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.

Family-Based Molecular Genetic Studies in Bipolar Affective Disorder

Ricardo Segurado

Ph.D.

University of Dublin
Trinity College

2002

TRINITY COLLEGE
13 OCT 2003
LIBRARY DUBLIN
THESIS
7496.
7372

Declaration of Author

I hereby declare that this work has not been concurrently submitted for nor accepted as an exercise for a degree at this or any other University. This degree is the result of entirely my own investigations, except where otherwise indicated in the Attribution of Work (page vi) or the text. A full bibliography is appended. I give my consent for the library of Trinity College to copy and lend this thesis upon request.

Candidate:.....Ramla Ismail.....

Date:.....26/8/2003.....

Summary

This thesis covers gene mapping experiments applied to bipolar affective disorder (BPD), a severe recurrent psychiatric condition with a lifetime prevalence of 1-2%, and a substantial familial component. The study sample consisted of affected sibling pairs (ASPs), where one sibling had a diagnosis of bipolar I disorder (BPI), and the other one of a range of broader diagnoses within the affective spectrum. Initial studies focused on the investigation of a candidate gene, tyrosine hydroxylase (TH), for association to BPD in an Irish pedigree sample (216 individuals). This gene revealed no evidence for association with the phenotype, possibly due to lack of statistical power of the sample, or other confounding factors.

As part of the collaborative Wellcome Trust UK/Irish BPD Sib-Pair Study a joint British/Irish sample (255 ASPs) was examined for genetic linkage with 398 microsatellite markers across the genome. This thesis includes genotyping and statistical analysis of markers on chromosomes 2, 5, 6, 9, 10, 11, 17, 18, 19, 20, 21 and 22. As part of a two-stage strategy, this study established a maximum likelihood score (MLS) threshold equivalent to a point-wise p-value of 0.05, to identify regions to be pursued in a second stage. This study represents the largest of its kind ever undertaken. Seventeen non-contiguous regions of the genome surpassed this threshold, however no region achieved statistical criteria for genome-wide significance. Regions exceeding the threshold lay on chromosomes 2, 3, 4, 6, 7, 9, 10, 12, 17, X, and chromosome 18 where the most significant MLS of the genome was observed. Second stage genotyping of an expanded sample set (total 399 ASPs) was performed on regions on chromosomes 2q37 and 10q11-q21, where modest increases in the MLS were observed.

Subsequent to an exploratory meta-analysis of 8 genome scans for BPD using data available from published reports, a collaborative effort was begun to assemble full linkage information from all 22 genome scans conducted to date in BPD, the first effort of its kind

with BPD. Meta-analysis was performed using the rank-based Genome Scan Meta-Analysis (GSMA) method. A comprehensive power study using empirical simulations of the data was also performed. No statistical significance was observed in the results, in contrast to a parallel study in schizophrenia. The small sample sizes used by many studies and the inconsistency of disease classification models applied, among other factors, may account for these results. Regions on chromosomes 18, 14, 8 and 9 were highlighted as being the most likely to harbour susceptibility loci for BPD.

The results of these studies confirm the view that the genetic architecture of BPD is complex. The largest samples collected to date, including the present one, are probably too small to identify replicable genetic linkage to any one locus. A meta-analysis of all genome-wide studies also failed to identify any loci at an appropriate statistical level. It is likely that the effects of individual genes are very small, but methodological refinements could yet assist in the effort to map susceptibility loci for BPD. Nevertheless chromosome 18 represents the most likely location of a locus or loci conferring vulnerability to BPD.

Acknowledgements

I would first and foremost like to thank Professor Michael Gill for allowing me to work toward this degree under his supervision, and for his support and patience throughout.

I would also like to thank co-workers on the Wellcome Trust project: Dr. Tanya Mulcahy, David Lambert, and Dr. Sylvia Bort, for help with genotyping and making me a better scientist, Professor Nick Craddock and Dr. Phil Bennett, formerly of the University of Birmingham, who performed half the genome scan study, and performed the power studies for the genome scan with Dr. Peter Holmans. Thanks to Iain Fenton for writing the MegaBase program. On the clinical side I thank Drs. Aiden Corvin and Ed O'Mahony, Anne Payne, Brian Larsen, Claire Comerford, Robert O'Connell in Ireland, and Dr. Ian Jones, Fiona McCandless and Fiona Middle in the UK. Thanks for help and insight on the meta-analysis study to Drs. Sevilla Detera-Wadleigh, Doug Levinson, Cathryn Lewis and Lesley Wise. Thanks also to Dr. Ziarih Hawi for invaluable advice and support on all aspects of genetics, and all who have passed through the Neuropsychiatric Genetics lab over the years, and those who are still there. Thanks to all the people of the Genetics Department, including Louis, Dave, Paul and Brenda, Profs. David McConnell and Kevin Devine for permitting me to study in their department, and to Geraldine Quinn in the Department of Psychiatry.

Thanks to my family, especially my parents and siblings Susana, Paulo and Sam for patience, financing, proofreading and moral and dietary support over the past five years. Thanks to Cian, Amy, Emmet, Mike and Tom for ignoring me for the past few years, and to Clive*Star for delaying their world tour to accommodate my studies.

Last but not least, I would like to extended sincere thanks to all the patients and their families that participated in these studies, to whom this thesis is dedicated; may we one day be able to help you.

Attribution of Work

Due to the large size and collaborative nature of the projects undertaken as part of this thesis, a great many individuals assisted in much of the work described below. In order to clarify which work was undertaken by the author, which was not, and which was a shared effort, a brief statement follows, listing colleagues and collaborators and their individual contributions.

DNA extraction from blood was performed by the author, Dr. Ziarih Hawi, Dr. Tanya Mulcahy, Susan Byrne and David Lambert, and DNA quantitation was performed mostly by the author, with assistance from Susan Byrne (chapter 2). The tyrosine hydroxylase study was performed by the author with advice from Drs. Hawi and Mulcahy (chapter 3). In the genome scan study (chapter 4), the power simulations were conducted entirely by Prof. Nick Craddock and Dr. Peter Holmans. The first stage genome scan genotyping was performed by the author for chromosomes 5, 6, 19, 20, 21 and 22, and panel 16 (approximately one quarter of the markers on chromosomes 9, 10 and 11), in addition to 7 drop-out markers. All other genotypes for Dublin chromosomes generated by Dr. Mulcahy, David Lambert, Johnathan Lowery and Dr. Sylvia Bort. In Birmingham first stage genotyping was performed by Dr. Phil Bennett. All data checking and linkage analysis on the twelve first stage Dublin chromosomes were performed by the author, including writing of Microsoft Excel macros and batch files for linkage analysis. DNA quantitation of the second stage sample was shared between the author, David Lambert and Dr. Bort, who also mainly performed second stage genotyping of the markers on chromosomes 2 and 10, with assistance from the author. The author alone performed genotyping of 3 markers on chromosome 10 in the expanded sample, and linkage analysis on the second stage chromosome 2 and 10 data.

The meta-analysis study involved collection of full genome linkage data from each research group, which was performed principally by Dr. Sevilla Detera-Wadleigh, with assistance from the author and Dr. Douglas Levinson. The power study was conducted in the main by Dr Levinson and Matthew Levinson with assistance from Dr. Cathryn Lewis and the author. The author had sole responsibility for conducting the meta-analysis proper on the bipolar disorder data, including the marker binning and bin ranking for each genome scan study, and calculation of empirical significance levels by computer simulations, with advice from Drs. Levinson and Lewis.

Table of Contents

DECLARATION OF AUTHOR II

SUMMARY III

ACKNOWLEDGEMENTS V

ATTRIBUTION OF WORK VI

TABLE OF CONTENTS VII

LIST OF FIGURES XIV

LIST OF TABLES XV

ABBREVIATIONS XVII

CHAPTER 1. INTRODUCTION 1

1.1 BIPOLAR AFFECTIVE DISORDER 1

1.1.1 History of Affective Disorders 1

1.1.2 Clinical Features and Diagnosis 1

1.1.3 Biological features 2

1.2 AETIOLOGY OF BIPOLAR DISORDER 3

1.2.1 Pharmacological clues 3

1.2.2 Cellular and Physiological Changes in Bipolar Disorder 9

1.2.3 Difficulties with Biochemical Research 11

1.3 EPIDEMIOLOGY OF BIPOLAR DISORDER 12

1.3.1 Familiality 12

1.3.2 Adoption and Twin Studies 13

1.3.3 Segregation Analysis 16

1.3.4 Diagnostic Subclassifications and Heritability 17

1.3.5 Summary of epidemiological research 19

1.4 LINKAGE AND ASSOCIATION STATISTICS 21

1.4.1 Introduction 21

1.4.2 Genetic Association 22
1.4.3 Genetic Linkage 23
1.4.5 Genome Scans 30
1.5 GENETIC STUDIES IN BIPOLAR DISORDER 32
1.5.1 Genetic Association to Bipolar Disorder 32
1.5.2 Genetic Linkage to Bipolar Disorder 34
1.6 GENETICALLY COMPLEX DISORDERS 40
1.6.1 Introduction 40
1.6.2 Genetic Heterogeneity 41
1.6.3 Gene-Gene Interactions 41
1.6.4 Gene-Environment Interactions 42
1.6.5 Anticipation and Trinucleotide Repeats 42
1.6.6 Assortative Mating 43
1.6.7 Imprinting 44
1.7 SUMMARY 45
1.8 AIMS 47
CHAPTER 2. MATERIALS AND METHODS 50
2.1 INTRODUCTION 50
2.2 PATIENT COLLECTION AND DIAGNOSIS 50
2.3 DNA EXTRACTION FROM WHOLE BLOOD 51
2.4 MICROSATELLITE GENOTYPING USING RADIO-ISOTOPE LABELS 53
2.5 SEMI-AUTOMATED MICROSATELLITE GENOTYPING USING FLUORESCENT LABELS 54
2.5.1 ABI 377 54
2.5.2 Polymerase Chain Reaction 55
2.5.3 Gel Electrophoresis 57

2.5.4 Processing of the Gel File 57
2.5.5 Genotype Data Management 58
2.5.6 Data Handling and Allele Binning 59
2.6 STATISTICAL METHODS 62
2.6.1 Pedcheck 62
2.6.2 SibError 62
2.6.3 RELATIVE 63
2.6.4 Extended Transmission Disequilibrium Test (ETDT) 63
2.6.5 Sibling Transmission Disequilibrium Test 64
2.6.6 Two-point Linkage Analysis 65
2.6.7 Multipoint Non-Parametric Linkage (NPL) Analysis 66
2.6.8 Multipoint Maximum Likelihood Linkage Analysis 67
2.7 BUFFERS AND MEDIA 69
CHAPTER 3. ASSOCIATION AND LINKAGE STUDIES OF THE TYROSINE HYDROXYLASE GENE 71
3.1 INTRODUCTION 71
3.1.1 Tyrosine Hydroxylase And Monoamines 71
3.1.2 Linkage Studies 71
3.1.3 Previous Association 72
3.1.4 Aims 73
3.2 METHODS 74
3.2.1 Sample Description 74
3.2.2 PCR 77
3.2.3 Statistical Analyses 78
3.3 RESULTS 79
3.3.1 Marker Allele Frequencies 79

3.3.2 ETDT 79
3.3.3 S-TDT 80
3.3.4 SPLINK 80
3.4 DISCUSSION 80
CHAPTER 4. GENOME SCAN STAGE I 83
4.1 INTRODUCTION 83
4.1.1 Background and Aims 83
4.1.2 Study Design 84
4.2 METHODS 87
4.2.1 Sample Description 87
4.2.2 Power Calculations and Threshold selection 88
4.2.3 Marker Map and Genotyping 89
4.2.4 Logistics and Quality Control 91
4.2.5 Statistical analyses 96
4.3 RESULTS 96
4.3.1 Power and Significance Level Simulations 96
4.3.2 Genotype Error Detection and Quality Control 97
4.3.2 Linkage Results and Tables 100
4.3.3 Exclusion Lod Analysis 106
4.4 DISCUSSION AND FUTURE WORK 107
4.4.1 Discussion of Significance 107
4.4.2 Overlaps with previous Studies 108
4.4.3 Summary 110
CHAPTER 5. GENOME SCAN STAGE II: PHASE 1 112
5.1 INTRODUCTION 112
5.1.1 Study rationale and Design 112

5.1.2 Candidate Areas for Follow-up 112
5.1.3 Selection of Follow-up Regions 117
5.2 MATERIALS AND METHODS 118
5.2.1 Sample Description 118
5.2.2 Marker Map 120
5.2.3 Genotyping 120
5.2.4 Statistical analyses 121
5.3 RESULTS 122
5.3.1 Scan IIA results 122
5.3.2 Error Rate and Quality Control 124
5.4 DISCUSSION 125
5.4.1 Evidence for Linkage 125
5.4.2 Summary 125
CHAPTER 6. GENOME SCAN META-ANALYSIS 127
6.1 INTRODUCTION 127
6.1.1 Overlap of Linkage Findings 127
6.1.2 Problems with Meta-Analysis 128
6.1.3 Types of Meta-analysis 130
6.2 PRELIMINARY GSMA 133
6.2 METHODS 134
6.2.1 Scan Selection 134
6.2.2 Phenotypic Models 138
6.2.3 Power 139
6.2.4 GSMA 140
6.2.5 Bin Positioning 142
6.3 RESULTS 143

6.3.1 Power Study 143
6.3.2 Sample Size 144
6.3.3 Bin-Wise significance 146
6.3.4 Rank Order Significance 152
6.3.5 Bin Position 154
6.4 DISCUSSION 155
CHAPTER 7. DISCUSSION 161
7.1 SUMMARY OF RESULTS 161
7.1.1 Association Study 161
7.1.2 First Stage Genome Scan 162
7.1.3 Second Stage Genome Scan 162
7.1.4 Meta-Analysis of Genome Scans 162
7.2 STUDY LIMITATIONS 163
7.2.1 Association Studies 163
7.2.2 Genetic Linkage 165
7.3 IMPACT OF THESE RESULTS 167
7.3.1 The Differences Between Schizophrenia and Bipolar Disorder 167
7.3.2 Uncertainty About the True Underlying Genetic Model 167
7.4 FUTURE DIRECTIONS 169
7.4.1 Familiality and Diagnostic Issues 169
7.4.2 Genome-Wide Association Studies 170
7.4.3 Linkage Studies 172
BIBLIOGRAPHY 175
1. ONLINE RESOURCES 175
2. REFERENCES 175
APPENDIX A. GENOME SCAN COMMA SEPARATED VALUE FILE 216

APPENDIX B. GENOME SCAN LINKAGE ANALYSIS BATCH FILES 221
1. SPLINK BATCH FILES 221
2. GENEHUNTER BATCH FILE 222
APPENDIX C. LINKAGE MAPPING SET V.2 224
APPENDIX D. GENOME SCAN EXCEL MACROS 232
1. MACRO 'RECLASS' 232
2. MACRO 'M3CLASS' 234
APPENDIX E. GENOME SCAN STAGE I RESULTS 235
1. SINGLE-POINT MLS RESULTS FOR DUBLIN CHROMOSOMES 235
2. MULTIPOINT NPL RESULTS FOR DUBLIN CHROMOSOMES 239
APPENDIX F. META-ANALYSIS BIN DESCRIPTIONS 243
APPENDIX G. META-ANALYSIS EXCEL MACROS 246
1. MACRO 'RANDCREATE' 246
2. MACRO 'FULLSHEETRC' 251
3. MACRO 'TOTALAUTO' 268
APPENDIX H. META-ANALYSIS RANKING RESULTS 276
LEGEND: 276
A) MODEL 1 277
A) MODEL 1 (CONT.) 278
B) MODEL 2 279
B) MODEL 2 (CONT.) 280
C) MODEL 3 281
C) MODEL 3 (CONT.) 282

List of Figures

Figure 1.1. Biosynthetic pathway of catecholamines from tyrosine....4

Figure 1.2. Biosynthetic pathway of serotonin from tryptophan....4

Figure 2.1. Spectral overlap of the ABI Filter Set D fluorescent dyes....55

Figure 2.2. Microsatellite stutter pattern....59

Figure 2.3. Screen capture from the MegaBase genotype management database program.60

Figure 2.4. Allele binning screen from the MegaBase database and genotype management program....61

Figure 4.1. Flowchart outlining dataflow through the data generation process from DNA samples to genetic linkage analysis....93

Figure 4.2. Multipoint MLS plots for chromosomes 2, 5, 6, 9, 10, 11, 17, 18, 19, 20, 21 and 22, from Mapmaker/SIBS....101

Figure 4.2. (cont.)....102

Figure 4.2. (cont.)....103

Figure 4.2. (cont.)....104

Figure 5.1(A-C) Positions of the nine regions on Dublin-genotyped chromosomes which exceeded the MLS threshold....115

Figure 5.1(D-F). (cont.)....116

Figure 5.2. Increase in Mapmaker/SIBS MLS from scan I to scan IIa....123

Figure 6.1. Weighted and Unweighted Summed ranks for disease model 1 (9 scans)....147

Figure 6.2. Weighted and Unweighted Summed ranks for disease model 2 (14 scans)....148

Figure 6.3. Weighted and Unweighted Summed ranks for disease model 3 (18 scans)....149

Figure 6.4. Weighted and Unweighted Summed ranks for disease model 4 (18 scans)....150

List of Tables

Table 1.1. Twin studies of bipolar disorder showing pairwise concordance rates for monozygotic (MZ) and dizygotic (DZ) twins.....15

Table 2.1 Thermal cycling conditions for the amplification of genome scan microsatellite markers.....56

Table 2.2 Components of the polymerase chain reaction for amplification of the Linkage Mapping Set v2 microsatellite markers.....56

Table 3.2. Thermal block cycling conditions for amplification of the TH gene polymorphism.....77

Table 3.3. Constitution of the polymerase chain reaction for amplification of the TH gene polymorphism.....78

Table 3.4. Published and observed allele frequencies of TH tetranucleotide repeat polymorphism.....79

Table 3.5. Transmissions and non-transmission of TH polymorphism alleles determined by the ETDT program.....79

Table 3.6. Results of the ETDT analysis of the TH gene polymorphism.....80

Table 3.7. Results of the TDT/S-TDT combined analysis of the TH gene polymorphism.80

Table 4.1. Genome scan disease affection status models.....87

Table 4.2. Individuals Genotyped for the stage I genome scan.....88

Table 4.3. Power study results from simulation studies.....97

Table 4.4 Details of genotyping for the 12 Dublin chromosomes.....99

Table 4.5. Marker details for the twelve Dublin chromosomes.....99

Table 4.6 Regions from first stage genome scan showing highest Mapmaker/SIBS maximum likelihood scores for regions passing the threshold.....105

Table 4.7 Percentage of each chromosome demonstrating a multipoint exclusion lod > –2 from first stage genome scan.....106

Table 5.1. Second stage genome scan sample description 119

Table 6.1. Description of genome scans included in meta-analysis 136

Table 6.3. Number of affected individuals contributed by and weighting applied to the eighteen genome scan studies participating in the meta-analysis 145

Table 6.4. Top ranked bins under the two primary meta-analysis models 151

Table 6.5. Meta-analysis bins which achieved BW $p < 0.05$ under diagnostic Models 1-3 153

Abbreviations

5-HIAA:	5-hydroxyindole acetic acid
ABI:	Applied Biosystems International
BW:	Bin-Wise (p-value)
BPD:	Bipolar Disorder
BPI:	Bipolar I Disorder
BPII:	Bipolar II Disorder
BPN:	Bipolar Disorder – Not Otherwise Specified
CEPH:	Centre d'Etude du Polymorphisme Humain
DNA:	Deoxyribonucleic Acid
DSM:	Diagnostic and Statistical Manual (for Mental Disorders)
FAM:	5-carboxyfluorescin
GSMA:	Genome Scan Meta-Analysis
HEX:	6-carboxy-1,4-dichloro-2',4',5',7'-tetra-chlorofluorescein
IBD/IBS:	Identity/Identical-By-Descent/State
LD:	Linkage Disequilibrium
LMSv2:	Linkage Mapping Set version 2
MHPG:	3-methoxy-4-hydroxyphenylglycol
MLS:	Maximum Likelihood Score
NED:	A proprietary fluor from Perkin Elmer/Applied Biosystems International
NPL:	Non-Parametric Linkage
RO:	Rank Order (p-value)
ROX:	6-carboxy-X-rhodamine
RUP:	Recurrent Unipolar Disorder
SAB:	Schizoaffective Disorder – Bipolar Subtype
SDS:	Sodium Dodecyl Sulphate
TDT:	Transmission/Disequilibrium Test
TH:	Tyrosine Hydroxylase
Triton X-100:	t-Octylphenoxypolyethoxyethanol

Chapter 1. Introduction

1.1 Bipolar Affective Disorder

1.1.1 History of Affective Disorders

The classification of mood disorder as a distinct clinical entity had its origins in the early 19th century. Bipolar disorder was probably first described by Baillarget, as "folie à double forme" after observations of alternating forms of mania and melancholy (Whybrow et al, 1984). The first empirical studies were performed by Emil Kraepelin, classifying psychoses into "manic-depressive insanity", and "dementia praecox" (Kraepelin, 1921). These classifications developed into the modern classification schemes used in psychiatry. They were based strongly on Cartesian concepts of scientific observation and mind-body duality, distinguishing somatic or hereditary aetiology from "psychogenic" or environmentally occasioned illness. This approach appears simplistic but was necessitated then, as now, partly by the lack of theories of mind which could successfully explain the interaction of environmental and biological causes of mental illness.

1.1.2 Clinical Features and Diagnosis

Bipolar affective disorder (BPD) as currently classified is a disorder of mood in which the sufferer undergoes alternating periods of clinical depression, and manic, euphoric "highs". Typically, untreated episodes of depression last a few months, with manic phases being shorter. Episodes of illness are usually separated by symptom-free periods. The depressive phase is typically characterised by an inability to experience pleasure and an overwhelming sadness that may evolve into despair and numbness of emotion. It is accompanied by a loss of motivation and sense of self-worth, inactivity, fatigue, and a decreased appetite. In mania the sufferer experiences euphoric happiness, often with an underlying irritability. There is excessive drive and energy and a loss of usual inhibitions.

The diagnosis of affective disorders is made by structured or semi-structured assessments according to operational diagnostic criteria, such as the DSM IV (Diagnostic and Statistical Manual for Mental Disorders, version IV; American Psychiatric Association, 1996) and ICD-10 (International Classification of Diseases, 10th edition; World Health Organisation, 1992) booklets. These are symptom checklists, and diagnosis is accomplished through interviews with patients, often including information from other family members. Diagnosis is principally based on a phenomenological approach assessing mood, duration of episode, severity (psychoticism) and response to environmental intervention. Other criteria, or axes, which are taken into account but given less weight, include patterns of personality or developmental dysfunction, psychosocial stressors, adaptive function, and other physical disorders (Whybrow et al, 1984). The latter is important as depression often accompanies other illness (Gelder et al, 1991). A distinction between endogenous and exogenous affective disturbance for the purposes of research may be a difficult one, as besides being causative in and of themselves, environmental stressors could act as precipitating agents which trigger affective disorders in people with an existing underlying genetic susceptibility (Winokur et al, 1993; Owen et al, 2000).

1.1.3 Biological features

Some of the symptoms occurring in affective disorders might be considered to have a clearly biological or somatic origin. Such features in depression include psychomotor retardation, reduced energy or lethargy, sleep disturbances (insomnia and early waking), appetite and weight loss, loss of libido, constipation and, in women, amenorrhoea (Gelder et al, 1991; Whybrow et al, 1984). These symptoms are not universally present, but seem to be more frequent in severe depression, and insomnia and loss of energy are particularly characteristic of more severe and more familial cases (Negrão, 1997). In mania, hyperactivity and agitation are common symptoms. Sleep disturbances (less need for

sleep), increased libido, and increased appetite (sometimes with weight loss) are also observed. However, given the lack of obvious underlying cause to such a wide range of symptoms, the leading theories of the underlying biology are principally derived from successful pharmacological interventions.

1.2 Aetiology of Bipolar Disorder

1.2.1 Pharmacological clues

Monoamine hypotheses

In the early 1930s amines were first established as neurotransmitters through the work of Henry Dale on adrenaline and noradrenaline (Kandel et al, eds. 1995). There are two predominant types of amines that serve as neurotransmitters, distinguished chemically as having a catechol or an indole core structure, and called catecholamines or indolamines. The former include dopamine, adrenaline and noradrenaline, and are synthesised from tyrosine in part by common biochemical pathways (figure 1.1). The indolamine serotonin is synthesised from tryptophan by a similar set of enzymes (figure 1.2), which are not normally co-expressed with catcholamine pathway elements in the same cell. It is these enzymes that in normal circumstances control production of the biogenic amines by limiting the rate of pathways' chemical reactions.

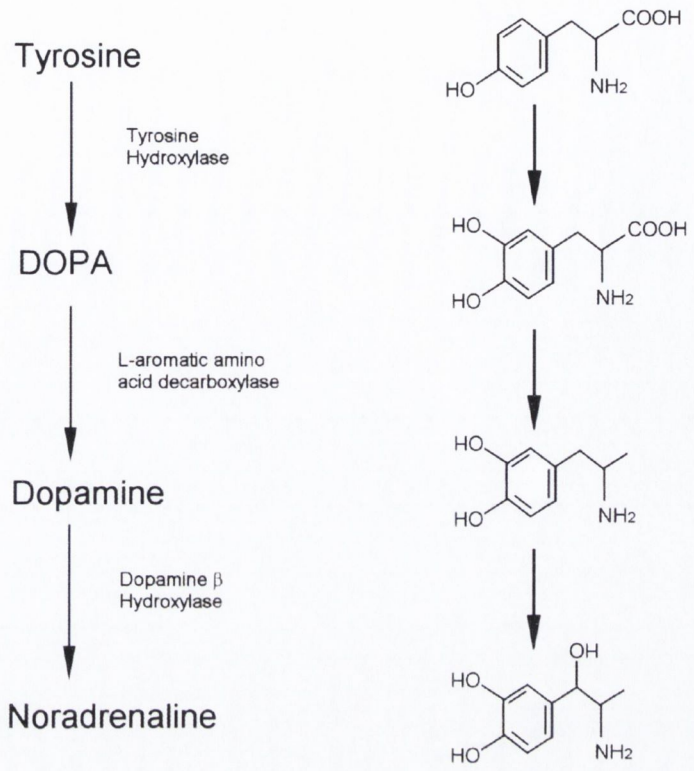


Figure 1.1. Biosynthetic pathway of catecholamines from tyrosine (adapted from Whybrow et al, 1984)

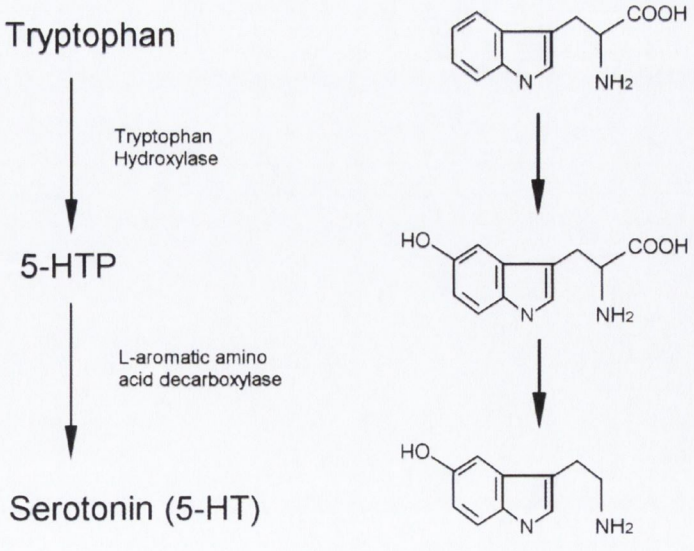


Figure 1.2. Biosynthetic pathway of serotonin from tryptophan (adapted from Whybrow et al, 1984)

Theories of catecholamine and indolamine disturbance leading to mental disorders were elaborated in the 1960s, as drugs targeting particular components of synaptic signalling revealed their usefulness in treating depression and/or mania. This led to the hypothesis that the apparent opposite nature of the manic and depressive episodes of BPD could be due to functional excess or depletion of one or more neurotransmitters in neurons (Schildkraut, 1965).

Classical antidepressant drugs have the effect of increasing amine availability at the synapse. Tricyclic antidepressants block re-uptake of amines from the synapse, and Monoamine Oxidase Inhibitors (MAOIs) prevent amine destruction. The antidepressant effect is also noted with the use of presynaptic receptor blockers, which increase amine availability by inhibiting the negative feedback of neurotramsmitter release (Whybrow et al, 1984). Thus it was hypothesized that the opposite symptoms of depression and mania could result from functional depletion of monoamines in depression versus functional excess in mania (Schildkraut, 1965). It has been shown that postsynaptic dopamine receptor blockade can improve mania, although the case for adrenergic receptors is more confused (Lecrubier et al, 1980; Petrie et al, 1982). Conversely, inhibiting the synthesis of noradrenaline can induce depression or improve mania. Precursor loading using either tryptophan or DOPA have been explored as possible treatments, although these do not seem to have specific mood-altering effects. There is evidence that serotonin may act in an opposite manner, as increasing its availability may improve mania (Whybrow et al, 1984). The picture is complicated by the observation that not all patients respond to these pharmacological agents, although it is unclear if this reflects a difference in underlying psychopathology or differences in pharmacokinetic factors related to metabolism of the drug (reviewed by van Praag, 1977). Similarly, it has not been established if these

pharmacological actions directly treat the biochemical disturbance causative of affective disorders, or if the destabilisation of monoaminergic neurotransmitters is secondary to a more fundamental pathogenesis.

Endocrine theories

Theories of endocrine involvement in affective disorders have been described in the literature over many years, based initially on mood disturbances often seen in thyroid and adrenal gland dysfunction. It is clear from many endocrinopathies that peripheral endocrine function can have effects on mood (Whybrow et al, 1984). Depression is occasionally seen in both hypo- and hyperadrenalism (Addison's disease and Cushing's syndrome), in hypopituitarism, acromegaly, and hyperparathyroidism (Gelder et al, 1991). However, increases in schizophrenic and other psychiatric symptoms are also observed. Endocrine disturbances may also be responsible for affective symptoms during menstruation, menopause and puerperally (*post partum*) in women.

Chronically increased serum cortisol levels are seen in about half of clinically depressed patients, although this seems to be the case in mania and schizophrenia also (Christie et al, 1986). About 20-40% of depressed patients exhibit a rapid recovery from the inhibitory effect which dexamethasone has on serum cortisol and adrenocorticotropic hormone (ACTH) levels, the so-called dexamethasone suppression test (Carroll et al, 1968). This effect seems to be more frequently observed in patients with more biological symptoms such as sleep, psychomotor and appetite disturbances, and disappears on recovery from the depression (Carroll et al, 1981). This has been hypothesised to be due to low noradrenaline levels, which exert a tonic inhibitory control of cortisol-releasing hormone (CRH) levels in the hypothalamus (Sacher et al, 1980). Other evidence for involvement of cortisol includes

triggering of euphoria or mania, occasionally followed by severe depression, during corticosteroid treatment.

Evidence of the involvement of thyroid hormones comes from the known anti-thyroid activity of the mood-stabilizing agent lithium (Rogers and Whybrow, 1971), observed increase in serum thyroxine during depression, and links between the thyroid and catecholamine systems (Whybrow et al, 1984). It has been reported that in approximately a third of patients there is a blunted response of thyroid-stimulating hormone (TSH) to thyrotropin-releasing hormone (TRH), which seems to correlate with better response to medication (Whybrow et al, 1984). Glucocorticoid alteration of adrenergic receptor sensitivity in affective disorders, possibly via serotonergic and noradrenergic systems, has also been observed, but a better understanding of how these systems interact is required (Horton and Katona, 1991).

Lithium

Lithium stabilises mood, treating both depression and mania, and has the unexpected effect of reducing amine neurotransmitter storage and release. However, it also has a range of other biological effects, including on ion homoeostasis and on second messenger pathways (Singer and Rotenberg, 1973).

An inositol depletion mechanism of action for lithium has been proposed (Berridge et al, 1982), which suggests a therapeutic action resulting from inhibition of inositol monophosphatase (Sherman et al, 1986). This would reduce cellular free inositol and block inositol trisphosphate(IP_3)-based second messenger signalling systems, which may be hyperactive in BPD (Horrobin and Bennett, 1999). However, this hypothesis remains unproven. The effects of lithium on inositol appear to occur well before clinical

improvement is seen, and may be limited to neurons with intrinsically low inositol levels or high phosphotidylinositide turnover (Jope, 1999b; Vawter et al, 2000). In addition, the antidepressant valproate, which can act synergistically with lithium in treating depression, seems to have very different effects on IP_3 (Dixon and Hokin, 1997). Changes have been observed in BPD in proteins downstream of this system such as protein kinase A (cAMP-dependent protein kinase) and specific isozymes of protein kinase C (Vawter et al, 2000).

The action of lithium, as with several antidepressants, could be explained by their alteration of many cellular mechanisms simultaneously (Popoli et al, 2000; Jope, 1999b). Lithium has been shown to affect almost every neurotransmitter system examined, at multiple receptor sites (Jope, 1999a). It has also been shown to have effects on several other components of cellular second messenger systems including intracellular calcium response, G-protein activity, cyclic guanosine monophosphate (cGMP), and nitric oxide production, as well as on the cytoskeleton (possibly impacting on neuronal plasticity) (Jope, 1999b). It has also been shown to affect expression of several genes, it has a well documented effect on the thyroid hormone, and it can affect the period of the circadian rhythm in rats (Hafen and Wollnik, 1994). Lithium has also been shown to have neuroprotective effects in the long term against glutamate excitotoxicity (Dixon and Hokin, 1997; Nonaka et al, 1998) via NMDA-class glutamate receptors. Similarly, antidepressant drugs have been shown to alter intracellular protein phosphorylation in a specific manner, again with implications for the cytoskeleton, and in the long term receptor sensitivity and number (Popoli et al, 2000). These varied effects are proposed to act synergistically to ameliorate affective disturbance via a neuroprotective effect, and stabilisation of multiple biochemical pathways (Jope, 1999b; Popoli et al, 2000).

1.2.2 Cellular and Physiological Changes in Bipolar Disorder

Protein alterations

Traditional theories which inferred depletion or functional excess of monoamines in affective disorders have been altered to take account of reports of alterations in post-synaptic receptor levels, particularly β1 and α1 adrenergic receptors (Hirschfeld, 2000). The discrepancy between biochemical changes observable within minutes of pharmacological intervention, and clinical improvement which can take 10-14 days, suggests that gene expression is involved, leading to changes in membrane protein levels or synaptoplastic alterations. As a result, research has recently begun to focus on protein and physiological changes.

It has been suggested that alterations in postsynaptic receptor sensitivity or number could be involved in affective disorders (Ashcroft et al, 1972; Bunney et al, 1977). A depletion of neurotransmitters presynaptically could lead to an increase in the sensitivity or amount of postsynaptic receptors in an attempt to compensate. This has been proposed to lie behind the phenomenon of mood switching in BPD, where a rapid recovery in neurotransmitter levels due to medication could induce a switch in mood from depressed to manic (Bunney et al, 1977).

Protein expression studies have suggested alterations in neural cell adhesion molecules such as N-CAM, and in proteins associated with neurotransmitter release at the synapse (Vawter et al, 2000). Also alterations have been observed in the levels of various G-proteins, vital enzymes which couple receptor signalling to intracellular pathways, specifically increases in $G_{\alpha s}$. Changes have been shown in the adenylate cyclase/cyclic adenosine monophosphate (cAMP) second messenger system, and in expression of various neuropeptides including vasopressin, oxytocin, corticotropin-releasing hormone and

neuropeptide Y. Such studies are typically conducted in small sample sizes, and are often not replicated.

Physiological changes

Several physiological alterations have been observed in BPD, principally from post mortem investigations, but also from neuroimaging studies, where the technology is undergoing rapid advances. On a neuroanatomical level, the most striking findings are decreases in glial densities in the cortex (Drevets et al, 1998), reductions in cortical thickness and possible cerebellar abnormalities. A decrease in the volume of the basal ganglia has been observed in major depression (Vawter et al, 2000). There is also evidence that hypercortisolaemia, observed in some cases of affective disorder, may cause abnormalities and inhibit neurogenesis in the hippocampus.

Some experiments have shown increases in intracellular calcium response to stimulation in peripheral cells of bipolar patients (Hough et al, 1999; Suzuki et al, 2001). If these changes are present in neurons, it could be of significance, as controlled intracellular calcium release is integral to release of neurotransmitters at the synapse, and postsynaptically is important in activation of second messenger systems, particularly IP_3/DAG-mediated stimulation of protein kinase C. There has not been much research in this area on neurons, or *in vivo*.

Serum cortisol, like many biological markers follows a circadian rhythm, which appears disturbed in depression (Whybrow et al, 1984), following a phase-shifted cycle. The aetiology of Seasonal Affective Disorder (SAD) which involves mostly depressive symptoms, is thought to result from the circannual shifts in daylight. In SAD 75% of

sufferers are female, and there is some evidence that exposure to simulated sunlight can improve mood (reviewed by Whybrow et al, 1984, pp-161-164).

Other physiological and biochemical theories include ionic changes in affective disorder due to reduction in the activity of the Na-K ATPase membrane protein, central to a theory of neuronal excitability in affective disorder (El-Mallakh and Wyatt, 1995), and alterations in the fatty acid cycle and phospholipid metabolism (Horrobin and Bennett, 1999).

1.2.3 Difficulties with Biochemical Research

There are experimental difficulties inherent in achieving specific pharmacological activities with drugs, and in making accurate measurements of brain neurotransmitter levels. Attempts at measuring neurotransmitter levels often focus on urinary and cerebrospinal fluid measurements of their metabolites. These measurements reflect mainly peripheral neurotransmitter levels, and are highly dependent on many factors including physical exertion (Ebert et al, 1972) and circadian rhythms (Whybrow et al, 1984). Measures of monoamine metabolite levels (e.g. MHPG and HIAA) may be taken as an indication of functional excess or depletion of these neurotransmitters. Inconsistent findings suggest these are unreliable, probably due to the difficulties in properly controlling for natural variation in levels. A more direct source of data on central nervous system neurotransmitter levels is post-mortem studies of brain matter. However, small and randomly ascertained samples, and the rapid metabolism of amines after death will often preclude studies of a specific pharmacological agent. Diagnosis is also a problem, as often suicide victims are used for studies in depression, and retrospective diagnosis of affective disturbance may be a source of experimental error (Whybrow et al, 1984).

