenni S, Aurich-Costa J, Cherif D, Gimalac A, Van Duijn C, Gauvereau D, Ouellette G, Fortier I, Raelson J, Sherbatich T, Riazanskaia N, Rogaev E, Raeymaekers P, Aerssens J, Konings F, Luyten W, Macciardi F, Sham PC, Straub RE, Weinberger DR, Cohen N, Cohen D. Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. Proceedings of the National Academy of Sciences, U S A. 2002; 15; 99(21): 13675-80.

Clark AG. Inference of haplotypes from PCR-amplified samples of diploid populations. Molecular Biology and Evolution 1990; 7:111-122.

Clouston TS. The Neuroses of Development. Oliver & Bond: Edinburgh, 1891.

Coffey CS, Hebert PR, Ritchie MD, Krumholz HM, Morgan TM, Gaziano JM, Ridker PM, Moore JH. An application of conditional logistic regression and multifactor dimensionality reduction for detecting gene-gene interactions on risk of myocardial infarction: the importance of model validation. BMC Bioinformatics 2004; 5: 49.

Coon H, Jensen S, Holik J, Hoff M, Myles-Worsley M, Reimherr F, Wender P, Waldo M, Freedman R, Leppert M, Byerley W. Genomic scan for genes predisposing to schizophrenia. American Journal of Medical Genetics 1994; 54(1):59-71.

Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Human Molecular Genetics 2002; 11: 2463-2468.

Corvin A, Morris DW, McGhee KA, Siobhan Schwaiger, Paul Scully, John Quinn, David Meagher, John L. Waddington, Michael Gill. Confirmation and refinement of an 'at risk' haplotype for SZ suggests the EST cluster, *Hs.*97362, as a potential susceptibility gene in addition to Neuregulin 1. Molecular Psychiatry; 9(2):208-13.

Coyle JT & Tsai G. The NMDA receptor glycine modulatory site: a therapeutic target for improving cognition and reducing negative symptoms in schizophrenia. Psychopharmacology 2004; 174(1):32-8.

Craddock N, O'Donovan MC, Owen MJ. The genetics of schizophrenia and bipolar disorder: dissecting phenotypes. Journal of Medical Genetics 2005; 42: 193-204.

Crocq MA, Mant R, Asherson P, Williams J, Hode Y, Mayerova A, Collier D, Lannfelt L, Sokoloff P, Schwartz JC, et al. Association between schizophrenia and homozygosity at the dopamine D3 receptor gene. Journal of Med Genetics 1992; 29(12): 858-60.

Crow T. Genes for schizophrenia. Lancet 2003; 361(9371):1829.

Cruickshanks KJ, Vadheim CM, Moss SE, Roth MP, Riley WJ, Maclaren NK, Langfield D, Sparkes RS, Klein R, Rotter JI. Genetic marker associations with proliferative retinopathy in persons diagnosed with diabetes before 30 yr of age. Diabetes 1992; 41: 879-85.

Culverhouse R, Klein T, Shannon W. Detecting epistatic interactions contributing to quantitative traits. Genetic Epidemiology 2004; 27:141-52.

Czika WA, Weir BS, Edwards SR, Thompson RW, Nielsen DM, Brocklebank JC, Zinkus C, Martin ER, Hobler KE. Applying data mining techniques to the mapping of complex disease genes. Genetic Epidemiology 2001; 21: S435-S440.

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. Nature Genetics 2001; 29:229-232.

Davis KL, Kahn RS, Ko G, Davidson M. Dopamine in schizophrenia: a review and reconceptualization. American Journal of Psychiatry 1991; 148: 1474-1486.

Dawson E, Abecasis GR, Bumpstead S, Chen Y, Hunt S, Beare DM, Pabial J, Dibling T, Tinsley E, Kirby S, Carter D, Papasypyridonos M, Livingstone S, Ganske R, Lohmussaar E, Zernant J, Tonisson N, Remm M, Magi R, Puurand T, Vilo J, Kurg A, Rice K, Deloukas P, Mott R, Metspalu A, Bentley DR, Cardon LR, Dunham I. A first generation linkage disequilibrium map of human chromosome 22. Nature 2002; 418: 544-548.

DeLisi LE, Shaw SH, Crow TJ, Shields G, Smith AB, Larach VW, Wellman N, Loftus J, Nanthakumar B, Razi K, Stewart J, Comazzi M, Vita A, Heffner T, Sherrington R. A genome-wide scan for linkage to chromosomal regions in 382 sibling pairs with schizophrenia or schizoaffective disorder. American Journal of Psychiatry 2002; 159(5):803-12.

Devin B, Roeder K. Genomic control for association studies. Biometrics 1999; 55: 997-1004.

Diagnostic and Statistical Manual of Mental Disorders. Fourth Edition. Text Revision. (DSM-IV-TR). American Psychiatric Association 2000, Washington DC.

Duchateau PN, Pullinger CR, Orellana RE, Kunitake ST, Naya-Vigne J, O'Connor PM, Malloy MJ, Kane JP. Apolipoprotein L, a new human high density lipoprotein apolipoprotein expressed by the pancreas. Identification, cloning, characterization, and plasma distribution of apolipoprotein L. Journal of Biological Chemistry 1997; 272(41):25576-82.

Egan MF, Goldberg TE, Kolachana BS, Callicott JH, Mazzanti CM, Straub RE, Goldman D, Weinberger DR. Effect of COMT Val 108/158 Met genotype on frontal lobe function and risk of schizophrenia. Proceedings of the National Academy of Sciences USA 2001; 98: 6917-22.

Erlenmeyer-Kimling L, Rock D, Roberts SA, Janal M, Kestenbaum C, Cornblatt B, Adamo UH, Gottesman II. Attention, memory and motor skills as childhood predictors of schizophrenia-related psychoses: the New York High Risk Project. American Journal of Psychiatry 2000; 157: 1416-1422.

Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Molecular Biology and Evolution 1995; 12:921-927.

Falls DL. Neuregulins: functions, forms, and signalling strategies. Experimental Cell Research 2003; 284: 14-30.

Feinberg I. Schizophrenia: caused by a fault in programmed synaptic elimination during adolescence? Journal of Psychiatric Research 1982-83;17(4):319-34.

Fischer M. Genetic and environmental factors in schizophrenia. Acta Psychiatrica Scandinavia 1973; 238: 1-158.

Fish B, Marcus J, Hans SL, Auerbach JG, Perdue S. Infants at risk of schizophrenia: sequelae of a genetic neurointegrative defect. A review and replication analysis of pandysmaturation in the Jerusalem Infant Development study. Archives of General Psychiatry 1992; 49: 221-235.

Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. Transcripts of The Royal Society Edinburgh 1918; 52: 399-433.

Flashman LA, Flaum M, Gupta S, Andreasen NC. Soft signs and neuropsychological performance in schizophrenia. American Journal of Psychiatry 1996; 153(4):526-32.

Freitas 2001

Frisse L, Hudson RR, Bartoszewicz A, Wall JD, Donfack J, Di Rienzo A. Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. American Journal of Human Genetics 2001; 69(4): 831-43.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D. The structure of haplotype blocks in the human genome. Science 2002; 296(5576): 2225-2229.

Gervin M, Browne S, Lane A, Clarke M, Waddington JL, Larkin C, O'Callaghan E. Spontaneous abnormal involuntary movements in first-episode schizophrenia and schizophreniform disorder: baseline rate in a group of patients from an Irish catchment area. American Journal of Psychiatry 1998;155(9):1202-6.

Gill M, Vallada H, Collier D, Sham P, Holmans P, Murray R, McGuffin P, Nanko S, Owen M, Antonarakis S, Housman D, Kazazian H, Nestadt G, Pulver AE, Straub RE, MacLean CJ, Walsh D, Kendler KS, DeLisi L, Polymeropoulos M, Coon H, Byerley W, Lofthouse R, Gershon E, Golden L, Crow T, Freedman R, Laurent C, BodeauPean S, d'Amato T, Jay M, Campion D, Mallet J, Wildenauer DB, Lerer B, Albus M, Acken heil M, Ebstein RP, Hallmayer J, Maier W, Gurling H, Curtis D, Kalsi G, Brynjolfsson J, Sigmundson T, Petursson H,m Blackwood D, Muir W, St Clair D, He L, Maguire S, Moises HW, Hwu HG, Yang L, Wiese C, Tao L, Liu XH, Kristbjarnason H, Levinson DF, Mowry BJ, DonisKeller H, Hayward NK, Crowe RR< Silverman JM, Nancarrow DJ, Read CM. A combined analysis of D22S278 marker alleles in affected sib-pairs: support for a susceptibility locus for SZ at chromosome 22q12. SZ Collaborative Linkage Group (Chromosome 22). American Journal of Medical Genetics. Part B: Neuropsychiatric Genetics. 16; 67(1):40-5.