The research on pharmacological targets, apart from problems with access to the relevant systems, has highlighted biochemical pathways that are often ubiquitous. If problems with the cAMP second messenger system, G-protein expression, or even alterations in dopamine or serotonin signalling were central to affective disorder, other symptoms might be expected, in the CNS or peripherally. The techniques used are not sensitive to very specific biochemical alterations, or those highly localised to particular regions of the brain. For these reasons other approaches have been increasingly favoured, which would be able to detect a specific alteration, without restricting the search on the basis of a prior hypothesis. The techniques of genetic mapping represent one such approach.

1.3 Epidemiology of Bipolar Disorder

1.3.1 Familiality

Rates of disease

Estimates of the lifetime risk for BPD in the general population are of approximately 1% (Robins et al, 1984). Worldwide, there is some reported variation in the rate of BPD, with higher incidence in far eastern countries, the USA and perhaps in central European nations (Weissman et al, 1996; Negrão, 1997). These variations may reflect local differences in diagnosis and treatment and/or differences in social or environmental factors involved in BPD aetiology. A birth cohort effect has been noted for BPD, with an increased lifetime risk observed in younger cohorts (Robins et al, 1984; Rice et al, 1987). This may be due to a decreased age of onset in the younger cohort, possibly involving genetic anticipation (see section 1.4.4), social factors, or a true increase in lifetime prevalence (Craddock et al, 1995).

Sex differences in rates of affective illness have been consistently reported, with females more at risk of unipolar illness and to a lesser extent, bipolar disorder (Gershon et al, 1975;

Gershon et al, 1976; Angst et al, 1980; Gershon et al, 1982; Winokur et al, 1993), although some studies have observed no difference between the sexes in rates of bipolar illness (e.g.: Robins et al, 1984; Weissman et al, 1984; Weissman et al, 1996). Some reports suggest relatives of female probands are more at risk of suffering from affective illness (Gershon et al, 1976; Angst et al, 1980; Weissman et al, 1984), although this observation is also not consistent.

Familiality

It is now well established that BPD is familial. Examining the rates of disorders in relatives of various degrees can help determine the extent of familial (and therefore potential genetic) versus non-familial factors, and possibly provide some insight into the genetic mechanism and mode of transmission (Risch, 1990a; Craddock et al, 1995). In addition, family studies can allow distinction between a model of continuity of aetiology for affective spectrum disorders, and a model of distinct genetic entities which "breed true" (Blehar et al, 1988). First degree relatives of probands with BPD appear to have between a 2-fold and a 15-fold increased morbidity for BPD, and are at increased risk of suffering from bipolar I or II disorder, cyclothymia, schizoaffective disorder and unipolar depression (Gershon et al, 1975; Angst et al, 1980; Gershon et al, 1982; Weissman et al, 1984; Rice et al, 1987; Andreasen et al, 1987; Winokur et al, 1993; Nelson et al, 1995; Coryell et al, 1984), assuming a population lifetime risk of between 0.5 and 1.5%.

1.3.2 Adoption and Twin Studies

Adoption studies attempt to examine the extent of shared environmental risk factors in familial traits. They generally examine rates of illness of the adoptive and true biological parents of adoptee probands, or rates of illness in the offspring of affected parents, adopted by healthy foster parents. If increased disease is seen in the biological parents in relation to

the adoptive parents, this would suggest a susceptibility to the disorder due to non-environmental factors. Comparing the rates in the biological and adoptive parents of unaffected adoptees, or of adoptees with ill adoptive parents, to population risks the extent of genetic versus environmental influences can be estimated. These studies can still be subject to biases, such as the reasons why the child was adopted, and the effect of raising an adopted child on foster parents, and the selection of potential foster parents by adoption agencies (Gelder et al, 1991). They also do not allow for the possible influence of pre-adoption or pre-natal factors on later affective illness.

Examination of biological and adoptive parents of affective disorder adoptee probands has shown an up to 15-fold increased risk of affective disorders in the biological parents of bipolar patients compared to the adoptive parents (Mendlewicz and Rainer, 1977; Wender et al, 1986). One study has found milder effects in general affective disorders (possibly due to a large excess of affectively ill adoptive fathers) (Von Knorring et al, 1983). Conversely, in a study of adopted-away offspring of affectively ill biological parents, excess incidence of affective illness has been seen in the children (Cadoret, 1978). Overall, adoption studies support a genetic hypothesis, in that increased risk is conferred by the biological parents rather than the adoptive parents.

The aim of twin studies is to estimate the magnitude of the heritability of a trait, by examining incidence in pairs who share 100% (identical or monozygotic) or 50% (fraternal or dizygotic) of their genetic material. Twin studies usually examine trait concordance rates between monozygotic (MZ) and dizygotic (DZ) pairs, where one of the twins was selected as a proband. For a fully penetrant genetic trait determined by a single dominant genetic locus, it would be expected that concordance would be 100% in MZ pairs, and 50% in DZ pairs. For a trait entirely determined by environment, concordance rates should

be the same on average in both groups. By examining a range of traits, the extent of shared familial factors can be estimated. Often a control group of twins ascertained through an unaffected proband is used to establish population prevalence and control for the influence of having a identical co-twin. Occasionally the study may be able to contrast pairs of twins raised together or apart (adopted twin study), which can serve as an additional control for environmental factors.

Studies based on the incidence of BPD in DZ and MZ twins of probands have revealed a significant risk component shared by identical twins (reviewed by Gershon et al, 1976). Table 1.1 summarises the results of ten twin studies conducted between 1928 and 1986 (Allen, 1976; Torgesen, 1986). The MZ twin concordance rate for BPD appears to be approximately 65-70%, with a wide range across studies, and for DZ twins between 0 and 25%, indicating a substantial heritable component to the disorder. Some studies found much lower DZ and MZ concordance rates (Kringlen, 1967; Pollin et al, 1969; Allen et al, 1974), which may be due to small samples, or other factors related to the study design which are not clear.

Reference	Diagnosis	No. MZ pairs	MZ concord (%)	No. DZ pairs	DZ concord (%)
Luxenburger, 1928; 1930	Bipolar	4	75	13	0
Rosanoff et al, 1935	Bipolar	23	70	67	16
Kallmann, 1953	Bipolar	27	96	58	26
Slater, 1953	Bipolar	3	67	13	23
Da Fonseca, 1959	BP+UP	21	71	39	38
Harvald and Hauge, 1965	Bipolar	15	67	40	5
Kringlen, 1967	Bipolar	6	33	20	0
Pollin et al, 1969	BP+UP	24	4	58	0
Allen et al, 1974	Bipolar	5	20	15	0
Bertelsen et al, 1977	Bipolar	27	74	36	17
Torgesen, 1986	Bipolar	4	75	6	0

Table 1.1. Twin studies of bipolar disorder showing pairwise concordance rates (not age-corrected) for monozygotic (MZ) and dizygotic (DZ) twins. [Adapted from Allen, 1976]

1.3.3 Segregation Analysis

Segregation studies examine the familial incidence and transmission patterns of affective disorders in an attempt to elucidate the underlying genetic model (Farrer and Cupples, 1998). This is a relatively easy exercise in animal studies where breeding can be controlled. In human populations it must be achieved by observation of natural matings. The sibling relative risk, λ_s, can give a rough estimation of the underlying genetic model in its simplest form: for an autosomal dominant disorder this would be $1/2x$, for autosomal recessive $1/4x$, and for a multifactorial disorder approximately $1/\sqrt{x}$, where x is the population prevalence of the disorder (Emory, 1986). For bipolar disorder, with an estimated population prevalence of about 0.01, the latter provides the best fit ($\lambda_s = 10$). Segregation studies try to fit a number of genetic, environmental and mixed models to the data and estimate a number of parameters such as disease allele frequencies and penetrances. The genetic models tested usually include a single major locus (SML) and multifactorial (epistatic or additive) models. Often a combined model is also fitted, comprising a large effect by a single locus, plus a polygenic background of small effect.

Segregation studies on some data sets have suggested that an SML (necessary and sufficient for developing BPD) model best fits familial transmission data (Gershon et al, 1976; O'Rourke et al, 1983; Nelson et al, 1995; Spence et al, 1995). Spence et al (1995) found that a single dominant locus best fitted their data, and this evidence appeared to be diluted by reclassification of unipolar patients as affected. Rice et al, (1987) found some evidence for an SML combined with background multifactorial transmission, and further analysis of the data set incorporating age of onset data suggested an additive polygenic model. A multiplicative model with 3 or more loci was found to best fit the range of estimated lifetime prevalence in the general population and in parent/offspring, sibling and

monozygotic twin relations of probands (Craddock et al, 1995). This study concluded that an SML model which confers risks to first degree relatives broadly compatible with those observed, could not also account for the large disease prevalence in monozygotic twins. Similarly no model of genetic heterogeneity was compatible with the lifetime prevalence estimates. The inconsistent findings may suggest different aetiologies at work in different sample populations due to either genetic or cultural/environmental differences, and are reflective of the variability in relative risk estimates from family studies.

Sex linkage has been extensively investigated for linkage to BPD, beginning with a report of cosegregation of X-linked colour blindness with manic-depression (Mendlewicz et al, 1972; see section 1.5.2). Linkage studies remain inconclusive, as repeated studies have found evidence for heterogeneity, or no evidence for linkage to this region. One segregation study found evidence for X-linked transmission with genetic heterogeneity, suggesting a proportion of families may carry an X-linked susceptibility gene (Risch et al, 1986). However, the presence of father-to-son transmission in many datasets precludes a dominant X-linked major locus, in the absence of heterogeneity (Hebebrand, 1992).

1.3.4 Diagnostic Subclassifications and Heritability

Many studies have addressed the issue of overlap of the genetic component of bipolar and unipolar (major depression) affective disorders. It had been suggested early on that bipolar and unipolar disorders should be considered as separate entities (Leonhard, 1979). Studies have consistently shown a larger familial component to bipolar illness (Perris, 1966; Angst, 1980). Unipolar disorder generally appears to have a lower concordance amongst monozygotic twins compared to BPD (Allen, 1976; Torgesen, 1986). Values for heritability (proportion of genetic contribution to phenotypic variance) for major depression range from 54 (Torgesen, 1986) to over 70 (McGuffin et al, 1996). High

heritability of unipolar disorder may be most correlated with episode duration, occurrence of multiple episodes, and endogenous depression as defined in ICD-9 (World Health Organisation, 1992; McGuffin et al, 1996).

Relatives of unipolar probands are at increased risk for depression, but their risk for bipolar disorders are no different from the population rates (Gershon et al, 1975; Angst et al, 1980; Gershon et al, 1982; Coryell et al, 1984; Weissman et al 1984; Rice et al, 1987; Andreasen et al, 1987). This has led to suggestions that unipolar relatives of BPD probands have a more genetic form of the illness (Blacker and Tsuang, 1993; Blacker et al, 1996). These patients appear to differ from unipolar relatives of control individuals in later age at onset, and higher incapacitation and psychoticism measures, and are more readily identified by modified RDC criteria (Blacker et al, 1996).

Bipolar II disorder (BPII) is defined as major depression with at least one episode of hypomania. It has been investigated in a family study (Coryell et al, 1984), which suggested that relatives of BPII probands have an increased morbid risk for BPII. While this study also presents evidence for a spectrum of disease severity, it suggests that BPII may be a distinct genetic entity.

Schizoaffective disorder represents an illness where the patient shows simultaneous schizophreniform and major depressive or manic disorder symptoms, and which cannot be classified as schizophrenia. The two forms, depressive and manic, have been separated by a family study showing that patients with the depressive subtype may have a form of the disorder more closely linked to schizophrenia than those with the bipolar subtype (Andreasen et al, 1987). Family studies have indicated that relatives of schizoaffective probands are at increased risk of both BPD and, to a lesser extent, schizophrenia (Gershon

et al, 1982), and that the genetic component of schizoaffective illness is at least as large as that of bipolar I disorder (reviewed by Bertelsen and Gottesman, 1995). There have been suggestions that schizoaffective disorder may breed true, however the indications from family studies are that relative risks for bipolar disorder or schizophrenia, or both, are increased significantly more than risk for schizoaffective disorder itself. As part of the continuum theory of major psychiatric disorders, it has also been represented as a "bridge" between schizophreniform and affective disorders (Crow, 1986). However, the familial evidence for this remains limited (Bertelsen and Gottesman, 1995).

Current theories of the classification of affective disorders include a spectrum theory of unipolar depression, bipolar II and bipolar I disorder (Coryell et al, 1984; Blacker et al, 1996), probably also including schizoaffective disorder at the higher end of severity (Bertelsen and Gottesman, 1995). It has been suggested that unipolar and bipolar affective illness may represent different thresholds of quantitative risk factors (Gershon et al, 1975), possibly including schizoaffective disorder as a third, higher threshold.

1.3.5 Summary of epidemiological research

It has been pointed out that many of the twin and other epidemiological studies were performed using diagnostic criteria and procedures that would be considered imprecise by today's standards (Reus and Freimer, 1997), affecting the estimates of heritability. Stricter methods and criteria in use today may reflect a more biological definition of BPD (Owen et al, 2000).

Studies concur that there is a heritable risk factor or factors for bipolar affective disorder, and other affective spectrum disorders, notwithstanding the variation between studies in the observed magnitude of the effect. The twin and adoption studies, taken together

suggest that the increased morbid risk to relatives of patients is mostly genetic in origin. A measure of heritability, h^2, can be made for the risk for affective disorder to relatives of probands, based on the observed correlation between first degree relatives (Gershon et al, 1982). These suggest an estimated heritability of 78% for a diagnosis of affective disorder, including schizoaffective bipolar (I and II) and unipolar disorders. Non-genetic familial factors may account for some of this proportion, so this represents a maximum value for the proportion of susceptibility to affective disorders attributable to genetic loci. The next question is, what is the nature of this genetic component? Segregation studies and mathematical models indicate that a single major locus conferring susceptibility to the disease is unlikely in most populations. An epistatic model of three or more loci appears to be more probable, although the number and nature of the individual genes and their likely interactions are unknown.

It is possible that the disorders in the affective spectrum share susceptibility loci, and linkage analyses (see section 1.5.2) are often performed under several disease classification models (Reus and Freimer, 1997). A more narrowly defined affection model might be expected to have greater power to detect a locus conferring susceptibility specifically for BPD due to the larger recurrence risks and, probably, genetic component. However, if a diagnosis spanning the spectrum of affective disorders can be applied, much larger numbers of affected individuals can be ascertained, giving higher power to detect common susceptibility loci (Blacker et al, 1996). The increased rates of major depression and other affective spectrum disorders in probands suffering from BPD suggests that combining these diagnoses into a single affected category may be a valid approach. However, the reverse is not necessarily true, e.g. depressive probands do not show an increased risk for bipolar I disorder in relatives. It has been estimated that approximately 70% of unipolar relatives of bipolar probands are themselves genetically "bipolar" (Blacker and Tsuang,

1993), and therefore the case for including these in a disease affection model is stronger. This is the approach that was taken in the current study (see section 1.8 and chapter 2, section 2.2).

1.4 Linkage and Association Statistics

1.4.1 Introduction

The science of gene mapping by genetic linkage analysis depends on Mendel's second law, of independent assortment. In modern terms this states that any two genetic traits should be passed on to offspring independently of each other. Genetic linkage refers to instances where this law is violated, i.e. where two genetic loci are in sufficient proximity to each other on a chromosome that different variants, or alleles, at the loci tend to stay together when passed to the next generation. This is due to the inverse relationship between the physical distance and the probability of a recombination event between two loci.

In linkage and association analyses a genetic polymorphism of unknown location that accounts for some phenotypic variation, known as the disease locus, is hypothesised. If certain assumptions concerning this locus are made then violation of Mendel's second law can be tested between it and a genetic marker of known location, which is genotyped by molecular methods.

In this context genetic linkage and genetic association can be seen as essentially the same phenomenon. In the former co-inheritance of the disease and marker loci over a small number of generations (usually two-three) is tested. The low probability of recombination occurring over a single generation allows widely spaced markers to be used, with several hundred covering the entire genome, for a test of the null hypothesis of no genetic locus conferring any susceptibility to the disease. In association analysis a case-control study

design is usual, where excess frequency of a marker allele in one sample versus the other is tested. This approach can be viewed as using the entire population as a large family, and trying to associate a particular marker allele with the disease allele despite the potential for many recombination events. Therefore this approach requires that the marker allele be either directly causative of the trait under observation, or in such close proximity to the disease allele that even over many thousands of generations few or no recombination events have occurred between the two, a phenomenon known as linkage disequilibrium (LD).

1.4.2 Genetic Association

LD between a genetic marker and a trait is normally detected as a difference in allele frequencies at a genetic locus between trait-bearing persons, and the general population as determined from a control sample. The standard statistical approach is to apply a chi-square analysis to the allele frequencies in the case versus control populations. These case-control studies are vulnerable to detecting association due to the presence of hidden population substructures. If the population from which the sample and controls are taken contains two recently admixed populations, with different allele frequencies and disease frequency, or if the cases and controls are not properly matched, it is possible to detect allelic association, even in the absence of a genetic component for the trait (Hamer and Sirota, 2000).

To avoid the problems of population stratification, methods using internal controls have been developed. Although these may be susceptible to certain types of bias for disorders with late age of onset (Bruun and Ewald, 1999), they ensure patients are ethnically and sociodemographically matched to the controls. The haplotype relative risk (HRR) test uses parents of the affected individual as controls (Falk and Rubinstein, 1987). The statistic is

based on the fact that alleles not transmitted to the affected individual reflect normal population allele frequencies, and is constructed in an equivalent way to the odds ratio of Woolf (1955). The transmission/disequilibrium test (TDT), formalised by Spielman et al (1993) evaluates the frequency of transmission of one allele versus the frequency of transmission of the other allele. A χ^2 test, also known as a McNemar Test, can then be applied which uses only transmission information from heterozygous parents, and therefore the statistic requires the presence of both linkage and linkage disequilibrium. Sham and Curtis (1995) extended the test (ETDT) to a multiallelic system using logistic regression. The TDT has been successfully applied to the fine-mapping of susceptibility genes for complex disorders, such as the insulin gene to insulin-dependent diabetes mellitus (Spielman et al, 1993). An application of this method to BPD is described in chapter 3.

1.4.3 Genetic Linkage

Lod score method

Genetic linkage analysis depends on the non-independent co-segregation of two loci in close physical proximity on transmission from one generation to the next. A simple measure of genetic linkage can be calculated by counting the frequency of recombination events between any two loci. Alleles at independently segregating loci should be co-transmitted 50% of the time. Deviation from a 50% recombination rate indicates that the loci are linked. The basis of linkage studies attempting to localise disease loci requires estimation of disease allele frequencies and penetrance (the probability that each genotype will lead to expression of the trait) of a hypothetical disease gene, and examination of linkage between this, and genetic markers of known location in the genome.

The magnitude of the linkage is estimated by the recombination fraction, θ. This is the observed probability of a recombination event, and is roughly representative of physical distance. It takes values between 0 (complete linkage), and 0.5 (independent segregation; no linkage). The detection of recombination events depends on inference of phase, i.e. the ability to trace an allele unambiguously from one generation to the next. Parents should be heterozygous at both (disease and marker) loci, so that it is apparent that a specific allele at one locus co-segregates with a specific allele at the other locus.

The lod (logarithm of the odds) score (Morton, 1955) is calculated from the likelihood ratio: the probability of the genotype data, given that θ equals a test value $x < 0.5$, divided by the probability of the data given the null hypothesis of no linkage, i.e. $\theta = 0.5$:

$$lod = \log_{10}\left[\frac{P(data \mid \theta = x)}{P(data \mid \theta = 0.5)}\right],$$

The log normalisation of this statistic means that it amenable to sequential calculation. That is, as families are added to the sample, the score can be cumulatively summed until it passes a given threshold reflecting the presence (traditionally a lod of 3) or absence (lod of -2) of linkage. Linkage is usually tested at a range of test values of θ between 0 and 0.5, and the maximum lod score, Z_{max} is quoted along with the corresponding estimate of the recombination fraction.

The probability or likelihood of the data given the null hypothesis ($\theta = 0.5$), and the alternative hypothesis ($0.5 > \theta > 0$) can be calculated from the number of observed recombinant (R) and parental, or non-recombinant (NR) offspring:

$$L = \theta^{R}(1-\theta)^{NR}.$$

This calculation can include data from phase-unknown pedigrees, where the genotypes are ambiguous with regard to recombination. In this case terms are inserted reflecting the likelihood under all possible phases, weighted either equally or using population allele frequencies.

The statistical significance of the lod score can be calculated, as the statistic is asymptotically distributed under the null hypothesis as χ^2, with degrees of freedom equal to the number of parameters maximised (usually θ), and is one tailed, as θ is only tested to be equal to or less than 0.5:

$$P(lod) = \frac{1}{2}\left[\chi^2 > (2\log_e 10) \times lod\right]$$

Therefore it can be shown that a lod of 3 corresponds to odds of approximately 20:1 in favour of linkage (significance level of 0.05).

This statistic is the most powerful method of measuring linkage, yet it functions best in Mendelian disorders, where the genetic model can be correctly defined. Traits where Mendelian inheritance cannot be established are susceptible to error in terms of model misspecification. If, for example, several genes of small effect are acting, there can be difficulties in attaining a significant score (Risch, 1990b).

It has been shown that model-dependence may not be a serious liability if approximately correct models are used, such as simple dominant or recessive (Durner et al, 1999), with

allowance for genetic heterogeneity. However, multiple testing inherent to this method requires an adjustment to the lod score threshold or significance level (Nyholt, 2000).

An approach that reduces dependence on the specified genetic model is the affecteds-only method. This method relies on a decreased estimate of the penetrance of the disease, so that only the trait genotype information from affected individuals is assumed to be known. In general, this approach is useful only for Mendelian disorders with unknown, or age-dependant penetrance values, and has been used successfully in identifying genes for both early- and late-onset Alzheimer's Disease (Pericak-Vance et al, 1991; Corder et al, 1993).

Non-parametric methods

Non-parametric analyses have potential advantages in the examination of common or complex genetic disorders, as they do not rely on the modelling of a disease gene. Deviation from the expected extent of alleles shared between affected family members at genetic loci can be statistically tested to give a measure of linkage to a trait.

Affected sibling pair analysis (Penrose, 1953) examining alleles shared identical-by-descent (IBD) is one of the most common methods in non-parametric analysis. Each person in a sibling pair inherits one allele paternally and one maternally. Siblings should share the same paternal allele 50% of the time, and the same maternal allele 50% of the time. Therefore they should share two alleles IBD 25% (z_2) of the time, one allele IBD 50% (z_1) of the time, and no alleles IBD 25% (z_0) of the time. A statistic is calculated from the deviation from this 25:50:25 ratio in a sample population, representing the extent of linkage to a disease locus. A maximum likelihood method was developed by Risch (1990c) for cases where there is incomplete inheritance information, for example if parents are missing or markers are incompletely polymorphic. This method maximises the likelihood

of the data with respect to the sharing probabilities z_0, z_1 and z_2, holding z_1 fixed at 0.5, and restricting $z_0 \leq 0.25$ (or $z_2 \geq 0.25$). Since only one parameter (z_0 or z_2) is maximised, the statistic has the same asymptotic distribution under the null hypothesis as a parametric lod score (Holmans, 1993). The restrictions on the allele sharing values represent an additive genetic model, or one of genetic heterogeneity. Dominance variance, where an allele can show a dominance effect over another is not accounted for (Nyholt, 2000).

Risch (1990b) also demonstrated that power to detect linkage with a perfectly informative marker, in a case where a single disease locus influences susceptibility depends solely on the relative risk, λ_R, which can be estimated from epidemiological studies as the ratio of the risk to a relative of type R of a patient, to the risk in the general population. This fact can be used to calculate the power of various sample sizes to detect linkage to a disease locus at different distances θ from the marker locus, given any relative risk λ_R. Thus a sample of 100 ASPs should have a power of 80% to detect linkage with a locus conferring a relative risk to a sibling, λ_S, of about 3 at $\theta = 0$, or about 6 at $\theta = 0.05$. Using 300 ASPs, 80% power can be achieved for a locus conferring a risk as small as 2.5 at $\theta = 0.1$ (Risch, 1990b).

The allele-sharing method was extended by Holmans (1993) to use the 'possible triangle' restrictions:

$$z_1 \leq 0.5$$

$$z_0 \leq \frac{z_1}{2},$$

These restrict the allele sharing values to biologically plausible values. These values allow for a multiplicative (or epistatic) genetic model, allowing dominance variance (Nyholt, 2000). Significance levels need to be adjusted, however, as the statistic is asymptotically distributed as a χ^2 with a mixture of 0, 1 and 2 degrees of freedom under the null hypothesis. This method has increased power for loci of small effect, or where information is incomplete due to low marker polymorphism or values of $\theta > 0$ (Holmans, 1993), but has slightly less power if the genetic model approximates an additive system, with z_1 close or equal to 0.5.

A non-parametric statistic for exclusion mapping was devised by Kruglyak and Lander (1995), which is based on the relationship of the sharing probabilities z_0, z_1 and z_2 to the relative risk, λ_R (Risch, 1990a) which can be calculated from epidemiological data. This calculates a likelihood ratio score based on comparison of the likelihood of the data assuming an underlying susceptibility locus with fixed λ_R, to the likelihood under the null hypothesis. A lod score can be examined for the traditional exclusion criterion of $Z < -2$, thereby excluding the existence of loci of a specified magnitude from regions of the genome. The use of negative tests such as exclusion mapping, even the non-parametric exclusion test, can be problematic as discussed by Terwilliger and Ott (1994), in the absence of knowledge about the genetic model, or in the case of genetic heterogeneity. The increased possibility of falsely excluding the location of a susceptibility gene of small effect should be cause for extreme caution in their interpretation.

In cases of incomplete inheritance information the use of identity-by-state (IBS) allele sharing may be applied, using known population frequencies of the alleles. This method is usefully applied when the genetic map may be sparse or incompletely polymorphic (Lander and Schork, 1994). Affected relative pairs can be examined for IBS sharing in the

same way as for IBD sharing, but without the assumption that shared alleles originate from the same ancestor, by weighting the occurrence of a shared allele with the inverse of its population frequency. For example a rare allele shared more often by an affected pair will be taken as more indicative of linkage than a shared allele with a higher population frequency. It becomes important to obtain an accurate estimate of the allele frequencies in the population, and the results can be sensitive to small variations in this information. These methods have shown their usefulness in the mapping of genes of minor effect for essential hypertension (Jeunemaitre et al, 1992) and late-onset Alzheimer's Disease (Pericak-Vance et al, 1991).

The NPL approach used by the computer program Genehunter (Kruglyak et al, 1996) uses a similar approach, but takes advantage of any available IBD information by examining the distribution of inheritance vectors. It also allows allele sharing to be examined between all affected individuals of a pedigree simultaneously (the NPL_{ALL} statistic), as opposed to examining each pair in turn. This increases power by giving greater weight to several affected members, or more distant relative pairs sharing the same allele, than would be given by examining them in a single pair-wise fashion. This gives greater power to detect dominant trait alleles, which would cluster in this fashion in extended pedigrees. This NPL is normally distributed in the exponent, but is unreliable in small sample sizes, or if varied pedigree structures are examined (Nyholt, 2000), and exact p-values must be obtained from sample-specific simulations. The programs Genehunter-Plus (Kong and Cox, 1997) and ALLEGRO (Gudbjartsson et al, 2000) incorporate likelihood ratio methods, producing an NPL equivalent to a standard parametric lod score.

Multipoint mapping can be used to improve the performance of linkage mapping, as information from adjacent markers can compensate in situations where there is a lack of

informative meioses (Risch, 1990b; Terwilliger and Ott, 1994). Maximum likelihood multipoint linkage is usually performed by evaluation of the ratio of likelihood of the data given a locus at fixed intervals across the map, to the likelihood with no underlying locus (Xu et al, 1998). The likelihoods are maximised and the ratio converted to a logarithmic score.

1.4.5 Genome Scans

Searching for genetic linkage to a disease, with no prior evidence for a susceptibility locus, and no information on the genetic model of the disease, can lead to certain difficulties. It has been suggested that examination of linkage to a single region or chromosome, without strong prior evidence for a susceptibility locus in the region, must be subject to the same statistical significance criteria as a genome-wide study, since a negative result will almost always lead to further regions being investigated (Lander and Kruglyak, 1995). In this context, evaluation of the relative importance of a linked locus cannot be made without knowledge of the extent of linkage at each point along the genome. Thus Lander and Kruglyak (1995) recommended that genome scans are used to extract full inheritance information from the outset, and subsequent evaluation of the most likely candidate region for a susceptibility locus be made on the basis of a p-value corrected for multiple testing in order to reflect genome-wide significance.

Statistical correction for multiple testing using the Bonferroni method (dividing significance by the number of tests) presupposes that the tests performed are independent: this is not the case in a genome scan using genetic markers in linkage with each other. Different genetic and diagnostic models are also non-independent. Generally, simulation studies are performed to determine the power and relevant significance levels for a genome scan.

Lander and Kruglyak (1995) demonstrated that a linkage significant at the 5% level at one marker is likely to occur by chance over 20 times in a 3 cM resolution genome scan of 100 affected sibling pairs, and some of these are likely to be erroneously "replicated" by type I error (false positive) in a follow-up study in a new sample set. For a genome-wide false positive level of 0.05 under these conditions, the threshold for significant linkage should be a point-wise p-value of 0.00002, equivalent to a lod score of 3.6. These strict criteria are likely, however to lead to erroneous exclusion of some true peaks (type II error). This is particularly true if weak linkage is expected, due to for example a polygenic disease model. Therefore proposed criteria, which have been widely adopted in genome scan studies, included a threshold for 'suggestive' linkage of a lod score of 2.2, corresponding to a p-value of 0.0007, expected to occur once in each genome scan. The disclosure of all loci with nominal p-values of 0.05 was also suggested to allow researchers to compare their results. Confirmation of linkage must then be established by extension of the study with a larger sample and increased marker density, and replication by independent groups.

Holmans and Craddock (1997) performed power calculations that show that these criteria may be too strict. They investigated power and efficiency of a two-stage study design employing grid-tightening, where suggestive regions are followed up in a larger sample set with a higher density of markers. They found that this design could be significantly more efficient in terms of number of genotypes required, than genotyping a full sample set for all loci at the higher marker resolution, without a significant loss in power. They also show that this strategy is relatively robust to loss of Hardy-Weinberg equilibrium due to population stratification, if parents are also genotyped in the first stage.

The use of genome scans in searching for genetic linkage has increased in recent years. It has been facilitated by technological advances in genotyping technology, principally the replacement of radiolabel-based genotyping techniques with semi-automated DNA sequencers using fluorescent technology. Furthermore, only within the past decade have high density, highly informative genetic maps suitable for genome screening been developed (Dib et al, 1996). Accurately positioned markers of high informativeness, such as microsatellites (Weber and May, 1989) have been crucial in this development.

1.5 Genetic Studies in Bipolar Disorder

1.5.1 Genetic Association to Bipolar Disorder

A review of association studies (Jones and Craddock, 2001) has shown around 50 genes have been investigated for association with BPD over the past decade. Although most of these are likely to have no effect on the phenotype, it is possible that many speculative studies on true genes have been unable to detect a small effect due to insufficient sample sizes, except in meta-analyses with large sample sizes (Risch and Merikangas, 1996).

Catechol-O-methyltransferase (COMT) is an enzyme involved in the breakdown of monoamines, and disturbances in its activity could lead to alterations in the levels of neurotransmitter present at the synapse. The gene maps to a chromosome 22q region associated with velo-cardio-facial syndrome and psychiatric disorders (Shprintzen et al, 1992; Lachman et al, 1996), and has shown a weak trend towards genetic association with female BPD probands (Mynett-Johnson et al, 1998). It may also be associated with rapid cycling in BPD patients (Jones and Craddock, 2001).

The monoamine oxidase A (MAOA) enzyme is also involved in the breakdown of monoamines, and two recent meta-analyses of association studies have shown a weak association with BPD, mainly in females (Furlong et al, 1999b; Preisig et al, 2000).

Tyrosine hydroxylase is the rate-limiting enzyme in the synthesis of the catecholamine family of neurotransmitters, and thus may be responsible for depletion or other functional disturbances in these systems. The tyrosine hydroxylase gene on chromosome 11p15 has shown some evidence of genetic association with BPD (Lim et al, 1993), however several other studies were mostly negative (Korner et al, 1994; Souery et al, 1999), including two meta-analyses (Turecki et al, 1997; Furlong et al, 1999a). Studies of the tyrosine hydroxylase gene are discussed in more detail in chapter 3.

Components of the serotonin (5-HT) signalling pathway have also been investigated due to the action of the selective serotonin reuptake inhibitor (SSRI) class of antidepressants. The serotonin transporter gene has not shown consistent evidence for association with BPD (Guttiérrez et al, 1998), and a meta-analysis has supported this (Furlong et al, 1998). Of the serotonin receptor genes, the 5-HT1A and 2A receptors have also been ruled out as a major gene by several studies (Curtis et al, 1993; Mahieu et al, 1997; Jones and Craddock, 2001)

Positional candidate studies have led to investigation of candidate genes in several regions where linkage with BPD has been found. The G-protein alpha subunit (GNAZ) gene (Saito et al, 1999) on chromosome 22q11, and the dopamine receptor 5 (DRD5) on chromosome 4p (Asherson et al, 1998) have been investigated, with negative results. The phospholipase A2 gene on chromosome 12 has been shown to be associated to BPD (Dawson et al, 1995). This is of interest given the linkage findings in the region (see section 1.5.2) and observed

co-segregation of BPD with Darier's disease (Craddock et al, 1994), the gene for which maps to this region (Craddock et al, 1993).

The gene for Brain-Derived Neurotrophic Factor (BDNF) has also been investigated for association to schizophrenia (Hawi et al, 1998) and BPD (Neves-Pereira et al, 2002; Sklar et al, 2002). Relatively strong associations between BPD and the val66met polymorphism were detected. BDNF is a growth factor expressed in foetal and adult brain, and represents both a candidate gene from a neurodevelopmental perspective, and a positional candidate for BPD due to its location on chromosome 11p13 (see section 1.5.2).

1.5.2 Genetic Linkage to Bipolar Disorder

X Chromosome

The first linkage analyses for BPD were based on serotyping of biological markers such as HLA markers and the ABO blood group (reviewed by Mendlewicz, 1994). Linkage was found to known X-linked biological markers such as Protan and Deutan Colour Blindness and glucose-6-phosphate dehydrogenase deficiency that map to Xq27, and to the Xg blood group which maps to Xp22 (Mendlewicz and Fleiss, 1974). Linkage to Xq27-q28 markers including protan and deutan colour blindness, glucose-6-phosphate dehydrogenase deficiency and blood coagulation factor IX were initially strongly positive (Mendlewicz et al, 1972; Mendlewicz and Fleiss, 1974; Mendlewicz et al, 1980; Del Zompo et al, 1984; Baron et al, 1987), but more recent research with highly polymorphic DNA markers has reduced the possible importance of any BPD locus in this region (Berrettini et al, 1990; Gill et al, 1992; Jeffries et al, 1993; Smyth et al, 1997; Vallada et al, 1998; Hawi et al, 1999), and some of the earlier studies have shown reduced evidence upon re-analysis (Baron et al, 1993; Mendelbaum et al, 1995).

It has been observed that depression may have a higher incidence in females (see section 1.3.1), however in the absence of distinct X-linked segregation patterns in BPD, it is now clear that this chromosome does not harbour a major gene for the illness (Hebebrand, 1992), although there may be a susceptibility gene acting in a sub-set of families (Risch et al, 1986; Bocchetta et al, 1994; Pekkarinen et al, 1995). In particular, evidence exists suggesting an X-linked gene may be associated with early age of onset or be subject to genetic anticipation (Risch et al, 1986; Hawi et al, 1999).

Chromosome 11

Interest in chromosome 11 was initiated by a finding of linkage between BPD and markers at the insulin (INS) and Harvey-*ras*-1 (HRAS1) loci on 11p (Egeland et al, 1987). However, attempts at replication by other groups were unsuccessful (Hodgkinson et al, 1987; Detera-Wadleigh et al, 1987; Gill et al, 1988; Mendlewicz et al, 1991) and reanalysis of the original data with updated diagnoses and an extended pedigree led to a reduction of the lod score (Kelsoe et al, 1989; Pauls et al, 1991). It emerged that individuals classified as unaffected in the original study later developed BPD, and re-classifying them as affected led to a reduction of the evidence for linkage. Weak evidence for linkage to this region has persisted in a minority of families (Lim et al, 1993) or when examined under assumption of genetic heterogeneity (Gurling et al, 1995; Smyth et al, 1996), and attention has shifted to nearby genes for Tyrosine Hydroxylase (TH) and Dopamine Receptor 4 (DRD4), which map to 11p15.5. Co-segregation with the hereditary β-thalassaemia minor gene: β-globin, which also maps to this area of the short arm of chromosome 11, has been observed in some families (Harada et al, 1995). Direct association studies with TH have yielded mixed results (see chapter 3, section 3.1.3). The DRD2 receptor on chromosome 11q (Holmes et al, 1991; Mitchell et al, 1992) has also failed to show consistent evidence of association.

Chromosome 18

Berrettini et al (1994) first reported a finding of linkage to the pericentromeric region of chromosome 18 in the course of a genome scan for linkage to BPD. Stine et al (1995) attempted replication of this finding and found excess allele sharing in the same region (centromeric 18p), and with markers on 18q12. The latter study also investigated parent-of-origin effects, due to previous findings of increased risk of BPD in maternal pedigrees (McMahon et al, 1995), and found increased paternal transmission of the 18q region. This finding has been examined in an expanded sample, and may be most pronounced in sibships where both patients have BPII (McMahon et al, 1997; McMahon et al, 2001), although parent-of-origin effects have not been found in family studies of BPII (Kato et al, 1998). Evidence for linkage has also been reported in the 18q21-q23 region by other groups (Coon et al, 1996; De Bruyn et al, 1996; Ewald et al, 1997), in some cases with an excess of paternal allele transmission (Badner and Goldin, 1997; Nöthen et al, 1999; Turecki et al, 1999), in others with maternal transmission (Gershon et al, 1996) However other studies have reported an absence of linkage to the chromosome (Pauls et al, 1995; Claes et al, 1997; Knowles et al, 1998; Bowen et al, 1999).

McInnes et al (1996) reported the finding of weak linkage to 18p, and stronger linkage to 18q in the course of a genome screen in two large Costa Rican pedigrees. The latter region overlaps many of the above findings on chromosome 18q, and was robust to a combined linkage and association analysis (Freimer et al, 1996), showing an 8 centimorgan (cM) haplotype highly shared amongst affected individuals, and to reanalysis with non-parametric statistics on genealogical combination of the two pedigrees (Garner et al, 2001). Many other genome scans have not found any evidence of linkage to chromosome 18, although the NIMH data set showed weak linkage to both 18p and 18q (Detera-Wadleigh

et al, 1997). Evidence for linkage to chromosome 18 centromeric region persists (Berrettini et al, 1997; Detera-Wadleigh et al, 1999).

As a result, and despite mixed findings, groups who have a sample showing robust linkage to chromosome 18 regions are pursuing the characterisation of candidate genes (e.g. Tsiouris et al, 1996; Berrettini et al, 1998; Kosaki et al, 2001), with interest in the G-protein sub-unit $G_{olf}\alpha$ and the human myo-inositol monophosphatase gene (IMPA2) on 18p11 (Vuoristo et al, 2000; Sjoholt et al, 2000). Fine mapping efforts are also underway on 18p (Rojas et al, 2000; McInnes et al, 2001) and two separate loci on chromosome 18q21-23 (Escamilla et al, 1999; Verheyen et al, 1999).

Chromosome 4

Two distinct candidate loci on chromosome 4 have been identified by linkage studies. A genome scan in twelve Scottish pedigrees segregating affective illness obtained a lod score of 4.1 under a dominant mode of transmission on chromosome 4p16 (Blackwood et al, 1996), which is significant by the criteria of Lander and Kruglyak (1995; see section 1.4.5). They also identified a common haplotype extending over 7 markers, spanning approximately 14 cM, in affected individuals. However, the linkage was due entirely to one extended pedigree. The criteria for confirmed linkage require replication of the finding in an independent sample set and this appears to have been achieved (Ewald et al, 1998), albeit under a recessive genetic model. A number of genome scans have found no evidence for linkage, although others have found modest linkage, to this region (Detera-Wadleigh et al, 1997; Detera-Wadleigh et al, 1999). The DRD5 receptor on chromosome 4p (Byerley et al, 1994; Asherson et al, 1998) has not yielded any evidence of linkage to BPD. On chromosome 4q35, Adams et al (1998) obtained their highest lod score in a genome scan in a large Australian pedigree. Support for the finding is sparse, but some groups have also

reported weak linkage to this general region (McInnes et al, 1996; Friddle et al, 2000). Of interest is a finding of linkage in a genome scan for mental health wellness to regions on chromosome 4p and 4q, amongst others (Ginns et al, 1998).

Chromosome 12

Chromosome 12 has become the subject of interest due to a report of co-segregation of BPD and Darier's disease, an autosomal dominant hereditary skin disorder (Craddock et al, 1994). The gene for Darier's disease has been mapped to chromosome 12q23-24.1 (Craddock et al, 1993) where linkage to BPD has also been described (Dawson et al, 1995; Ewald et al, 1998; Jacobsen et al, 1999; Morissette et al, 1999).

Chromosome 21

Non-parametric APM statistics were used to demonstrate evidence for linkage of BPD to the chromosome 21q22 PFKL (Phosphofructokinase, Liver) marker as part of an on-going genome scan (Straub et al, 1994), with strongest results from a single family. An analysis performed under a model of heterogeneity in another set of families yielded considerable lod scores when a subset of the families were considered to be linked to either the chromosome 11 TH locus, or PFKL (Gurling et al, 1995). Several other groups have confirmed the presence of weak or suggestive evidence for linkage in this region (Detera-Wadleigh et al, 1996; Vallada et al, 1996; Detera-Wadleigh et al, 1997; Detera-Wadleigh et al, 1999; Kwok et al, 1999; Morissette et al, 1999; Kelsoe et al, 2001) with suggestions of genetic heterogeneity, although one group has reported completely negative results (Byerley et al, 1995).

Chromosome 10

A genome scan for BPD has highlighted a region on chromosome 10q25-26 as showing linkage (Cichon et al, 2001a, 2001b). This region overlaps an area of linkage in a separate genome scan (Kelsoe et al, 2001), and has also been identified as a possible location for a susceptibility gene to schizophrenia (Wildenauer and Schwab, 1999). Chromosome 10p has also been identified as a possible locus of a susceptibility gene for BPD (Rice et al, 1997; Foroud et al, 2000) and schizophrenia (Schwab et al, 1998; Straub et al, 1998). These two regions are sufficiently separated as to represent distinct susceptibility loci.

Chromosome 13

Chromosome 13q32 contains a region identified as a strong candidate for a susceptibility gene for schizophrenia (Blouin et al, 1998). It may therefore be important that one genome scan for BPD has found this region to be their most significant, a finding supported in a follow-up series of pedigrees from the NIMH bipolar initiative (Stine et al, 1997; Detera-Wadleigh et al, 1999). Linkage to this same region was replicated by Kelsoe and co-workers (2001) in another genome scan.

Chromosome 22

Kelsoe et al (2001) found linkage to a region of chromosome 22q11, with a lod score of 3.8. This area has also been implicated in schizophrenia (Pulver et al, 1994; Myles-Worsley et al, 1999). Reports of suggestive linkage of BPD to this arm of chromosome 22 has also been reported by other groups (Edenberg et al, 1997; Detera-Wadleigh et al, 1999; Turecki et al, 2001).

Other Regions

Investigation of genetic linkage to BPD has also been attempted by purely speculative approaches, investigating regions adjacent to candidate genes such as the DOPA decarboxylase gene on chromosome 7p (Ewald et al, 1995), the β2 adrenergic, serotonin 1A and dopamine 1 receptors on chromosome 5 (Mirow et al, 1994; Curtis et al, 1993; Mitchell et al, 1992), and the dopamine 3 receptor gene on chromosome 3q (Mitchell et al, 1993). None of these have produced any significant evidence of linkage.

1.6 Genetically Complex Disorders

1.6.1 Introduction

Several factors can complicate or alter the familial transmission of a trait from the expected patterns of single gene traits. If a single gene is neither necessary nor sufficient for a given phenotype to be expressed, the pattern of transmissions will deviate from those expected with a simple autosomal dominant or recessive gene. In the most basic case, a single gene may be responsible for a trait, but incomplete penetrance may be observed, where the genotype does not always give rise to the expected phenotype, due to environmental or chance factors. A similar confounding factor is the presence of phenocopies, where the phenotype arises entirely due to non-genetic factors. A number of separate genes may be necessary for the expression of the phenotype, and these can interact in various ways. Other factors such as trinucleotide repeat expansion (giving rise to anticipation), bilineal transmission (arising from non-random mating), sex-specific transmission (due to X-linkage or imprinting), and any combination of gene-environment interactions can mask or skew simple Mendelian transmission patterns, and some of these are hypothesised to be acting in BPD.

1.6.2 Genetic Heterogeneity

Genetic heterogeneity is a feature whereby several different polymorphisms are associated with the same phenotype. One form of this is allelic heterogeneity, where different polymorphisms or mutations at the same locus separately affect disease risk. This usually will not affect linkage analysis, as the markers utilised will not be able to differentiate between the separate alleles. However, it can have an effect on association studies (Terwilliger and Weiss, 1998). Locus heterogeneity, where different genes underlie the trait in different families or sample populations, can dilute the power of the linkage statistic. This problem might be circumvented by carefully selecting a ethnically homogeneous sample population, collecting single large, extended pedigrees segregating the disorder, or ensuring an even distribution in the probands for as many other risk factors as possible. However, the problem remains that any susceptibility locus detected may only be important in the population sub-section examined. Linkage statistics have been developed which allow the examination of linkage where only a proportion, α, of the families are assumed to be linked (Ott, 1999). An example of extreme locus heterogeneity is retinitis pigmentosa (RP), where at least 22 separate loci can cause the same phenotype (Inglehearn, 1998). Several reports have suggested genetic heterogeneity may be present in BPD (Morissette et al, 1999; Friddle et al, 2000), however other investigations have indicated this may be the exception rather than the rule in the majority of pedigrees (Craddock et al, 1995)

1.6.3 Gene-Gene Interactions

Genes of small effect, which individually are neither necessary nor sufficient to cause the disease, can interact in two ways. An additive effect involves several distinct loci cumulatively increasing susceptibility to the disease. A multiplicative, or epistatic, effect involves more complex interactions, where genetic loci or gene products interact to

collectively increase risk beyond their individual contributions. An example of a disease demonstrating gene-gene interactions is insulin-dependant diabetes mellitus (IDDM; Cordell and Todd, 1995). Statistical methods for detecting linkage or association to several genes simultaneously have been developed (e.g. Cordell et al, 2000; Koeleman et al, 2000). Additive and epistatic models can be distinguished on the basis of the relative risk alone (Risch, 1990a), and these seem to indicate epistatic interactions in BPD (see section 1.3.3).

1.6.4 Gene-Environment Interactions

Environmental factors interacting with a genetic trait are difficult to examine without a characterised genetic locus. Known examples of gene-environment interactions include phenylketonuria (PKU), which requires both a genetic mutation and an environmental factor (phenylalanine in the diet) to produce a phenotype, and the increased risk of Alzheimer's disease due to the apolipoprotein E ε4 allele, which can be augmented by head trauma (Mayeux et al, 1995). Alternatively a genetic factor alone may not have any effect, but may interact with an environmental risk factor to further increase risk (Schildkraut, 1998, p381). A combination of epidemiological and genetic investigations can help unravel the effects of gene-environment interactions. There have been no studies examining environmental factors that may influence the familial risk rates for BPD.

1.6.5 Anticipation and Trinucleotide Repeats

Anticipation describes an increase in disease severity and a decreased age at onset from one generation to the next. It is known to be associated with the expansion of trinucleotide repeat motifs in the genome. Extended triplet repeats have been shown to cause Fragile X syndrome, Huntington's disease, spinocerebellar ataxias, and several other neurological diseases (Ross et al, 1993; Cummings and Zoghbi, 2000). These disorders demonstrate clinical anticipation as a repeat increases in length from generation to generation due to

errors during DNA replication (Margolis et al, 1999). This effect is often seen to be dependant on parent of origin, with maternal transmission conferring increased risk for Fragile X syndrome, and only male transmission of Huntington's disease demonstrating anticipation (Ross et al, 1993). Other factors can also contribute to the observation of non-Mendelian transmission in trinucleotide repeat expansion disorders (Margolis et al, 1999).

Anticipation has been reported as a feature in BPD, with an earlier age of onset and increase in number of episodes in successive generations (Ross et al, 1993; Margolis et al, 1999). The Repeat Expansion Detection technique (RED; Schalling et al, 1993) is a PCR-based method of detecting the average length of genomic trinucleotide repeats. Using this technique, some studies have found that in BPD patients there are increases in the average triplet repeat length in their genome (Lindblad et al, 1998) which may contribute towards susceptibility for BPD, although results are conflicting (Vincent et al, 1996; Vincent et al, 1999; Guy et al, 1999; Verheyen et al, 1999). The RED technique has been shown to reflect expansion mainly of the two loci ERDA1 and CTG18.1, in some populations (Lindblad et al, 1998) although not others (Guy et al, 1999).

1.6.6 Assortative Mating

Spouse similarity exists for traits such as religion, ethnicity, age, education and social class. True assortative mating can result in distortion of Hardy-Weinberg equilibria for the genetic factors contributing to that phenotype, leading to increased homozygosity at neutral loci, and an increase in the variance of a trait or disorder in the population. The latter is effectively due to a tendency toward polarization of the population distribution of the trait into two or more classes that preferentially mate within their class, and is predicted to be more severe if the trait is genetically multifactorial (Merikangas, 1982).

Similarity for psychological traits or psychiatric illness could result from factors other than assortative mating, including common environmental factors or induction of illness in the spouse of a patient (Du Fort et al, 1994). In addition selection biases can affect studies investigating assortative mating, such as ascertainment bias in clinical samples, and increased likelihood of diagnosis of the spouse of a patient.

Such an effect would be relevant to molecular genetic studies of BPD, whether it reflects preference for personality traits associated with psychiatric morbidity, or which increase susceptibility for psychiatric illness later in life. There is evidence that it occurs with affective disorders although most reports are subject to certain biases (Du Fort et al, 1994). Genetic linkage analysis has not attempted to address the issue of assortative mating, mainly due to uncertainty about its magnitude or importance.

1.6.7 Imprinting

Genetic imprinting involves inactivation of genes through DNA modification. The DNA sequence is unaltered, but genes are silenced, usually through methylation of cytosine residues, in a way that is heritable, and often sex-specific. Thus the same marker locus variants inherited from the father and the mother can have different effects. In bipolar disorder, observations of increased or more severe paternal transmission of the illness have been made (Grigoroiu-Serbanescu et al, 1995; Kornberg et al, 2000), as well as maternal (McMahon et al, 1995; Gershon et al, 1996). Evidence of a parent of origin effect linked to chromosome 18 loci have also been made (see section 1.5.2). Imprinting has been investigated in one genome-wide linkage study (Cichon et al, 2001b), which found suggestive evidence for imprinted loci on chromosome 2, 14 and 16 linked to BPD.

1.7 Summary

Bipolar affective disorder (BPD) is a disorder of mood where the patient undergoes alternate bouts of depression and mania. Patients have varied symptoms, and a spectrum of affective disorders has been hypothesised, ranging in severity from single episode depression to bipolar I disorder (BPI), and possibly including schizoaffective disorders. BPD has an incidence of about 1%, distributed evenly between the sexes, and similarly across the world. First degree relatives of patients demonstrate a five- to ten-fold increased risk of also suffering from BPD, and are at increased risk of suffering from depression. Family studies indicate the risk to relatives of suffering from unipolar depression, bipolar II disorder, schizoaffective disorder and BPI increase with the severity of diagnosis of the proband, but distinct patterns are not clear from the research.

Aetiological theories have mainly focused on the monoamine neurotransmitters, and although differences in BPD patients as compared with controls in biochemical research have been found, there is still no clear biological factor associated with the disorder. Disturbances have also been noted in endocrine systems, which may reflect a sub-form of BPD, and in several biochemical pathways and levels of protein expression. Some neuroanatomical disturbances have been noted, but experimental difficulties preclude any conclusive findings. Antidepressants of the MAOI and tricyclic variety target monoamine systems, but are effective in only a portion of "responder" patients, and they and the mood-stabilising agent lithium often also have non-specific and varied biochemical effects, and no clear mode of action can be determined. In the absence of a definite lead from biochemical or other sources (see section 1.2), the development of new or improved treatments for BPD remains difficult. Research is hindered without an unambiguous aetiological model as a starting point.

Twin and adoption studies indicate a substantial genetic component to BPD. If a disorder can be traced to a heritable biological factor, this would represent a useful starting point for further research. For this reason studies have sought to determine extent of genetic influence on BPD, which is known to show familial factors. For these reasons research has turned to what used to be termed 'reverse genetics', which can use anonymous genetic markers in affected individuals to detect the genomic location of a genetic susceptibility factor.

No clear mode of familial transmission can be seen, although claims of a single major locus, X-linkage and of multifactorial inheritance have been made. It now appears that a polygenic model is most likely in the majority of familial cases. Complex genetic disorders by nature are difficult to characterise, as their polygenic or heterogeneous nature reduces the statistical power of the studies. Gene-environment interactions may also complicate the situation.

Classification of BPD is an important factor in genetic studies, as these generally assume a diagnosis reflects an aetiologically pure entity. The definition of BPD is useful from a clinical perspective, but there is no proven causal relationship between any physiological disturbance and the diagnosis. A "strict" diagnosis of BPI appears to have most credibility from epidemiological evidence, for a genetic component (Reus and Freimer, 1997). However, the possibility of susceptibility loci common to the range of affective disorders cannot be excluded, and the examination of linkage to at least a narrowly defined disease classification and a broad classification including major depression is justified.

Genetic association studies have examined differences in allele frequencies between cases and a control sample at candidate loci. In BPD, association studies have examined a wide

range of different genes, most commonly candidate genes involved in monoaminergic neurotransmitter systems. However, no consistent positive results have been reported. Despite this the possibility remains that genes including COMT and MAOA (see section 1.5.1) may be weakly associated with BPD. Positional candidate genes have also been examined with mixed results.

Genetic studies have also examined pedigrees where the disorder segregates, for linkage between the trait and genetic loci. Traditional lod score analysis has been superseded by non-parametric linkage methods expected to be more robust in examining linkage to complex genetic traits. Power is still a cause for concern, as very weak genetic effects may require unfeasibly large sample sizes. Statistical advances have tried to maximise the extraction of full inheritance information from genetic markers, and the use of maximum likelihood methods and multipoint linkage analysis are common. Full-genome scans for linkage have been used to try to determine the regions of the genome most likely to harbour genetics susceptibility loci. Few studies appear to have concordant results, although regions on chromosomes 4, 10, 12q, 13q, 18, 21q, 22q and the X-chromosome are of current interest.

1.8 Aims

This thesis aimed to test for the presence of genetic linkage to BPD in a combined British and Irish sample of affected sibling-pair (ASP) based families. The sample and statistical methods were selected for maximum power and minimal assumptions of the underlying genetic model. For the purpose of developing the laboratory techniques required for the genome scan project and examining the integrity of the sample set, a preliminary study was undertaken to examine a marker within a candidate gene for association to BPD, in the families ascertained in Ireland (chapter 3).

The mode of transmission of BPD is largely unknown, but is likely to be complex, possibly involving several genes of small individual effect (section 1.3.3). Non-parametric statistics such as ASP and allele-sharing based methods are expected to be robust in the detection of genes for complex genetic traits, or where the underlying genetic model is unclear (section 1.4.3). The ascertainment and collection of samples that meet the requirements of ASP linkage methods is expected to be efficient with respect to the time and cost invested, and therefore larger samples are more easily attained.

The investigation of the genetic aetiology of BPD is complicated by the range of diagnoses which exist within the affective disorder spectrum. There is some evidence for a continuum of severity from bipolar I disorder (BPI) and possibly schizoaffective disorder – bipolar subtype (SAB) at one extreme, to recurrent unipolar disorder (RUP) which may have a more modest genetic component (section 1.3.4). Family studies have concluded that there is a shared familial component to these diagnoses, but that this component is greater in relatives of a severely affected proband (section 1.3.5). With this in mind the pedigrees in the main study sample were ascertained on the basis of a proband diagnosed with BPI, and at least one sibling affected with any diagnosis from BPI to RUP. Thus, the ASPs were ascertained in a manner to ensure the maximum shared genetic component to the disorder, according to current knowledge.

Previous linkage studies to BPD have highlighted a number of chromosomal regions, however none of these has demonstrated consistent evidence for linkage at levels considered statistically significant. This influenced aspects of the study to be undertaken in two ways. Firstly, the main study was designed to examine linkage to BPD across the whole genome (chapter 4), thus removing potential bias and ensuring an objective measure of linkage was obtained for all genetic loci. Improvements in genotyping technology have

allowed even small research labs to acquire the ability to generate large numbers of genotypes relatively rapidly, permitting the completion of genome scan studies in larger samples within feasible periods of time. The collaborative nature of the genome scan project allowed efficient use of a pre-developed strategy for the study, including a pre-designed marker map, use of commercially supplied reagent sets and integrated semi-automated genotyping technology. Second, an attempt was made to determine the overall evidence for linkage to BPD, from published studies. This investigation resulted in a meta-analysis study of genome scans for bipolar disorder (chapter 6), in collaboration with various groups around the world.

Cost considerations were also taken into account in order to achieve the maximum efficiency in the main study (chapter 4), while minimally compromising statistical power. For this reason, the work by Holmans and Craddock (1997) was considered, and a split sample design was adopted. The first stage study consisted of a genome-wide linkage study, while second stage involved prioritisation of regions of the genome for examination with a tighter marker grid on the basis of the first stage results, and examination of these in an expanded sample (chapter 5).

Chapter 2. Materials and Methods

2.1 Introduction

The methods described below are general methods applied in chapters 3, 4 and 5. Detailed and specific methods are described in more depth in the Materials and Methods sections of each chapter.

2.2 Patient Collection and Diagnosis

Sample ascertainment and collection was performed by clinicians associated with the UK-Irish Wellcome Trust Bipolar Affected Sibling Pair Genome Scan study, at St. James' Hospital, Dublin, and Trinity College Dublin, and by collaborators at the University of Birmingham. Probands were ascertained from media advertisements for the project, and contact with general practitioners, consultant psychiatrists, community nurses, social workers and patient groups throughout Ireland and Britain.

Diagnosis was reached by interview with a clinical psychiatrist using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN; Wing et al, 1990) instrument, a review of the past medical history, and if possible interviews with other family members. Case vignettes were prepared and these were reviewed by an independent psychiatrist. If these three diagnoses concurred then a consensus diagnosis was reached under DSM-IV (Diagnostic and Statistical manual of Mental Disorders version IV, American Psychiatric Association, 1996) and ICD-10 (International Classification of Diseases 10^{th} edition, World Health Organisation, 1992) criteria. Otherwise, the material was fully reviewed and either consensus reached, or the case recorded as unknown diagnosis. Diagnoses and clinical notes were recorded for each interviewed person, along with unaffected individuals and details of pedigree structure, in an adapted version of the MegaBase database (Fenton & Sandkuijl, 1992).

The equivalence of diagnosis between and within the two centres was established by using this well-defined process for diagnosis, and communication between the clinicians. Regular meetings were held, and cases were cross-rated by different clinicians. Inter-rater reliability calculations were performed by Professor Nick Craddock, and found to be high, with a mean weighted-kappa value (Cohen, 1968) of 0.88.

The sample population consisted of a joint British and Irish sample of affected sibling pairs where the proband had a DSM-IV diagnosis of bipolar I disorder (BPI), and at least one subsequent sibling had a DSM-IV diagnosis of BPI, bipolar II disorder (BPII), bipolar disorder- not otherwise specified (BPN), Schizoaffective disorder - bipolar subtype (SAB), or Recurrent Unipolar Disorder (RUP).

On interview a blood sample was taken from consenting family members, consisting of 10-30ml samples in EDTA vacutainers, and one 3-10ml coagulated sample for the separation and storage of serum for any future protein expression or biochemical analysis. Samples for DNA extraction were frozen at -70°C, or processed immediately.

2.3 DNA Extraction from Whole Blood

DNA extraction from whole blood was performed by a modification of the standard phenol/chloroform method (Sambrook et al, 1982). Six millilitre samples of anticoagulated blood were thawed at 4°C for 2-4 hours. Samples were treated to lyse erythrocytes with an equal volume of sterile deionised water, and double volume of lysis buffer (section 2.7), and mixed on a rotary platform at 4°C for 30min. Samples were centrifuged at 6000 rpm for 15 minutes at 4°C. The supernatant was removed, and the pellet gently resuspended in 25ml lysis buffer at 4°C, and again spun at 6000 rpm for 15 minutes. The supernatant was

poured off, and the lymphocyte pellet was resuspended in 1.5ml of suspension buffer (section 2.7). Sodium Dodecyl Sulphate (SDS) and Proteinase K (Sigma) were added to each sample, to 1% (w/v) and 0.4 mg/ml respectively. Samples were left at 37°C overnight on a rotary platform to denature and digest enzymes and other proteins in the solution.

The DNA was extracted once with Tris-saturated phenol (GibcoBRL), then again with phenol:chroroform/isoamylalcohol (50:50 v/v). Finally two extractions with chloroform/isoamylalcohol (1:24 v/v) were performed to remove all protein traces. The DNA was precipitated after the final extraction with 50µl 3M sodium acetate solution and two volumes of 100% ethanol at –20°C. Precipitated genomic DNA was removed with a pipette and washed twice in 70% ethanol to remove salt residue. It was air-dried, and resuspended in 200-300µl of Tris/EDTA (TE) buffer, depending on the original blood volume. The samples were stored at 4°C for 72 hours, with occasional shaking to fully resuspend the DNA, then stored at –20°C. All supernatants were preserved until visual verification was made of DNA precipitation.

The stock DNA solutions were analysed by spectrophotometry (Shimadzu UVmini 1240 UV-Vis spectrophotometer) at 280 nm and at 260 nm, under appropriate dilutions, to determine concentration of DNA (260 nm). Beer's Law for the relationship between absorbance at 260nm (A_{260}) and DNA concentration was used to derive the formula:

[DNA] (µg/ml) = A_{260} x 50 (µg/ml) x (Dilution Factor),

where Dilution Factor represents the dilution made of the test aliquot taken from the stock solution.

The ratio of A_{260}:A_{280} was used as a control for the presence of RNA. If this value exceeded 1.7 the sample was not deemed to be of sufficient purity, and re-extraction was performed on the pellet or stock blood samples.

Physical condition of the extracted DNA was assessed by running 100ng of genomic DNA solution on a 1% agarose gel in 1x Tris-Acetate EDTA (TAE) buffer. Suitability of the extracted DNA samples for polymerase chain reaction (PCR) amplification was investigated in more detail as part of the study of candidate gene Tyrosine Hydroxylase (chapter 4) for all Irish DNA samples.

Aliquots of approximately 600µl of DNA at 20ng/µl was stored in sealed 96 well boxes at 4°C for use in polymerase chain reaction amplifications. It was considered that the continual use of this DNA would not warrant storage at –20°C as repeated freeze-thaw cycles may damage it. Continuous use of this DNA stored at 4°C in polymerase chain reactions revealed no perceptible degradation of the signal over time.

2.4 Microsatellite Genotyping Using Radio-Isotope Labels

Genotyping methods using the labelling of amplified fragments with a radioactive isotope were used in chapter 3. All Polymerase Chain Reactions (PCR) incorporated a positive control of known genotype and at least one negative control containing no template DNA to detect contamination. DNA from individual 1347.02 of the *Centre d'Etude du Polymorphisme Humain* (CEPH) DNA bank, supplied by Perkin Elmer/Applied Biosystems International, was used as the positive control.

Reactions were optimised by adjusting primer and $MgCl_2$ concentration, and varying the annealing temperature to determine the optimal amplification conditions. A set of control

DNA samples known to be polymorphic for the microsatellite were used for optimisation. Amplification of the marker region was performed with incorporation of dCTP labelled at the alpha position with the phosphorus isotope ^{32}P (Amersham) to allow detection of the product. Amplification was performed on a Hybaid Omingene HBTR3SM2 thermal cycler, in 25 µl volumes.

Amplified products were mixed with loading dye in a 2:1 ratio of dye to product. A volume of 2.5µl was loaded onto 6% polyacrylamide gels, for separation of the products by denaturing gel electrophoresis. The gels were run using 1x Tris-Borate-EDTA (TBE) Buffer (GibcoBRL) at 40 volts for 3 hours, at a temperature of 55°C. Empty lanes and positive and negative controls were included in each gel, to aid in determining sample positioning and assist calling of alleles. The glass plates were then dismantled, and the gel transferred to Whatman paper (3mm Chr, Whatman) and dried on a vacuum apparatus at 80°C for 45 minutes.

X-ray film (Curix RP1, Agfa) was exposed to the gel at –70°C, for an average exposure time of 12 hours before being developed (Fuji RGII X-Ray Film Processor). The product sizes were visualised, and alleles independently called by two investigators.

2.5 Semi-Automated Microsatellite Genotyping Using Fluorescent Labels

2.5.1 ABI 377

The ABI 377 DNA sequencer (Perkin Elmer/Applied Biosystems International) was used for genotyping in both the Dublin and Birmingham centres, as part of the genome scan project (chapters 4 and 5). The ABI 377 is a semi-automated vertical gel electrophoresis system, which reads fluorescence from chemical tags with a charge-coupled device (CCD) detector as they are excited by an argon-ion laser. The machine was controlled by an Apple Macintosh Power PC 4400 via a serial connection, running proprietary PE ABI

Collection™ software (version 2.6). PCR amplifications were loaded into a polyacrylamide gel, and fluorescence at four wavelength channels read in real time as fragments passed the end of the gel. The four wavelengths examined were determined by a multicomponent software matrix specific to the dye set used. A matrix file is used to compensate for spectral overlap between the dyes (figure 2.1), preventing excess bleed-through of one dye's fluorescence into other channels. The fluorescence data for each gel is stored in a gel image file, which can then be further analysed by ABI software. This system had the advantage of the existence of several commercial genome-wide marker sets specifically designed and pre-optimised for use with it (chapter 4, section 4.2.3).

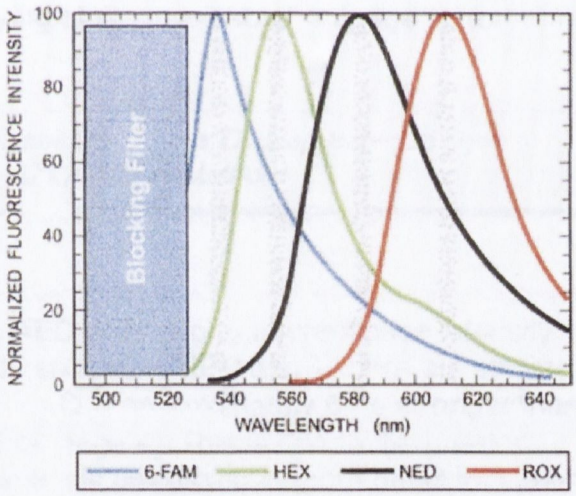


Figure 2.1. Spectral overlap of the ABI Filter Set D fluorescent dyes

2.5.2 Polymerase Chain Reaction

One of each pair of primers was end-labelled with one of three fluorescent dyes, which fluoresce at peak wavelengths of 560 (HEX: green), 580 (NED: yellow), and 540 (FAM: blue) nanometers when excited by the argon-ion laser. Fluorescence at these wavelengths

was detected by the CCD and translated by the collection software using a matrix file into green, yellow and blue respectively.

The primers were used with the ABI PRISM True Allele™ PCR pre-mix, which contained a Taq polymerase and dNTPs in a buffer. The thermal cycling conditions recommended by ABI were used throughout (see table 2.1), except where otherwise indicated.

Number of cycles	Temperature (°C)	Time (min:sec)
1	95.0	12:00
10	94.0	0:15
	55.0	0:15
	72.0	0:30
20	89.0	0:15
	55.0	0:15
	72.0	0:30
1	72.0	10:00

Table 2.1 Thermal cycling conditions for the amplification of genome scan microsatellite markers.

PCR reactions were performed in 96 well polypropylene plates on an MJ Research PTC-200 thermal cycler, in 10µl reactions (table 2.2), unless otherwise indicated. One microlitre of 20 ng/µl genomic DNA samples was dried in the 96 well plate, and the reaction mix simply aliquoted on top of them. In a comparison, the amplicon peaks from PCR using the dried DNA method was shown not to amplify differently from that using liquid DNA solution.

Component	Volume (µl)
True Allele™ PCR Premix	6.0
DNA	-
Primer pairs	4.0
Total	10.0

Table 2.2 Components of the polymerase chain reaction for amplification of the Linkage Mapping Set v2 microsatellite markers.

All markers were typed for the CEPH individual 1347.02 cell-line DNA (PE-ABI), which was used as a positive control for each reaction, and for inter-gel allele alignment. All reactions also included a negative control reaction omitting genomic DNA.

2.5.3 Gel Electrophoresis

The PCR products were run out on 64-lane 4% polyacrylamide denaturing gels containing 6M urea. Electrophoresis was performed at 3000 Volts (60 milliamperes; 200 Watts), at a temperature of 51 °C, using 1x TBE buffer.

The internal size standard ROX 400HD (PE-ABI) was included in each lane, consisting of a 50-400 bp DNA ladder with a specific fragment pattern, labelled with a fourth fluorescent dye, ROX, with emission peak at 615 nm (red).

2.5.4 Processing of the Gel File

Processing of the ABI 377 sequencer gel image was performed by the ABI GeneScan™ version 3.1.2. It detected the lanes on the gel by means of a neural net-based tracking program, the results of which can be adjusted manually. For each lane, it then extracted fluorescence intensity data over time for the four dyes used. Each lane was then visualised as a line-plot of intensity of fluorescence on the ordinate, against scan number, or time, on the abscissa. The software used a peak recognition algorithm to detect fluorescence peaks. Using fluorescence from the ROX dye, the Local Southern method was used to calculate the length of the fragments corresponding to the peaks detected (Ghosh et al, 1997). This method used Southern's reciprocal relationship between fragment length and mobility:

$$L = \left(\frac{c}{m - m_0} \right) + L_0 \quad ,$$

where L is length in base pairs, m is mobility, and c is a constant. L_0 and m_0 are the sample length and mobility, respectively. A size curve was automatically fitted to each set of three adjacent size standard peaks, allowing an accurate size to be estimated for each fragment peak in the other dye colours. Fluorescence and peak size data from each lane were then output into sample files.

2.5.5 Genotype Data Management

The ABI Genotyper™ software package was used to call alleles for each marker genotyped with fluorescent semi-automated techniques. Genotyper™ is based on a template file containing details of markers to be analysed (colour, fragment size range, peak labelling algorithm parameters). A set of GeneScan™ sample files from a gel was then loaded into the template, allowing the alleles for each marker of the panel, labelled with size in base pairs, to be inspected manually. Amplification of dinucleotide microsatellite repeat markers typically led to a characteristic pattern of stutter fragments of smaller size, differing by the length of the repeat unit (see figure 2.2).

Tentative allele-calling by the Genotyper™ software, based on the microsatellite stutter pattern was performed for each individual for each marker. The software peak labelling was based on an algorithm which would correctly label the peaks only if the peak shape was clean and clear. Often allele assignment was incorrect if the amplification or gel conditions resulted in excess or insufficient fluorescence levels, leading to a sub-optimal peak shape or height. However, an experienced operator could usually easily recognise the allelic pattern, and all alleles were visually screened, and either corrected or marked as unreadable.

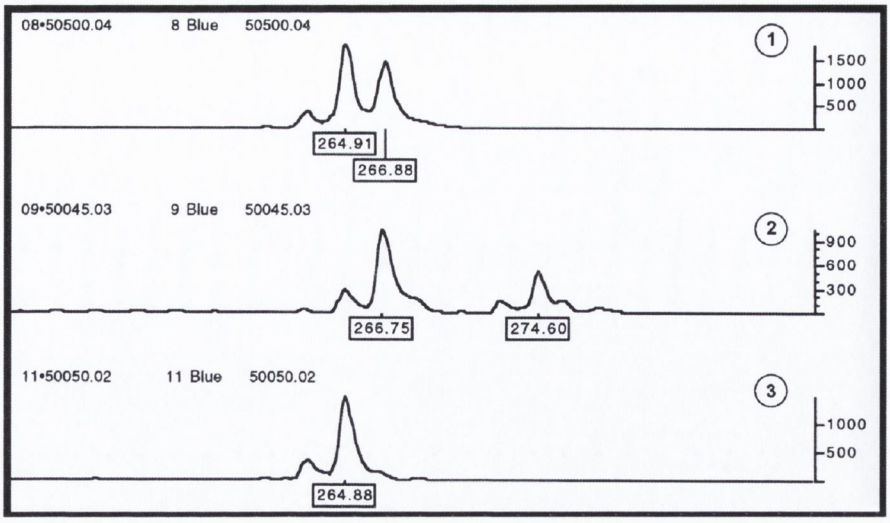


Figure 2.2. Microsatellite stutter pattern; from top to bottom: 1) heterozygous genotype where alleles differ by 2 base pairs; 2) heterozygous genotype with widely spaced alleles and differential amplification of the larger product; 3) homozygous genotype

Genotypes for each marker were outputted in table format listing the sample identifier, marker name, the genotype, and an "overflow" field. The genotypes were represented as numbers in "ABI units" which approximately equate to fragment length in base pairs. They were calculated to a precision of 2 decimal places, as a linear extrapolation from the positioning of the allele peaks relative to six flanking size standard peaks. An "overflow" column served to notify if three or more peaks were mistakenly labelled for one sample by the software or human operator. Samples with unreadable genotype were given allele size of –1. This table was then directly exported to a Microsoft Excel™ spreadsheet specific to that marker.

2.5.6 Data Handling and Allele Binning

As all genotypes for each marker were entered into the Microsoft Excel™ spreadsheet, these were saved as plain-text comma separated value (CSV) files. A control value (CON), representing a marker-specific and gel-specific standard was entered on a separate line at the head of the information from each 64-lane electrophoretic gel. This value was

calculated as the allele of the CEPH individual 1324.02 genotype which is closest to the midpoint of the allele size range (ASR) for that marker. Thus it represents the gel-specific migration of a mid-sized PCR product relative to the migration of the size standard fragments, for each marker. The finalised CSV file (see Appendix A) was then imported into the pedigree management database, MegaBase (figure 2.3).

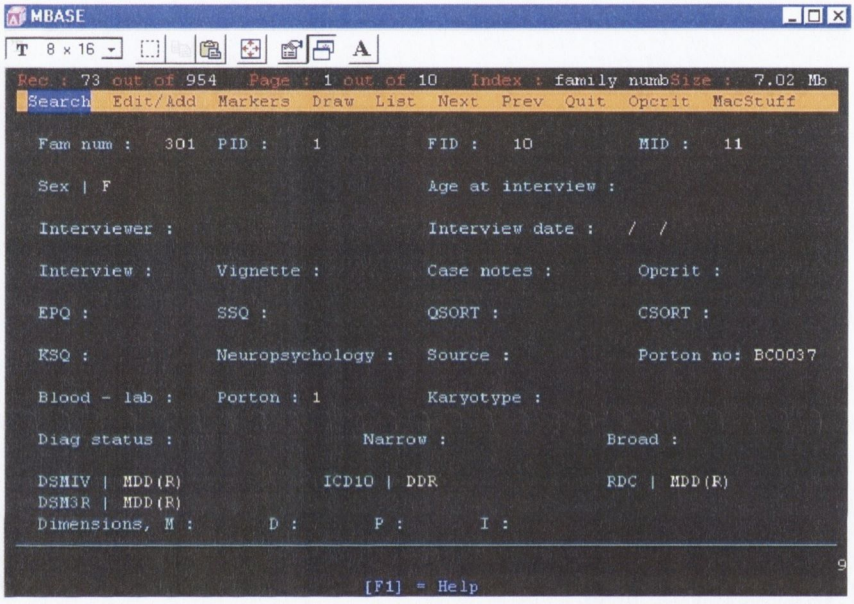


Figure 2.3. Screen capture from the MegaBase genotype management database program (Fenton and Sandkuijl, 1992).