Goff DC & Coyle JT. The emerging role of glutamate in the pathophysiology and treatment of schizophrenia. American Journal of Psychiatry 2001; 158: 1367-1377.

Goring HH, Terwilliger JD, Blangero J. Large upward bias in estimation of locus-specific effects from genomewide scans. American Journal of Human Genetics 2001; 69: 1357-1369.

Gottesman II & Bertelsen A. Confirming unexpressed genotypes for schizophrenia. Archives of General Psychiatry 1989; 46: 867-872.

Gottesman II & Shields J. A polygenic theory of schizophrenia. Proceedings of the National Academy of Sciences, USA 1967; 58: 199-205.

Gottesman II. Schizophrenia Genesis: The Origins of Madness. Freeman, New York, 1991.

Green EK, Raybould R, Macgregor S, Gordon-Smith K, Heron J, Hyde S, Grozeva D, Hamshere M, Williams N, Owen MJ, O'Donovan MC, Jones L, Jones I, Kirov G, Craddock N. Operation of the schizophrenia susceptibility gene, neuregulin 1, across traditional diagnostic boundaries to increase risk for bipolar disorder. Archives of General Psychiatry 2005; 62(6): 642-8.

Gretarsdottir S, Sveinbjornsdottir S, Jonsson HH, Jakobsson F, Einearsdottir E, Agnarsson U, Shkolny D, Einarsson G, Gudjonsdottir HM, Valdimarsson EM, EInarsson OB, Thorgeirsson G, HAdzic R, Jonsdottir S, Reynisdottir ST< Bjarnsdottir SM, Gudmundsdottir T, Gullaugsdottir GJ, Gill R, Lindpaintner K, Sainz J, Hannesson HH, Sugurdsson GT, Frigge ML, Kong A, Gudnason V, Stefansson K, Gulcher JR. Localization of a susceptibility gene for common forms of stroke to 5q12. American Journal of Human Genetics 2002; 70(3):593-603.

Gross C, Zhuang X, Stark K, Ramboz S, Oosting R, Kriby L, Santarelli L, Beck S, Hen R. Serotonin 1A receptor acts during development to establish normal anxiety-like behaviours in the adult. Nature 2002; 416: 396-400.

Gudbjartsson DF, Jonasson K, Frigge ML, Kong A. Allegro, a new computer program for multipoint linkage analysis. Nature Genetics 2000; 25(1):12-3.

Gurling H, Kalsi G, Brynjolfson J, Sigmundsson T, Sherrington R, Mankoo BS, Read T, Murphy P, Blaveri E, McQuillin A, Petursson H, Curtis D. Genomedwide genetic linkage analysis confirms the presence of susceptibility loci for schizophrenia, on chromosomes 1q32.2, 5q33.2, and 8p21-22 and provides support for linkage to schizophrenia, on chromosome 11q23.3-24 and 20q12.1-11.23. American Journal of Human Genetics 2001; 68: 661-673.

Guy W. ECDEU Assessment Manual for Psychopharmacology. Washington DC: Department of Health, Education and Welfare, 1976.

Hakak Y, Walker JR, Li C, Wong WH, Davis KL, Buxbaum JD, Haroutunian V, Fienberg AA.Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia. Proceedings of the National Academy of Sciences, U S A. 2001; 98(8):4746-51.

Hanson DR, Gottesman II, Heston LL. Some possible childhood indicators of adult schizophrenia inferred from children of schizophrenics. British Journal of Psychiatry 1976; 129: 142-154.

Harrison PJ, Lewis DA. Neuropathology of schizophrenia. In: Schizophrenia, II, Biological Aspects. Eds: Hirsch SR & Weinberger DA, pp 310-325. Blackwell Publishing, Oxford, UK, 2002.

Harrison PJ, Weinberger DR. Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. Molecular Psychiatry 2005; 10: 40-68.

Hays P. The causation of schizophrenia. In: New Horizons in Psychiatry. Cox & Wyman: London, 1967.

Hedrick PW. Gametic disequilibrium measures: proceed with caution. Genetics 1987; 117: 331-341.

Helgason A, Hickey E, Goodacre S, Bosnes V, Stefansson K, Ward R, Sykes B. mtDna and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. American Journal of Human Genetics 2001; 68(3): 723-737

Helgason A, Yngvadottir B, Hrafnkelsson B, Gulcher J, Stefansson K. An Icelandic example of the impact of population structure on association studies. Nature Genetics 2005; 37: 90-5.

Hennessy RJ, Lane A, Kinsella A, Larkin C, O'Callaghan E, Waddington JL. 3D morphometrics of craniofacial dysmorphology reveals sex-specific asymmetries in schizophrenia. Schizophrenia Research 2004; 67(2-3):261-8.

Heston LL. Psychiatric disorders in foster home reared children of schizophrenic mothers. British Journal of Psychiatry 1966; 112: 819-825.

Hill EW, Jobling MA, Bradley DG. Y-chromosome variation and Irish origins. Nature 2000; 404: 351-2.

Hoh J, Ott J. Mathematical multi-locus approaches to localizing complex human trait genes. Nature Reviews Genetics 2003; 4: 701-709.

Holmans P, Craddock N. Efficient strategies for genome scanning using maximum-likelihood affected-sib-pair analysis. American Journal of Human Genetics 1997; 60: 657-66.

Hoogendoorn B, Norton N, Kirov G, Williams N, Hamshere ML, Spurlock G, Austin J, Stephens MK, Buckland PR, Owen MJ, O'Donovan MC.Cheap, accurate and rapid allele frequency estimation of single nucleotide polymorphisms by primer extension and DHPLC in DNA pools. Human Genetics 2000; 107(5):488-93.

Horrevoets AJ, Fontijn RD, van Zonneveld AJ, de Vries CJ, ten Cate JW, Pannekoek H. Vascular endothelial genes that are responsive to tumor necrosis factor-alpha in vitro are expressed in atherosclerotic lesions, including inhibitor of apoptosis protein-1, stannin, and two novel genes. Blood. 1999; 93(10):3418-31

Iafrate AJ, Feuk L, Rivera MN, Listwnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. Nature Genetics 2004; 36: 949-51.

ICD-10: The ICD-10 classification of mental and behavioural disorders. World Health Organization 1992: Geneva.

International Human Genome Sequencing Consortium. A physical map of the human genome. Nature 2001; 409: 931-941.

Ismail B, Cantor-Graae E, McNeil TF. Minor physical anomalies in schizophrenic patients and their siblings. American Journal of Psychiatry 1998; 155:1695-1702.

Ismail B, Cantor-Graae E, McNeil TF. Neurological abnormalities in schizophrenic patients and their siblings. American Journal of Psychiatry 1998; 155:84-89.

Javitt DC & Zukin SR. Recent advances in the phencyclidine model of schizophrenia. American Journal of Psychiatry 1991; 148: 1301-1308.

Jeffreys AJ, Kauppi L, Neumann R. Intense punctate meiotic recombination in the class II region of the major histocompatibility complex. Nature Genetics 2001; 29: 217-222.

Jeffreys AJ, May CA. Intense and highly localized gene conversion activity in human meiotic crossover hot spots. Nature Genetics 2004; 36(2): 151-6.

Jeffreys AJ, Ritchie A, Neumann R. High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. Human Molecular Genetics 2000; 9: 725-733.

Johnstone EC, Cooling NJ, Frith CD, Crow TJ & Owens DG. Phenomenology of organic and functional psychoses and the overlap between them. British Journal of Psychiatry 1988; 153: 770-776.

Jones PB. Longitudinal approaches to the search for the causes of schizophrenia: past, present and future. In : Search for the causes of schizophrenia, Vol. IV Balance of the century, Gatta WF, Haffner H eds. Darmstadt, Steinkopf and Berlin: Springer, 1999.

Kaufmann CA, Suarez B, Malaspina D, Pepple J, Svrakic D, Markel PD, Meyer J, Zambuto CT, Schmitt K, Matise TC, Harkavy, Friedman JM, Hampe C, Lee H, Shore D, Wynne D, Faraone SV, Tsuang MT, Cloninger CR. NIMH Genetics Initiative Millennium Schizophrenia Consortium: linkage analysis of African-American pedigrees. American Journal of Medical Genetics 1998; 81(4):282-289

Kauppi L, Jeffreys AJ, Keeney S. Where the crossovers are: recombination distributions in mammals. Nature Reviews Genetics 2004; 5: 413-424.

Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophrenia Bulletin 1987; 13: 261-76.