MegaBase used the CON value to align marker fragment sizes from each gel with those of the first gel. For any allele A in gel N, of size a_N and control value c_N, the size of A is adjusted to be:

$$A = a_N + (c_1 - c_N) \quad ,$$

where c_1 is the control value of the first gel.

MegaBase also allowed organisation of the continuous allele sizes to integer categories, or "bins". This was achieved by eye on a graphical histogram of the allele sizes in ABI units for all individuals typed for that marker (figure 2.4). Vertical lines delimiting the bins were placed manually, with alleles which did not fall into a clear "bin" identified for investigation.

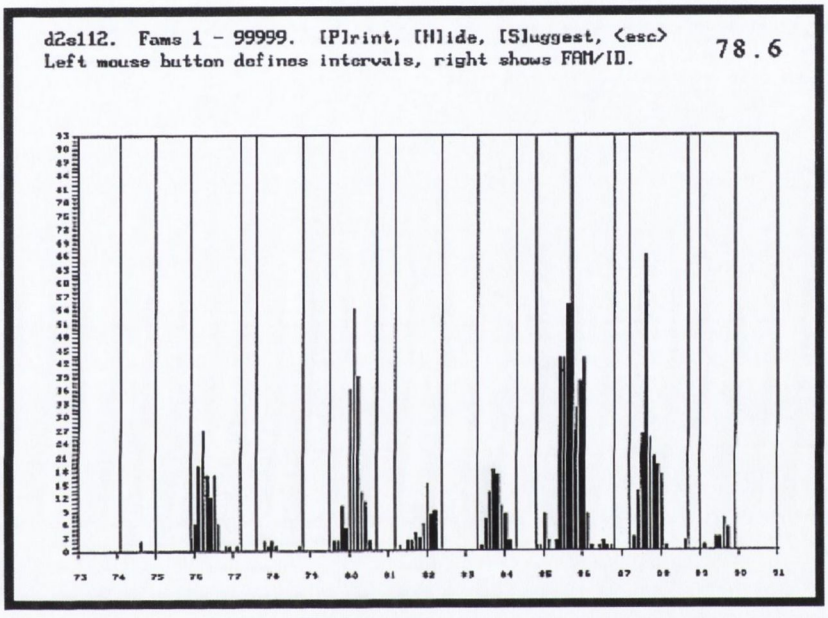


Figure 2.4. Allele binning screen from the MegaBase database and genotype management program.

MegaBase outputted files of genotype data in standard LINKAGE format, compatible with most linkage analysis programs.

Issues of data quality control and data flow management of the large volumes of data produced by the genome scan project are detailed in chapter 4.

2.6 Statistical Methods

2.6.1 Pedcheck

The Pedcheck program (version 1.1; O'Connell and Weeks, 1998) was applied to marker data before examination for linkage, to identify Mendelian inconsistencies. The program was run at level 4, using the '–4' flag, which is guaranteed to identify all Mendelian inconsistencies within each pedigree using a checklist of conditions (level 1), and the Lange-Goradia algorithm for genotype elimination (level 2). The Levels 3 and 4 analyses are also applied, if there were no level 1 errors, to identify the critical individuals most likely to be responsible for an inconsistency, and to calculate an associated odds ratio using the allele frequencies provided. Pedcheck takes two standard LINKAGE format input files:

A pedigree file (plain text) in LINKAGE format containing on each line family and personal identifiers, sex, affection status, and genotypes, for each individual;

A data file containing marker order and allele frequencies.

2.6.2 SibError

The SibError program (Ehm and Wagner, 1998) was applied to Affected Sibling Pair (ASP) genotype data in order to identify affected pairs which were not true siblings. The SibError algorithm generates a test statistic based on the sum, over many markers, of the number of alleles shared Identical-by-State (IBS) by siblings. Thus large negative values of the statistic will identify half-siblings, or unrelated individuals, and large positive values identify monozygotic twin pairs.

The program takes as input two files in standard LINKAGE format: a pedigree and a data file.

2.6.3 RELATIVE

The RELATIVE program (Göring and Ott, 1997) was applied to ASP data to determine the probability of relatedness of ASPs where both parents were absent from the sample. When the 'sib' option is specified, the program computes likelihoods for the probability of siblings with one or no parents genotyped, being full sibs, half-sibs or unrelated.

The program takes a control file as input, with lines containing:

1 or 2, specifying the 'sib' or 'general' options respectively;

name, of output file;

name, of LINKAGE format data and pedigree files;

prior probabilities that a sib pair are full sibs, half-sibs, and unrelated;

prior probabilities that a sib pair are full sibs, half-sibs when the mother is typed;

prior probabilities that a sib pair are full sibs, half-sibs when the father is typed.

2.6.4 Extended Transmission Disequilibrium Test (ETDT)

The ETDT (version 2.2; Sham and Curtis, 1995) program performs a family-based association analysis (TDT) with a multiallelic marker (chapter 1, section 1.4.2). It tests for linkage disequilibrium in the presence of linkage, using the parents as internal controls, which eliminates any potential problems caused by population stratification. It produced three main statistics: a genotype-wise TDT statistic which examines each heterozygous parent separately and determines if the observed transmission rate of each allele deviates from the null hypothesis of 50%; and an allele-wise TDT statistic examining transmission of a particular allele across all parental genotypes. It also generates a 'goodness-of-fit' statistic, which is a comparison of the genotype- and allele-wise models, and should provide a measure of how well the allele-wise statistic fits the data.

It is subdivided into four programs: the 'ped2tdt' program treats a standard LINKAGE format pedigree file to extract only the genotype information useful for the TDT. The 'Calctdt' program then totals the numbers of allele transmissions from the modified pedigree file. Then the 'Lrtdt' program performs a logistic regression in order to fit a model to the data and compute a log-likelihood. Finally the 'Mctdt' program performs a Monte Carlo modelling permutation test to determine empirical significance levels for the test. The ETDT program itself runs each of the above programs in succession.

The ETDT program takes two standard LINKAGE format files as input: a pedigree file, modified by the 'ped2tdt' program; and a data file.

2.6.5 Sibling Transmission Disequilibrium Test

The S-TDT algorithm (version 1.1; Spielman and Ewans, 1998) compares marker genotypes of affected and unaffected children, calculating significance by repeated simulation of randomised replicates of the data inputted. The TDT/S-TDT program combines the TDT and S-TDT into an overall z score, by summing the number of test alleles transmitted in pedigrees for which both parents are available, and the number of test alleles shared between affected siblings in pedigrees in which both parents are not available. The combined allele count is the test statistic used and is converted to a standard z score, which can be tested against statistical tables. If multiple (>2) alleles are present at the marker locus, the largest absolute z score generated, z(max), is taken, and significance levels can be derived by simulation. A Bonferroni correction for the number of alleles examined can be applied, and will be only slightly conservative.

TDT/S-TDT requires a single input file in standard LINKAGE pedigree file format, containing pedigree details and marker genotypes.

2.6.6 Two-point Linkage Analysis

Two-point Maximum Likelihood Scores were calculated using the SPLINK program (version 1.09; Holmans, 1993; Holmans and Clayton, 1995) by remote access to the HGMP-RC computing facilities. SPLINK calculates a maximum likelihood score (MLS) for linkage between a marker locus and a disease locus using Identity-by-descent (IBD) probabilities between all possible affected sibling pairs, weighted for non-independence, and subject to the 'possible triangle' restrictions (Holmans, 1993; chapter 1, section 1.4.3). It uses pseudo-maximum-likelihood methods to estimate haplotype probabilities to compensate for missing or ambiguous genotype data. In addition, the SPLINK program was used to estimate population allele frequencies from our sample, using the '-asp-' flag which uses data from individuals classed as diagnosis 'unaffected' or 'unknown'. SPLINK also provided the heterozygosity for each marker, calculated as:

$$H = 1 - \sum_{i=1}^{n} p_i^2 \ ,$$

where n is the number of alleles at the locus, and p_i is the frequency of the i^{th} allele. Polymorphism information content (PIC) was also calculated as the heterozygosity value adjusted with the probability of informative parental chromosomes, and it represents the proportion of fully informative matings. It is calculated as:

$$PIC = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n} \sum_{j=1}^{n} p_i^2 p_j^2 \ .$$

Two input files in LINKAGE format were required for SPLINK: a pedigree file in LINKAGE format containing on each line family and personal identifiers, sex, affection status, and genotypes, for each individual; a data file also in LINKAGE format, from which SPLINK deduces marker order.

As the SPLINK program can only input data from one genetic marker at a time, batch files were prepared to expedite the analysis of whole-chromosome input files. These were plain text files executable under the UNIX environment of the HGMP-RC, which directed the sequential analysis of each marker in turn (see Appendix B).

2.6.7 Multipoint Non-Parametric Linkage (NPL) Analysis

Multipoint NPL scores were calculated with the Genehunter (version 1.3; Kruglyak et al, 1996) program, accessed remotely at the HGMP-RC. The Genehunter estimates the distribution of inheritance vectors using Hidden Markov Models, and from this calculates a Non-Parametric Linkage (NPL) statistic (chapter 1, section 1.4.3). Thus full inheritance information is extracted, using a combination of identity-by-descent (IBD) information when available, and identity-by-state (IBS) information from other affected pedigree members, if present. Significance levels are calculated by comparison with a standard normal distribution, which should be slightly conservative. Marker allele frequencies were taken from the output of the SPLINK two-point linkage program (2.6.6).

We used the 'score all' option, applying the S_{all} statistic (Whittemore and Halpern, 1994) which utilises information from every all affected relatives simultaneously within each pedigree. This statistic applies extra weight to a shared allele, proportional to the number of individuals who share it IBD. The 'off end' option was used in each full-chromosome analysis to extend the measure of linkage 10cM off either end of the map. This allows for full multipoint linkage information to be extracted at the ends of each chromosome. The default 'increment' option of 'step 5' was used, which instructs the NPL to be calculated at each marker, and at 5 steps within each inter-marker interval.

The 'count recs' option was used in the GENEHUNTER program to instruct recombination counts to be made, and provide observed and estimated values for the inter-marker genetic distance, using the Kosambi map function. This can permit the identification of markers prone to excess genotyping error, where an inflated intermarker distance is likely to be observed.

Two LINKAGE format files were required by Genehunter: a pedigree file; and a data file from which Genehunter also extracts marker allele frequencies. Batch files were also prepared to perform full-chromosome analyses with Genehunter (see Appendix B). These consisted of plain text files, executed from within the Genehunter environment with the 'run' command.

2.6.8 Multipoint Maximum Likelihood Linkage Analysis

The Mapmaker/SIBS program (version 2.0; Kruglyak and Lander, 1995) was used on the HGMP-RC to perform the principal method of linkage analysis. It applies a non-parametric multipoint Maximum Likelihood Score (chapter 1, section 1.4.3). Mapmaker/SIBS is subject to the restriction that only two-generation, nuclear families can be examined. Prior to analysis with Mapmaker/SIBS, the linkage files were edited to nuclearise larger pedigrees, using the FORTRAN program Makenuc on the HGMP computers. We performed an MLS analysis using all available sibling pairs, weighted if more than two affected siblings exist in the pedigree, and allowing for dominance variance by applying the 'possible triangle' restrictions to the IBD sharing probabilities (Holmans, 1993; Kruglyak and Lander, 1995). We also used the 'off end' option in full-chromosome analyses to examine evidence for linkage 10 cM past either end of the map. The MLS was calculated at 1 cM intervals from the start of the map, the default value.

The 'infomap' option was used to generate an information-content score, expressed as the inheritance information extracted from the data as a proportion of the complete inheritance information. The 'exclude' option in Mapmaker/SIBS was also used to perform exclusion lod analysis. For this analysis the presence of dominance variance (implying a multiplicative polygenic locus model) was assumed, and an exclusion lod score was calculated for each of several values of λ_S, the relative risk to siblings attributable to the locus under examination.

Mapmaker sibs takes as input two LINKAGE format files: a pedigree file, modified with the Makenuc program; and a data file. Batch files were not used with the Mapmaker/SIBS program, due to the requirement to restart the program between the analysis of each disease model.

2.7 Buffers and Media

Gel Loading dye:

95% formamide

2% 0.5 M EDTA

3% xylene cyanol/bromophenol blue suspension.

Lysis buffer:

1.6M Sucrose

5mM Tris-HCl pH 7.5

2.5mM MgCl2

1:200 Triton X-100 (Sigma)

Suspension buffer:

10mM Tris-HCl pH 7.5

10mM EDTA pH 8.0

10mM NaCl

TE (Tris/EDTA) buffer:

10mM Tris-HCl pH7.5

10μM EDTA pH 8.0.

Tris-Borate-EDTA (TBE) 1x:

45mM Tris base

45mM Boric acid

1mM EDTA pH 8.0

Tris-Acetate-EDTA (TAE) 1x:

40mM Tris base

40mM Acetic acid

1mM EDTA pH 8.0

Phenol, buffer-saturated; GibcoBRL, Cat #15513-039

Used as supplied

Polyacrylamide solution, LongRanger Gel Solution (50%); FMC Bioproducts, Cat #50611

Used as supplied

Chapter 3. Association and Linkage Studies of the Tyrosine Hydroxylase gene

3.1 Introduction

3.1.1 Tyrosine Hydroxylase And Monoamines

Tyrosine Hydroxylase (TH) is the rate-limiting enzyme in the biosynthesis of the catecholamine family of neurotransmitters, responsible for the production of brain supplies of dopamine and noradrenaline. Disturbances in the activity or expression of this enzyme will affect the availability of these neurotransmitters at the synapse. Because the activity of TH determines the rate of synthesis of these important neurotransmitters, it has been the focus of considerable interest in the investigation of monoaminergic theories of affective disorder (described in chapter 1). In theory, altered expression or activity of the enzyme could result from sequence variations in or near the gene for tyrosine hydroxylase, which has been mapped to chromosome 11p15.5 (Xue et al, 1988).

3.1.2 Linkage Studies

In the context of the still-emerging theories of chemical and anatomical basis for even the normal emotional mind, and of certain experimental difficulties, candidate gene analysis represents an opportunity to dissect a portion of the aetiology using established laboratory methods, of proven worth in other disorders. The genetic component of bipolar disorder (BPD) has been established and was outlined in chapter 1, section 1.3.

Interest in chromosome 11p was stimulated by the finding, in 1987, of evidence of genetic linkage to markers close to the TH gene in an extended Old Order Amish pedigree segregating BPD (Egeland et al, 1987; chapter 1, section 1.5.2). The evidence for linkage in this pedigree was later revised in the light of the onset of illness in two family members previously determined to be unaffected and the extension of the pedigree to include a

further section of the family (Kelsoe et al, 1989). However, interest in this region remained due to reports of cosegregation of BPD and β-thalassaemia minor, the gene for which maps to the same area of chromosome 11p (Joffe et al, 1986; Harada et al, 1995), and due to localisation of two potential candidate genes for affective disorders, tyrosine hydroxylase (TH) and the dopamine receptor 4 (DRD4), to this region. Almost all studies investigating linkage to TH and related markers in this area have been negative. One possible explanation for this is the presence of genetic heterogeneity, with some families showing evidence of linkage and others not (Hodgkinson et al, 1987; Lim et al, 1993).

3.1.3 Previous Association

Several genetic association studies have been conducted on markers in or close to this gene, with mixed results. Serretti et al (1998a) present a summary of studies conducted up to 1997, and it is apparent from this, and a perusal of studies published since then, that the majority of studies report no association between TH markers and BPD. Interestingly, studies appear to be consistent within national populations. Hence studies on French, Italian, Belgian and Spanish sample populations are all positive, while no evidence for association has been found in Japanese, German, British, North American or Croatian populations. This may reflect population heterogeneity in the aetiology of BPD, but this is probably unlikely. An alternative explanation is that studies on any one national population are usually conducted by the same research group using similar ascertainment or statistical methodologies, and this may account for the above observation.

More recent studies also include unipolar illness in the diagnostic spectrum, or perform analyses on separate diagnoses. Association has been detected to unipolar depression only in a Belgian sample (Souery et al, 1996) and to depressive symptomatology in an Italian sample (Serretti et al, 1998b).

Two formal meta-analyses have been performed, and these failed to identify any association between tyrosine hydroxylase markers and BPD (Turecki et al, 1997; Furlong et al, 1999a). However, Furlong and colleagues found a weak association between an RFLP (restriction fragment length polymorphism) marker at the 3' end of the gene and unipolar depression. In addition a large multi-centre European study (Souery et al, 1999) found no evidence for association of BPD with markers at Tyrosine Hydroxylase.

3.1.4 Aims

The attractiveness of the TH gene as a positional candidate merited its examination in this large sample of BPD pedigrees. The aims of this project were to examine the TH gene for evidence of association, and in doing so to develop the researcher's skills in DNA handling, amplification, and genotyping in advance of the main genome scan project. The sample consisted of small nuclear families based upon an affected sibling pair and this made it somewhat unsuitable for association studies. For this reason the transmission/disequilibrium test (TDT) was used (Spielman et al, 1993), which tests for association in the presence of linkage, by examining transmission from heterozygous parents to affected offspring (see chapter 1, section 1.4.3).

The sample is not expected to provide much information to the TDT, due to the low expected proportion of affected children with heterozygous parents. The sample was ascertained to consist mostly of affected sibling pairs, and parents were not always available. In addition the polymorphism examined had only 5 alleles, reducing the expected number of heterozygous individuals. For this reason the Sib-TDT (S-TDT; Spielman and Ewens, 1998) was also applied (see chapter 2, section 2.6.5). This test applies a standard TDT in addition to the S-TDT, which is based on the sharing of alleles

between affected and unaffected siblings, and combines the evidence from both tests. As the sample often contained unaffected siblings in pedigrees where the parents were not available, this test may be expected to have increased power.

The marker used has been extensively investigated in the past for association with bipolar disorder, and unipolar depression. The majority of these sutdies did not find any evidence for association (see Serretti et al, 1998), however those that did identified allele 2 (Cauli et al, 1994), allele 5 (De Castro et al, 1995), or genotype 2/5 (Meloni et al, 1995). Serretti et al (1998) themselves found a significant increase in allele 2, and concomitant decrease in allele 1.

In addition, the sample was examined for genetic linkage to the TH gene. When LD is present in an old population between two loci it is usually the case that tight linkage can also be found, as LD will usually rapidly dissipate over the space of a few generations, while linkage can still be detected. As genetic linkage may be expected to detect a disease susceptibility locus some distance from the polymorphism under examination, this would be expected to detect loci more distant from the gene which affect the phenotype, even if the association analysis does not reveal the presence of LD. Thus the presence of promoter polymorphisms, or adjacent genes, which increase susceptibility to BPD can be tested.

3.2 Methods

3.2.1 Sample Description

The collection, nature and structure of the first stage genome scan pedigrees in the sample have been described in chapter 2, section 2.2. Briefly, affected sibling pairs were collected on the basis of a proband suffering from DSM-IV (American Psychiatric Association, 1996) bipolar I disorder (BPI). The sample used consisted of Irish pedigrees containing one

sibling suffering from BPI, and at least one other sibling suffering from BPI, bipolar II disorder (BPII), Schizoaffective disorder (SAB), bipolar disorder - not otherwise specified (BPN) or recurrent unipolar disorder (RUP). Parents, if possible, and any unaffected siblings available were also collected. DNA was isolated from blood samples as described in chapter 2, section 2.3. Irish pedigrees only were used, as the British samples were still undergoing quantification in Birmingham, and we had not yet received the DNA samples. In addition the intended method of analysis, the TDT, is unsuitable for examining pooled samples from geographically or genetically separate populations, as this violates the assumption of random mating in the sample population. Although the TDT was initially intended to be robust to the influence of population stratification in the study design, the deliberate use of a stratified population is not advisable.

In total 216 individuals were genotyped. Table 3.1 describes the sample in detail. The full sample contained 141 individuals affected with BPI, of whom 113 were successfully genotyped. This relatively high drop-out rate (19.9%) resulted from the training nature of the project, and it's timing immedeately before the commencement of the genome scan project (chapter 4). Although a number of additional PCR reactions were performed in an attempt to bring this rate down, the main project took precedence when the sample was finalised and the fluorescent genotyping instrumentation (see chapter 2, section 2.5) was installed.

Three of genotyped affected individuals were not included in the analysis as they were parents of affected individuals; however several other affected parents who were part of extended pedigrees remained in the analysis due to the presence of affected and/or unaffected siblings of theirs (for example see below). A number of lone parents and unaffected individuals were genotyped, in the expectation that the ongoing sample

collection would complete additional pedigrees, or add further affected individuals in extended pedigrees. This did not always come to pass; these individuals remained superfluous, and were not counted in table 3.1. Several individuals affected with broader diagnoses were classified as 'unknown', as the full consensus diagnosis had not yet been reached for all non-proband members of the pedigrees. This led to a reduced sample compared with the full Irish cohort used in the genome scan (see chapter 4).

The genotyped individuals included in the analysis came from a total of 60 pedigrees. These mainly consisted of affected siblings or singletons, although there were also two affected half-siblings and seven affected trios. In three cases of extended pedigrees (as mentioned above), affected relations other than siblings were included: an affected sibling pair (ASP) with an affected mother, an ASP with an affected mother and affected maternal uncle, and a singleton with an affected paternal uncle. The use of these affected individuals led to a total of 63 singleton or sibships available for family-based association tests (table 3.1).

Number of:		
Singletons:	23 *(36.5%)*	
Sibling pairs:	33 *(52.4%)*	66 individuals
Sibling trios:	7 *(11.1%)*	21 individuals

Affected singletons/sibships with:	
0 parents genotyped:	40 *(63.5%)*
1 parent genotyped:	14 *(22.2%)*
both parents genotyped:	9 *(14.3%)*
0 UA siblings genotyped:	27 *(42.9%)*
1 UA sibling genotyped:	21 *(33.3%)*
2 UA siblings genotyped:	10 *(15.9%)*
3 or more UA siblings genotyped:	5 *(7.9%)*
total:	63

Table 3.1. Description of the effective association study sample, comprised of pedigrees with at least one genotyped affected individual; UA=unaffected.

The power of a sample of 100 pedigrees, consisting of two affected siblings, with no parents or unaffected siblings genotyped (representative of the majority of our sample),

3.2.2 PCR

The polymorphism examined was a tetranucleotide $(TCAT)_n$ repeat motif (HUMTH01) located in intron 1 of the tyrosine hydroxylase gene, beginning at base pair 1170 (Polymeropoulos et al, 1991). This polymorphism has been shown to have a quantitative effect on expression of the TH gene, possibly by binding transcription factors such as ZNF191 and HBP (Albanese et al, 2001), and thus may theoretically have a direct causative effect on the phenotype. The published primer sequences were used:

Forward strand: 5'-CAGCTGCCCTAGTCAGCAC-3'

Reverse strand: 3'-GCTTCCGAGTGCAGGTCACA-5'

The PCR product was expected to have a length ranging from 244 to 260 base pairs. Polymerase chain reaction (PCR) amplification of the TH tetranucleotide repeat locus was performed on the 216 individuals, under the conditions outlined in tables 3.2 and 3.3. Genotyping using radio-isotope labelling, and gel electrophoresis were performed as detailed in chapter 2, section 2.4.

Number of cycles	Temperature (°C)	Duration
1x	94.0	5m00s
1x	84.0	held for Taq addition
27x	94.0	1m30s
	61.0	1m00s
	72.0	1m00s
1x	72.0	10m00s

Table 3.2. Thermal block cycling conditions for amplification of the TH gene polymorphism.

	Stock concentration	Volume added for 1x
TH(F)	20 ng/μl	2.5 μl
TH(R)	20 ng/μl	2.5 μl
Mg^{2+} (pH 8.9)	15.0 mM	2.5 μl
dNTPs	3.3 mM	1.32 μl
gDNA	20 ng/μl	2.5 μl
α^{32}P-dCTP		0.05 μl
H_2O		11.13 μl
Taq polymerase	0.5 units/μl	2.5 μl
Total		25 μl

Table 3.3. Constitution of the polymerase chain reaction for amplification of the TH gene polymorphism.

3.2.3 Statistical Analyses

The tyrosine hydroxylase genotype data was analysed using the Extended Transmission Disequilibrium Test (ETDT) statistical program (Sham and Curtis, 1995; see chapter 2, section 2.6.4). The TDT/S-TDT program (Spielman and Ewens, 1998) was also applied to the data in order to maximise information from our sample (see chapter 2, section 2.6.5). As the S-TDT statistic uses sibling pairs discordant for the phenotype, certain siblings unaffected with BPI were capable of introducing a bias into the analysis, as they may be suffering from SAB, BPII, BPN, or RUP. For this reason these individuals, who would have been classed as affected under a broader disease classification model, were removed prior to the S-TDT analysis.

Subsequent to the TDT analyses, the pedigrees were subjected to non-parametric linkage analysis using the SPLINK program (Holmans, 1993; see chapter 2, section 2.6.6), to test for the presence of genetic linkage between the marker and disease loci. Allele frequencies in our population were also obtained from the SPLINK program

3.3 Results

3.3.1 Marker Allele Frequencies

Published allele frequencies are shown in table 3.4, which were based on the frequencies of 70 chromosomes of unrelated individuals of U.S. origin (Polymeropoulos et al, 1991). These were found to differ slightly from the frequencies observed in our sample for two of the alleles (1 and 2).

Allele	Size (bp)	Published frequencies	Observed frequencies
1	260	0.21	0.338
2	256	0.26	0.108
3	252	0.13	0.133
4	248	0.14	0.153
5	244	0.26	0.268

Table 3.4. Published and observed allele frequencies of TH tetranucleotide repeat polymorphism (Polymeropoulos et al, 1991).

3.3.2 ETDT

Transmission and non-transmission counts from the ETDT program are presented in table 3.5. Results of the association are outlined in table 3.6. Twenty-seven parent-child transmissions were revealed to be informative for the ETDT. No significant association was detected. There was a slight trend observed with the genotype-wise statistic ($p = 0.058$). The goodness-of-fit statistic compares the likelihoods of the allele-wise and genotype-wise statistics to determine how well the allele-wise model fits the data, and therefore this significant p-value was not relevant.

Allele	*1*	*2*	*3*	*4*	*5*
Transmitted	7	5	4	4	7
Not Transmitted	11	1	3	3	9
χ^2	0.889				0.250
p value	0.349				0.617

Table 3.5. Transmissions and non-transmission of TH polymorphism alleles determined by the ETDT program

	χ^2	d.f.	p value
Allele-wise	3.488	4	0.480
Genotype-wise	15.075	8	0.058
Goodness-of-fit	11.587	4	0.021

Table 3.6. Results of the ETDT analysis of the TH gene polymorphism; d.f.= degrees of freedom.

3.3.3 S-TDT

The S-TDT test (chapter 2, section 2.6.5) gave a combined z(max) value of 0.5, for allele 2 (table 3.7). According to the significance levels reported by Spielman and Ewens (1998), this lies below the threshold considered significant at a 5% level either uncorrected (2.55) or corrected for multiple testing using the Bonferroni method (2.58), for a five-allele locus.

Allele	Transmissions (combined TDT/S-TDT)	z' score
1	12	-0.136
2	3	0.500
3	7	0.000
4	7	0.086
5	14	0.205

Table 3.7. Results of the TDT/S-TDT combined analysis of the TH gene polymorphism

3.3.4 SPLINK

The results of the SPLINK program were not significant: the log-likelihood ratio gave a χ^2-value of 0.002 ($p = 0.567$). This indicates that there was no detectable linkage between this polymorphism and BPD in the sample examined.

3.4 Discussion

Family-based association analysis (ETDT) similarly did not show any evidence to support linkage or linkage disequilibrium (LD) between the disorder and this marker. The TDT has been described as a test for linkage and LD, as it can only deviate from the null distribution

when the loci are linked, i.e. $\theta<0.5$, and is likely to be significant only in the presence of very tight linkage ($\theta=0$) (Sham and Curtis, 1995). Similarly the Sib-TDT did not reveal any significant association, despite the additional information provided from discordant sibling pairs with no informative parents. The trend in allele frequencies appears to be the reverse of those previously reported (Serretti et al, 1998), with an increase in allele 1 at the expense of allele 2. Linkage analysis of the sample subsequent to the TDT analysis did not reveal any evidence for linkage of this gene to BPAD.

The goodness-of-fit statistic can be seen as a measure of how well the allele-wise model fitted the data, and how much importance can be attached to the allele-wise χ^2 statistic. It tends to be significant if there is non-random mating present in the population, or if the recombination fraction between the marker and the disease exceeds 0, as appears to be the case with this marker.

Perhaps of interest is the slight trend toward significance of the genotype-wise (saturated) statistic ($p=0.058$). This statistic is calculated by taking each parental genotype separately, and is held to provide less power than the allele-wise statistic, which examines marker alleles separately, and allows each to be associated with a disease allele to a different extent. This distinction is comparable to the difference between the genotype-based haplotype relative risk and the haplotype-based haplotype relative risk statistics. This genotype-wise TDT result may conceivably reflect weak LD between the marker and a susceptibility locus, with no individual allele associated with the phenotype. However the allele-wise statistic would be expected to detect a subgroup of alleles in association with the phenotype, and this result is most likely due to random variation. Despite its apparent functional relevance (Albanese et al, 2001) this polymorphism, or one in strong linkage disequilibrium with it do not appear to be major causative loci for BPD in our sample.

As with all experiments in complex genetic disorders, these results cannot be taken as excluding an effect entirely. In the case of a complex disorder it is rare to find a strong effect with any one locus, and it may be that the absence of linkage here reflects the polygenic nature of BPD. Calculating the power of the TDT requires an estimation of the disease allele frequency and the genotype penetrances. Our sample, containing 86 nuclear families, with 216 genotyped individuals was in fact quite small. The number of genotyped affected children was only 110, and the number with at least one heterozygous parent also typed reduced the number of transmissions informative for the ETDT to only 27, which would be expected to have very poor power to detect a susceptibility gene of small effect (Knapp, 1999). According to Risch and Merikangas (1996), association studies attempting to identify a gene with a relatively large Genotypic risk ratio, $\gamma = 4.0$ (the increased chance that an individual with a particular allele has the disease) would require 222 parent-child trios, at the observed proportion of heterozygous parents. In addition, the power studies performed by Sham and Curtis (1995) indicate that our sample size would have poor power to detect an association under any genetic model.

Chapter 4. Genome Scan Stage I

4.1 Introduction

4.1.1 Background and Aims

The Wellcome Trust UK-Irish Sibling-Pair Bipolar Affective Disorder genome scan study was a collaborative genome scan project supervised jointly by Professors Michael Gill and Nick Craddock to search for susceptibility genes for bipolar disorder (BPD) in a pooled Irish and British sample. The aim of the project was to identify regions of the genome containing susceptibility loci for BPD by genetic linkage analysis. This was judged to be best achieved by studying as large a sample population as possible, achieving maximum statistical power.

An important aspect of the genome scan with respect to laboratory experiments was consistency of genotyping and analysis between the two centres. Modern semi-automated fluorescent genotyping technology was adopted with this in mind, and identical pre-optimised commercial genotyping kits and integrated software were used by each group. The combined sample of pedigrees was shared between the groups, and the genome evenly divided, by chromosome.

The management of data generated was coordinated between Dublin and Birmingham, with the use of commercial genotyping sets and allele calling software. A version of the MS-DOS database program MegaBase (Fenton and Sandkuijl, 1992) adapted for psychiatric disorders was used as the data management database (chapter 2, section 2.5.6). Using this database, individual patient files can be browsed, fields can be directly edited, searches can be performed by family and individual number, and a graphical display of the pedigrees can be obtained. The program also stores marker genotypes, and allows markers

to be listed on a pedigree diagram, written to an external file, or alleles to be displayed as a histogram for genotype binning.

4.1.2 Study Design

The study was conducted using a split sample and grid-tightening design which has been shown to be an effective approach to genome screening, with good cost and time efficiency for the statistical power obtained (Holmans and Craddock, 1997). The initial sample of affected sibling pairs, their parents and unaffected siblings, was genotyped for a ~10 centimorgan (cM) genome-wide genetic map. A lod score threshold of 0.74 (corresponding to a point-wise p-value of 0.05) was then applied to the results, identifying regions to be followed up in an expanded sample of about twice the size, with a finer marker map, to extract full inheritance information.

The strategy of a two-stage study using grid-tightening for increased efficiency compromises power slightly, compared to what could be achieved by typing the whole sample at a higher density from the start (Holmans and Craddock, 1997; see chapter 1, section 1.4.5). If a conservative threshold for follow-up is taken, the probability of missing a true peak (false negative, type II error) may become unacceptable. A lax threshold favours type I error, which can be exposed by the larger sample set and denser marker map at the follow-up stage. If the number of regions passing the threshold exceeds expectations, one of these regions may contain a true susceptibility locus, despite the reduced power. The benefits of genotyping parents in the first stage depend on the required efficiency (Holmans and Craddock, 1997). However, if there is the possibility that population stratification exists, initial typing of parents will reduce the type I error probability (false positive). As our sample consists of a combined Irish and British sample, the presence of stratification was considered a possibility.

The principal statistical package in the genome scan, Mapmaker/SIBS, obtains a maximum likelihood score (MLS) using a statistic based on alleles shared identical-by-descent (IBD) between affected siblings (chapter 1, section 1.4.3). Incompletely informative marker data is compensated for by maximising sharing probabilities with an E-M (expectation-maximisation) algorithm (Kruglyak and Lander, 1995). In addition the program applies the 'possible triangle' restrictions to limit allele sharing to biologically feasible values (see chapter 1, section 1.4.3). It has been shown that this possible triangle MLS corresponds to a mixture of χ^2 distributions with mostly 0 and 1 degrees of freedom (Nyholt, 2000). This statistic is best suited to the sample available, which consists mainly of nuclear families based on an affected sibling pair, if highly polymorphic markers are used where IBD status can be unambiguously determined in most cases. The restriction to the possible triangle models an epistatic (multilocus multiplicative) locus model, increasing the power of the maximum likelihood statistic (Nyholt, 2000), and may therefore be suited to BPD (Craddock et al, 1995). Nyholt (2000) calculated that a nominal p-value of 0.05 is obtained at an MLS of 0.74 on the autosomes, or 1.18 on the X chromosome. These levels were considered an appropriately low threshold for follow-up in the second stage of the genome scan (Lander and Kruglyak, 1995).

In addition to this statistic the data were analysed with a single point MLS (using the SPLINK computer program) and a multipoint non-parametric lod score analysis using inheritance information from all affected pairs, thus taking advantage of extra information from the few extended pedigrees in the sample (using the GENEHUNTER program). The results of these statistics were examined to verify if any region would show large differences in linkage between analytical methodologies, which would point to a possible spurious positive or negative finding.

Many highly polymorphic markers are available through on-line databases (e.g. the Genome Database; www.gdb.org) with heterozygosities of greater than 0.7. When conducting a genome screen using several hundred markers, the cost and efficiency of identifying markers, constructing a map and compiling and optimising amplification conditions must be taken into account. There are commercially available sets of polymorphic markers which span the genome at regular intervals, and are pre-optimised and supplied with appropriate reagents. These linkage marker sets were therefore ideal for the requirement of consistency of methodology between the two centres.

The primary analysis method was predetermined to be the multipoint allele-sharing program Mapmaker/SIBS, performed on the narrowest diagnostic model (Model 1; see section 4.2.1), consisting of BPI - BPI affected sibling pairs. This was expected to yield the most robust statistical measure of allele sharing, with the additional advantage of extracting full multipoint information from the data (see section 4.2.5).

Genotyping of the microsatellite markers were performed by myself and co-workers. This thesis presents results of genotyping for chromosomes 5, 6, 19, 20, 21 and 22, and panel 16 (markers on chromosomes 9, 10 and 11), in addition to 7 markers which were problematic and required individual reamplification (including D22S420 and D22S423 on chromosome 22), as detailed in Appendix C. It also contains the results of full statistical analyses for chromosomes 2, 5, 6, 9, 10, 11, 17, 18, 19, 20, 21, and 22, performed on remotely accessed computers of the Human Genome Mapping Project (HGMP) resource centre.

4.2 Methods

4.2.1 Sample Description

The patient sample consisted of a collection of 150 nuclear pedigrees containing at least one affected sibling pair (ASP), as detailed in chapter 2, section 2.2. Affected sibling pairs were classified into five disease affection status models based on the Diagnostic and Statistical Manual for Mental Disorders – version IV (DSM-IV) diagnosis of the second sibling, the first sibling always having a diagnosis of Bipolar I Disorder (BPI). These five models corresponded to a series of increasingly broadly defined classifications of affective disorder as outlined in table 4.1.

Model	Diagnostic criteria
1	Bipolar I disorder
2	Model 1 + Bipolar II disorder
3	Model 2 + Schizoaffective Disorder (bipolar)
4	Model 3 + Bipolar disorder - not otherwise specified
5	Model 4 + Recurrent unipolar depression

Table 4.1. Genome scan disease affection status models (DSM-IV).

The sample set is described in table 4.2. The full sample to be genotyped included 541 individuals, 391 of whom were affected. This figure included 37 affected individuals who did not form part of an ASP, such as parents and singleton children.

Diagnostic model:	1	2	3	4	5
Pedigrees with:					
Affected sibling pair only	59	70	71	78	89
ASP + 1 parent	20	21	22	22	23
ASP + both parents	30	31	33	34	37
no extra unaffected siblings	67	81	84	91	111
1 UA sibling	30	30	31	32	27
2 or more UA siblings	12	11	11	11	11
Total (*pedigrees*)	109	122	126	134	149
Total ASPs (*all pairs*)	154	193	201	216	260
Total ASPs (*weighted*)	129	153	159	170	198
Number of pairs:	93	98	100	106	111
trios:	13	19	21	22	29
quads or greater:	3	5	5	6	9

Siblings with:	**Total (*M/F*)**
Bipolar I disorder	278 (*123/155*)
Schizoaffective (bipolar) disorder	22 (*7/15*)
Bipolar II disorder	6 (*1/5*)
Bipolar disorder (NOS)	11 (*6/5*)
Major depressive disorder (recurrent)	31 (*9/22*)
Classed as unknown/other	123 (*62/61*)

Table 4.2. Individuals Genotyped for the stage I genome scan

Legend: ASP- Affected Sibling Pair; UA- Unaffected; M- Male; F- Female; diagnoses as referred to in text (American Psychiatric Association, 1996); unweighted sibling pairs calculated as the sum over all pedigrees using n(n-1)/2, where n is the number of affected siblings in the pedigree; weighted sibling pairs calculated similarly, using (n-1).