Kendler KS, Diehl SR. The genetics of schizophrenia: a current, genetic-epidemiologic perspective. Schizophrenia Bulletin 1993; 19(2): 261-85.

Kendler KS, Gruenberg AM, Kinney DK. Independent diagnoses of adoptees and relatives as defined by DSM-III in the provincial and national samples of the Danish Adoption Study of Schizophrenia. Archives of General Psychiatry 1994; 51(6): 456-68.

Kendler KS, Karkowski-Shuman L, O'Neill FA, Straub RE, MacLean CJ, Walsh D. Resemblance of psychotic symptoms and syndromes in affected sibling pairs from the Irish Family Study of high-density schizophrenia families: evidence for possible etiologic heterogeneity. American Journal of Psychiatry 1997; 154: 191-198.

Kendler KS, MacLean CJ, Ma Y, O'Neill FA, Walsh D, Straub RE. Marker-to-marker linkage disequilibrium on chromosomes 5q, 6p, and 8p in Irish high-density schizophrenia pedigrees. American Journal of Medical Genetics: Neuropsychiatric Genetics 1999:5; 88(1): 29-33

Kendler KS, MacLean CJ, O'Neill FA, Burke J, Murphy B, Duke F, Shinkwin R, Easter SM, Webb BT, Zhang J, Walsh D, Straub RE. Evidence for a schizophrenia vulnerability locus on chromosome 8p in the Irish Study of High-Density Schizophrenia Families. American Journal of Psychiatry 1996; 153(12): 1534-1540.

Kendler KS, McGuire M, Gruenberg AM, Walsh D. Clinical heterogeneity in schizophrenia and the pattern of psychopathology in relatives: results from an epidemiologically based family study. Acta Psychiatrica Scandinavia 1994; 89: 294-300.

Kendler KS, Myers JM, O'Neill FA, Martin R, Murphy B, MacLean CJ, Walsh D, Straub RE. Clinical features of schizophrenia and linkage to chromosomes 5q, 6p, 8p, and 10p in the Irish Study of High-Density Schizophrenia Families. American Journal of Psychiatry 2000; 157: 402-408.

Kendler KS. Schizophrenia genetics and dysbindin: a corner turned? American Journal of Psychiatry 2004; 161(9): 1533-6.

Keshaven MS. Development, disease and degeneration in schizophrenia: a unitary pathophysiological model. Journal of Psychiatric Research 1999; 33:513-521

Kety SS, Wender PH, Jacobsen B, Ingraham LJ, Jansson L, Faber B, Kinney DK. Mental illness in the biological and adoptive relatives of schizophrenic adoptees. Replication of the Copenhagen Study in the rest of Denmark. Archives of General Psychiatry 1994; 51(6): 442-55.

Kety SS. Schizophrenic illness in the families of schizophrenic adoptees: findings from the Danish national sample. Schizophrenia Bulletin 1988; 14(2): 217-22.

Kimble GA. Evolution of the nature-nurture issue in the history of psychology. In: Plomin R, McClearn GE Nature Nurture and Psychology. Washington DC: American Psychological Association, 1993.

Kleinjan DA, van Heyningen V. Long range control of Gene Expression: Emerging mechanisms and disruption in disease. American Journal of Human Genetics 2005; 76: 8-32.

Knowler WC, Williams RC, Pettitt DJ, Steinberg AG. Gm3; 5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. American Journal of Human Genetics 1988; 43: 520-26.

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K. A high-resolution recombination map of the human genome. Nature Genetics 2002; 31: 241-7.

Kooperberg C & Ruczinski I. Identifying interacting SNPs using Monte Carlo Logistic Regression. Genetic Epidemiology 2005; 28: 157-170.

Kraepelin E (1919). Dementia praecox and paraphrenia (trans R M Barclay). Facsimile edition, Kreiger, New York, 1971.

Krakow D, Robertson SP, King LM, Sebald ET, Bertolotto C, Wachsmann-Hogiu S, Acuna D, Shapiro SS, Takafuta T, Aftimos S, Kim CA, Firth H, Steiner CE, et al. Mutations in the gene encoding filamin B disrupt vertebral segmentation, joint formation and skeletogenesis. Nature Genetics 2004; 36(4):405-10.

Kringlen E & Cramer G. Offspring of monozygotic twins discordant for schizophrenia. Archives of General Psychiatry 1989: 46: 873-877.

Lander ES, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nature Genetics 1995; 11: 241-7.

Lander ES. The new genomics: global views of biology. Science 1996: 274; 536-539.

Lane A, Kinsella A, Murphy P, Byrne M, Keenan J, Colgan K, Cassidy B, Sheppard N, Horgan R, Waddington JL, Larkin C, O'Callaghan E. The anthropometric assessment of dysmorphic features in schizophrenia as an index of its developmental origins. Psychological Medicine 1997; 27: 1155-1164.

Laruelle M. Dopamine transmission in the schizophrenic brain. In: Schizophrenia. Ed: Hirsh S & Weinberger DR. Blackwell Publishing. Oxford, UK, 2003.

Levin ML. The occurrence of lung cancer in man. Acta Unio Int Contra Cancer 1953; 9:531-541.

Levinson DF, Lewis CM, Wise LH. Meta-analysis of genome scans for schizophrenia. American Journal of Medical Genetics: Neuropsychiatric Genetics 2002; 114: 700.

Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I, Williams NM, Schwab SG, Pulver AE, Faraone SV, Brzustowicz LM, Kaufmann CA, Garver DL, Gurling HM, Lindholm E, Coon H, Moises HW, Byerley W, Shaw SH, Mesen A, Sherrington R, O'Neill FA, Walsh D, Kendler KS, Ekelund J, Paunio T, Lonnqvist J, Peltonen L, O'Donovan MC, Owen MJ, Wildenauer DB, Maier W, Nestadt G, Blouin JL, Antonarakis SE, Mowry BJ, Silverman JM, Crowe RR, Cloninger CR, Tsuang MT, Malaspina D, Harkavy-Friedman JM, Svrakic DM, Bassett AS, Holcomb J, Kalsi G, McQuillin A, Brynjolfson J, Sigmundsson T, Petursson H, Jazin E, Zoega T, Helgason T. Genome scan meta-analysis of SZ and bipolar disorder, part II: SZ. American Journal of Human Genetics 2003; 73(1): 34-48.

Lewis SW & Murray RM. Obstetric complications, neurodevelopmental deviance, and risk of schizophrenia. Journal of Psychiatric Research 1987; 21(4):413-21.

Lewis SW. Computed tomography in schizophrenia fifteen years on. British Journal of Psychiatry 1990; 157:16-24.

Lewontin RC. The interaction of selection and linkage I General considerations: heterotic models. Genetics 1964; 49: 49-67.

Lin DY. An efficient Monte Carlo approach to assessing statistical significance in genomic studies. Bioinformatics 2005; 21: 781-7.

Lindholm E, Ekbolm B, Shaw S, Jalonen P, Johansson G, Pettersson U, Sherrington, R, Adolfsson R, Jazin E. A schizophrenia-susceptibility locus at 6q25, in one of the world's largest reported pedigrees. American Journal of Human Genetics 2001; 69(1): 96-105.

Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nature Genetics 2003; 33(2): 177-82.

Luby E, Cohen B, Rosenbaum G, Gottlieb J, Kelley R. Study of a new schizophrenomimetic drug: seryl. American Medical Associations Archives of Neurology and Psychiatry 1959; 81: 363-369.

Malaspina D, Goetz RR, Yale S, Berman A, Friedman JH, Tremeau F et al. Relation of familial schizophrenia to negative symptoms but not to the deficit syndrome. American Journal of Psychiatry 2000; 157: 994-1003.

Malhotra AK, Kestler LJ, Mazzanti C, Bates JA, Goldberg T, Goldman D. A functional polymorphism in the COMT gene and performance on a test of prefrontal cognition. American Journal of Psychiatry 2002; 159(4): 652-4.

Maniatis N, Collins A, Xu CF, McCarthy LC, Hewett DR, Tapper W, Ennis S, Ke X, Morton NE. The first linkage disequilibrium (LD) maps: delineation of hot and cold blocks by diplotype analysis. Proceedings of the National Academy of Sciences, U.S.A. 2002: 19; 99(4): 2228-33.

Marchini J, Donnelly P, Cardon LR. Genome-wide strategies for detecting multiple loci that influence complex diseases. Nature Genetics 2005; 37(4): 413-7.

Matsumoto M, Weinberger DR, Straub RE. Molecular cloning, sequencing, and characterization of a novel 500 kilobase gene (MRDS1) from 6p24, a schizophrenia candidate region. American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics 2002; 114: 288.

Maziade M, Bissonette L, Rouillard E, Martinez M, Turgeon M, Charron L, Pouliot V, Boutin P, Cliché D, Dion C, Fournier JP. 6p24-22 region and major psychoses in the Eastern Quebec population. Le Groupe IREP. American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics 1997; 74(3):311-8.

McEvoy B, Richards M, Forster P, Bradley DG. The Longue Duree of genetic ancestry: multiple genetic marker systems and Celtic origins on the Atlantic facade of Europe. American Journal of Human Genetics 2004; 75(4): 693-702.

McGuffin P, Owen MJ, Farmer AE. The genetic basis of schizophrenia. Lancet 1995; 346: 678-682.

McGhee KA, Morris DW, Schwaiger S, Nangle JM, Donohoe G, Clarke S, Meagher D, Quinn J, Scully P, Waddington JL, Gill M, Corvin A. Investigation of the apolipoprotein-L (APOL) gene family and schizophrenia using a novel DNA pooling strategy for public database SNPs. Schizophrenia Research 2005; 76: 231-238.

McGhee KA, Morris DW, Murphy K, Schwaiger S, Nangle J, Donohoe G, Clarke S, Scully P, Quinn J, Meagher D, Baldwin P, Waddington JL, Gill M, Corvin A. Positive association of G72/G30 and DAAO with schizophrenia in an Irish case-control

sample. American Journal of Medical Genetics, Part B; Neuropsychiatric Genetics, 2004; 130B: 138.

McInnes MG, Dick DM, MacKinnon DF, McMahon FJ, Potash JB, Nurnberger JI, Gershon ES, Edenberg HJ, Reich T, Rice JP, Simpson SG, Willour VL, DePaulo JR, Zandi PP, Foroud T. Genome-wide scan of 153 bipolar multiplex pedigrees ascertained through the NIMH initiative for Bipolar Disorder. American Journal of Human Genetics 2002; 71: 216.

Meng Z, Zaykin DV, Xu CF, Wagner M, Ehm MG. Selection of genetic markers for association analyses, using linkage disequilibrium and haplotypes. American Journal of Human Genetics 2003; 73(1): 115-30.

Mimmack ML, Ryan M, Baba H, Navarro-Ruiz J, Iritani S, Faull RL, McKenna PJ, Jones PB, Arai H, Starkey M, Emson PC, Bahn S. Gene expression analysis in SZ: reproducible up-regulation of several members of the apolipoprotein L family located in a high-susceptibility locus for SZ on chromosome 22. Proceedings of the National Academy of Sciences U S A 2002; 99(7): 4680-5.

Mirnics K. Functional genomics of schizophrenia: Alterations of specific pathways. European Neuropsychopharmacology 2003; 13: S110-S111.

Mirsky AF, Kugelmass S, Ingraham LJ, Frenkel E, Nathan M. Overview and summary: twenty-five-year followup of high-risk children. Schizophrenia Bulletin 1995; 21(2): 227-39.

Moghaddam B, Krystal JH. The Neurochemistry of Schizophrenia. In; Schizophrenia. Ed: Hirsh S & Weinberger DR. Blackwell Publishing. Oxford, UK, 2003.

Moises HW, Yang L, Kristbjarnarson H, Wiese C, Byerley W, Macciardi F, Arolt V, Blackwood D, Liu X, Sjorgren B et al. An international two-stage genome-wide search for schizophrenia susceptibility genes. Nature Genetics 1995; 11(3):321-4

Monajemi H, Fontijn RD, Pannekoek H, Horrevoets AJ. The apolipoprotein L gene cluster has emerged recently in evolution and is expressed in human vascular tissue. Genomics 2002; 79(4): 539-46.

Moore JH. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. Human Heredity 2003; 56: 73-82.

Moore JH, Williams SM. New strategies for identifying gene-gene interactions in hypertension. Annals of Medicine 2002; 34: 88-95.

Morris DW, Rodgers A, McGhee K, Schwaiger S, Scully P, Quinn J, Meagher D, Waddington JL, Gill M and Corvin A. Confirming RGS-4 as a susceptibility gene for schizophrenia. American Journal of Medical Genetics: Neuropsychiatric Genetics 2004; 125B: 50-3.

Morton NE, Collins A. Tests and estimates of allelic association in complex inheritance. Proceedings of the National Academy of Sciences USA 1998; 95: 11389-93.

Mowry BJ, Ewen KR, Nancarrow DJ, Lennon DP, Nertney DA, Jones HL, O'Brien MS, Thornley CE, Silverman JM, Endicott J, Sharpe L, Hayward NK, Gladis MM, Foote SJ, Levinson DF. Second stage of a genome scan of schizophrenia: study of five positive regions in an expanded sample. American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics 2000; 96(6): 864-9.

Mueller JC, Lohmussaar E, Magi R, Remm M, Bettecken T, Lichtner P, Biskup S, Illig T, Pfeufer A, Luedemann J, Schreiber S, Pramstaller P, Pichler I, Gaddi A, Testa A, Wichmann HE, Metspalu A, Meitinger T. Linkage disequilibrium patterns and tagSNP transferability among European populations. American Journal of Human Genetics 2005; 76(3): 387-98.

Murphy KC, Owen MJ. Minor physical anomalies and their relationship to the aetiology of schizophrenia. British Journal of Psychiatry 1996; 168:139-142.

Murphy KC. SZ and velo-cardio-facial syndrome. Lancet 2002; 359(9304): 426-30.

Murtagh A, McTigue O, Hegarty AM, Stallings RL, Green AJ, Ramsay L, Corvin A. Interstitial deletion of chromosome 21 and schizophrenia. Schizophrenia Research (in press).

Nachman MW. Single nucleotide polymorphisms and recombination rate in humans. Trends in Genetics 2001; 17(9): 481-5.

Neale BM, Sham PC. The future of association studies: gene-based analysis and replication. American Journal of Human Genetics 2004; 75(3): 353-62.

Nelson MR, Kardia SLR, Ferrell RE, Sing CF. A combinatorial repartitioning method to identify multilocus genotype partitions that predict quantitative trait variation. Genome Research 2001; 11: 458-470.

Niethammer R, Weisbrod M, Schiesser S, Grothe J, Maier S, Peter U, Kaufmann C, Schroder J, Sauer H. Genetic influence on laterality in schizophrenia? A twin study of neurological soft signs. American Journal of Psychiatry 2000; 157(2): 272-4.

Niu T, Qin ZS, Xu X, Liu JS. Bayesian haplotype inference for multiple linkage single-nucleotide polymorphisms. American Journal of Human Genetics 2002; 70:159-169.

Norton N, Moskvina V, Morris DW, Bray NJ, Zammit S, Williams NM, Williams HJ, Preece AC, Dwyer SL, Wilkinson JC, Kirov G, Buckland P, Waddington JL, Gill M, Corvin A, Owen MJ, O'Donovan MC. Interaction between neuregulin 1 and its receptor ERBB4 increases susceptibility to schizophrenia. American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics 2004; 130:18.

Norton N, Williams NM, Williams HJ, Spurlock G, Kirov G, Morris DW, Hoogendoorn B, Owen MJ, O'Donovan MC. Universal, robust, highly quantitative SNP allele frequency measurement in DNA pools. Human Genetics 2002; 110(5): 471-8.

Nuamkawa T, Yagasaki Y, Ishimoto T, Okada T, Suzuki T, Iwata N, Ozaki N, Taguchi T, Tatsumi M, Kamijima K, Straub RE, Weinberger DR, Kunugi H, Hashimoto R. Evidence of novel neuronal functions of dysbindin, a susceptibility gene for schizophrenia. Human Molecular Genetics 2004; 13(21): 2699-708.

O'Donovan MC, Williams NM, Owen MJ. Recent advances in the genetics of SZ. Human Molecular Genetics 2003; 12 Spec No 2:R125-33.

Owen MJ, Williams NM, O'Donovan MC. The molecular genetics of schizophrenia: new findings promise new insights. Molecular Psychiatry 2004; 9(1): 14-27.

Owen MJ. Molecular genetic studies of SZ. Brain Research Reviews 2000; (2-3): 179-86.

Ozaki S, Sasner M, Yano R, Lu HS, Buonanno A. Neuregulin-beta induces expression of an NMDA-receptor subunit. Nature 1997; 18-25; 390(6661): 691-4.

Padhukasahasram B, Marjoram P, Nordborg M. Estimating the rate of gene conversion on human chromosome 21. American Journal of Human Genetics 2004; 75(3): 386-97.

Pagani F, Baralle FE. Genomic variants in exons and introns: identifying the splicing spoilers. Nature Reviews Genetics 2004; 5(5): 389-96.