4.2.2 Power Calculations and Threshold selection

Dr. Peter Holmans and Prof. Nick Craddock, using computer simulations, investigated the statistical properties of our sample for detecting a vulnerability locus for BPD. They simulated 1000 replicates of a genome scan, identical to our own with respect to family structure, affection status and availability of DNA for genotyping. Genotypes were generated under the hypothesis of no underlying susceptibility locus, at 400 markers with 4 equifrequent alleles, and even 10cM spacing, in order to obtain the distribution of the MLS under the null hypothesis, and estimate genome-wide significance levels

In addition, a disease susceptibility locus conferring a λ_s of 3, 2, 1.5 or 1.25 located midway between two markers was modelled in a simulated sample containing an identical number of ASPs to our study. A 10cM marker grid was simulated, with 5 equifrequent alleles at each marker locus, and the power and average MLS expected under each locus model were calculated.

4.2.3 Marker Map and Genotyping

The genetic marker map used was the ABI PRISM Linkage Mapping Set version 2 (LMSv2; now renamed Medium Density 10cM mapping set: MD-10). This set of markers comprised 400 dinucleotide repeat microsatellite markers, selected from the Généthon linkage map (Dib et al, 1996). Reverse primers were altered by the addition of a poly-A tail ('pig-tailed') to improve amplified product stability and resolution with the genotyping method (Brownstein et al, 1996). Inter-marker distances and order were established from the published 1996 Généthon CEPH genetic map (Dib et al, 1996).

Primer pairs from the LMSv2 were supplied pre-optimised for use with the ABI Prism 377 DNA sequencer and True Allele PCR Premix® (PE-ABI; part No. 403061; see chapter 2, section 2.5.2). The markers were pre-selected so that PCR products of several amplification reactions could be run simultaneously on the ABI 377 DNA sequencer. They were supplied in 28 panels of ten to twenty markers that could be co-electrophoresed due to discrepant product fragment lengths, and differential labelling with one of the three dyes. The 12 Dublin chromosomes were covered by 14 panels (see chapter 2, section 2.5.2; Appendix C). Product sizes ranged from 70 to 395 bp. Average marker spacing was 9.43 centimorgans over the 22 autosomes and the X chromosome. Polymerase chain reaction (PCR) was performed as detailed in chapter 2, section 2.5.2

In order to maximise efficiency and minimise the cost of the genotyping process we used the 14 panels assigned to Dublin as the basis to select groups of up to five primers to be co-amplified in a multiplex PCR. How efficiently the individual marker sequences amplified, reflected primer sequence and composition, and each primer pair altered the kinetics of amplification for the other primers in the reaction. As the exact sequences of the primer pairs were proprietary information, the amplification efficiency was best estimated from the size of the amplified product, and by observation of the intensity and sharpness of the bands produced in test reactions. Due to absence of this critical information and the large number of markers, optimisation of the multiplex reactions was limited to empirical experimentation. Each multiplex reaction was composed of 3 to 5 primer pairs of similar product lengths, in equal concentration; reactions containing shorter products often tolerated a greater number of simultaneous amplifications. Optimisation of the initial multiplex reaction proceeded by variation of the relative proportions of the primers. While multiplex PCR was used to attempt to reduce the cost and time of genotyping, this was necessarily compromised by the need to preserve the advantages of the pre-optimised reagents. Thus, when efficiency of primer annealing led to complete failure of amplification at one or more marker loci, these were removed for individual amplification. The successful multiplex reactions constructed will be made available to the wider scientific community on the internet. The primers were pre-optimised for use with the ABI PRISM True Allele PCR pre-mix, which contained Taq polymerase, buffer and dNTPs. This pre-mix was therefore used for the multiplex reaction, and alteration of buffer, dNTP or magnesium chloride concentrations, as recommended for optimisation of multiplex PCR (Vance and Othmane, 1998), was foregone.

Amplification products were pooled into the panels of 10-15 markers for multiplex gel electrophoresis (see chapter 2, section 2.5.3). Each panel of markers was used to amplify

DNA from the total of 544 individuals making up the first stage sample. The amplified products, along with positive and negative controls were electrophoresed in nine 64-lane polyacrylamide gels. The order of the samples, along with the position of the controls permitted the detection of errors in sample arrangement and the recognition of inter-gel variation. Before each electrophoresis, one of nine custom sample sheets was loaded into the collection software assigning sample identifiers to each lane. A GENOTYPER™ template file for each panel was composed, detailing each marker in that panel. The final gel file was processed as detailed in chapter 2, sections 2.5.4 and 2.5.5. Briefly, the sample lanes were identified, fluorescence data extracted for each sample, and genotypes called, using the GENESCAN™ and GENOTYPER™ software. Alleles were binned using the MegaBase software, as detailed in chapter 2, section 2.5.6.

4.2.4 Logistics and Quality Control

A consistent and structured flow of data through the genotyping process was essential given the large amounts of data being processed, and the variety of forms through which it passed (see chapter 2, section 2.5). A flow-chart was composed outlining the flow of data, and procedure to be followed in order to ensure data integrity, minimisation of human error, and safety considerations (figure 4.1). The plan outlined in the chart provided for data back-up at specific points, which could be restored in case of data loss or corruption later in the process. All intermediate files were kept on the computer until the end of the project.

The sample list, corresponding to the samples stored in nearly six complete 96-well deep-well storage boxes, was used to compile nine sample sheets containing 62 samples in addition to a negative control and the CEPH positive control. A PCR was performed for with 64-sample groups, and these were run on polyacrylamide gels with the corresponding

sample sheet identifying the individual samples. From these gels image files were generated, which were backed-up to 1 Gb Iomega JAZ disks. The gel file, containing embedded sample sheet information, was imported into the GENESCAN™ program, processed, and 64 sample files were output for each gel. The details from each LMSv2 panel were entered into GENOTYPER™ template files, which were used to call alleles for sample files containing PCR product fluorescence data for markers in the relevant panel. The alleles called for each marker were displayed in a table, and copied directly into the Microsoft Excel spreadsheet, thus accumulating, gel by gel, the alleles for one marker into a single spreadsheet. When completed these Excel sheets were saved as CSV files, which were transferred to a PC computer running a master version of the MegaBase genotype management database. The master database contained clinical diagnoses and pedigree information for every person in the sample list. After allele-binning (see chapter 2, section 2.5.6) each marker, the complete data for a chromosome was output as a LINKAGE file.

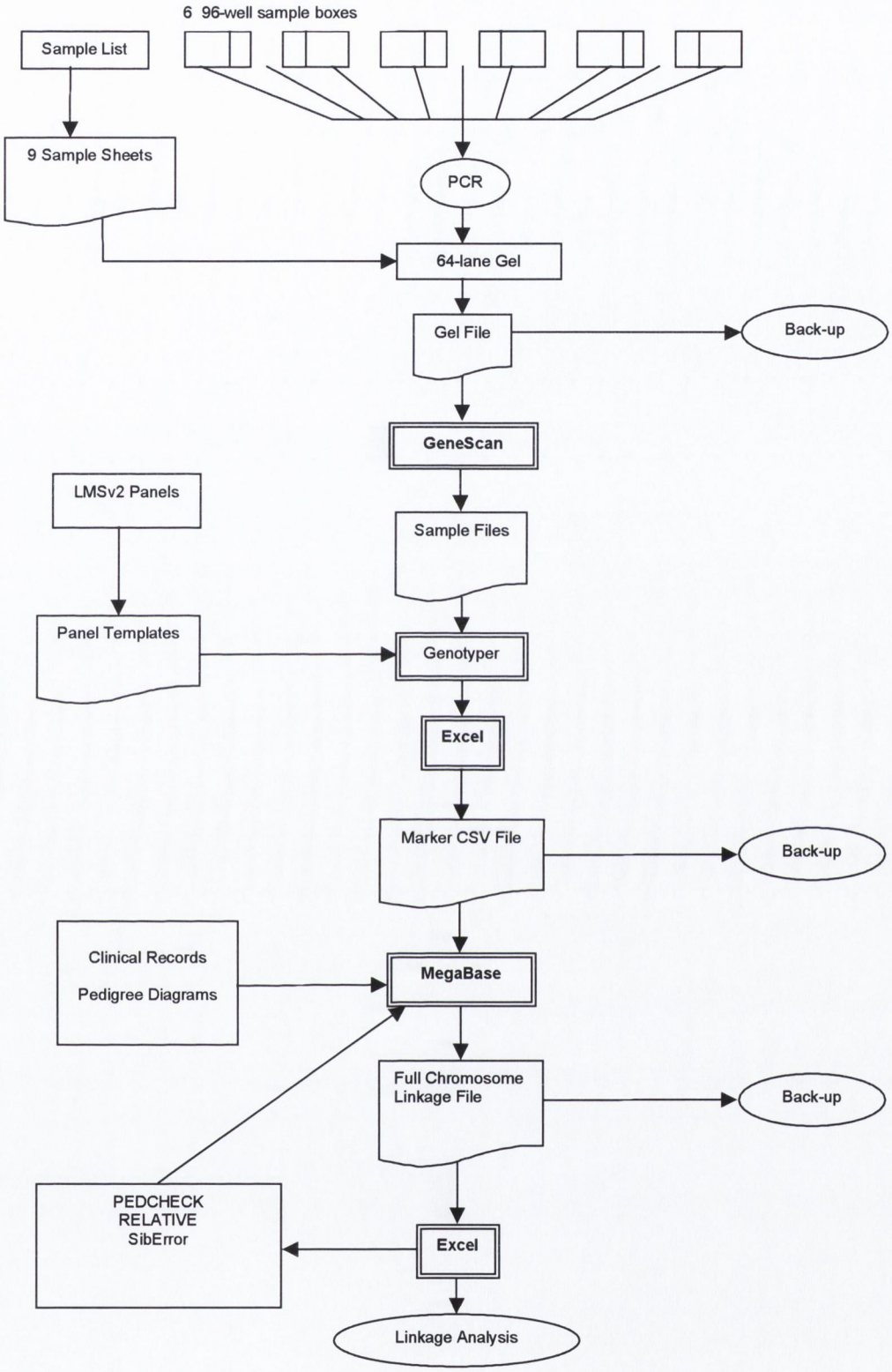


Figure 4.1. Flowchart outlining dataflow through the data generation process from DNA samples to genetic linkage analysis.

The LINKAGE file was opened in Excel, and edited to remove liability classes and other irrelevant information appended by the MegaBase program. The file was then edited to reclassify diagnostic status to conform to the five diagnostic models: all persons with a broad diagnosis reclassified to status unknown for the narrow model 1, etc. Excel Visual Basic macros (see Appendix D) were written to perform this editing rapidly and consistently. The use of macros to edit these files was expected to greatly reduce the human error inevitable in handling such a large amount of data. The final LINKAGE files were first transferred to the HGMP computers by File Transfer Protocol (FTP), and examined with the Pedcheck program for Mendelian inheritance errors (see below, and chapter 2, section 2.6.1). A file containing errors for each chromosome was retrieved, and examined with reference to the MegaBase allele-binning screen. If this did not help resolve the error, the sample files for the relevant persons and markers were re-examined in GENOTYPER™. On resolution (see paragraph below), the process of generating LINKAGE files was repeated. As all errors for each marker were resolved, the CSV file for that marker was backed up to ZIP disk from the PC. On completion of a final LINKAGE file without errors, these were also stored on ZIP disk. Copies of this back-up ZIP disk were kept in the laboratory, and two other separate locations to ensure the data could not be lost.

Due to the need to maximise power in our sample, it was vital to ensure a low genotype failure rate. Genotyping targets for the genome scan were set at 95% successful genotypes per gel. This was to be achieved in one initial amplification reaction, and if necessary one repeat reaction performed on a custom gel, with failed samples picked by hand from the stock DNA boxes. If over 50% of a gel could not be genotyped all 64 samples were re-amplified. As approximately 40% of our sample consisted of members of an affected sibling pair, and assuming random variation in DNA quality is responsible for failure to

genotype, the 95% target corresponds to a pedigree dropout rate of approximately 2% for each marker.

Detection of erroneously genotyped individuals was performed by the Pedcheck program (O'Connell and Weeks, 1998), which detects errors in Mendelian inheritance at multiallelic markers (see chapter 2, section 2.6.1). Familial genotype inconsistencies for a marker were traced back to the gel, and re-scored. If irresolvable, genotypes for the entire family at that locus were erased. Otherwise the errors were corrected in the CSV file, which was re-imported into MegaBase, and re-examined with Pedcheck. Patterns of inconsistencies from genotypes read from a specific gel resulted in that gel being re-run and re-scored. Due to the possible presence of undetected MZ twins in the sample, the SibError program (Ehm and Wagner, 1998) was run on the genotype data to detect monozygotic twins, which could lead to an inflated allele-sharing score at all markers (see chapter 2, section 2.6.2). The RELATIVE program (Göring and Ott, 1997) was run by Peter Holmans on the sample with half the genotype data generated from Dublin and Birmingham to detect any potential half-siblings in the sample, which could falsely deflate the ASP statistics (see chapter 2, section 2.6.3). Where any pedigree consistently yielded Mendelisation errors across many markers sample mix-up was assumed to have occurred, and the entire family or, if apparent, the responsible individual was removed from the study.

These safeguards were not expected to detect all errors in genotyping, because many genotype errors do not result in non-Mendelisations, and are therefore not detected by the Pedcheck program. This was unavoidable, but was not expected to result in a large number of erroneous genotypes in the final data, due to the high heterozygosity of the microsatellite markers. Six monozygotic twins were identified by the SibError program,

and the pair was eliminated in subsequent linkage analyses. However, these genotypes gave an opportunity to directly examine the genotyping error rate.

4.2.5 Statistical analyses

Three analytical programs: SPLINK, Genehunter and Mapmaker/SIBS, were used to perform linkage analysis on the final data files for each chromosome (see chapter 2, sections 2.6.6, 2.6.7 and 2.6.8 respectively). Mapmaker/SIBS was predetermined to be the principal statistical method, which should allow extraction of maximum multipoint allele-sharing information from the data. In addition each analysis was performed under the five disease affection models (see section 4.2.1), with the narrowest (Model 1) defined as the primary model.

Exclusion lod scores (chapter 1, section 1.4.3) were calculated by the Mapmaker/SIBS program (see chapter 2, section 2.6.8), under the narrow disease locus model modelling a susceptibility locus conferring λ_s of 1.5, 2 and 3, equivalent to a proportion of 2 alleles shared IBD (z_2) of 0.333, 0.375 and 0.417, respectively (Risch, 1990c). Exclusion lods were calculated under the assumption of no dominance variance.

4.3 Results

4.3.1 Power and Significance Level Simulations

The simulation studies under the null hypothesis of no underlying susceptibility locus suggested that a multipoint MLS of 1.7 under the narrow disease model would be obtained on average once per genome scan, and therefore corresponds to a 'suggestive' linkage finding (Lander and Kruglyak, 1995). A genome-wide score >3 could be expected every 20 genome scans, corresponding to a significant genome-wide p-value of 0.05 (Lander and Kruglyak, 1995).

Results of the power calculations performed by Peter Holmans are shown in table 4.3. Shown is the power to obtain an MLS > 2 or > 3, with the Model 1 and Model 5 samples, and an underlying locus-specific relative risk to siblings, λ_S, of 2 or 3. Also shown is mean expected MLS in a genome scan in the Models 1 and 5 samples with an underlying locus conferring a λ_s of 2 or 3.

Locus effect size	MLS	Power under:	
		Model 1	Model 5
$\lambda_s = 2$	MLS > 2	0.66	0.80
	MLS > 3	0.34	0.59
	Expected average MLS	*2.7*	*3.5*
$\lambda_s = 3$	MLS > 2	0.88	0.98
	MLS > 3	0.77	0.93
	Expected average MLS	*4.0*	*5.8*

Table 4.3. Power study results from simulation studies (PH, personal communication).

4.3.2 Genotype Error Detection and Quality Control

Where markers consistently failed to amplify they were excluded, and replaced with markers selected from the Généthon linkage map by myself, and co-workers. In total 6 markers were removed from LMSv2 (D2S396, D5S428, D6S1610, D10S192, D11S901, D11S4175, D22S423). Five replacement markers were inserted where the intermarker distance was left unacceptably high on chromosomes 2, 5, 6 and 9 (D2S2354, D5S618, D6S1616, D6S1650, D9S1818). The final adjusted map had an average inter-marker distance of 9.45 cM.

In Dublin genotypes were generated for 194 markers in 541 individuals. Six monozygotic twin pairs were detected by the SibError program (pedigrees 347, 5015, 5023, 5050, 50502, 50512). The genotype agreement was similar within all six pairs, and indicated a

genotype discordance rate of 1.62%, where both twins were typed. This error rate may be higher for markers for which genotyping failed in one or both co-twins.

In four pedigrees (families 608, 5249, 50021 and 50530) the affected sibling pairs were confirmed as half-siblings. One of these (pedigree 608) was maintained in the sample due to the presence of additional true siblings. The RELATIVE program detected genotyping inconsistencies in four pedigrees (families 332, 6010, 50005 and 50020), in two of which the problem could be isolated to a single individual (family 6010, individual 02; family 50005, individual 2). All clinical and genotype entries for the monozygotic twins, and problem individuals and families were removed from the database before linkage analysis was performed.

A total of 89,268 genotypes were generated, corresponding to a mean genotyping success rate of 85.1%. This fell short of the target of 95% (section 4.2.4), most likely due to problems with the multiplex PCR reactions, and human error in, for example pipetting. The Birmingham genotypes approximately achieved the target rate, and their use of a robot in all pipetting was deemed to have been an important factor in reducing errors made at the bench (Dr. Phil Bennett, personal communication). The genotype failure rate for the chromosomes genotyped in Dublin are presented in table 4.4A.

Table 4.5 summarises further details for each of the Dublin chromosomes, and table 4.6 shows the information content across the regions that exceeded the MLS threshold for follow-up in the second stage. Mean information content was calculated across the Dublin chromosomes as 0.64. Observed recombination counts were 10.6% greater on average than the number expected from the LMSv2 map distances indicating the possibility of some genotyping error. SPLINK output indicated that each marker was genotyped for 47-146

weighted sibling pairs (mean = 125.7) under the narrow disease model 1, and for 91-247 (mean = 212.6) under the broad disease model 5.

Chromosome	2	5	6	9	10	11	
Number of markers	30	22	21	21	19	16	
No. genotypes generated	14399	10212	9811	9689	8894	7334	
Genotype success rate	88.7%	85.8%	86.4%	85.3%	86.5%	84.7%	
Chromosome	17	18	19	20	21	22	Mean
Number of markers	15	14	12	13	5	6	
No. genotypes generated	6579	6617	5456	5775	2073	2429	
Genotype success rate	81.1%	87.4%	84.0%	82.1%	76.6%	74.8%	85.1%

Table 4.4 Details of genotyping for the 12 Dublin chromosomes; mean genotyping rate corresponds to the average number of successful genotypes per marker.

Average polymorphism information content of the whole genome map was 0.747 and mean heterozygosity, *H*, was 0.775, calculated by the SPLINK program (see chapter 2, section 2.6.6).

Chromosome	Markers	Mean Spacing (cM)	Expected Rec	Observed Rec	Info Content	Het	PIC	Weighted ASPs	
								Model 1	Model 5
2	30	9.06	2070.64	2316.90	0.68	0.79	0.77	113.87	225.73
5	22	9.65	1598.79	1750.49	0.66	0.81	0.78	108.59	215.86
6	21	9.86	1524.12	1622.32	0.66	0.77	0.74	111.24	220.62
9	21	8.13	1252.72	1461.19	0.68	0.78	0.75	105.90	212.52
10	19	9.62	1335.54	1470.75	0.64	0.76	0.73	106.89	214.84
11	16	9.78	1130.82	1295.69	0.62	0.78	0.75	103.38	205.25
17	15	9.76	1084.91	1197.81	0.65	0.77	0.74	106.40	209.53
18	14	10.29	1050.66	1146.82	0.65	0.76	0.74	111.71	221.36
19	12	8.45	747.16	820.93	0.67	0.79	0.76	104.83	206.92
20	13	8.44	809.73	876.09	0.66	0.80	0.77	101.92	202.00
21	5	10.33	330.89	354.97	0.54	0.73	0.69	89.20	174.20
22	6	9.94	392.36	421.66	0.56	0.76	0.74	87.67	173.33
Total:	194								
Mean:		9.35	1110.69	1227.97	0.64	0.78	0.75	125.69	212.60

Table 4.5. Marker details for the twelve Dublin chromosomes. Expected and observed recombinations (Rec) from GENEHUNTER; mean information (info) content from Mapmaker/SIBS; mean heterozygosity (Het) and PIC from SPLINK; number of weighted ASPs calculated as the sum of (n-1), where n is the number of affected siblings within each family.

4.3.2 Linkage Results and Tables

Figure 4.2 shows the graphs of genome-wide linkage results from the multipoint MLS analysis (left-hand axis) for the Dublin chromosomes, under disease models 1 (red trace), 3 (green trace) and 5 (blue trace). Intermediate disease models 2 and 4 showed scores lying between those of the next broadest and narrowest disease models, and are not shown for clarity of presentation. Also plotted on the same graphs is the information content obtained across the chromosome (right-hand axis). This data was obtained from the Mapmaker/SIBS program, and averaged 0.64 across the chromosome; mean information content for each chromosome are given in table 4.5. Points of highest multipoint information content coincide with the positions of the microsatellite markers (indicated by an X on the graphs). Also indicated on each plot (dotted trace) is the MLS theshold of 0.74 for follow-up in stage II. The regions selected for follow-up are those where the model 1 (red) score exceed this threshold. The results of the single point MLS and multipoint NPL analyses, considered secondary, are presented in Appendix E.

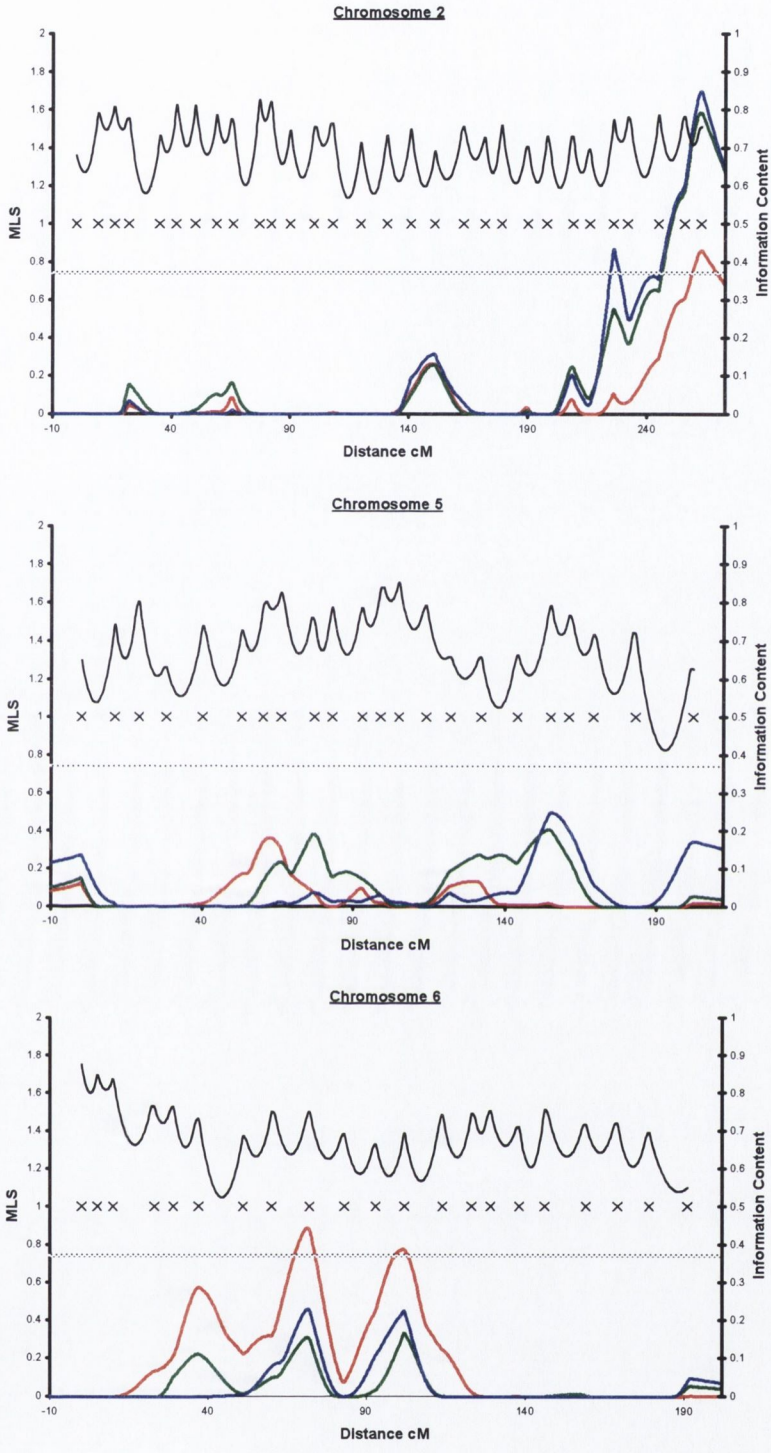


Figure 4.2. Multipoint MLS plots for chromosomes 2, 5, 6, 9, 10, 11, 17, 18, 19, 20, 21 and 22, from Mapmaker/SIBS. Shown are results under Models 1 (red trace), 3 (green trace) and 5 (blue trace); also depicted are the multipoint information content (black trace, right hand axis), and marker positions (X) (see table 4.4 and text); the dotted trace indicates the MLS threshold of 0.74 for follow-up in stage II (see text).

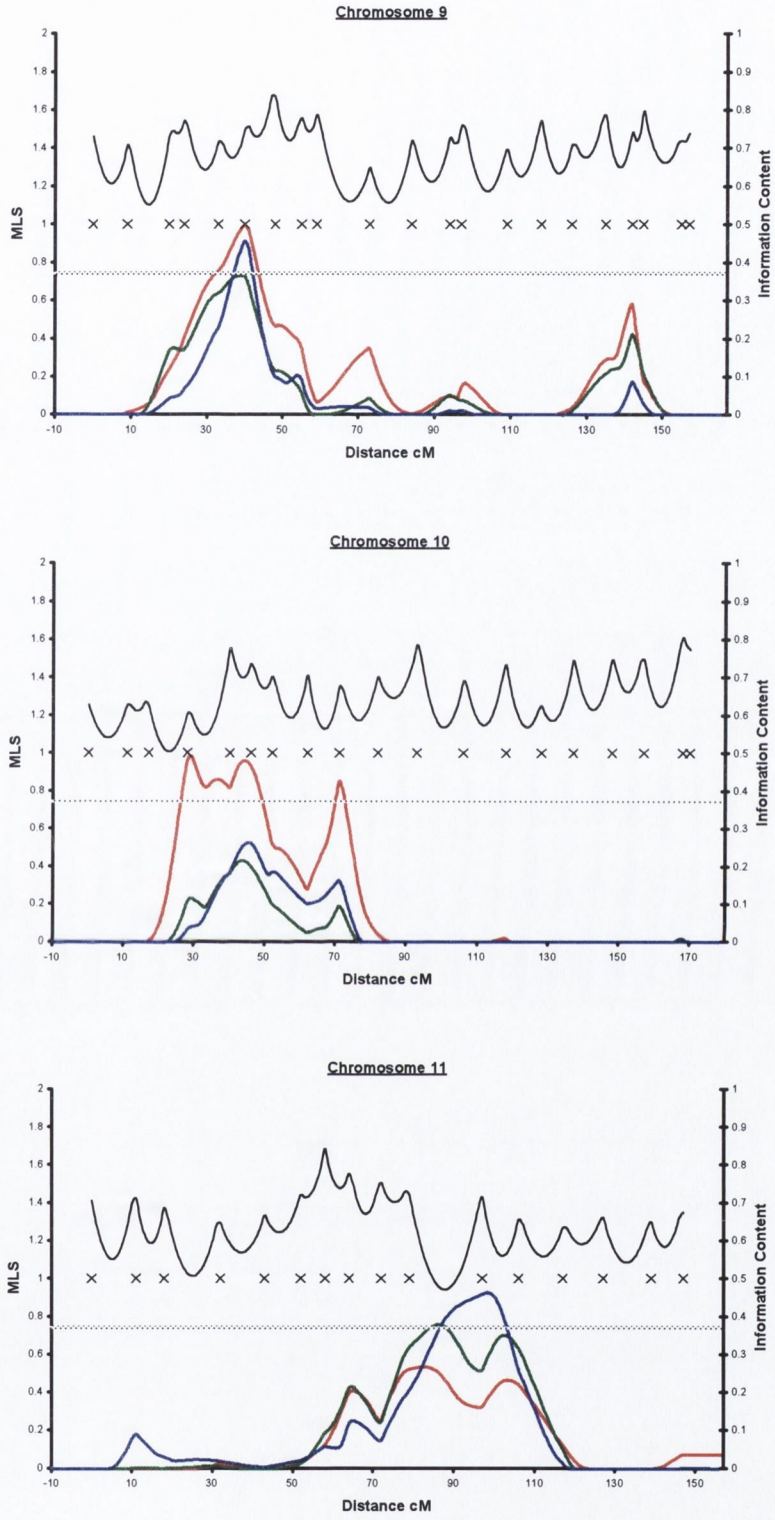


Figure 4.2. (cont.)

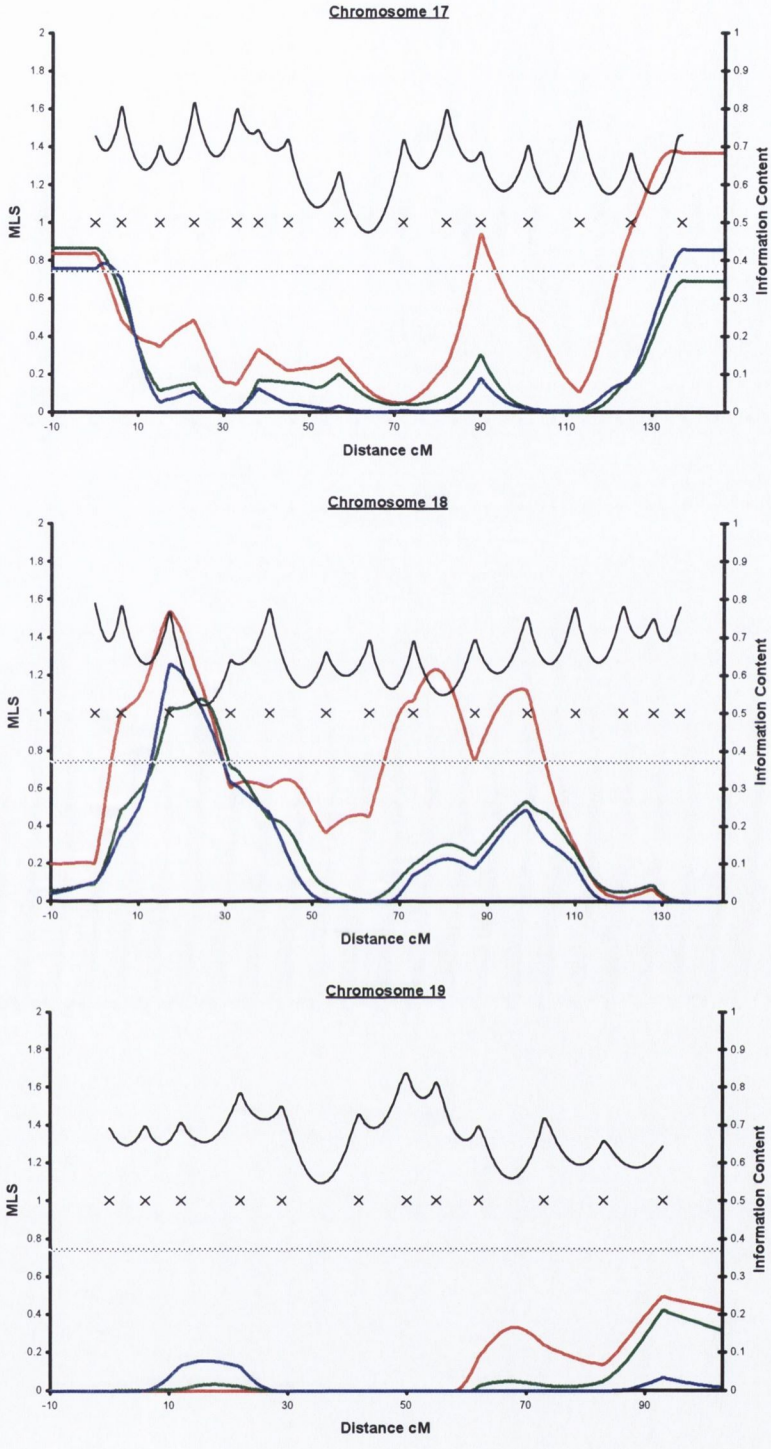


Figure 4.2. (cont.)

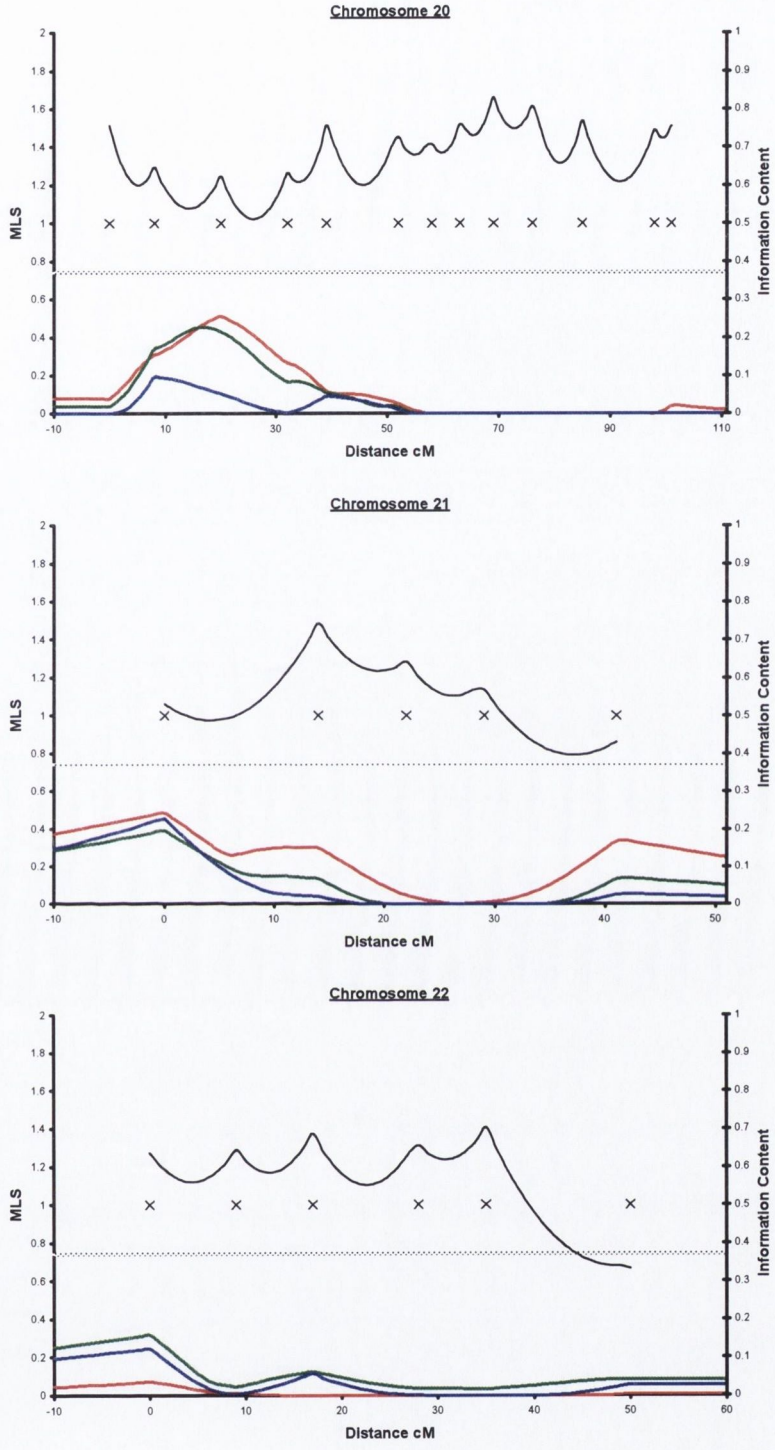


Figure 4.2. (cont.)

Seventeen non-contiguous regions across 11 chromosomes produced scores that exceeded our threshold of a nominal point-wise p-value of 0.05 under disease model 1 (table 4.6). These lay on chromosomes 2q37, 3p26-24, 4p15-p14, 4q12-q13, 6p21, 6q13-q14, 7p15-q21, 9p21, 10p15-p12, 10q11-q21, 12q21, 17p13, 17p15-q21, 17q23-q24, 18p11, 18q21-q22, and Xq21.1. The proximity of some of regions to each other and the nature of the peaks (see figure 4.2) are such that they may represent a single linkage signal, which passes below the threshold at a marker, or between markers, due to random fluctuation. This was taken into consideration when selecting regions to pursue in the second stage (see chapter 5, section 5.1.3).

Chromosome	Position (cM)	MLS (model 1)	Mean Info Content
18	4-29	1.54	0.637
7	97-112	1.40	0.736
17	122-146	1.38	0.633
18	67-103	1.23	0.625
X	97-98	1.18	0.797
9	33-43	0.99	0.712
10	27-49	0.98	0.648
17	89-93	0.94	0.646
6	68-73	0.89	0.704
2	260-269	0.86	0.735
10	70-72	0.85	0.657
17	-10-2	0.84	0.707
3	34-36	0.79	0.852
4	57-58	0.79	0.764
4	81-83	0.78	0.780
6	99-103	0.78	0.64
12	90-92	0.77	0.749

Table 4.6 Regions from first stage genome scan showing a Mapmaker/SIBS maximum likelihood score under disease affection status model 1 (bipolar I sib-pairs only) of > 0.74, or > 1.18 for the X-chromosome; positions are distances from pter in centimorgans over which the score exceeds this threshold; mean information content for the relevant regions is indicated; Dublin-genotyped regions are shaded.

4.3.3 Exclusion Lod Analysis

Table 4.7 shows detailed exclusion analysis per chromosome under the narrow disease model, expressed as a percentage of the length in centimorgans which demonstrated a multipoint exclusion lod above –2, and therefore not excluded as a possible locus with the tested effect size. Most of the genome, approximately 94%, was shown to have an exclusion lod < –2 under a locus conferring a $\lambda_s = 3$. The exclusion was less extensive for weaker loci: about 70% for a λ_S of 2, and only 9% for λ_S of 1.5.