Pagani F, Buratti E, Stuani C, Bendix R, Dork T, Baralle FE. A new type of mutation causes a splicing defect in ATM. Nature Genetics 2002; 30(4): 426-9.

Page NM, Butlin DJ, Lomthaisong K, Lowry PJ. The human apolipoprotein L gene cluster: identification, classification, and sites of distribution. Genomics 2001; 74(1): 71-8.

Pastinen T, Partanen J, Syvanen AC. Multiplex, fluorescent, solid-phase minisequencing for efficient screening of DNA sequence variation. Clinical Chemistry 1996; 42(9): 1391-7.

Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BT, Norris MC, Sheehan JB, Shen N, Stern D, Stowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SP, Cox DR. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. Science 2001; 294: 1719-1723.

Patterson N, Hattangadi N, Lane B, Lohmueller KE, Hafler DA, Oksenberg JR, Hauser SL, Smith MW, O'Brien SJ, Altshuler D, Daly MJ, Reich D. Methods for high-density admixture mapping of disease genes. American Journal of Human Genetics 2004; 74(5):979-1000.

Phillips MS, Lawrence R, Sachidanandam R, Morris AP, Balding DJ, Donaldson MA, Studebaker JF, Ankener WM, Alfisi SV, Kuo FS, Camisa AL, Pazorov AL, Scott KE, Carey BJ, Faith J, KAtari G, Bhatti HA, Cyr JM, Derohannessian V, Elousa C, Forman AM, Grecco NM, Hock CR, Kuebler JM, Lathrop JA, Mockler MA, Nachtman EP, Restine SL, Varde SA, Hozza MJ, Gelfand CA, Broxholme J, Abecasis GR, Boyce-Jacino MT, Cardon LR. Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. Nature Genetics 2003; 33: 382-387.

Plomin R, Hill L, Craig IW, MCGuffin P, Purcell S, Sham P, Lubinkski D, Thompson LA, Fisher PJ, Turic D, Owen MJ. A genome-wide scan of 1847 DNA markers for allelic associations with general cognitive ability: a five-stage design using DNA pooling. Behavioural Genetics 2001, 31: 497-509.

Potash JB, Buervenich S, Rathe JA, Steele CJM, Akula N, Detera-Wadleigh SD, Gershon ES, DePaulo JR, Feinberg AP, McMahon FJ. Gene-based SNP mapping of a psychotic bipolar disorder linkage region on 22q12.3 implicates a large haplotype block including HMG2L1 and TOM1. American Journal of Medical Genetics, Part B Neuropsychiatric Genetics 2004; 130B: 27-28.

Pritchard JK, Donnelly P. Case-control studies of association in structured or admixed populations. Theoretical Population Biology 2001; 60: 227-237.

Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. American Journal of Human Genetics 1999; 65: 220-228.

Pritchard JK. Are rare variants responsible for susceptibility to complex disease? American Journal of Human Genetics 2001; 69: 124-37.

Przeworski M, Wall JD. Why is there so little intragenic linkage disequilibrium in humans? Genetic Research 2001; 77(2): 143-51.

Ptak SE, Przeworski M. Evidence for population growth in humans is confounded by fine-scale population structure. Trends in Genetics 2002; 18(11):559-63.

Pulver AE, Karayiorgou M, Wolyniec PS, Lasseter VK, Kasch L, Nestadt G, Antonarakis S, Housman D, Kazazian HH, Meyers D, Ott J, et al. Follow-up of a Lamacz M, Liang KY, Hanfelt J, Ullrich G, Demarchi N, Ramu E, McHugh PR, Adler L, Thomas M, Carpenter WT, Manschreck T, Gordon CT, Kimberland M, Babb R, Puck J, Childs B. report of a potential linkage for schizophrenia on chromosome 22q12-q13.1: Part 2. American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics 1994 ;54(1):44-50.

Pulver AE, Lasseter VK, Kasch L, Wolyniec P, Nestadt G, Thornquist M, Ullrich G, McGrath J, Kasch L, Lamacz M, Thomas MG, Gehrig C, Radhakrishna U, Snyder SE, Balk KG, Nuefeld K, Schwartz KL, DeMarch N, Papadimitriou GN, Dikeos DG, Stefanis CN, Chakrvarti A, Childs B, Housman DE, Kazazian HH, Antonarakis S, Pulver AE. Schizophrenia: a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics 1995: 60: 252-260.

Purcell S, Cherny SS, Sham P. Genetic power calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics 2003; 19:149-150.

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES. Linkage disequilibrium in the human genome. Nature 2001; 411: 199-204.

Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996; 273: 1516-1517

Risch N. Linkage strategies for genetically complex traits: 1. Multilocus models. American Journal of Human Genetics1990; 46: 222-228.

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH. Multifactor dimensionality reduction reveals high-order interactions among oestrogen metabolism genes in sporadic breast cancer. American Journal of Human Genetics 2001; 69: 138-147.

Rosenthal D, Wender PH, Kety SS, Welner J, Schulsinger F. The adopted-away offspring of schizophrenics. American Journal of Psychiatry 1971; 128: 307-311.
Rossi A, De Cataldo S, Di Michele V, Manna V, Ceccoli S, Stratta P, Casacchia M. Neurological soft signs in schizophrenia. British Journal of Psychiatry 1990;157:735-9.

Ryan MM, Huffaker SJ, Webster MJ, Wayland M, Freeman T, Bahn S . Application and optimization of microarray technologies for human postmortem brain studies. Biological Psychiatry 2004;55(4):329-36.

Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A laboratory manual (2nd edition). Cold Springs Harbour Laboratory Press, New York, 1989.

Schizophrenia Collaborative Linkage Group. Additional support for schizophrenia linkage on chromosomes 6 and 8: a multicenter study. Schizophrenia Linkage Collaborative Group for Chromosomes 3, 6 and 8. American Journal of Medical Genetics: Part B, Neuropsychiatric Genetics 1996;67(6):580-94.

Schneider K. Clinical psychopathology (trans M W Hamilton). Grune and Stratton, New York, 1959.

Schumacher J, Abon Jamra R, Becker T, Klopp N, Franke P, Jacob C, Sand P, Fritze J, Ohlraun J, Schulze TG, Rietschel M, Illig T, Propping P, Cichon S, Deckert J, Nothen MM. Investigation of the DAOA/G30 locus in panic disorder. Molecular Psychiatry 2005; 10(5):428-9.

Schumacher J, Abon Jamra R, Freudenberg J, Becker T, Ohlraun S, Otte ACJ, Tullius M, Kovalenko S, Van Den Bogaert A, Maier W, Rietschel M, Propping P, Nothen M, Cichon S. Examination of G72 and D-amino acid oxidase as genetic risk factors for schizophrenia and bipolar affective disorder. Molecular Psychiatry 2004; 9: 203-207.

Schwab SG, Hallmayer J, Albus M, Lerer B, Eckstein GN, Borrmann M, Segman RH, Hanses C, Freymann J, Yakir A, Trixler M, Falkai P, Rietschel M, Maier W, Widenauer DB. A genome-wide autosomal screen for schizophrenia susceptibility loci in 71 families with affected siblings: support for loci on chromosome 10p and 6. Molecular Psychiatry 2000; 5(6):638-49.

Schwab SG, Knapp M, Mondabon S, Hallmayer J, Borrmann-Hassenbach M, Albus M, Lerer B, Rietschel M, Trixler M, Maier W, Wildenauer DB. 2003 Support for association of SZ with genetic variation in the 6p22.3 gene, dysbindin, in sib-pair families with linkage and in an additional sample of triad families. American Journal of Human Genetics;72(1):185-90.

Schwartz JE, Fennig S, Tanenberg-Karant M, Carlson G, Craig T, Galambos N, Lavelle J, Bromet EJ. Congruence of diagnoses 2 years after a first-admission diagnosis of psychosis. Archives of General Psychiatry 2000;57(6):593-600.

Seaman SR, Muller-Myhsok B. Rapid simulation of P values for product methods and multiple-testing adjustment in association studies. American Journal of Human Genetics 2005; 76(3):399-408.

Segal M. Rapid plasticity of dendritic spine: hints to possible functions? Progress in Neurobiology 2001; 63, 61-70.

Sham P, Bader J, Craig I, O'Donovan M & Owen MJ. DNA pooling: a tool for large-scale association studies. Nature Reviews Genetics 2002; 3: 862-871.

Sham PC, Cherney SS, Purcell, Hewitt JK. Power of linkage versus association analysis of quantitative traits, by use of variance-components models, sibship data. American Journal of Human Genetics 2000; 66: 1616-1630.

Sham PC, Curtis D. Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. Annals of Human Genetics 1995; 59: 97-105.

Slager SL, Huang J, Vieland VJ. Effect of allelic heterogeneity on the power of the transmission disequilibrium test. Genetic Epidemiology 2000; 18: 143-56.

Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). American Journal of Human Genetics 1993; 52(3): 506-16.

Spitzer RL, Endicott J, Robins E. Research diagnostic criteria: rationale and reliability. Archives of General Psychiatry 1978; 35(6): 773-82.

Spitzer RL, Williams JBW, Gibbons M, First MB. Structured clinical interview for DSM-IIIR-Patient Edition (SCID-P, Version 1.0). American Psychiatric Press, Washington DC, 1990.

St Clair D, Blackwood D, Muir W, Carothers A, Walker M, Spowart G, Gosden C, Evans HJ. Association within a family of a balanced autosomal translocation with major mental illness. Lancet 1990; 336(8706): 13-6.

Stefansson H, Sarginson J, Kong A, Yates P, Steinthorsdottir V, Gudfinnsson E, Gunnarsdottir S, Walker N, Petursson H, Crombie C, Ingason A, Gulcher JR, Stefansson K, Clair DS. Association of neuregulin 1 with schizophrenia confirmed in a Scottish population. American Journal of Human Genetics 2003; 72:83-87.

Stefansson H, Sigurdsson E, Steinthorsdottir V, Bjornsdottir S, Sigmundsson T, Ghosh S, Brynjolfsson J, Gunnarsdottir S, Ivarsson O, Chou TT, Hjaltason O, Birgisdottir B, Jonsson H, Gudnadottir VG, Gudmundsdottir E, Bjornsson A, Ingvarsson B, Ingason A, Sigfusson S, Hardardottir H, Harvey RP, Lai D, Zhou M, Brunner D, Mutel V, Gonzalo A, Lemke G, Sainz J, Johannesson G, Andresson T, Gudbjartsson D, Manolescu A, Frigge ML, Gurney ME, Kong A, Gulcher JR, Petursson H, Stefansson K. Neuregulin 1 and susceptibility to schizophrenia. American Journal of Human Genetics 2002; 71(4): 877-92.

Stephens M, Donnelly P. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. American Journal of Human Genetics 2003; 73: 1162-1169.

Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. American Journal of Human Genetics 2001; 68(4):978-89.

Straub RE, Jiang Y, MacLean CJ, Ma Y, Webb BT, Myakishev MV, Harris-Kerr C, Wormley B, Sadek H, Kadambi B, Cesare AJ, Gibberman A, Wang X, O'Neill FA, Walsh D, Kendler KS. Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with SZ. American Journal of Human Genetics 2002; 71(2): 337-48.

Straub RE, MacLean CJ, O'Neill FA, Burke J, Murphy B, Duke F, Shinkwin R, Webb BT, Zhang J, Walsh D, Kendler KS. A potential vulnerability locus for schizophrenia on chromosome 6p24-22: evidence for genetic heterogeneity. Nature Genetics 1995;11(3):287-93.

Sutcliffe JG, Thomas EA. The neurobiology of apolipoproteins in psychiatric disorders. Molecular Neurobiology 2002; 26(2-3):369-88.

Syvanen AC. Accessing genetic variation: genotyping single nucleotide polymorphisms. Nature Reviews Genetics 2001; 2: 930-942.

Takahashi S, Cui YH, Kojima T, Han YH, Zhou RL, Kamioka M, Yu SY, Matsuura M, Matsushima E, Wilcox M, Arinami T, Shen YC, Faraone SV, Tsuang MT. Family-based association study of markers on chromosome 22 in SZ using African-American, European-American, and Chinese families. American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics 2003; 120B(1):11-7.

Talkowski M, Bassett AS, Brzustowicz LM, Chen X, Chowdari KV, Collier DA, Cordeiro Q, Corvin A, Deshpande S, Egan MF, Gill M, Kendler KS, Kirov G, Levitt P, Lewis DA, Li, Mirnics K, Morris DW, O'Donovan MC, Owen MJ, Seltman H, Semwall P, Sobell JL, Thelma BK, Weinberger DA, Williams NM, Wood J,

Nimgaonkar VL. Evaluating an association of RGS4 polymorphisms with schizophrenia: mega-analysis of over 12,000 individuals from 11 international sites. American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics 2004; 130B (1); 141.

Templeton AR, Clark AG, Weiss KM, Nickerson DA, Boerwinkle E, Sing CF. Recombinational and mutational hotspots within the human lipoprotein lipase gene. American Journal of Human Genetics 2000; 66: 69-83.

Terwilliger JD, Weiss KM. Linkage disequilibrium mapping of complex disease: fantasy or reality? Current Opinion Biotechnology 1998; 9(6):578-94.

Terwilliger JD, Goring HH. Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. Human Biology 2000; 72(1):63-132.

The International HapMap Consortium. The International HapMap Project. Nature 2003; 426: 789-96.

The US-Venezuela Collaborative Research Project. Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. Proceedings of the National Academy of Sciences, USA 2004; 101(10):3498-503.

Tomas DC, Haile RW, Duggan D. Recent developments in genomewide association scans: a workshop summary and review. American Journal of Human Genetics 2005; 77: 337-345.

Thomson G, Esposito MS. The genetics of complex diseases. Trends in Cell Biology 1999; 9: 17-20.

Tienari P, Wynne LC, Laksy K, Moring J, Nieminen P, Sorri A, Lahti I, Wahlberg KE. Genetic boundaries of the schizophrenia spectrum: evidence from the Finnish Adoptive Family Study of Schizophrenia. American Journal of Psychiatry 2003; 160(9): 1587-94.

Tkachev D, Mimmack ML, Ryan MM, Wayland M, Freeman T, Jones PB, Starkey M, Brewster MJ, Yolken RH, Bahn S.Oligodendrocyte dysfunction in SZ and bipolar disorder. Lancet 2003; 362(9386):798-805

Toivonen HT, Onkamo P, Vasko K, Ollikainen V, Sevon P, Mannila H, Herr M, Kere J. Data mining applied to linkage disequilibrium mapping. American Journal of Human Genetics 2000 ;67(1):133-45.

Toro CT, Kasher PR, Deakin JFW. Altered D-serine metabolism in schizophrenia? A post-mortem study using the Stanley Consortium brains. Schizophrenia Research 2004; 67: 125-126S.

Tsuang MT, Woolson RF, Winokur G, Crowe RR Stability of psychiatric diagnosis. Schizophrenia and affective disorders followed up over a 30- to 40-year period. Archives of General Psychiatry 1981; 38(5):535-9.

Van den Oord EJ, Neale BM. Will haplotype maps be useful for finding genes? Molecular Psychiatry 2004; 9 (3): 227-36

Van Os J & Marcelis M. The ecogenetics of schizophrenia: a review. Schizophrenia Research 1998; 32(2): 127-35.

Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. Journal of the National Cancer Institute 2000; 92: 1151-58.

Waddington JL, Scully PJ, Youssef HA. Developmental trajectory and disease progression in schizophrenia: the conundrum, and insights from a 12-year prospective study in the Monaghan 101. Schizophrenia Research 1997;23(2):107-18.

Wang N, Akey JM, Zhang K, Chakraborty R, Jin L. Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutations. American Journal of Human Genetics 2002; 71: 1227-34.

Wang X, He G, Gu N, Yang J, Tang J, Chen Q, Liu X, Shen Y, Qian X, Lin W, Duam Y, Feng G, He L. Association of G72/G30 with schizophrenia in the Chinese population. Biochemical Biophysical Research Communications 2004; 319(4): 1281-6.

Weickert CS, Straub RE, McClintock BW, Matsumoto M, Hashimoto R, Hyde TM, Herman MM, Weinberger DR, Kleinman JE. Human dysbindin (DTNBP1) gene expression in normal brain and in schizophrenic prefrontal cortex and midbrain. Archives of General Psychiatry 2004;61(6):544-55.

Weinberger DR. From neuropathology to neurodevelopment. Lancet 1995; 346 (8974):552-557

Weiss KM, Clark AG. Linkage disequilibrium and the mapping of complex human traits. Trends in Genetics 2002; 18(1):19-24.

Weiss KM, Terwilliger JD. How many diseases does it take to map a gene with SNPs? Nature Genetics 2000; 26: 151-7.

Wender PH, Rosenthal D, Kety SS, Schulsinger F, Welner J. Crossfostering. A research strategy for clarifying the role of genetic and experiential factors in the etiology of schizophrenia. Archives of General Psychiatry 1974;30(1):121-8.