λ_S=	3	2	1.5
Chromosome 1	10	53	93
Chromosome 2	4	19	85
Chromosome 3	7	29	79
Chromosome 4	6	68	100
Chromosome 5	3	38	98
Chromosome 6	11	51	76
Chromosome 7	19	63	97
Chromosome 8	16	97	100
Chromosome 9	11	57	94
Chromosome 10	27	64	76
Chromosome 11	22	64	100
Chromosome 12	0	18	87
Chromosome 13	0	25	71
Chromosome 14	3	48	100
Chromosome 15	0	33	98
Chromosome 16	0	3	72
Chromosome 17	24	71	100
Chromosome 18	67	81	98
Chromosome 19	7	36	82
Chromosome 20	26	60	96
Chromosome 21	38	95	100
Chromosome 22	19	57	100
Genome-wide Mean:	6	30	91

Table 4.7 Percentage of each chromosome demonstrating a multipoint exclusion lod > –2 from first stage genome scan, assuming an underlying susceptibility locus conferring a relative risk to siblings of 3, 2 and 1.5.

4.4 Discussion and Future Work

4.4.1 Discussion of Significance

The purpose of the first screen of the genome was to identify regions, on the basis of a low significance level, to be investigated in more detail. Seventeen regions exceeded our threshold of a multipoint MLS of 0.74. Thresholds corresponding to strict genome-wide significance levels (Lander and Kruglyak, 1995) for the Mapmaker/SIBS maximum likelihood MLS under the possible triangle restrictions, are 3.93 (significant), 2.45 (suggestive), and 0.74 (nominal point-wise p = 0.05) for the autosomes (Nyholt, 2000). The sample-specific power studies suggested that an MLS of 1.7 would occur once per genome scan, and therefore correspond to a "suggestive" linkage finding. Also an MLS > 3 would be expected to occur less than once in every twenty genome scans, corresponding to genome-wide 5% significance. However no region in the study met these levels, nor the strict genome-wide significance levels required to suggest the presence of a susceptibility gene to BPAD. This was not unexpected due to the increasing evidence that genes of small effect confer susceptibility to this disorder. The power calculations for our sample suggested the sample, especially under the narrow disease model, would have reduced power to detect loci conferring a λ_S of 2 or less at a genome-wide significance level (see section 4.3.1). This strongly supports the idea that BPD susceptibility genes are of small effect, λs of 2 or less. This was supported by Peter Holmans' estimation of the effect size of $\lambda_S = 1.72$ underlying our highest MLS score, at marker D18S452.

The exclusion lod analysis excluded 94% of the genome from the presence of a locus conferring a λ_s of 3, at a lod score of –2. Of the autosomal regions which exceeded the MLS threshold, eleven of seventeen were excluded as locations of a locus of $\lambda_s = 3$, lower than the genome-wide exclusion of 94%. It is arguable whether this threshold for the absence of linkage can be usefully applied to a complex disorder, and genome-wide

criteria. On one hand it is dangerous to definitively exclude any region of the genome, especially where the genetic model is unknown, and underlying loci may be of weak effect, and therefore undetectable in the sample being examined (chapter 1, section 1.4.3). In particular the presence of heterogeneity can lead to a true disease locus being falsely excluded (Ott, 1985, pp 262-263). On the other hand the exclusion lod score of –2, representing a likelihood of 100 to 1 against an underlying susceptibility locus, was based on a specific locus effect size, the λ_s, and should be robust in the case of a monogenic or multiplicative polygenic locus model (Kruglyak and Lander, 1995).

4.4.2 Overlaps with previous Studies

Once the regions exceeding the threshold were identified it became of interest to determine which of these regions should be given highest priority in the second stage. An intuitive way of evaluating this is to contrast the results with linkage findings in previous studies. This can be hampered by lack of a fully integrated marker map, but linkage findings are usually broad enough for this to be done in a relatively accurate if not precise manner. Often upon examining overlap of linkage results at a genomic region, studies that have found no evidence are overlooked. This discounts the potential difference between, for example, two studies which examine the region and both find linkage, and two studies which find linkage, from a total of twenty studies which examine the region. Direct comparison of the evidence for linkage in different studies is also hampered by the use of different linkage statistics. The evaluation of coincidence of results across studies in a formal statistical sense, is dealt with in more detail in chapter 6. Comparisons of the locations of our highest scores with those of previous scans reveal several regions where weak or suggestive evidence of linkage to BPAD and schizophrenia overlap.

Our broad model peak on chromosome 2q lies about 32 cM centromeric of a region where, previously, a lod score of 1.03 at marker D2S159 was reported (Morissette et al, 1999), and the NIMH Bipolar Initiative (Stine et al, 1997) reported excess allele sharing at D2S427 about 27 cM centromeric of our MLS, with a nominal $p < 0.05$. These findings are relatively distant from our peak. This region is known to harbour several neurodevelopmental genes, the ATSV (axonal transport of synaptic vesicles) gene and the serotonin 2B receptor gene.

Centromeric chromosome 18p has been the focus of much interest since the first report of suggestive evidence of a susceptibility gene by Berrettini et al (1994) centred on D18S21, on the centromeric side of our maximal peak (see chapter 1, section 1.5.2). Detera-Wadleigh and colleagues (1999) also reported interesting results in the more centromeric region of this peak. In schizophrenia, Schwab et al (1998) found a possible susceptibility locus, evidence for which was enhanced when individuals with affective disorder were included in the disease affection model. Stine et al (1995) reported a lod of 3.51 at D18S41 on 18q, which overlaps our q-arm peaks. Also a study of BPI disorder patients in a large Costa Rican family found evidence for linkage to a region overlapping ours (McInnes et al, 1996). This region is also reported to show a parent-of-origin effect (Stine et al, 1995), which the Dublin group is currently investigating.

Our peak on chromosome 9p22 at markers D9S171 and D9S161 coincides with the finding of Morissette et al (1999) of lod scores of 1.37 and 1.14, respectively, at these markers. Modest linkage at these same markers, with lod scores of 1.27 and 1.29 respectively, was also reported by Cichon et al (2001b) in a genome scan study.

On chromosome 10 our peak covers the potential schizophrenia susceptibility region (Straub et al, 1998; Foroud et al, 2000). However, this region has been reported to exhibit segregation distortion (Paterson and Petronis, 1999a), which has not been excluded in our sample.

Neither chromosome 17 nor chromosome 6 have been of much past interest for linkage to bipolar disorder. Chromosome 17 is the location of the serotonin transporter gene (SERT) at 17q11-q12, where Murphy et al (2000) found evidence for linkage to bipolar disorder. The gene lies about 50cM from the p-terminus, and about 40cM from our peaks, which lie at each telomere and centrally, at 90cM from the pter. Our peak at the pter overlaps a finding of a lod score over 1 in one of the families of McInnes et al (1996). On chromosome 6 Kelsoe et al (2001) reported a lod score of 1.2 at chromosome 6q23. Our peaks at 6p12 and 6q16 are at least 20 cM away from this finding. However, our q-arm peak lies only 10 cM from modest linkage identified in the NIMH collaborative genome scan (Rice et al, 1997). Our p-arm peak lies 20-30 cM distal of findings in schizophrenia on chromosome 6p22 (e.g. Straub et al, 1995).

4.4.3 Summary

A genome scan was performed on a joint British and Irish population of affected sibling pairs, plus parents and unaffected siblings, where available. The sample consisted of 541 individuals classified into five diagnostic models, with 150 ASPs under the narrowest, and 255 under the broadest models. As part of a collaborative study, half the genotyping was performed in Dublin, and the remainder was performed by Prof. Nick Craddock's group in the University of Birmingham. For efficiency and cost, the project was undertaken within a split sample paradigm, with marker grid tightening employed in the second stage (Holmans and Craddock, 1997). Statistical non-parametric linkage analysis did not reveal any

markers or regions significantly linked to BPD at a genome-wide level under any diagnostic model. However, seventeen regions exceeding the threshold MLS of 0.74 (1.18 for the X chromosome) were selected for follow-up in the second stage, on chromosomes 2, 3, 4, 6, 7, 12, 17, 18, and the X chromosome. These regions were selected for inclusion in the initial phase of the second stage genome scan (chapter 5).

Chapter 5. Genome Scan Stage II: Phase 1

5.1 Introduction

5.1.1 Study rationale and Design

The second stage of the genome scan for bipolar disorder (BPD) is underway separately in the Dublin and Birmingham centres. The expanded sample of 291 British and Irish pedigrees was collected, and the aim was to examine the expanded set of pedigrees for linkage to regions 'suggestive' of linkage from the first stage data. In the long term the project will fully genotype the extended sample at an average density of 5 centimorgans. This is to avoid the possibility that regions of the genome not reaching the significance level for follow-up might be false negatives.

The second stage was divided into two working phases. First the extended sample was genotyped with the 10 cM resolution marker map used in the first stage (chapter 4, section 4.2.3). Second, the grid will be tightened to a 5 cM resolution with the genotyping of extra markers in the full sample. Linkage analysis was performed at the intermediate point in the second stage genome scan in order to determine whether our selection of initial follow-up regions fulfil their potential as higher-priority regions. This approach may not be strictly statistically appropriate, but was necessary on a pragmatic level, with a view to maximising efficiency in the use of limited resources. Thus the significance levels of these results are not evaluated in a genome-wide sense, and only observations of increased, decreased or unchanged evidence for linkage to these regions in the extended sample were made.

5.1.2 Candidate Areas for Follow-up

For convenience, each collaborative group was given first refusal for the follow-up work on chromosomes they had typed in the first stage, minimizing the need for redistribution of

reagents. The eleven regions on Dublin-genotyped chromosomes which exceeded the first stage MLS threshold are shown in figure 5.1(A-G).

Several formal methods exist for evaluation of the true location of a linkage finding, or confidence interval for a lod score. For example, the "lod minus 1" support interval for a maximum likelihood lod score peak is given by the width of the peak at a point, *m* units below the maximum lod score, where *m* is 1, or 2 for a 95% confidence interval (see Terwilliger and Ott, 1994). However, due to the low threshold scores, and lax significance level we adopted in the first stage, these methods are not applicable, or would recommend too wide an area for follow-up. The variation in location of lod peaks in relation to the underlying susceptibility locus can be large, especially for incompletely penetrant loci or in the presence of heterogeneity, and are in inverse proportion to the peak score (Roberts et al, 1999). Therefore linkage findings that are even tens of centimorgans away from our peak cannot be ruled out as representative of signals from the same locus.

In order to assign a priority to each of the follow-up regions, previous indications of linkage to BPD were taken into account, as well as the linkage scores from the three diagnostic models examined in the first stage. The overlap between the peaks on Dublin chromosomes and other linkage studies are described in chapter 4, section 4.4.2. The Model 1 scores are summarised again here, and indicated on figures 5.1(A-G). On chromosome 2, two studies have found weak linkage to BPD about 20cM proximal of our peak (Figure 5.1A; Stine et al, 1997; Morissette et al, 1999). Chromosome 6 has not been of much interest from previous linkage studies in BPD, however Turecki et al (2001) found weak linkage about 25cM from our peak in a sample ascertained using excellent response to Lithium treatment in probands (Figure 5.1B). Both Morissette et al (1999) and Cichon et al (2001b) found linkage to the same two markers, D9S161 and D9S171, on chromosome 9

which exceeded our follow-up threshold (Figure 5.1C). On chromosome 10 our larger peak coincided with a finding in the NIMH genome scan (Figure 5.1D; Rice et al, 1997; Foroud et al, 2000). Their region of interest extends beyond our peak towards the centromere for another 15cM, coming within 10cM of our smaller peak on chromosome 10q. Two studies have highlighted linkage to chromosome 11 near our peak (Figure 5.1E). Coon et al (1993) reported weak linkage (lod score of 1.0) about 5cM distally, and Detera-Wadleigh et al (1999), and Cichon and colleagues (2001a, 2001b) each reported weak linkage 10cM from the other side of our peak. Chromosome 17 has been investigated for association at the locus of the serotonin transporter gene, but suggestive linkage findings have not been reported in the vicinity of our peak, which lies some distance from this candidate gene. The nearest report of linkage was made by Curtis et al (in press) 10-20 cM towards the q-terminus (Figure 5.1F).

Chromosome 18 has also been the focus of several linkage studies, and the regions of positive linkage are too numerous to depict in a simple diagram (see chapter 1, section 1.5.2). They extend over much of the chromosome, but our two peaks (Figure 5.1G) could be said to correspond to two clusters of findings. Studies have found linkage in the chromosome 18 pericentromeric region (Berrettini et al, 1994; Stine et al, 1995; Coon et al, 1996; Gershon et al, 1996; Detera-Wadleigh et al, 1999) and over an area spanning 18q21-25 (Stine et al, 1995; De Bruyn et al, 1996; McMahon et al, 1997; Morissette et al, 1999). However some studies have found no linkage to chromosome 18 (Pauls et al, 1995; Claes et al, 1997; Bowen et al, 1999).

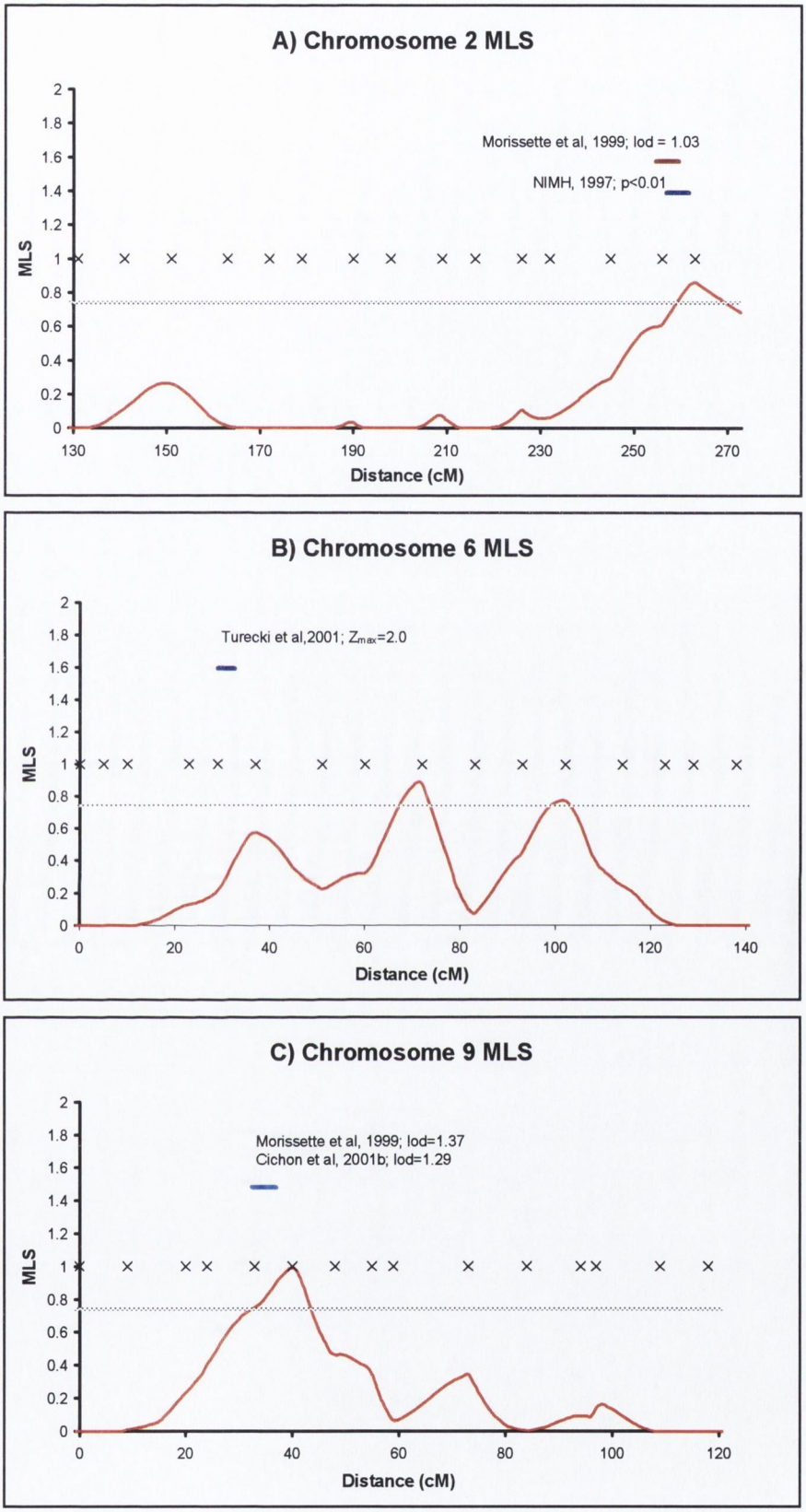
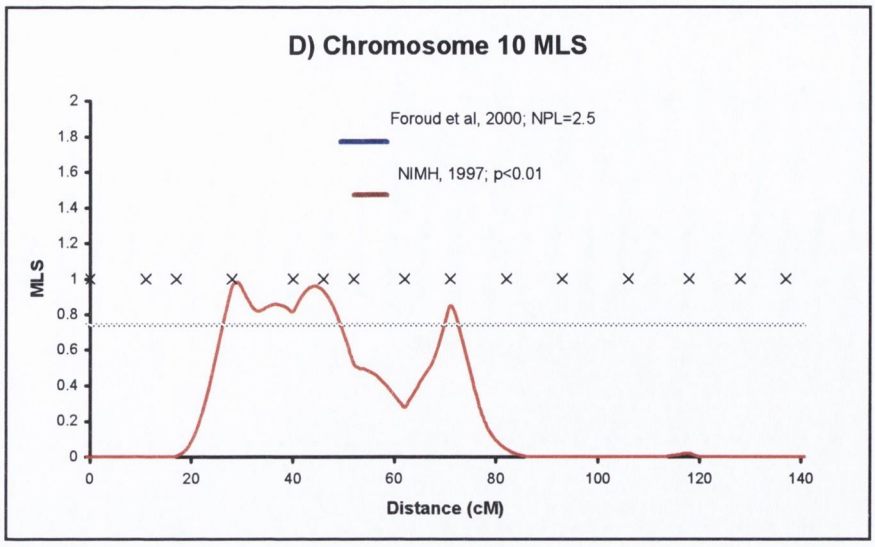
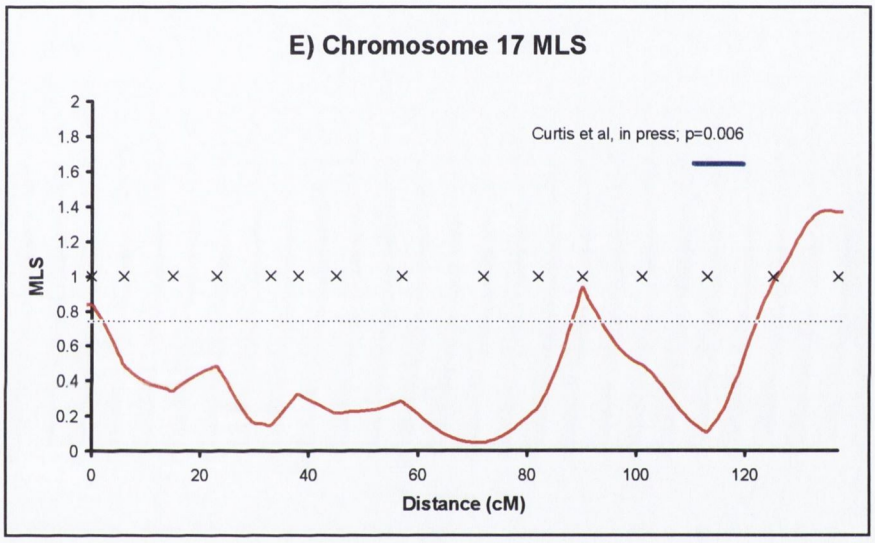


Figure 5.1(A-C) Positions of the nine regions on Dublin-genotyped chromosomes which exceeded the MLS threshold of a 5% nominal significance level under the narrow disease model; X indicates marker positions; also marked are positions of linkage findings in bipolar disorder close to these regions (see text for details).





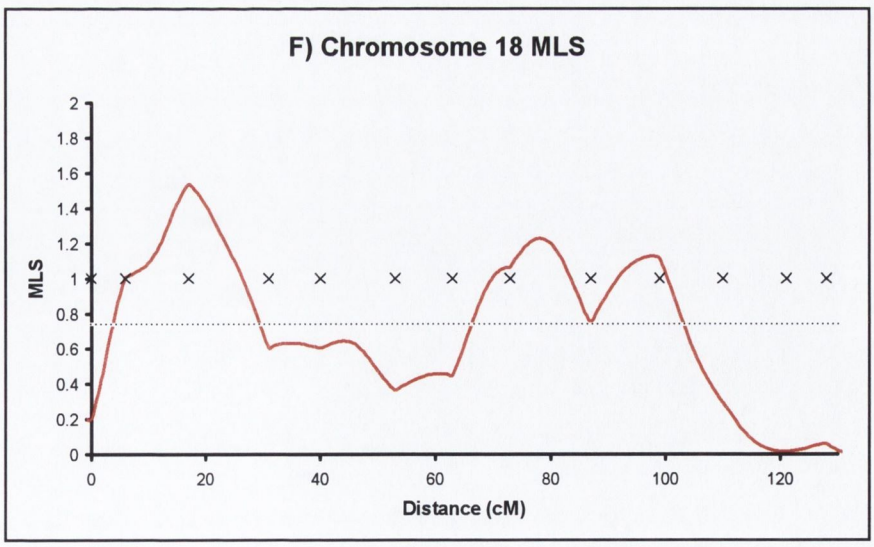


Figure 5.1(D-F). (cont.)

5.1.3 Selection of Follow-up Regions

On the basis of the various criteria outlined above, three regions were given the highest priority for examination in the second stage genome scan:

Chromosome 2q

The Model 1 MLS peak of 0.86 on chromosome 2q37 was chosen as a promising follow-up region because the results of analysis of disease Model 5 provided the highest overall MLS score of the first stage scan (Model 3 MLS = 1.58; Model 5 MLS = 1.70). In addition, modest evidence for linkage has been published for markers D2S159 (Morissette et al, 1999), and D2S427 (Stine et al, 1997), both about 20 cM proximal of our peak.

Chromosome 10

Interesting linkage findings at the distal edge of our chromosome 10 MLS peak have been published by the large NIMH collaborative genome scan project (NIMH, 1997), and represent their most promising candidate region (Foroud et al, 2000). In schizophrenia this region has previously been of interest (Faraone et al, 1998; Schwab et al, 1998; Straub et al, 1998). The presence of weak segregation distortion in this area of chromosome 10p has been suggested (Paterson and Petronis, 1999a), but has been excluded in the NIMH schizophrenia sample (Faraone et al, 1999). The peak on chromosome 10 was selected for follow-up as it also presented a broad peak, exceeding the threshold across an interval of 22 cM. Another 20 cM distal from this peak the score exceeded the threshold again at marker D10S196. A total genetic distance of 65 cM was selected for further genotyping, overlapping the NIMH region.

Chromosome 18

Although uncertainty exists in the literature about the exact locations of linkage evidence on chromosome 18, general consensus in the field suggests the possibility of a locus on the p-arm and one or more loci on the q-arm (Sklar, 2002). These approximate to the locations of our peaks, and these were chosen due to the strength of our linkage finding, and the evidence from the literature.

The author genotyped and performed linkage analysis on regions from chromosomes 2 and 10, in conjunction with David Lambert and Dr. Sylvia Bort. The two peaks on chromosome 18 are under investigation by David Lambert, who will also investigate the possibility of a parent-of-origin effect. To complete the first phase of the second stage follow-up, the Birmingham group gave priority to regions on chromosomes 4, 7, and the X chromosome.

5.2 Materials and Methods

5.2.1 Sample Description

Sample collection and ascertainment were as described in chapter 2, section 2.2. Briefly, patient diagnoses were made according to DSM IV criteria (American Psychiatric Association, 1996). New pedigrees were ascertained as before, on the basis of a proband with a diagnosis of bipolar I disorder (BPI), and subsequent ascertainment of siblings with affective illness. Each affected sibling pair was classified into one of three disease models on the basis of the second and subsequent diagnoses, omitting the Model 2 and Model 4 classifications of the first scan. The omitted models did not provide any more information than the most narrow, the most broad and an intermediate model, almost invariably yielding an MLS midway between the flanking models. At the same time, the use of five models increased the number of tests, and hence decreased the significance of the linkage

results, although this was not taken into account under the first scan. Therefore only a narrow "Model 1" incorporating BPI, an intermediate "Model 3" adding Bipolar II disorder (BPII) and Schizoaffective disorder – bipolar subtype (SAB), and a broad "Model 5" extending the disease affection model to Bipolar Disorder – not otherwise specified (BPN) and Recurrent Unipolar Disorder (RUP), were examined.

Diagnostic model:	**Narrow**	**Intermediate**	**Broad**
Pedigrees with:			
Affected sibling pair only	89	105	138
ASP + 1 parent	36	40	45
ASP + both parents	34	39	49
no extra unaffected siblings	79	104	154
1 UA sibling	58	60	56
2 or more UA siblings	22	20	22
Total (*pedigrees*)	159	184	232
Total ASPs (*all pairs*)	218	283	395
Total ASPs (*weighted*)	186	229	305
Number of pairs:	136	146	174
trios:	20	33	45
quads or greater:	3	5	13

Siblings with:	**Total (*M/F*)**
Total Bipolar I disorder	463 (*204/259*)
Total Schizoaffective (bipolar) disorder	44 (*12/32*)
Total Bipolar II disorder	13 (*2/11*)
Total Bipolar disorder (NOS)	23 (*14/9*)
Total Major depressive disorder (recurrent)	76 (*21/55*)
Total classed as unknown/other	263 (*122/141*)

Table 5.1. Second stage genome scan sample description

Legend: ASP- Affected Sibling Pair; M- Male; F- Female ; UA- Unaffected; diagnoses as referred to in text (APA, 1996); unweighted sibling pairs calculated as the sum over all pedigrees using n(n-1)/2, where n is the number of affected siblings in the pedigree; weighted sibling pairs calculated similarly, using (n-1).

The extended sample consisted of the addition of new families to the sample, and in a few cases the addition of affected and/or unaffected individuals to existing families, as the clinical sample collection for the first and second stage samples was a continuous process. The total sample to be genotyped consisted of 891 individuals (623 affected) from 239

families. The details of the extended sample under each diagnostic category is given in table 5.1.

5.2.2 Marker Map

The markers to be first used in the second stage were taken from the Linkage Mapping Set version 2 (LMSv2, now re-named Medium Density 10cM set) marker set (Appendix C), as described in chapter 4. These consisted of dinucleotide repeat microsatellite markers spaced at ~10cM, for which the additional individuals of the second scan were typed. The markers to be genotyped in the extended sample were selected by visual inspection of the MLS peak and adjacent areas. All markers adjacent to the peak for which the MLS remained above zero, were adopted. In addition, the other disease models were examined in order to ensure that the true location of the peak was consistent across these. As a result an additional chromosome 2 peak at about 226 cM was observed under model 5, and the marker D2S126 included in the follow-up map.

5.2.3 Genotyping

Polymerase chain reaction was performed using identical conditions to the first stage (chapter 2, section 2.5.2). Briefly, the primer pairs were obtained from Perkin Elmer-Applied Biosystems International (ABI), and amplified with the pre-optimised True Allele™ Pre-mix (containing Taq Polymerase, dNTPs and appropriate buffer) according to the manufacturer's recommendations. Gel electrophoresis was performed on the ABI 377 DNA sequencer, and alleles were called with ABI GeneScan and GenoTyper software (see chapter 2, section 2.5).

The MegaBase database software required some adjustment before use in the second stage. Additional individuals were added to the database. Due to a limitation on the maximum

number of genotyped markers, a 'cleaned' version of the program was obtained from Iain Fenton, with all the LMSv2 markers deleted except those on our three follow-up chromosomes. Genotype data was imported from plain text comma separated value (CSV) files as before, and the alleles were binned (see chapter 2, section 2.5.6). LINKAGE format files were exported for linkage analysis.

The data flow and error control procedures from the first stage genome scan were maintained for stage II genotyping (chapter 4, section 4.2.4). An overall genotyping target rate of 95% was maintained. In the case of failed genotypes, the sample identifiers were verified to ensure that as many genotypes as possible from affected members of a sibling pair were obtained. Computer copies of all data files were made to CD-ROM disks, several separate computer hard disks, and from there to a magnetic tape back-up system kept at a physically separate location.

Before linkage analysis, the data was subjected to investigation for Mendelisation errors with the PedCheck program, and examined with the RELATIVE and SibError programs to identify potential monozygotic twin pairs and half-sibships (chapter 2, sections 2.6.1, 2.6.2 and 2.6.3).

5.2.4 Statistical analyses

Remote computers at the Human Genome Mapping Project research centre (HGMP-RC) were used to perform the linkage analyses. Genotype data used in these analyses were limited to the markers which had been genotyped for the full sample. If markers genotyped only in the first stage had been used, it would decrease the information content, and may artificially deflate the overall multipoint linkage score. The SPLINK program provided updated allele frequencies from our dataset for use in the linkage analysis (chapter 2,

section 2.6.6). The principal statistical method remained the multipoint MLS score generated by the Mapmaker/SIBS package (chapter 2, section 2.6.8). The GENEHUNTER program was used to obtain recombination counts to estimate genotyping error rates (see chapter 2, section 2.6.7).

5.3 Results

5.3.1 Scan IIA results

The Mapmaker/SIBS MLS are depicted by the darker colour traces in Figure 5.2. The statistic was increased from the first stage scan at both candidate regions. The increase with the addition of extra affected sibling pairs was more striking at the chromosome 2 locus. On chromosome 2, a maximum score of 2.52 was achieved (green plot) under the Intermediate model (BPI, schizoaffective, and bipolar II disorder). The Narrow model (red plot) attained a maximum score of 1.88, and the Broad model (blue plot) a maximum score of 2.39. Chromosome 10 registered a maximum score of 1.36 under the broadest disease model at marker D10S197. Under the narrower models, the linkage peak broadened, with the intermediate model yielding a score of 1.30 also at D10S197, and the narrow model a score of 1.29 at D10S547.

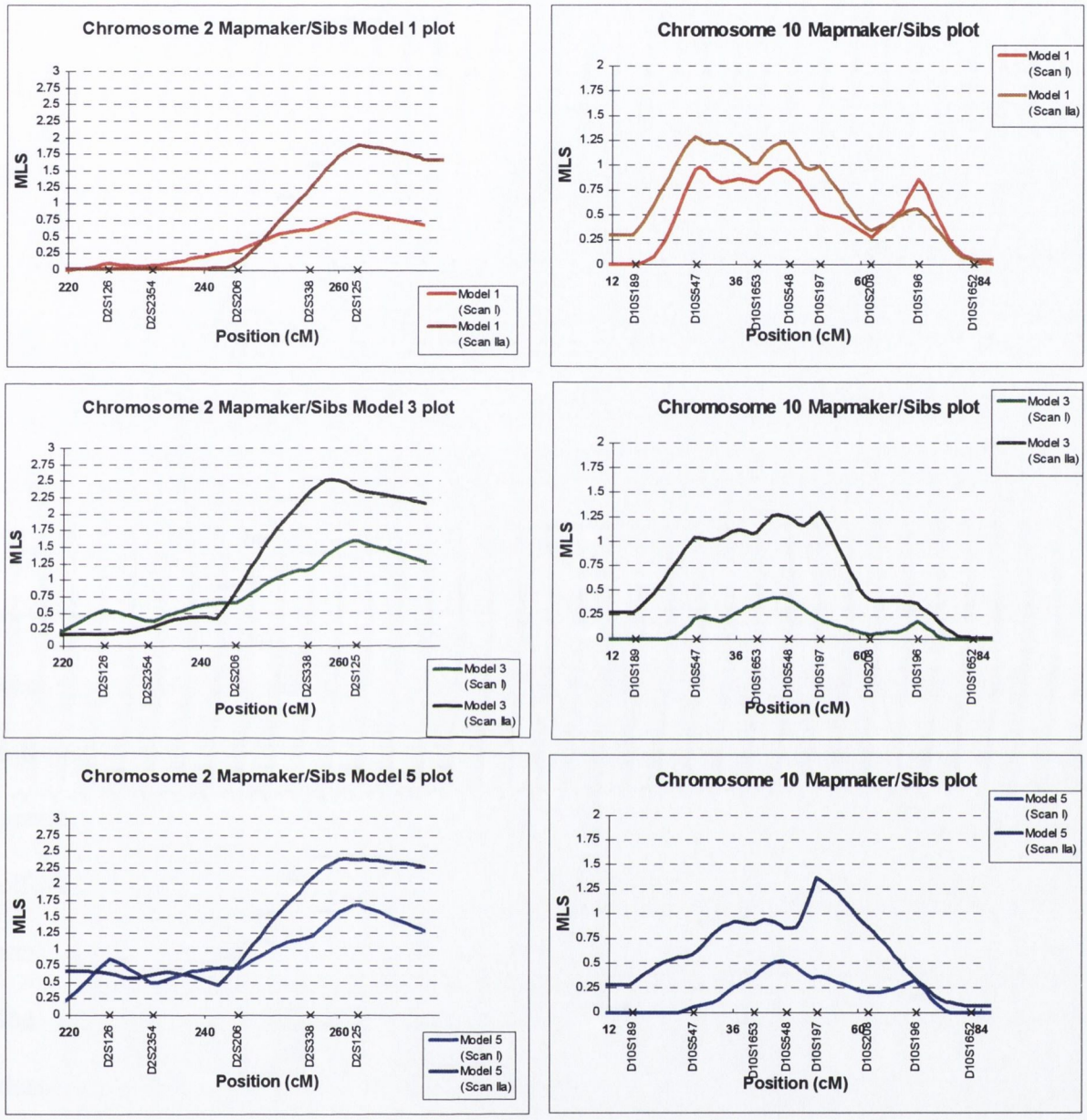


Figure 5.2. Increase in Mapmaker/SIBS MLS from scan I to scan IIa. Intermediate results of the second scan are represented by the darker lines, first scan results by the lighter traces.

5.3.2 Error Rate and Quality Control

Some individuals were omitted due to, amongst other factors, failure to obtain DNA samples for the full sib-pair, confirmation of a sib-pair as monozygotic twins or half-siblings, or detection of unexplained Mendelisation errors in the pedigree which led to its removal from the sample (as described in chapter 4). The added individuals presented no new problems in this regard.

The Genehunter program reported expected and observed recombination counts for each marker pair from the genotype data. The mean increase in observed recombination counts above what would be expected from the genetic map used, was 11.1%. A total of 1414 meioses were informative on chromosome 2, and 1426 on chromosome 10. The interval between markers 4 and 5 (D10S548 and D10S197) on chromosome 10 was of particular concern, as the observed recombination count was 35% larger than expected, and the intervals between the four flanking markers were all above 10%. This may reflect an high genotyping error rate, which would tend to increase the observed recombination count. The effect of this on the linkage score is difficult to determine, but it is likely to deflate the multipoint score. Douglas et al (2000) have shown that the effect of genotyping error on the probability of detecting linkage in a sample of ASPs can be dramatic. they demonstrated that the expected maximum lod score can be decreased by around 50% for an underlying locus corresponding to a recurrence risk ratio, $\lambda = 1.25$, or by 20-30% with a λ between 2 and 4. Thus the genotyping error rate estimate of 1.6% (section 4.3.2) could have had an important effect on deflating our linkage scores, as suggested by the inflated recombination rates observed.

5.4 Discussion

5.4.1 Evidence for Linkage

Evidence for linkage of BPD to chromosomes 10p and 2q in our sample was increased upon expansion of the sample. The increase was most striking under the narrow disease model on chromosome 2q37, with an increase in the MLS of over 1. The broader models continue to show more evidence for linkage than the most stringent, with the maximum score of 2.52 now shown with the intermediate model, between markers D2S338 and D2S125. On chromosome 10, more modest increases in the MLS were seen under the narrow model. Increases in the MLS of about 0.3 were seen at the main peak under the narrow model at markers D10S547 and D10S548. The increases were larger under the intermediate model (0.65), and under the broad model (0.8), with a shift of the peak toward the distal end under the latter. Both broader models now show MLS in excess of 1, whereas neither showed a nominal 5% significance level (MLS > 0.74) after the first stage.

5.4.2 Summary

These results represent the initial phase of the second stage of the genome scan project. Therefore these results could not be used as the basis of any conclusions. These regions are committed to being investigated at the higher 5 cM marker resolution, which should extract close to full linkage information from the sample. Nevertheless, these preliminary results are encouraging, on the basis that increasing the power of the sample set does not reduce or eliminate the evidence for linkage. The results on chromosome 2 are of interest, where a large increase in the MLS was observed in a region which has not previously been identified as a candidate locus for a susceptibility gene for BPD. This region may represent a specific locus in the Anglo-Irish population, although there have been reports of weak linkage in other datasets (Stine et al, 1997; Morissette et al, 1999).

The chromosome 10p results are equally encouraging, especially since there has been another report of linkage to this area in BPD (NIMH, 1997), and also in schizophrenia (Straub et al, 1998) and alcoholism (Paterson and Petronis, 1999b). The suggestion of a pleiotropic 10p locus may be reinforced by the increased MLS under the broader diagnostic categories at this stage of the follow-up study, but definite conclusions cannot be drawn. However, weak transmission distortion may exist in this area (Paterson and Petronis, 1999a), in the form of reported linkage to the female sex. This could in theory inflate linkage scores if there is an elevated number of females with the disorder, as was the case in this sample (see table 5.1).

The attempts to prioritise first stage linkage peaks highlighted an important issue in genetic linkage studies. The examination of coincidence of linkage findings across studies is frequently made, yet is an *ad hoc* procedure. A true evaluation of the extent to which studies have highlighted the same genomic region is difficult to make objectively, yet is often the basis for further investigations. It has been shown that the standard error in location of linkage score peaks is a function of the expected lod score of the study, and can be of the order of tens of cM for smaller studies and susceptibility genes of small effect (Roberts et al, 1999). I have addressed this issue further in chapter 6.

Chapter 6. Genome Scan Meta-Analysis

6.1 Introduction

6.1.1 Overlap of Linkage Findings

Many genetic linkage studies in bipolar disorder (BPD) have been conducted (see chapter 1, section 1.5.2). However, evidence for linkage to particular chromosomal regions reported in any one study, have proven difficult to replicate in independent samples. Genome scans are in principle comparable tests for linkage in independent samples, yet consistent evidence for linked genes has not emerged. Every linkage report is necessarily accompanied by an evaluation of where the results fit in with previous findings. This has led to the identification of "hot-spots" for linkage, where *ad hoc* comparisons have led to an accepted view about the likelihood of linkage within a region. However, these comparisons are not objective and can be subject to subtle biases.