Wiehe T, Mountain J, Parham P, Slatkin M. Distinguishing recombination and intragenic gene conversion by linkage disequilibrium patterns. Genetics Research 2000; 75: 61-73.

Williams J, Spurlock G, Holmans P, Mant R, Murphy K, Jones L, Cardno A, Asherson P, Blackwood D, Muir W, Meszaros K, Aschauer H, Mallet J, Laurent C, Pekkarinen P, Seppala J, Stefanis CN, Papadimitriou GN, Macciardi F, Verga M, Pato C, Azevedo H, Crocq MA, Gurling H, Kalsi G, McGuffin P, Owen MJ. A meta-

analysis and transmission disequilibrium study of association between the dopamine D3 receptor gene and schizophrenia. Molecular Psychiatry 1998;3(2):141-9.

Williams J, Spurlock G, McGuffin P, Mallet J, Nothen MM, Gill M, Aschauer H, Nylander PO, Macciardi F, Owen MJ. Association between schizophrenia and T102C polymorphism of the 5-hydroxytryptamine type 2a-receptor gene. European Multicentre Association Study of Schizophrenia (EMASS) Group. Lancet 1996; 11;347(9011):1294-6.

Williams NM, Norton N, Williams H, Ekholm B, Hamshere ML, Lindblom Y, Chowdari KV, Cardno AG, Zammit S, Jones LA, Murphy KC, et al. A systematic genomewide linkage study in 353 sib pairs with schizophrenia. American Journal of Human Genetics 2003; 73(6):1355-67.

Williams NM, Preece A, Morris DW, Spurlock G, Bray NJ, Stephens M, Norton N, Williams H, Waddington JL, Gill M, Corvin A, Zammit S, Kirov G, Owen MJ, O'Donovan MC. Confirmation in two independent samples that dystrobrevin binding protein gene (DTNBP1) is a susceptibility locus for schizophrenia. Archives of General Psychiatry 2004; 61: 336-44.

Williams NM, Preece A, Spurlock G, Norton N, Williams HJ, Zammit S, O'Donovan MC, Owen MJ. Support for genetic variation in neuregulin 1 and susceptibility to SZ. Mol Psychiatry 2003; 8(5): 485-7.

Wing JK, Cooper JE, Sartorius N. Measurement and classification of Psychiatric Symptoms. Cambridge: Cambridge University Press, 1974.

Wright A, Charlesworth B, Rudan I, Carothers A, Campbell H. A polygenic basis for late onset disease. Trends in Genetics 2003; 19: 97-106.

Wright AF, Carothers AD, Pirastu M. Population choice in mapping genes for complex diseases. Nature Genetics 1999; 22: 139-144.

Wright P, Takei N, Murray RM, Sham PC. Seasonality, prenatal influenza exposure and schizophrenia. In: Prenatal Exposures in Schizophrenia, Susser ES, Brown AS, Gorman JM, eds Washington, DC: American Psychiatric Press, 1999 (pp 89-112).

Yau HJ, Wang HF, C Lai, Liu FC. Neural development of the neuregulin receptor ErbB4 in the cerebral cortex and the hippocampus: preferential expression by interneurons tangentially migrating from the ganglionic eminences Cerebral Cortex 2003; 13 (3): 252-264

Zangenberg G, Huang MM, Arnheim N, Erlich H. New HLA-DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. Nature Genetics 1995; 10(4):407-14.

Zee RY, Hoh J, Cheng S, Reynolds R, Grow MA, Silbergleit A, Walker K, Steiner L, Zangenberg G, Fernandez-Ortiz A, Macaya C, Pintor E, Fernandez-Cruz A, Ott J, Lindpainter K. Multi-locus interactions predict risk for post-PTCA restenosis: an

approach to the genetic analysis of common complex disease. Pharmacogenomics Journal 2002;2(3):197-201.

Zhang K, Calabrese P, Nordborg M, Sun F. Haplotype block structure and its applications to association studies: power and study designs. American Journal of Human Genetics 2002;71(6):1386-94.

Zhao JH, Lissarrague S, Essioux L, Sham PC. GENECOUNTING: haplotype analysis with missing genotypes. Bioinformatics 2000; 18(12); 1694-1695.

Zhao JH, Sham PC. Faster allelic association analysis using unrelated subjects. Human Heredity 2002; 53 (1): 36-41.

Zhou F, Li C, Duan S, Zheng Y, Gu N, Feng G, Xing Y, Shi J, He L. A family-based study of the association between the G72/G30 genes and schizophrenia in the Chinese population. Schizophrenia Research 2005;73(2-3):257-61.

Zondervan KT, Cardon LR. The complex interplay among factors that influence allelic association. Nature Reviews Genetics 2004;5(2):89-100.

Trinity Centre for Health Sciences
St. James's Hospital,

James's St.,
Dublin 8
Phone: (01) 6082462
Email: **acorvin@tcd.ie**

Genetic Study

Of
Psychotic Disorders



Information Brochure

Dept. of Psychiatry,

Who are we?

We are a Research team based at the Dept. of Psychiatry at Trinity College Dublin and St. John of Gods Hospital, Stillorgan. We are involved in a collaborative study with the doctors, staff and patients of St. John of God's Hospital. This study aims to increase our understanding of the causes of psychotic disorders.

What is the study?

There is strong evidence that psychotic disorders are in part caused by something that is inherited within families. It is thought that many genes contribute to these disorders, each having a small effect. The study we are conducting aims to identify the genes involved, which will increase our understanding of these disorders. This is important, and may eventually lead to better treatment or even prevention of these common disorders.

What is involved?

If you agree to be involved in the study you will be asked to give a blood sample. This will be collected at a time and location convenient for you and will take a few minutes.
With your agreement we will
You will:

1) be asked a series of questions about your medical history
2) complete a brief medical check up (of co-ordination, reflexes etc.)
3) be asked to give a blood sample.

This study is strictly confidential and no one else will have access to your personal details. No individual results will be generated by this study, but by being involved you are helping researchers understand the causes of psychotic disorders.

Is participation voluntary?

Absolutely. You can participate, refuse to participate, or withdraw at any time over the course of this study. This will in no way effect your treatment, now or in the future.

Thank you!

For taking the time to read this leaflet. If you wish to participate in this study or would like more information, please contact Aiden or Siobhan at:

Dept. of Psychiatry,
Trinity Centre for Health Sciences
St. James's Hospital,

James's St.,
Dublin 8
Phone: (01) 6082462

APPENDIX II

PANSS

Name _____ Age_____ Sex_____

Diagnosis_____ Obs. Period_____

Ward_____ Date_____ Rater_____

Positive Scale		ABS	MIN	MIL	MOD	SEV	SEV	EXT
P1 . Delusions	P1	1	2	3	4	5	6	7
P2 . Conceptual Disorganisation	P2	1	2	3	4	5	6	7
P3 . Halluncinatory Behaviour	P3	1	2	3	4	5	6	7
P4 . Excitement	P4	1	2	3	4	5	6	7
P5 . Grandiosity	P5	1	2	3	4	5	6	7
P6 . Suspiciousness/Persecution	P6	1	2	3	4	5	6	7
P7 . Hostility	P7	1	2	3	4	5	6	7
Negative Scale								
N1 . Blunted Affect	N1	1	2	3	4	5	6	7
N2 . Emotional Withdrawal	N2	1	2	3	4	5	6	7
N3 . Poor Rapport	N3	1	2	3	4	5	6	7
N4 . Passive/Apathetic Withdrawal	N4	1	2	3	4	5	6	7
N5 . Difficulty in Abstract Thinking	N5	1	2	3	4	5	6	7
N6 . Lack of Spontaneity/Flow	N6	1	2	3	4	5	6	7
N7. Stereotyped Thinking	N7	1	2	3	4	5	6	7
General Psychopathology Scale								
G1 . Somatic Concern	G1	1	2	3	4	5	6	7
G2 . Anxiety	G2	1	2	3	4	5	6	7
G3 . Guilt Feelings	G3	1	2	3	4	5	6	7
G4 . Tension	G4	1	2	3	4	5	6	7
G5 . Mannerisms and posturing	G5	1	2	3	4	5	6	7
G6 . Depression	G6	1	2	3	4	5	6	7
G7 . Motor Retardation	G7	1	2	3	4	5	6	7
G8 . Uncooperativeness	G8	1	2	3	4	5	6	7
G9 . Unusual Thought Content	G9	1	2	3	4	5	6	7
G10.Disorientation	G10	1	2	3	4	5	6	7
G11.Poor Attention	G11	1	2	3	4	5	6	7
G12.Lack of Judgement/Insight	G12	1	2	3	4	5	6	7
G13.Disturbance of volition	G13	1	2	3	4	5	6	7
G14.Poor Impulse Control	G14	1	2	3	4	5	6	7
G15.Preoccupation	G15	1	2	3	4	5	6	7
G16.Active Social Avoidance	G16	1	2	3	4	5	6	7

PANSS RATING FORM

Scale	Total	Percentile	Range
Syndrome Scores*			
Positive	_____	_____	_____
Negative	_____	_____	_____
Composite	_____	_____	_____
General Psychopathology	_____	_____	_____
Cluster Scores**			
Anergia	_____	_____	_____
Thought Disturbance	_____	_____	_____
Activation	_____	_____	_____
Paranoid Belligerence	_____	_____	_____
Depression	_____	_____	_____

* Positive Syndrome = Sum of P1 through P7
Negative Syndrome = Sum of N1 through N7
Composite Index = Positive syndrome minus negative syndrome
General Psychopathology = Sum of G1 through G16

** Anergia = N1+ N2 + G7 + G10
Thought Disturbance = P2 + P3 + P5 + G9
Activation = P4 + G4 + G5
Paranoid Belligerence = P6 + P7 + G8
Depression = G1 + G2 + G3 + G6

ABNORMAL INVOLUNTARY MOVEMENT SCALE (AIMS)

Name: GASP Date Rater Score

MOVEMENT RATINGS: Rate highest severity observed. Rate movements that occur upon activation one less than those observed spontaneously.

Facial and Oral Movements	none	minimal	mild	moderate	severe
1. Muscles of facial expression e.g., movements of forehead, eyebrows, periorbital area, cheeks; include frowning, blinking, smiling, grimacing.	! 0	! 1	! 2	! 3	! 4
2. Lips and Perioral Area e.g. puckering, pouting, smacking.	! 0	! 1	! 2	! 3	! 4
3. Jaw e.g. biting, clenching, chewing, mouth opening, lateral movement.	! 0	! 1	! 2	! ! 3	 4
4. Tongue Rate only increase in movement both In and out of the mouth. NOT inability to sustain movement.	! 0	! 1	! 2	! ! 3	 4

Extremity Movements	none	minimal	mild	moderate	severe
5. Upper (arms, wrists, hands, fingers) Include choreic movements (i.e. rapid objectively purposeless, irregular, spontaneous): athetoid movements(i.e. slow, irregular, complex, serpentine). Do NOT include tremor (i.e. repetitive, Regular, rhythmic).	! 0	! 1	! 2	! 3	! 4

ABNORMAL INVOLUNTARY MOVEMENT SCALE (AIMS) Cntd.

Extremity Movements (continued)

	none	minimal	mild	moderate	severe
	!	!	!	! !	
	0	1	2	3	4
6. Lower (legs, knees, ankles, toes) e.g. lateral knee movement, foot tapping, heel dropping, foot squirming, inversion and eversion of foot.					

Trunk Movements

	!	!	!	!	!
	0	1	2	3	4
7. Neck, shoulders, hips e.g. rocking, twisting, squirming, pelvic gyrations.					

---

8. Severity of abnormal movements	!	!	!	!	!
	0	1	2	3	4

---

9. Incapacitation due to abnormal movements.	!	!	!	!	!
	0	1	2	3	4

---

10. Patients awareness of abnormal movements (rate only patient's report)

No awareness	Aware no distress	Aware mild distress	Aware moderate distress	Aware severe distress
!	!	!	!	!
0	1	2	3	4

Dental Status

11. Current problems with teeth and/or dentures? ! YES ! NO

---

12. Does patient usually wear dentures? ! YES ! NO

Soft signs assessment **Total Score:**_____

Name: _____ **Study Number:** _____ **Rater:** _____ **Date:** _____

1) Handedness

a. Throw	f. Jar		R	L
b. Kick	g. Draw			
c. Knife	h. Pen	Dominant		
d. Spoon	i. Eye			
e. T/brush	j. Brush			

2) Heel toe walk (15 steps) rated 0, 1 (if one or two misteps), 2 _____

Rombergs (0, 1, 2) _____

\+ adventitious overflow (0, 1 if irreg. finger movements only, 2 if arms) _____

\+ tremor (0, 1, 2 if a coarse tremor) _____

	R	L
3) Fist ring (0, 1, 2- major motion breakdown or 4 hesitations)	_____	
Fist-edge-palm (0, 1, 2)	_____	
Ozeretski [fist to flat, opposites with each hand] (0,1,2)	_____	
Dysdiadochokinesis [knee] (0-2)	_____	
Finger-thumb opposition (0,2)[watch for mirror movements](0,2)	_____	

Memory task: apple, chair, bicycle, pen

4) Audio-visual integration[which of these am I tapping? .. ... . or ... .(4)] _____
Eyes closed, test to make sure they understand. Scoring 0, 1 if one mistake, 2.

Rhythm tapping test [i. Reproduce loud and soft sounds in sequence] (0-2) _____

[ii. Tap out heard sequence .. ... .] (0-2) _____

5) Finger-nose [for past pointing and perseveration](0, 1=mild tremor/pp, 2) _____

Synkinesis [not moving head] 0, 1=movehead first,2 if more) _____

Nystagmus (scored 0-2, or 1 point if present on each side) _____

Gaze impertinence (0, 1=deviation from fixation after 20 sec., 2) _____

Convergence (0, 1=one or both converge more than halfway, 2) _____

Glabellar tap (0=3 or less blinks, 1=four or five blinks 2=more) _____

Memory task *(0= all words, 1=3 words, 2=less than three)* _____

		R	L
6) Graphaestesia [Dorsum r/l (mark 0-3: number is 1-5), Pulp r/l (mark 0-2)]	d	_____	
	p	_____	

Astereognosis [identifying keys etc.] (0-2, each side start with dominant hand) _____

Blunt-sharp discrimination [r/l dorsum] (score 0-5 each side) _____

Right/left confusion (point to own l/r (1) , find examiners hands scores (1)) _____

SBTE ? extinction rc/lh lc/lh rc/lh rc/rh lc/rh rc/lc rh/rh (0-6) _____

7) Reflexes

	r	l		r	l
a) Grasp(0-5)	____	____	c) Suck (0-1 each)	___	___
b) Snout (0-1)	____		d) Palmomental (0-1 each)	___	___

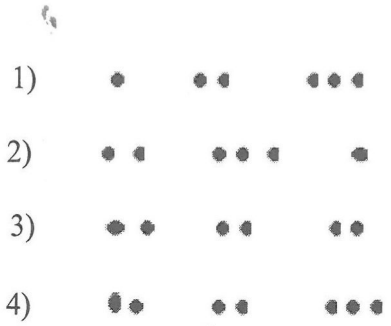
8) Complex motor act [3 parts] score 0-5
(pick up paper, fold, place on lap. Score abnormal effort positive, difficulty exhibited 1-5) _____

Imaginary act [like a children's tea party, talk through it] score 0-5 _____
(hold pot/kettle, pour, replace, get spoon, stir)

2 object test [two end of pen together] score 0-4 _____

Memory test *(again score 0-2)*

_____



M.P.A. Score

Rater
GASP No: _____
Name: _____
Date: _____

1. High palate: Y N (Inverted U: Y N Inverted V: Y N)

2. Increased palate ridges (>5) Y N

3. Supraorbital ridges protruding Y N

4. Bifid tongue Y N

5. Ear lobe size (hypoplastic) Y N

6. Epicanthus present: Y N

	Minor	Y	N
(extends, but doesn't cover commisure)	Moderate	Y	N
(covers commisure)	Major	Y	N

7. Mouth width: _____ Zygion to zygion: _____

8. Ear protrusion: R _____ L_____

9. Biocular diameter: _____

10. Skull base width: _____

11. Eye fissure inclination:

normal/horizontal	Y
upslanting	Y
downslanting	Y

12. Anterior ear shape:

normal rolled helix	Y
flat and unfolded	

M.P.A. Score

Rater
GASP No: _____
Name: _____
Date: _____

1. High palate: Y N (Inverted U: Y N Inverted V: Y N)

2. Increased palate ridges (>5) Y N

3. Supraorbital ridges protruding Y N

4. Bifid tongue Y N

5. Ear lobe size (hypoplastic) Y N

6. Epicanthus present: Y N
   Minor Y N
   (extends, but doesn't cover commisure) Moderate Y N
   (covers commisure) Major Y N

7. Mouth width: _____ Zygion to zygion: _____

8. Ear protrusion: R _____ L_____

9. Biocular diameter: _____

10. Skull base width: _____

11. Eye fissure inclination: normal/horizontal Y
    upslanting Y
    downslanting Y

12. Anterior ear shape: normal rolled helix Y
    flat and unfolded