Firstly, the phenomenon of publication bias may arise from the reluctance of researchers and journals to publish findings with low or no statistical significance. These results are therefore absent from any subsequent evaluation of mildly positive linkage findings. Similarly, over-eagerness to publish positive findings, or findings of any sort in regions of great interest, may lead to bias, if appropriate genome-wide significance levels are not applied (Lander and Kruglyak, 1995), or if the findings are not presented in the intended genome-wide context. This makes it difficult to identify regions consistently highlighted across studies, if individual linkage results are weak, or not localised to the most fashionable candidate regions.

Another complication in the identification of co-localisation of findings are the wide confidence intervals which apply to modest linkage score peaks, as obtained with underlying loci of weak effect (Roberts et al, 1999). Coincidental random fluctuations in

the linkage score should not produce false positive results if stringent statistical criteria are followed. However, there is a danger in assuming findings as replicated in independent studies on the basis of the location of linkage peaks. The probability of such spurious coincidence is highly dependent on the strength of the underlying genetic effect. This point was made clear by Lander and Kruglyak (1995), when they simulated a genome scan with 3 cM marker resolution and no underlying susceptibility gene, and observed one locus demonstrating "suggestive" linkage and 22 other loci exceeding a nominal p-value of 0.05, as would be expected from power calculations. A simulated "follow-up" scan, again with no underlying locus, "replicated" the suggestive peak and identified several other regions of interest, some containing candidate genes. Low power to detect an underlying locus would leave a study in this situation, where the random noise rivals the signal from a true linkage, possibly leading to spurious replications.

Therefore, one of the possible explanations for the dearth of replicated linkage findings to BPD is a lack of power to detect loci of small effect. Combining studies should increase power, due to both a reduction of the number of independent tests performed, and an increase in the size of the sample examined. This should allow an increase in the size of a true 'signal', while reducing the 'noise' of random fluctuations in the linkage score. Thus, if a measure of the evidence for linkage across all studies can be obtained, and this may allow a true underlying peak to emerge from background noise, while at the same time reducing the importance of spurious positive results, in the same way that an event-related potential is detected by repeated sampling of scalp EEG (electroencephalograph) readings.

6.1.2 Problems with Meta-Analysis

The ideal approach to meta-analysis of genetic linkage is to combine raw genotypes, and reanalyse the entire dataset with a single statistic (Lander and Kruglyak, 1995). However

analysis of raw genotype data with a unified statistic would involve a large amount of work and make this unattractive for many researchers. Therefore, a more pragmatic method would allow the combination of evidence for linkage, as published, rather than raw genotypes. The lod score was originally designed to be a measure of linkage which could be sequentially sampled (Morton, 1955). Lod scores from separate families can be summed until a threshold for statistical significance is passed. In practice, the differences between linkage statistics complicate this procedure. Often, differences in the assumed genetic model, the diagnostic classifications, and the genetic markers examined means they cannot be directly combined in this manner.

The issue of marker coverage is problematic, as many published linkage studies are limited to candidate regions. If only some or a few markers are identical across some of the studies, highly variable information content will result, with most of the families not genotyped for most of the markers in a pooled sample. In addition, statistical considerations of highly variable marker density may make the analysis prohibitively complex. A meta-analysis would therefore ideally be limited to genome-wide linkage data, which should be directly comparable and are typically conducted in large samples.

A potential stumbling block in BPD studies are the diagnostic criteria used. There is often variation in the constitutive diagnoses of disease affection models, although most modern studies have very similar or identical diagnostic criteria through the use of modern operational criteria (see chapter 1, section 1.1.2). Depending on the extent of genetic heterogeneity across the affective diagnostic spectrum, narrower or more broad disease models may have different power to detect linkage. It is generally accepted that Bipolar I disorder (BPI) is the most stringent and potentially most genetic classification of the diagnostic spectrum, as it confers the highest relative risk in family studies (chapter 1,

section 1.3). However, few research groups agree on exactly how other diagnoses relate to each other and to the underlying, and unknown, genetic component.

Phenotypic or locus heterogeneity will lead to reduced power to detect linkage to any trait. Despite the relatively uniform incidence and familial measures of BPD across countries, the possibility remains that different genes are acting in different populations, as some researchers have suggested (Morissette et al, 1999; Friddle et al, 2000). In the case of complete genetic locus heterogeneity for BPD, a sample pooled using any methodology would not provide increased ability to find any one locus. However the latest segregation studies in BPD (Craddock et al, 1995; chapter 1, section 1.3.3) suggest that this model is not probable.

6.1.3 Types of Meta-analysis

Multiple Scan Probability

Fisher's (1932) method for combining p-values can be applied to linkage data (Allison and Heo, 1998). Badner and Gershon (2002a) implemented this method for meta-analysis of regional evidence for linkage in the Multiple Scan Probability (MSP) statistic. The MSP is applied by obtaining p-values from the linkage statistics used, and combining them with a correction for the size of the region examined. The MSP has been applied to BPD (Badner and Gershon, 2002b), and is discussed further below (section 6.4).

Genome Scan Meta-Analysis

Another method, applied here, is the Genome Scan Meta-Analysis (GSMA), which requires considerably less manipulation of the data (Wise et al, 1999). It is based on the ranking of regions of the genome according to the linkage score, and subsequent summation of the ranks across scans. The minimum information required from each study

is therefore a list of the markers used, and the ranking of these markers by the evidence of linkage. The GSMA method allows the possibility of a comprehensive meta-analysis from either published linkage reports, or if available, full genome data sets. It allows each participating scan to use different diagnostic criteria, if largely comparable, exemplified by the GSMA for multiple sclerosis and auto-immune disorders (Wise et al, 1999). It allows weighting of individual scans on the basis of a discriminatory factor such as diagnostic criteria, or statistical power. It also allows each scan to have used a completely separate set of genetic markers, as long as they can be placed relative to each other on a genetic map.

The basis of the method is the analysis of defined 'bins' of fixed width across the genome. The marker set used in each genome scan is split into the bins, and the maximum linkage score in each bin is noted. The bins are then ranked for each scan, and a summed rank is calculated for each bin across all the scans participating in the analysis. The summed rank, R, obtained for each bin has a known distribution under the null hypothesis of no susceptibility locus within that bin. For m genome scans and n bins the probability that the sum of ranks X_i is equal to R is given by:

$$P\left(\sum_{i=1}^{m} X_i = R\right) = 0, \qquad \text{for} \quad R < m \quad \text{and} \quad R > mn,$$

$$= \frac{1}{n^m} \sum_{k=0}^{d} (-1)^k \binom{R - kn - 1}{m - 1} \binom{m}{k}, \qquad \text{for } m \le R \le mn$$

where d is the integer part of $(R\text{-}m)/n$. This distribution allows exact p-values to be calculated for given values of m and n. However, an assumption is made that the maximum lod scores within bins are independent and identically distributed, which means markers within a bin should be evenly spaced and have equal informativeness. This should be well

approximated if each bin contains at least one marker (Wise et al, 1999). A sufficiently wide bin should allow a greater degree of independence of lod scores, although a multipoint score will tend to reduce this. However, the bin must be sufficiently narrow to reduce the probability of two distinct linkage peaks falling within it. A bin width of 30 cM was adopted as the best compromise between these requirements. This allows 120 bins across the 22 autosomes, with 10 bins on the largest, and 2 bins on the smallest chromosomes (see Appendix F). It became clear that the bin positions may have an effect on the results of the ranking. For example, if a true susceptibility locus lies on the boundary of two bins, the evidence for linkage may be diluted between these bins. Therefore, an exploration of the effect of bin positioning on the ranks was performed.

The X chromosome was omitted as it was not clear if linkage statistics applied to the X chromosome are comparable to those on the autosomes. In addition a number of genome scans for BPD (section 6.2.1) did not include it in their linkage analyses. It therefore represented a complicating factor, which was unfortunate due to the interest in this chromosome in BPD (see chapter 1, section 1.5.2).

The GSMA method is also robust to missing information, and so can be used for studies which publish only lod scores exceeding a threshold, or where linkage scores of zero occur. In cases where data is being combined from a statistic defined in the positive range such as the Maximum Likelihood Score (MLS), and from a parametric lod or NPL score which can have negative scores, scores less than zero are reclassified as zero to ensure consistency (Wise et al, 1999). In these cases bins which have no information, or which have equal linkage scores, are assigned rank equal to the mean of the ranks they would have occupied.

The mathematical distribution described above is correct for situations where scans are not missing much information, and have similar, even marker coverage of the bins. If this assumption does not hold computer simulations are used to derive more accurate p-values. Similarly the distribution no longer holds when weighting factors are introduced; simulation methods incorporating the weights must be applied to obtain appropriate significance levels.

6.2 Preliminary GSMA

As part of a preliminary study I performed a GSMA on data available from published reports of genome scans (Segurado and Gill, 2001). The limited amount of information available restricted the study to the investigation of 8 scans (Coon et al, 1993; McInnes et al, 1996; NIMH, 1997; Detera-Wadleigh et al, 1999; Kelsoe et al, 2001; Radhakrishna et al, 2001; chapter 4: Bennett et al, 2002) using only eight regions from 121 across the genome. Data from more regions were available from some of these studies, and the decision to include only eight was somewhat arbitrary. The results highlighted six of these regions with summed ranks significantly larger than would be expected by random permutation of the ranks: chromosomes 22q ($p<0.005$), 13q31-34 ($p<0.005$), 7q353 ($p<0.006$), 11q21-22.2 ($p<0.009$), 2q37 ($p<0.009$) and 5q35 ($p<0.01$). Six bins from 121, with a nominal significance level less than 0.05 were no more than would be expected by chance. However they all showed bin-wise p values less than 0.01, which would be expected for only 1.2 bins on average.

The presentation of these results at the IXth World Congress on Psychiatric Genetics in St. Louis in 2001, led to a collaborative effort between the author, Prof. Doug Levinson of Pennsylvania University, Prof. Sevilla Detera-Wadleigh of the NIH, and Prof. Cathryn Lewis of Guy's King's and Thomas' School of Medicine, to perform a GSMA on full data

from all available studies. Therefore the preliminary analysis was superseded, and is referred to for comparative purposes only (section 6.4). The bins applied were altered to match those in use by Drs. Lewis and Levinson in an ongoing parallel GSMA for Schizophrenia (Lewis et al, submitted).

6.2 Methods

6.2.1 Scan Selection

Genome scans were identified by perusal of the literature on the online NCBI PubMed database, the scanning of conference presentations, and by personal communication with research groups. The samples sets identified were scrutinised to ensure that each was independent, and in cases of doubt, the group was queried. Each research group was contacted to notify them of the study, and request their participation. Each group was also asked to provide information on the diagnoses used in their sample, with the numbers of affected under each category, and the linkage statistics applied. In the case of parametric lod scores, the parameters used were also requested. Finally, a complete list of the genetic markers used and the lod score at each marker was requested. If they chose not to reveal the lod scores, they could alternatively provide a list of the markers, ranked by the linkage score obtained at each one.

The criterion of completed genome scans which use independent samples was rigorously applied and several studies were omitted due to sample overlap between them, or an inability to gauge the extent of independence of the samples. This was the case with studies using the Old Order Amish (OOA) pedigrees (Detera-Wadleigh et al, 1994; Gerhard et al, 1994; LaBuda et al, 1996; Polymeropoulos and Schaffer, 1996; Berrettini et al, 1997; Ginns et al, 1996), from which the study with the largest number of affected individuals was taken (Detera-Wadleigh et al, 1999). A study which examined the OOA pedigree for

linkage to mental health wellness (Ginns et al, 1998), and a modified genome scan (Inglehearn, 1997) which did not examine full-genome marker data (Murphy et al, 2000) were also excluded. A total of 22 eligible studies were identified, and data was obtained for these (table 6.1). Fifteen of these studies were published and seven were unpublished, or in press at the time of selection.

A number of genome scans (Blackwood et al, 1996; Adams et al, 1998; Radhakrishna et al, 2001; Ewald et al, 2002) were deemed to have too few affected individuals (N<20) to provide any useful linkage information for a locus common to many studies. In addition, two of these studies had sparse marker maps (Blackwood et al, 1996; Radhakrishna et al, 2001), further reducing their usefulness. However, these four studies were included in an exploratory analysis examining a broadest phenotype (see section 6.2.2).

The eighteen remaining scans were included in the principal and secondary analyses (section 6.2.2). The linkage statistics, and disease classification models applied are detailed in tables 6.1 and 6.2. The selected genome scans predominantly applied a parametric lod score analysis. Some scans assumed the presence of an autosomal dominant (or near-dominant) locus. Some performed the lod score analysis under 2-5 genetic models, in which case the lod score statistic was maximised over these. Non-parametric NPL or MLS statistics were also used. In situations where more than one linkage statistic was used, the groups were asked to indicate the principal method, to be used in the ranking.

Reference	Ethnicity	N. Markers	Empty Bins	Linkage Method			N. Aff
				Statistic	Models	Type	
Adams et al, 1998	Eur (Austr)	212	3	lod	(4)	2p	11
Blackwood et al, 1996	Br	206	25	lod	D	2p	11
Radhakrishna et al, 2001	Turkish	189	16	lod	D+R	2p	11
Ewald et al, 2002	Danish	605	0	lod/NPL‡	D+R	2p	17
McInnes et al, 1996	CR isol	473	1	lod	D	2p	24
Blackwood et al †[1]	Br	372	0	lod	D+R	2p	41
VanBroeckhoven et al 2 †	Swedish	361	0	NPL	NP	mp	47
Coon et al, 1993	N Am (Eur)	328	19	lod	D	2p	51
VanBroeckhoven et al 1 †	Belgian	372	0	lod	D+R	2p	56
Morissette et al, 1999	Queb isol	314	0	lod	D+R	2p	56
Schofield et al †[2]	Eur (Austr)	382	0	lod	(4)	2p	63
Badenhop et al, 2002	Eur (Austr)	298	0	lod	(4)	2p	69
Curtis et al †	Br+Icl	365	0	lod	max	2p/3p	74
Kelsoe et al, 2001	N Am	428	1	lod	(3)	2p	76
Turecki et al, 2001	Eur (Can)	363	1	lod	(5)	2p	106
Ekholm et al †[3]	Finnish	368	0	lod	D+R	2p	132
Detera-Wadleigh et al, 1999	N Am+OOA	584	0	MLS	NP	mp	160
Cichon et al, 2001(b)	Gr+Is+It	359	0	NPL	NP	mp	245
Liu et al (in press) †	N Am	333	0	lod	D+R	2p	297
Friddle et al, 2000	N Am	823	0	NPL	NP	mp	301
Bennett et al, 2002	Ir+Br	398	0	MLS	NP	mp	367
NIMH, 1997	N Am	355	3	NPL	NP	mp	424

Table 6.1. Description of genome scans included in meta-analysis in reverse order of sample size (number of affected individuals (N. Aff) under broadest disease classification); shaded studied were not included in principal analyses (see text).

†: unpublished data; for methods see: [1] - Blackwood et al, 1996; [2] - Badenhop et al, 2002; [3] - Ekholm et al, 2002;

‡: Study used different statistical methods for each disease model: multipoint non-parametric NPL statistic, and parametric lod (genetic model as indicated);

Ethnicity: Eur= Caucasian individuals of European descent; N Am= North American ethnicity (not otherwise specified); Gr, Is, It= German, Israeli, Italian ethnicity; Austr= Australian; Can= Canadian Ir= Irish; Br= British; Isl= Icelandic CR= Costa Rican; Quebl= Quebecan; isol= genetic isolate; OOA= Old Order Amish genetic isolate;

Empty Bins= number of bins containing no markers in each scan;

Linkage statistics: LOD= parametric lod score; NPL= Non-Parametric Linkage score (Z_{ALL}); MLS= Maximum Likelihood Score (non-parametric IBD linkage statistic subject to 'possible triangle' restrictions);

Genetic models: D= Dominant; R= Recessive; D+R= scores maximised over a dominant and a recessive genetic model; (*n*)= scores maximised over *n* genetic models; NP= non-parametric analysis;

Statistic type: 2p= two-point lod score; mp= multipoint lod score; 2p/3p= maximised two/three-point score.

| Study Site | Reference | Dx Crit | Disease Classification Models Applied | | |

6.2.2 Phenotypic Models

A decision with regard to the most appropriate diagnostic models to use in the meta-analysis had to be made, given the variety of disease classification models applied in the set of genome scans. Table 6.2 illustrates the diversity in the number, composition and stringency of the models applied by all 22 genome scans. The most narrow diagnosis of Bipolar I disorder (BPI) would be expected to yield most information in a genetic linkage study, as this diagnosis shows the largest relative risk in family studies (see chapter 1 section 1.3). On this basis a stringent disease model was adopted incorporating analyses examining linkage to BPI and Schizoaffective disorder - bipolar subtype (SAB). An intermediate disease model also including Bipolar II disorder (BPII) was also performed, as some groups favour a core BPD phenotype including this diagnosis (see chapter 1, section 1.3.4). Individuals diagnosed with Bipolar Disorder - not otherwise specified (BPN) under RDC from one study (Cichon et al, 2001b) were included in this model, after reassurance from the researchers that these patients met DSM criteria for BPII. These two disease models were considered primary analyses.

Secondary analyses included a Broad disease Model 3, which added cases of BPN and Recurrent Unipolar Disorder (RUP). This model also incorporated a very small number of patients with other diagnoses, namely Cyclothymia and Single Episode Depression (see section 6.3). A Very Broad Model 4 was also examined, which extended diagnoses to include significant numbers of patients diagnosed with Single Episode Depression and Cyclothymia. In order to present an unweighted analysis, which would perform best under a model of genetic locus heterogeneity, the four genome scans with samples smaller than 20 affected individuals were included in an expanded Very Broad analysis, Model 5. In a situation where loci may be present in only one or a few studies, the scans in smaller samples, including several of the studies in extended pedigrees, would have increased

power. Therefore a model including such studies would best allow identification of these loci. It must be noted that meta-analysis techniques are not best suited to identifying loci under a model of classic genetic locus heterogeneity, they are rather designed to identify loci which are common across many or all samples. The disease Models 1 and 2 included linkage data from nine and fourteen of the genome scans, respectively. The Models 3 and 4 included data from the same set of eighteen genome scans.

6.2.3 Power

A power study of the GSMA method was conducted following representations from some researchers in the field concerning the statistical power of the method. This study examined a simulated sample proportional in size to the BPD meta-analysis Model 1 and Model 2 samples, and a 20-study GSMA in schizophrenia (Lewis et al, submitted) which was also underway (Levinson et al, submitted). A pool of 100,000 simulated chromosomes was generated and sets of nine and fourteen "genome scans" were constructed in the form of affected sibling pair studies with identical or very close number of affected individuals to the original studies. Twenty chromosomes of 180 cM length (6 bins) with simulated genotypes at a 10 cM resolution (3 markers per bin), were sampled randomly from the pool to reconstruct individual studies. The chromosomes contained either no linked locus, or in the ASPs:

An "edge" linked locus: in the first (edge) bin of the chromosome (*ed*);

A "mid" linked locus: in the third (middle) bin of the chromosome (*md*);

The simulated linkages assumed dominant loci with effect size corresponding to a λ_s of 1.15, 1.2, 1.3 or 1.4, which reflect the range over which a sample of 500-1000 ASPs would have good to negligible power to detect linkage (Levinson et al, submitted). Each simulated genome scan modelled sets of ASPs with either zero, one or two parents genotyped, in equal proportions. Each set of simulated genome scans was then analysed for

linkage with Genehunter 2 (Kruglyak et al, 1996), and GSMA performed on the results (see section 6.2.4). A total of 10,000 GSMA replicates were performed with no linked locus to establish p-value thresholds (1000 replicates). Simulations containing linked loci was also performed under each of four locus models (100 replicates):

1 "edge" linked locus (*ed1*);

1 "mid" linked locus (*md1*);

5 loci: one "edge" and four "mid" (*ed1md4*);

10 linked loci: two "edge" and eight "mid" (*ed2md8*).

Power was also examined with weighting factors applied to each genome scan (see section 6.2.4), with datasets simulating λ_s = 1.15 and 1.3, and with no bin containing a susceptibility locus.

6.2.4 GSMA

One hundred and twenty bins were established by defining 98 anchor markers (detailed in Appendix F) spaced approximately every 30cM, across the 22 autosomes. Markers from each scan were binned using their relative position to the anchor markers primarily using the Marshfield Clinic genetic map. Other genetic maps from the Genome Database (GDB), or from the CEDAR Genetic Location Database, or the UCSC Golden Path Genome Browser were used if the marker could not be positioned on the Marshfield map. If the marker could still not be positioned, the map used by the individual researchers was used. A reference table of positioned markers was cumulatively constructed to aid in this process. The maximum score in each bin was established as described above and by Wise et al (1999), from the principal linkage statistic used by each study (table 6.1). Summed ranks for each of the disease classifications were calculated and tabulated in a Microsoft Excel spreadsheet, and ranks were summed across bins. In this instance a rank of 120 indicated the most evidence for linkage within each scan, and a rank of 1, the minimum.

Computer simulations were conducted in Microsoft Excel to derive a distribution for the summed rank under the null hypothesis of no underlying susceptibility gene. A series of Visual Basic macros (Appendix G) were written in order to expedite and automate this process. These macros generated replicates of randomly ranked data by randomly shuffling the ranks of the observed bins for each study, and re-calculating the summed ranks. This was repeated 5,000 times. The range and distribution of the summed ranks across replicates was calculated, and empirical significance values for the summed rank, and thresholds of significance at a 10%, 5% and 1% level were determined. Significance was determined in the form of an empirical bin-wise (BW) p-value from the distribution of simulations, representing the probability that any bin would achieve a summed rank higher than the observed value by chance. This was performed for each of the diagnostic models, under weighted and unweighted conditions

In order to provide a significance for the magnitude of each summed rank, the distribution of the highest, second-highest, etc. summed ranks were determined from simulations as above. These distributions of the 120^{th}, 119^{th}... n^{th} summed ranks were used to obtain p-values, representing a rank order (RO) probability that the n^{th} summed rank should reach at least its observed level by chance (Levinson et al, submitted). The p-values were calculated as (r+1)/(n+1), where r = the number of bins exceeding a particular summed rank in 5,000 GSMA replicates, and n is the number of replicates, i.e. 5,000 (North et al, 2002). Further Visual Basic macros were written to allow calculation of the range and distributions under each of the 120 ranks, from the simulated data derived for the BW simulations (see above). RO significance levels and thresholds were then calculated as before.

Weighting factors were applied to reflect the relative power of each scan to detect linkage. Several weights were explored using the number of pedigrees and of affected individuals. The latter should correlate strongly with the number of informative meioses, but has the advantage of being easily obtainable for each study. The square root of the number of affected individuals was used as the weighting factor, on the basis of a simulation study conducted by Douglas Levinson (Levinson et al, submitted). This study indicated that so weighted, the summed rank had an average Pearson correlation of 0.856 with the location of an underlying susceptibility gene in a simulation set of 100 GSMA replicates of 20 genome scans. The weighting factor was standardised to normalise its distribution about unity: the square root of the number of affected individuals was calculated for each scan, and then divided by the mean of its value across all scans. The factors thereby obtained were multiplied by the ranks of the genome scan before summation across scans. This was also incorporated into the simulation studies to determine significance levels. As the weighting scheme was shown to confer marginally higher power, it was considered the principal analysis. Therefore the results mentioned below will focus on the weighted analysis only, unless otherwise indicated.

6.2.5 Bin Positioning

In order to examine the effect of bin positioning, two exploratory analyses were conducted. In each of these a 60 cM bin width was used. In the first, the first 30 cM bin of each chromosome was combined with the one to its right; in the second with the one to its left. Thus two sets of 60 cM bins were constructed, 60R and 60L, for which every bin had 30 cM of overlap with each of two bins in the other set. For disease Models 1 and 2, the ranks of each bin were re-calculated as above, using the 60R and 60L bins. Bin-wise significance levels were calculated, and the two sets of results examined for discordance in the genomic regions highlighted.

6.3 Results

6.3.1 Power Study

The power study (Levinson et al, submitted) established a bin-wise (BW) p-value threshold for genome-wide level of significance as that expected in only 5% of GSMAs (p=0.000417), and "suggestive" linkage as that expected only once per GSMA (p=0.00833). These thresholds should be considered somewhat conservative as they utilise the Bonferroni correction, which assumes the tests (BW p-values) are independent of each other, which they are not, due to the non-independent nature of ranked bins. Interestingly, the power of the GSMA method to detect linkage at these thresholds was equal or greater than that of the maximum Genehunter 2 Non-Parametric Linkage (NPL) scores within each bin.

The study indicated that the Model 2 BPD sample (14 genome scans) would have good power to detect at least one susceptibility locus at a genome-wide level of significance. This was true under a one, five or ten-locus model, with a locus-specific λ_s of 1.3, or under a ten-locus model with a λ_s of 1.15. The Model 1 sample (9 scans) was shown to have poor to modest power to detect at least one locus under these conditions.

Power to detect linkage with nominally significant (5%) BW p-values was moderate to poor for the BPD sample simulations. For the Model 1 sample, power was less than 70% for all locus models and effect sizes. The Model 2 sample had power >70% for the 1-locus models, but this fell for the models with more loci when the locus effect size was less than a λ_s of 1.3. An effect typical of linked simulations, but occurring less than 5% of the time in unlinked replicates was also noted. This indicated that the presence of 10 or more bins with a BW p-value <0.05 was suggestive of linkage, without identifying particular bins.

The rank order (RO) p-values were shown not to be independent. This is due to the expected distribution of the n^{th} summed rank, which substantially overlaps the $(n-1)^{th}$ and $(n+1)^{th}$ summed rank distributions. Therefore, if the n^{th} summed rank is larger than would be expected by chance, the distribution of the $(n+1)^{th}$ summed rank (which is by definition larger than the n^{th} summed rank) is substantially truncated, and therefore it is also more likely to be larger than expected by chance. However, the presence of 4 or more bins with an RO $p<0.05$ in the top 10 ranks was shown to be indicative of linkage, without identifying any particular bin.

As the number of linked bins increased, the number of bins with both a BW and RO p-value <0.05 increased, even for very small locus effect sizes ($\lambda_s=1.15$). In unlinked simulations, more than 5 such bins were seen in less than 5% of the replicates. On the other hand, these bins both were more common and typically contained linked loci, in linked simulations. If only one linked locus was simulated, only the magnitude of the p-values could identify linkage. A genome-wide significance level would correspond to a combination of both BW and RO $p < 0.014$, for example.

6.3.2 Sample Size

The eighteen genome scans had an average of 406 markers over the 22 autosomes, with an average of 3.38 markers (range 0-22) in each bin. Four studies had 1-3 bins covered by no markers. One scan (Coon et al, 1993) had 19 empty bins. Bins on chromosomes 18, 5, 21 and 22 were covered by over 4.6 markers per study, which might be expected to increase the probability of attaining a larger lod score (Levinson et al, submitted).

Table 6.3 details the contribution of each scan to the two primary and two secondary disease models examined. Under the Narrow Model 1 disease classification nine genome scans were included contributing 948 affected individuals from 347 pedigrees. The Intermediate Model 2 included 14 genome scans encompassing data from 1733 affected individuals in 512 pedigrees. The Broad Model 3 analysis included a total of 2437 affected individuals in 593 pedigrees from 18 genome scans. This model included a small number of patients with Cyclothymia (N=7, VanBroeckhoven et al (1); Ekholm et al, unpublished data) and Single Episode Depression (N=2, Ekholm et al, unpublished data). Model 4 used the same sample as Model 3, but with a significant number of individuals under broader diagnoses from two studies (Cichon et al, 2001b; Curtis et al, in press), bringing the total sample to 2589 affected individuals in 617 pedigrees.

	Number of Affected				Scan Weights			
Model:	1	2	3	4	1	2	3	4
Studies in Model 1:								
Bennett et al, 2002	288	325	367	367	1.83	1.77	1.80	1.74
NIMH, 1997	264	336	424	424	1.75	1.80	1.93	1.87
Liu et al (in press) †	115	208	297	297	1.15	1.42	1.62	1.56
Ekholm et al †	107	107	132	132	1.11	1.02	1.08	1.04
Morissette et al, 1999	42	42	56	56	0.70	0.64	0.70	0.68
Schofield et al †	41	46	63	63	0.69	0.67	0.75	0.72
Badenhop et al, 2002	40	44	69	69	0.68	0.65	0.78	0.75
Blackwood et al †	27	36	41	41	0.56	0.59	0.60	0.58
McInnes et al, 1996	24	24	24	24	0.53	0.48	0.46	0.44
Studies added under Model 2:								
Friddle et al, 2000		232	301	301		1.50	1.63	1.58
Cichon et al, 2001b		128	128	245		1.11	1.06	1.42
Detera-Wadleigh et al, 1999		118	160	160		1.07	1.19	1.15
Kelsoe et al, 2001		48	76	76		0.68	0.82	0.79
Curtis et al †		39	39	74		0.61	0.59	0.78
Studies added under Model 3:								
VanBroeckhoven et al (2) †			47	47			0.64	0.62
Coon et al, 1993			51	51			0.67	0.65
VanBroeckhoven et al (1) †			56	56			0.70	0.68
Turecki et al, 2001			106	106			0.97	0.93
Total:	948	1733	2437	2589	9	14	18	18

Table 6.3. Number of affected individuals contributed by and weighting applied to the eighteen genome scan studies participating in the meta-analysis, under Models 1-4; † unpublished data (see table 6.1).

6.3.3 Bin-Wise significance

Figures 6.1-6.4 show scatter-plots of the summed ranks for each bin under Models 1-4 respectively. Weighted summed ranks are represented by solid symbols, unweighted by open symbols. The 95% and 99% significance thresholds are shown. Appendix H contains the full, tabulated contributions of each genome scan to the four weighted analyses. The Model 5 analysis did not reveal any different results, nor greater statistical significance, and is therefore not discussed further.

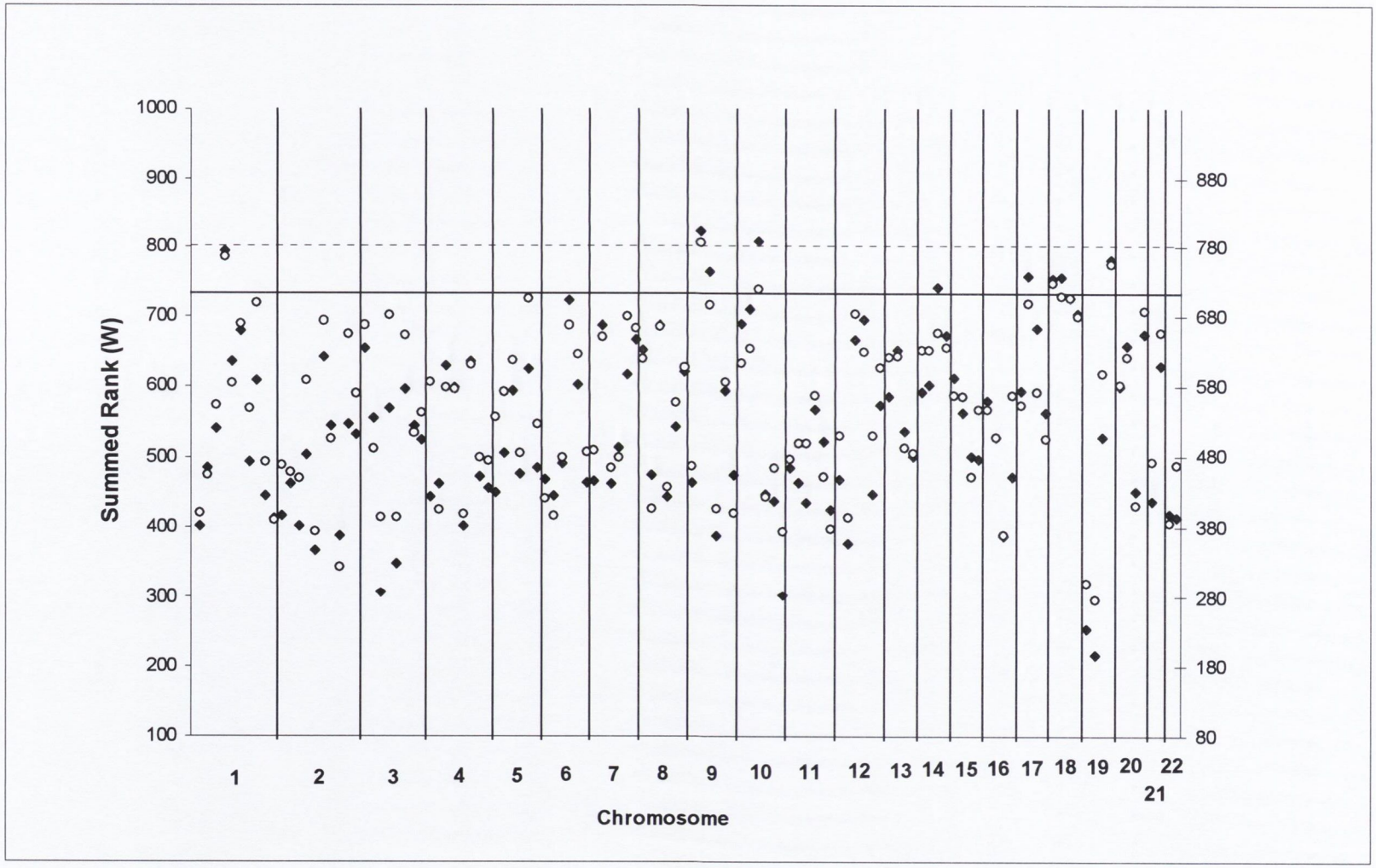


Figure 6.1. Weighted and Unweighted Summed ranks for disease model 1 (9 scans)

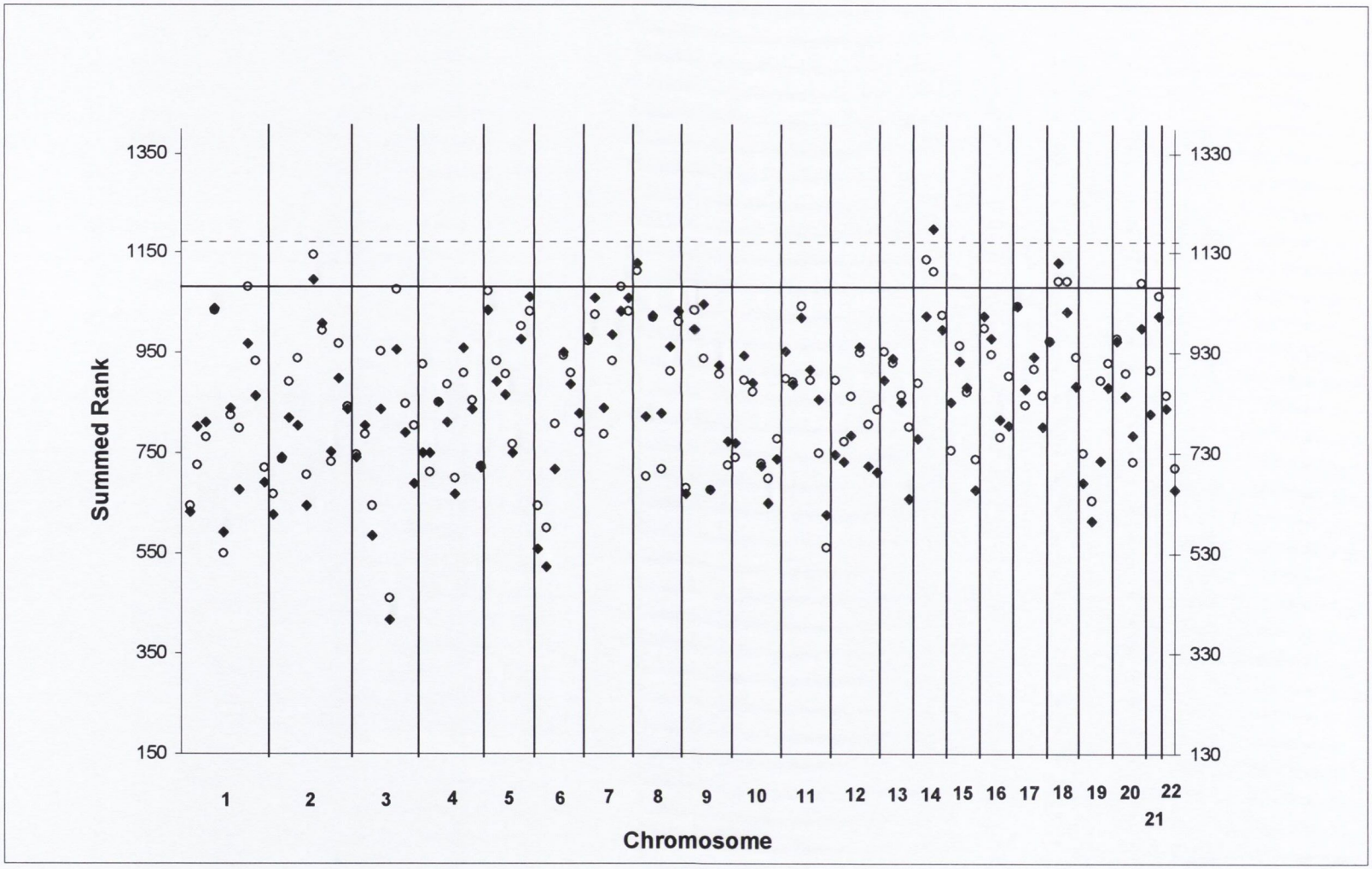


Figure 6.2. Weighted and Unweighted Summed ranks for disease model 2 (14 scans)

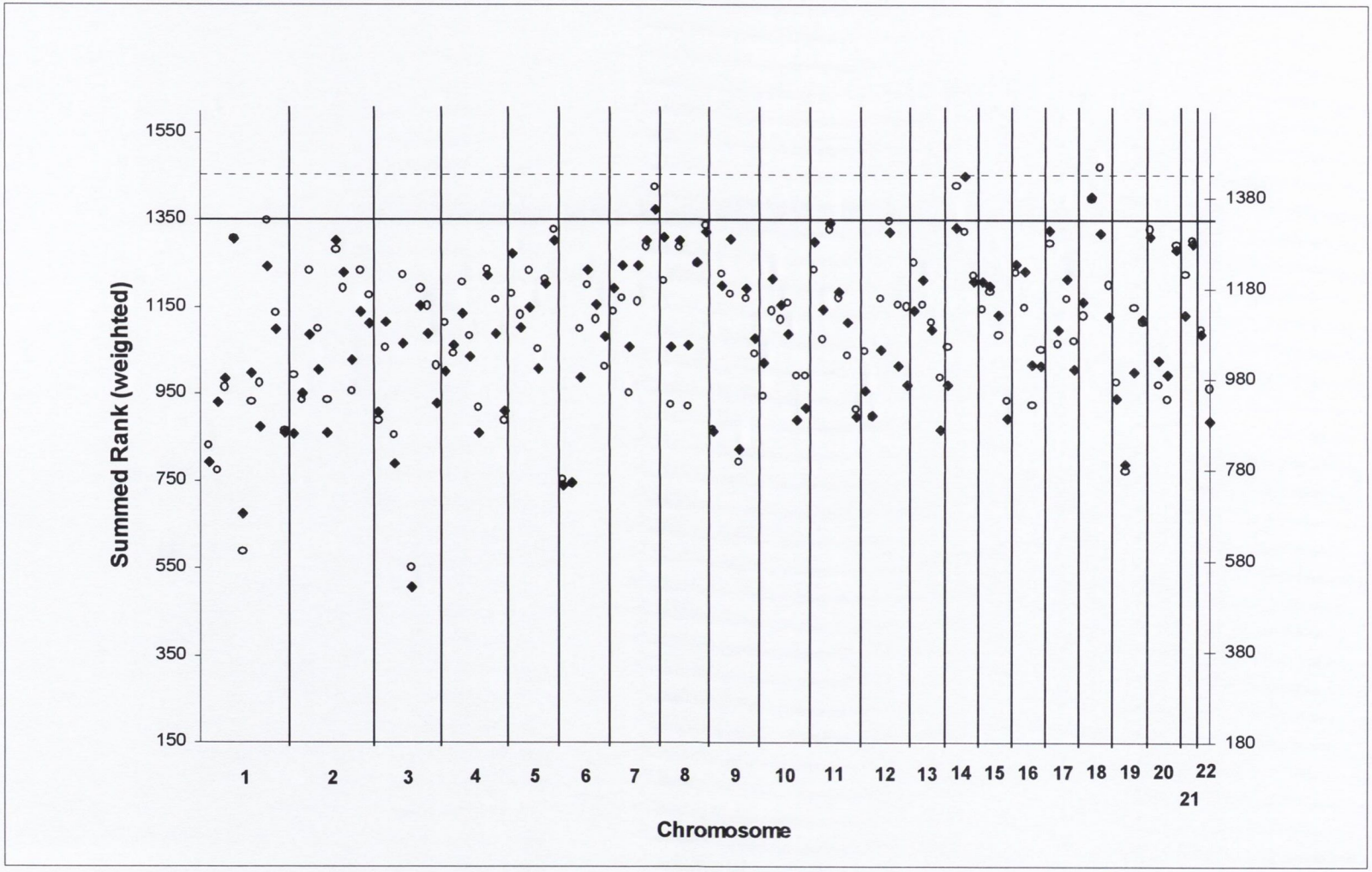


Figure 6.3. Weighted and Unweighted Summed ranks for disease model 3 (18 scans)

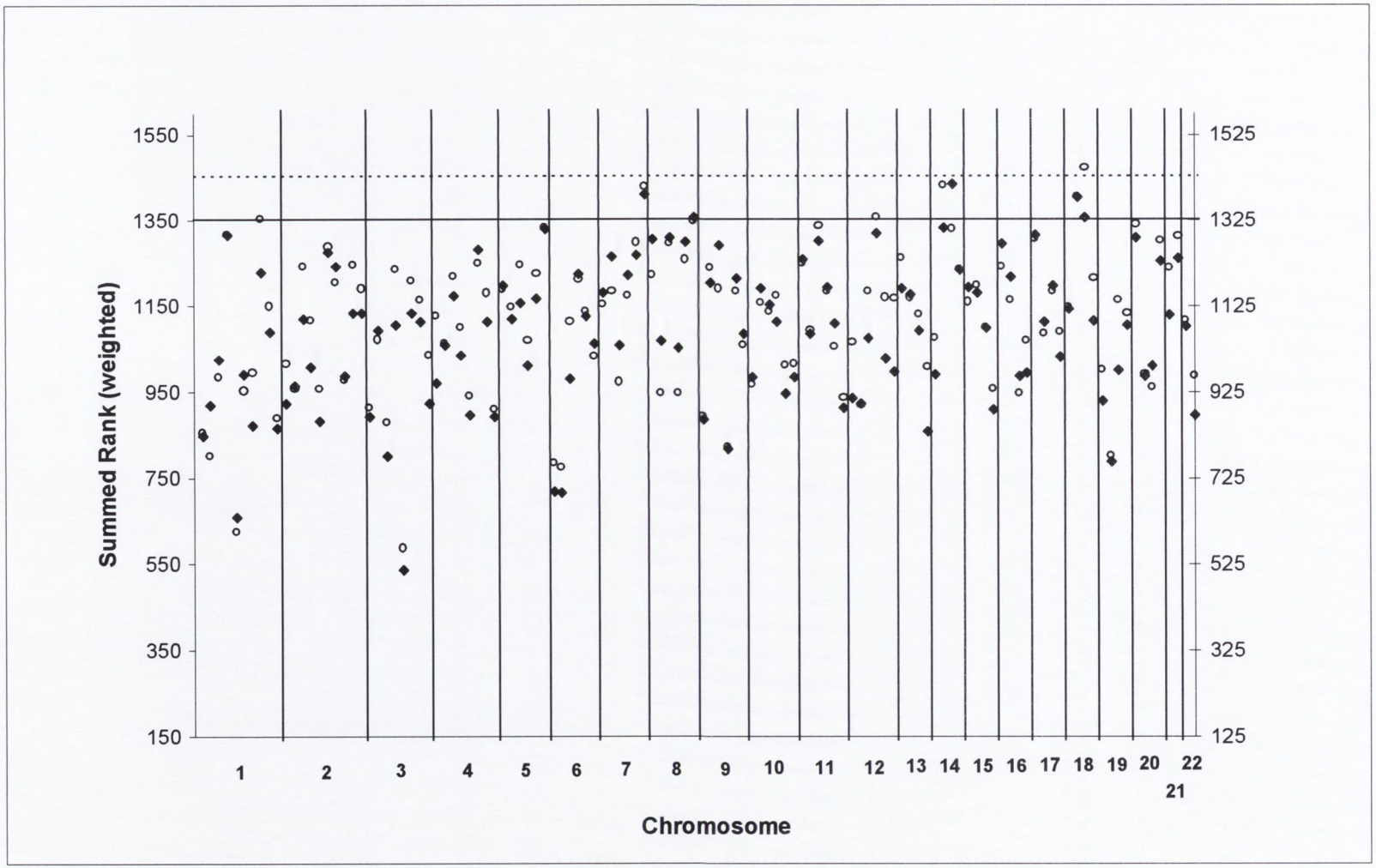


Figure 6.4. Weighted and Unweighted Summed ranks for disease model 4 (18 scans)

The bins which achieved a bin-wise (BW) significance level of 0.1 or better under the two primary models are displayed in table 6.4. No bin achieved a BW p-value considered to be significant at a genome-wide level (section 6.3.1).

Weighted Analysis

Model 1			Model 2		
Bin	Bin-Wise p	Rank Order p	Bin	Bin-Wise p	Rank Order p
9.2	**0.006**	0.497	14.3	**0.003**	0.309
10.3	**0.008**	0.267	18.2	**0.013**	0.478
1.4	**0.013**	0.191	2.6	**0.036**	0.841
19.4	**0.017**	0.144	8.1	**0.044**	0.809
9.3	**0.026**	0.195	8.6	**0.045**	0.672
17.2	**0.031**	0.143	5.6	0.050	0.588
18.2	**0.031**	0.057	18.3	0.061	0.631
18.1	**0.033**	**0.026**	5.1	0.065	0.547
14.3	**0.043**	**0.049**	21.2	0.075	0.582
18.3	0.055	0.091	7.6	0.079	0.483
6.4	0.060	0.070	9.3	0.080	0.342
10.2	0.075	0.153			
18.4	0.082	0.138			
12.4	0.094	0.200			

Unweighted Analysis

Model 1			Model 2		
Bin	Bin-Wise p	Rank Order p	Bin	Bin-Wise p	Rank Order p
9.2	**0.009**	0.685	2.6	**0.014**	0.829
1.4	**0.016**	0.583	14.2	**0.016**	0.605
19.4	**0.021**	0.465	14.3	**0.017**	0.334
18.1	**0.038**	0.698	18.3	**0.023**	0.294
10.3	**0.046**	0.685	18.2	**0.027**	0.209
18.2	0.057	0.719	5.1	**0.038**	0.293
18.3	0.059	0.595	1.8	**0.041**	0.203
5.5	0.062	0.468	21.2	**0.042**	0.107
17.2	0.070	0.469	3.6	**0.047**	0.088
1.8	0.070	0.325	8.1	**0.049**	**0.047**
9.3	0.073	0.236	7.5	0.059	0.065
20.4	0.083	0.269	8.6	0.069	0.081
12.3	0.091	0.259	20.4	0.074	0.067
3.4	0.092	0.170	5.6	0.080	0.060
7.5	0.095	0.117	11.3	0.085	0.051
18.4	0.097	0.070	9.2	0.087	0.029
			2.9	0.091	0.021

Table 6.4. Top ranked bins under the two primary meta-analysis models; all bins which achieved a BW p<0.1 are shown; bins are represented as Chromosome.Bin, e.g. the second bin on chromosome 9, is 9.2.

One bin under Model 1 (9.2, 9p22.3-p21.1), and one bin under Model 2 (14.3, 14q24.1-q32.12) attained a significance level which could be considered "suggestive". However, this threshold corresponds to that expected by chance at least once per GSMA. In addition no analysis, under any diagnostic scheme, produced ten bins or more at a nominal significance level, which would be suggestive of the presence of linked bins. Eight bins under Model 1, and nine under Model 2 (unweighted) attained a BW significance level of 0.05.

6.3.4 Rank Order Significance

The relationship between BW and RO p-values was expected to be problematic (see section 6.3.1). Bins with a nominally significant RO p-value showed substantial clustering, and did not correlate well to the summed rank. Under no analysis were four or more RO p-values <0.05 observed in the top 10 ranked bins.

The presence of both BW and RO nominally significant p-values was expected to have good power to highlight linked bins. However, no combination of BW and RO p-values was observed which correspond to a genome-wide significant result (section 6.3.1). The amount of overlap of the statistics at a nominal significance level was less than would be expected (5 bins or more) in the presence of several linked bins. The Model 1 weighted analysis came closest, with two such bins: 18.1 (18pter-p11) and 14.3 (14q24.1-q32.12). These bins, and bins adjacent to them also featured well across all phenotype models.

Table 6.5 shows bins which achieved nominal BW significance under the two principal disease models and Model 3. Both weighted and unweighted results are shown for comparison of the two schemes. Bins 14.3 and 18.2 (18p11-q12.3) were the only bins to achieve BW $p < 0.05$ under all three models. Both these bins also achieved BW $p < 0.05$

under the unweighted Models 2 and 3. Bin 14.3 and bin 18.1 were the only bins to achieve both BW and RO $p < 0.05$ in the weighted analysis (Model 1), and bin 14.3 surpassed the threshold for "suggestive" linkage (BW $p = 0.003$). Adjacent bins were also highly ranked on chromosomes 14 and 18.

Range (Marsh cM)	Cytogenetic position	Bin	Weighted Model			Unweighted Model		
			1	2	3	1	2	3
83.07-113.69	1p32.2-p31.1	1.4						
201.58-231.11	1q31.1-q32	1.8						
128.41-154.48	2q22.1-q23.3	2.6						
146.6-173.34	3q22.1-q25.31	3.6						
0-31.78	5pter-p15.1	5.1						
148.11-181.97	7q34-qter	7.6						
0-27.4	8pter-p22	8.1					⊙	
137.92-167.9	8q24.21-qter	8.6						
27.32-53.6	9p22.3-p21.1	9.2						
53.6-84.9	9p21.1-q21.32	9.3						
62.23-91.13	10q11.21-q22.1	10.3						
47.06-72.82	11p12-q13.3	11.3						
40.11-74.96	14q13.1-q24.1	14.2						
74.96-105	14q24.1-q32.12	14.3	⊙					
25.14-63.62	17p12-q21.33	17.2						
0-24.08	18pter-p11	18.1	⊙					
24.08-62.84	18p11-q12.3	18.2						
62.84-96.48	18q12.3-q22.1	18.3						
75.41-105.02	19q13.33-qter	19.4						
0-21.15	20pter-p12.3	20.1						
25.26-57.77	21q21.3-qter	21.2						

Table 6.5. Meta-analysis bins which achieved BW $p < 0.05$ under diagnostic Models 1-3; also shown are bins with BW $p < 0.1$ to demonstrate whether bins achieved high ranks under other models; Red/Deep Blue: $p < 0.01$, Orange/Medium Blue: $p < 0.05$, Yellow/Light Blue: $p < 0.1$; ⊙: indicates bin also achieved RO $p < 0.05$; Range is the start and end position of the bin in centimorgans according to the Marshfield map; cytogenetic positions obtained from UCSC Genome Browser, April 2002 Human Genome Freeze.

The second bin on chromosome 14 (14q13.1-q24.1) was nominally significant under unweighted Models 2 and 3. Most of chromosome 18 was highlighted. The first bin (18pter-p11) was nominally significant under Model 1 (weighted and unweighted), and the third bin (18q12.3-q22.1) under Model 3 and unweighted Models 2 and 3. Bin 4 (18q22.1-qter) showed a BW $p < 0.1$ under Model 1 (weighted and unweighted), but was not nominally significant under any disease model.

Adjacent bins were also both nominally significant on chromosome 9 (bins 2 and 3, 9p22.3-q21.32) under Model 1. Bin 2 attained genome-wide "suggestive" level ($p = 0.006$) under Model 1, and the third bin showed nominal significance under Model 1 and high ranks under Models 2 and 3 ($p < 0.1$). The only other bin to achieve both BW and RO $p < 0.05$ was the first bin on chromosome 8 (8pter-p22) under unweighted Model 2. The last bin (bin 6, 8q24.21-qter) on this chromosome was also the only other region to show nominal significance under more than one model (Models 2 and 3).

6.3.5 Bin Position

The results of the Model 1 and Model 2 analyses with 60 cM bins, eliminated many of the bins significant under the 30 cM analysis. However, no bins emerged as newly significant with the exception, under Model 2, of bins 5 and 6 on chromosome 7 (7q31.1-qter), and bins 2 and 3 on chromosome 9 (9p22.3-q21.32) under both weighted and unweighted analyses. The chromosome 7 region became nominally significant with the 60R bins, and high-ranked but not significant with the 60L bins (section 6.2.5). The opposite occurred with the chromosome 9 region, which became significant under the 60L scheme. This raises the possibility that these regions were overlooked under Model 2 in the main 30 cM analysis, due to the bin positioning. However, bins from both regions are ranked high under the 30 cM-bin Model 2 (weighted and unweighted), with BW $p < 0.1$.

Although crude, these results indicate that bin positioning has an effect on the rank which can alter statistical significance of the results. In addition, it was observed in the 60 cM analyses that significant bins under one bin scheme were often not significant under the other scheme, although usually highly ranked. In one extreme case, bins 8 and 9 on chromosome 1 under unweighted model 2, significance was observed in one scheme, but the overlapping bins under the other scheme were not present in the top 20 ranks. The large bin size probably exacerbated this effect.

6.4 Discussion

In order to objectively examine the extent of overlap of evidence for linkage across studies, meta-analysis of genome scans has been conducted on all studies known to date, using the rank-based GSMA method. Rigorous criteria for sample size and overlap were applied (section 6.2.1), which should ensure an objective measure of evidence for linkage. In addition, the nature of the method freed the analysis from confronting the problem of different marker maps and linkage statistics (section 6.1.2).

The results are in contrast to the preliminary study (section 6.2) applied to a limited portion of published data (Segurado and Gill, 2001). The results from the preliminary study of eight genome scans overlapped those highlighted by another meta-analysis of published data from eleven genome scans (Badner and Gershon, 2002b). Both studies reported their most significant results for regions of chromosomes 13q and 22q. The genome scans used in each overlapped considerably, the only difference being the use of re-analysed data (Garner et al, 2001) from one of the scans (McInnes et al, 1996). Also, as both were based on published data, they were both limited to examination of the same genomic regions.

Compared to the datasets included in the principal GSMA study, the preliminary GSMA included one study which was eventually dropped from the present study due to the small sample size (Radhakrishna et al, 2001). The MSP study also included that genome scan, and another also ultimately considered too small a sample for our investigation (Blackwood et al, 1996). The MSP study included two studies with overlapping sample sets (Ginns et al, 1996; Detera-Wadleigh et al, 1999), of which only the one in the larger sample was included herein, and one study for which we obtained more recent data from an extended sample (Friddle et al, 2000). Therefore the samples only partially overlap, which probably accounts for the differences in the results.

No genome-wide significance was observed for any bin in this study, leaving the possibility that the top ranked bins are merely the product of random variation. Several sources of variation may account for this result. Perhaps most importantly, the size of the sample was not as large as might be hoped. The BPD sample covered a maximum of 14 genome scans under the primary disease models, encompassing 1733 cases from 948 pedigrees. In contrast the Schizophrenia analysis conducted in parallel (Lewis et al, submitted) examined 20 studies, with almost 3000 cases from over 1200 pedigrees, and identified a number of regions significant at a genome-wide level, which coincided well with regions previously considered the most likely candidate loci. The power studies indicated that the 14-scan BPD sample should have good power to detect at least one locus at a genome-wide significance level (section 6.3.1), while power would be more modest for the 9-scan sample. However, under Model 1 the most bins at a nominal significance level, and the most bins with both BW and RO p-values < 0.05, were observed. This can be interpreted as making a case for the collection of larger samples of more strictly defined BPD, which should *a priori* have the most power to detect linkage. The reanalysis of many of the present genome scans using a strict BPI, or BPI and SAB, disease model should be

an easy first step in this direction, which could almost double the number of cases available for the meta-analysis.

Other sources of variation, which cannot be controlled, include the size and nature of the underlying genetic effect. This method, as any meta-analysis is expected to perform best in the presence of susceptibility genes common across most or all populations. Any amount of genetic locus heterogeneity would weaken the power, although the power studies did not examine this possibility. It may be for this reason that several regions where the most significant linkage scores have been found, in extended pedigrees or population isolates (e.g. Blackwood et al, 1996; McInnes et al, 1996; Adams et al, 1998), were not highlighted here. Any number of methodological factors may also have contributed to weakening the evidence for linkage, both specific to the individual studies (e.g. genotyping error, marker coverage, choice of linkage statistic) or related to the methodology adopted in the meta-analysis (e.g. data-reduction decisions, bin positioning). One criticism which has been made of the GSMA method is that, as it uses ranking of evidence for linkage, it does not take into account the magnitude of the lod score. This means that a bin which achieves a significant lod score of 6 in one study, and another bin which achieves a lod score of 1.5 in another study, could potentially end up with the same rank, and ascribed the same amount of evidence for linkage. This effect should over many studies, average out under the weighted analysis, as the studies with largest samples should have the largest expected linkage scores. However, this effect is central to the method's poor performance under a classical genetic heterogeneity model.

Given the lack of statistical significance, the most interesting results are the trends in certain regions of the genome. In particular the observation of bins nominally significant across disease models, such as on chromosomes 8, 14 and 18. Also of interest may be the

presence of nominal significance in adjacent bins on the same chromosome, such as on chromosomes 9, 14 and 18. Of these, only chromosome 18 has been of much previous interest to researchers in the field. The candidate region on chromosome 18p11.2 spans the boundary of bins 1 and 2 (Berrettini et al, 1994), and bin 18.2 extends onto the q-arm to cover the 18q12 candidate region first reported by Stine et al (1995). The studies which contributed most to the results on this chromosome, included two larger scans (Friddle et al, 2000; Bennett et al, 2002) at bins 2 and 3 under the broader models, and some of the smaller studies (McInnes et al, 1996; Blackwood et al, unpublished data; Schofield et al, unpublished data) under disease Model 1. The other candidate locus on this chromosome, at 18q22 (McInnes et al, 1996), lies within bin 4 and did not achieve nominal significance in the meta-analysis. A possible confounding factor is the coverage of bins 2, 3 and 4 on chromosome 18 by more markers than any other bins. High marker density can increase the summed rank under some conditions, and for the Model 3 sample a significant correlation was observed between the average number of markers per bin and the summed rank (C. Lewis and D. Levinson, personal communication). However, other regions with high marker density, on chromosomes 5, 21 and 22 did not achieve high summed ranks, and it is impossible to determine if increased marker coverage is the cause of the increased evidence for linkage, or *vice versa*.

Chromosome 14 was the other region where cross-diagnostic support, and evidence from adjacent bins were observed. Several of the larger studies contributed to the high ranks in bins 2 and 3 (Stine et al, 1997; Bennett et al, 2002; Liu et al, in press). It was also previously noted as a candidate region by two of the participant studies: Detera-Wadleigh et al, 1999, who observed a lod score of 3.19 on the boundary of bins 3 and 4 in a single large pedigree, and Cichon et al (2001b). It was also identified by Murphy et al (2000) in a

"modified genome scan" limited to gene-rich chromosomal regions. It has, however, received much less attention than other candidate regions in BPD.

The sixth bin on chromosome 8 was the only other bin where nominal significance was observed across two disease models (Models 2 and 3), contributed to mainly from Friddle et al (2000), Cichon et al (2001b), Kelsoe et al (2001), and Ekholm et al (unpublished data). However the experiment investigating the bin positions (sections 6.2.5 and 6.3.5) identified the 60 cM bin, 7.5+7.6, with nominally significant p-values under Model 2, in addition to bin 7.6 under Model 3. The combined bins 9.2+9.3 also demonstrated nominal significance under Model 2, in addition to the nominal significance under Model 1 on bins 9.2 and 9.3. In the second bin of chromosome 9, two markers (D9S171 and D9S161) have shown modest lod scores in three of the participating genome scans, including the Wellcome Trust UK/Irish study (Morissette et al, 1999; Cichon et al, 2001b; Bennett et al, 2002, chapter 4).

Another region which may be of interest is on chromosome 21 in bin 2 (21q21.3-qter), which lies over a reported linkage to BPD (Straub et al, 1994). On chromosome 10 the third bin (10q11.21-q22.1) attained the second highest weighted rank under model 1. This was mainly due to the effects of the largest genome scans, with the smaller ones showing no evidence for linkage to this region (McInnes et al. 1996; Douglas Blackwood, personal communication). However, this area does not coincide with the putative schizophrenia susceptibility locus on chromosome 10q for which there is some evidence for linkage to BPD (Cichon et al. 2001a; Kelsoe et al. 2001).

Many regions considered "hot-spots" by researchers in the field found no support in the meta-analysis, including on chromosomes 4p16 (Blackwood et al, 1996), 4q35 (Adams et

al, 1998), 13q (Detera-Wadleigh et al, 1999; Kelsoe et al, 2001; Badner and Gershon, 2002b), and 22q12 (Kelsoe et al, 2001). It is important to note that these results do not exclude linkage to these regions. There was also little evidence to support any genetic pleiotropy between BPD and schizophrenia. No bins with genome-wide significance in the schizophrenia GSMA (Lewis et al, submitted) achieved nominal significance in the BPD analysis.

In summary, the meta-analysis of genome scans for BPD produced no results with genome-wide significance. This may be due to a lack of power, particularly in the most stringent disease classification model, or other methodological factors. Regions which attained nominal significance include a candidate chromosome of great interest, 18, and other regions to which there has been much less attention in the past. Nevertheless, from the present analysis regions on chromosomes 14, 18, 8 and 9 appear to be the most likely locations of susceptibility loci common to many studies. It would be of use to the field to re-apply the meta-analysis, as more data becomes available, particularly under a narrow diagnostic model. The computer files and macros established for the analysis make the addition of new genome scans as they become available and re-analysis of the data a relatively easy process. The meta-analysis thus becomes a cumulative exploratory tool, showing the current amount of evidence for linkage to all regions of the genome. In addition, it has been suggested that the full ranking data of each scan be placed in an open on-line database, easily accessible by any researcher who wishes to explore the data further.

Chapter 7. Discussion

7.1 Summary of Results

The studies in this thesis cover a wide range of molecular genetic and statistical methodologies as applied to the search for susceptibility genes for Bipolar Affective Disorder (BPD) and related conditions. This chapter seeks to discuss and integrate the findings as a package. It will not discuss aspects of the individual studies covered in the discussion sections of those chapters. The studies, in particular the genome scan (chapter 4) and the meta-analysis (chapter 6) have contributed to the overall advancement of the field, albeit without producing strongly positive or 'headline' results.

7.1.1 Association Study

The first work undertaken toward the present thesis involved a candidate gene study to examine the tyrosine hydroxylase (TH) gene for evidence of association with BPD (chapter 3). This study was a preliminary study principally for the purposes of evaluation and preparation of the Irish sample and of some of the techniques to be applied in the genome scan study. However, in addition, tyrosine hydroxylase was one of the candidate genes that had been most investigated in this disorder, and hence represented an ideal starting point for examination in the present sample. The results showed no evidence for association in our sample, which is in keeping with the results of recent meta-analyses (Furlong et al, 1999a; Ho et al, 2000), and it seems likely that this gene does not confer a significant increase in vulnerability to BPD. However, several factors may be responsible for these results besides a true lack of involvement in this gene, as discussed below (section 7.2.1).

7.1.2 First Stage Genome Scan

This study was, and still remains one of the largest genome-wide linkage studies in BPD. Seventeen regions were identified that exceeded the set threshold of a point-wise p-value of 0.05 (MLS of 0.74) under a narrowly defined disease model, in a collaborative genome wide scan of a set of sibling pairs affected with Bipolar disorder, and related disorders (chapter 4). The sample size included 150 affected sibling pairs under the narrow bipolar I disorder (BPI) disorder disease classification model, and 255 under a broader disease model which included the range of affective spectrum diagnoses to recurrent unipolar disorder. Interestingly, the regions identified included chromosome 18, one of the most prominent candidate regions for a BPD susceptibility locus on the basis of previous linkage studies. Several of the regions overlapped previous findings of linkage to BPD, and indeed schizophrenia, including chromosomes 3, 7, 9, 10, 12, and the X chromosome (chapter 4, section 4.4.2; Bennett et al, 2002).

7.1.3 Second Stage Genome Scan

The second stage genome scan involved expansion of the sample to approximately twice the size, and prioritisation of the regions to be examined for linkage with a finer genetic map. Preliminary analyses on the second stage genome scan included extra affected individuals, examined for linkage with the first stage genetic markers (chapter 5). The follow-up regions on chromosomes 2 and 10 revealed modest increases in the MLS, suggesting the evidence for linkage to these regions persisted in the larger sample.

7.1.4 Meta-Analysis of Genome Scans

In facing up to the likely presence of many loci, each of very small effect, it is necessary to try to extract the maximum information from existing data. For this reason, a meta-analysis study using the GSMA (Genome Scan Meta-Analysis) method was conducted on a

combined data-set, including all known genome scans completed to date. The sample, which included almost 950 affected individuals under the most narrow (Model 1) disease classification model and almost 2600 under the broadest (Model 4), represented the first large scale multi-centre collaboration of its kind in BPD. Although the meta-analysis did not find any evidence for linkage at a genome-wide significance level, some candidate regions were highlighted due to their high rank under various disease models, and across disease classification schemes. These included 30 cM 'bins' (see chapter 6, section 6.1.3) on chromosome 18 which, as mentioned, is one of the most interesting candidate chromosomes for a BPD vulnerability locus. Bins on chromosome 14, 8 and 9 were also highlighted, which have received much less attention in the past, due to weak lod scores.

7.2 Study Limitations

7.2.1 Association Studies

Limitations of the TH Study

The lack of significant association between BPD and the tyrosine hydroxylase gene (chapter 3) may be due to a true lack of involvement of this gene in the disorder. However, a number of other factors must be considered in the interpretation of the result. The small effective sample size, and relatively low polymorphism of the marker, limited the parent-child transmissions used by the transmission/ disequilibrium test (TDT) statistic to 27. Other factors possibly affecting power were low levels of linkage disequilibrium (LD) between the marker and a putative disease locus due, for example, to distance from the disease locus, or allelic heterogeneity (Terwilliger and Weiss, 1998). The power of association studies (McGinnis et al, 2002) is also affected by other factors such as small relative risks, and the disease allele frequencies.

Are Association Study Samples Large Enough?

Power of association studies is usually examined in terms of the Genotypic Relative Risk (GRR), the increased risk to an individual with a certain genotype. The number of ASPs required for a TDT statistic to detect an association with a GRR of 4 is generally feasible, although dependent on the factors outlined above (Knapp, 1999). For a GRR as low as 2 the number of ASPs rises into the hundreds, and for a GRR of 1.5 can be from many hundred to several thousand. Therefore, the present study (chapter 3) could not exclude even modest effect sizes attributable to tyrosine hydroxylase. This may be the case for many association studies with bipolar disorder, including meta-analyses with hundreds of affected individuals (Bellivier et al, 1998) if, as is becoming apparent, individual susceptibility genes confer very small increases in risk (see section 7.2.2).

Conclusions

It has been suggested that large association studies are the best way to proceed in the identification of susceptibility genes for BPD, as association has more statistical power than linkage to detect genes of small effect (Risch and Merikangas, 1996). As reviewed by Jones and Craddock (2001), over 50 genes have been examined for association with BPD, with very little consistency of results. It may be the case that these studies are consistently examining genes irrelevant to the underlying pathology of BPD. However, as mentioned above, some of these genes may be true vulnerability loci, but of minor effect, undetectable with current sample sizes. The relative success of some collaborative studies with much larger sample populations (Furlong et al, 1999b; Craddock et al, 2001) in detecting small gene effects for the MAOA and COMT genes (Odds Ratios of 1.55 and 1.18, respectively) suggests that the latter case may be true.

7.2.2 Genetic Linkage

Interpretation of Scan I Results

The results of the genome scan must be interpreted in the context of the two-stage strategy, in which regions are not expected to attain genome-wide statistical significance in the first stage. Regions most likely to reflect underlying susceptibility loci are pursued in a larger sample, with higher marker density, and stricter statistical criteria with the expectation that some will turn out to be spurious. The expected power of the first stage sample was reasonable to detect a locus conferring a λ_s of 3 or more, or of 2 or more under the broadest disease model (see chapter 4, section 4.3.1 and table 4.3). However, the expected average MLS in the presence of loci of these effect sizes was not attained under any disease model, suggesting the underlying loci may be of smaller effect.

Are Linkage Study Samples Large Enough?

Efforts to increase statistical power to detect genes of small effect have characterised genetic studies to bipolar disorder in the recent past. Traditional parametric linkage analysis has increasingly been supplemented or replaced by non-parametric methods focusing on more easily ascertained affected relative pairs, which allows larger sample sizes. Power to detect linkage to a locus conferring susceptibility to a complex trait is a serious consideration, which goes beyond a simple case of sample size. Risch (1990b) pointed out that affected sibling pairs (ASPs) are not the most powerful relative pair on which to base a study, but are the easiest to ascertain. For a trait with a relative risk, λ_R, of 3, one hundred ASPs are marginally sufficient to achieve 80% power. This assumes that there is complete marker polymorphism, and zero recombination (θ=0) between the risk locus and the marker. Full marker polymorphism is rare in practice. For the highly polymorphic microsatellite loci used in the present genome scan, an average polymorphism information content of 0.75 was achieved, which would lead to an estimated decrease of

25% in the 'expected' MLS, which represents the available linkage information (Risch, 1990c). Alternatively, if recombination between the risk locus and marker rises to 0.05, the power of 100 ASPs is reduced to 60% for a λ_R of 4. Other factors such as errors in the marker order or size, genotyping errors, the presence of phenocopies (Bishop and Williamson, 1990) or other misspecifications of affection status have not been taken into account in these studies. All these factors may contribute to a substantial decrease in power. If individual risk loci in BPD confer an increase in risk of 2 or less, then power of the largest linkage studies could still be insufficient to reproducibly identify these loci.

Is the Meta-Analysis Sample Large Enough?

The meta-analysis (chapter 6) would be expected to have increased power to detect loci of smaller effects than would otherwise be the case in individual studies. Indeed the power estimates of the GSMA in the BPD samples (chapter 6, section 6.3.1) indicated that the model 2 analysis of 14 genome scans with 1733 affected individuals, should be able to detect at least one locus of effect size 1.3, if five or more such loci exist. For ten or more loci with smaller effect sizes (λ_s of 1.15) it should again be expected to detect at least one of these. It was striking that none of the results of the meta-analysis detected enough bins at nominal significance levels to suggest the presence of underlying linkage. A range of confounding factors may have affected the results of the meta-analysis study. Several of the genome scan studies examined relatively small samples, especially compared to the average sample size of the schizophrenia genome scans (Levinson et al, in press). For this reason the weighted analysis was deemed the principal one, although a case could be made for a more stringent weighting scheme. Furthermore, the differences between genome scans in classification of the phenotype resulted in a reduced data-set, and therefore reduced power, of the most narrow and possibly most genetic disease model analysis. These problems also relate to the inability of the method, or any meta-analytic method, to

examine a sample for heterogeneity. The presence of substantial locus heterogeneity could possibly account for the failure to identify a region significant at the genome-wide level.

7.3 Impact of these Results

7.3.1 The Differences Between Schizophrenia and Bipolar Disorder

Several chromosomal regions exceeding the statistical threshold in the genome scan study, appeared to overlap regions previously linked to schizophrenia (see chapter 4, section 4.4.2; Bennett et al, 2002). However, a better idea of the extent of genetic pleiotropy between the affective and schizophrenic phenotypes is possible by comparing the meta-analysis studies in each (chapter 7; Lewis et al, in press). Despite similar sample sizes the meta-analyses of bipolar disorder, under the broad model, and of schizophrenia showed very different results. The regions showing genome-wide significance in schizophrenia did not overlap with any of the regions showing nominal significance in any of the BPD disease models. Although no definite conclusions can be drawn, as the BPD results could change when more data is added, this suggests that underlying loci for the two disorders do not overlap to a large extent, despite previous suggestions of pleiotropy (Berrettini, 2000). This observation may also suggest a more fundamental difference in the underlying genetics of the two disorders, that schizophrenia may have fewer genes, of larger effect than bipolar disorder, or has less locus heterogeneity.

7.3.2 Uncertainty About the True Underlying Genetic Model

In BPD the results of the meta-analysis suggest that vulnerability loci are of very small effect. The genome scan results (chapter 4) indicate that loci conferring relative risks of 3 or greater were probably not present in the sample. The meta-analysis, which tested for loci common to all or most sample populations indicated that five or more loci conferring a λ_s greater than 1.3 were unlikely in a disease model including bipolar I, II and schizoaffective

disorders. Similarly ten or more loci conferring a λ_s greater than 1.15 were probably not present. Barring loci of effect size smaller than 1.15, and given the other caveats of the meta-analysis method (section 7.2.2), these findings allow an tentative upper limit of ten underlying loci in this disease model. Power was poor for the more narrow disease model, and so inferences of the genetic model for bipolar I and schizoaffective disorder alone were not possible. Thus there may be true loci showing very weak levels of linkage to BPD in individual studies, which can only be detected with appropriate statistical significance levels in samples of the magnitude of thousands of affected individuals. The idea of a polygenic mode of transmission for BPD is now more accepted in the field. However, such a small effect size, if true, has deep implications for the field, and calls into question whether many previous linkage and association studies had sufficient statistical power.

The alternative explanation of these results involves a model of considerable locus heterogeneity, combined with other confounding factors (section 7.2.2). It is clear that there is at least some heterogeneity, from the studies of rare, large pedigrees showing very significant linkage to individual loci (e.g. Blackwood et al, 1996; Morissette et al, 1999). However, single gene models or genetic heterogeneity are not consistent with the familial recurrence risks for BPD (Craddock et al, 1995), suggesting they cannot be responsible for the majority of cases of BPD, while a multiplicative model involving three or more loci has been shown to be plausible.

Substantial uncertainty about the nature of the genetic aetiology of BPD remains (see chapter 1, section 1.3.3). Several segregation studies that identified a single major locus (SML) model as that which best fits family data for BPD, have had questions raised about their reliability (Ott, 1990; Owen et al, 2000). The examination of pedigrees segregating BPD, and increases in risk to relatives of probands do not, in general, indicate an obvious

Mendelian transmission pattern. It may be that environmental risk factors, the presence of phenocopies, or other factors such as age cohort (Blangero and Elston, 1989) are obscuring such a pattern, and there appears to be a need for new studies to re-appraise the segregation of affective disorders in the population at large (see section 7.4.1).

7.4 Future Directions

7.4.1 Familiality and Diagnostic Issues

There is a need for modern, large and comprehensive family and twin studies of BPD. Such studies could allow for the evaluation of the familiality of modern diagnostic criteria directly. They would also allow a new opportunity for detailed examination of the role of sub-phenotypes, diagnostic dimensions, specific symptoms, and gender. The measurement of as many other factors as may be considered relevant may also turn out to provide valuable information in phenotype refinement, including blood biochemistry, psychological and cognitive tests, physical measurements, accompanying illnesses, life histories, etc.

Heterogeneity in the transmission patterns can also be investigated. A sample can be examined for presence of sporadic pedigrees segregating the illness at a high-density, potentially with a Mendelian transmission pattern (e.g. Blackwood et al, 1996), and therefore 'background' rates and familial risks for illness can be more accurately measured. Phenotypic measures and environmental risk factors can then be investigated in relation to pedigrees where a major locus appears to be transmitted, and in relation to other pedigrees where there are lower rates of the disorder, or where there is no apparent transmission pattern.

7.4.2 Genome-Wide Association Studies

Advances in molecular genetic techniques have recently made available a large number of very tightly spaced Single Nucleotide Polymorphisms (SNPs). These are usually biallelic, and are estimated to be spaced about every 1 kilobase-pair (kb). This high density could make these markers more informative than microsatellites, despite their low individual polymorphic content. It would also permit full-genome linkage disequilibrium (LD) scans, as has already been attempted for BPD in a population isolate using microsatellite markers at a 3.5 cM resolution (Ophoff et al, 2002).

Marker Selection

It has been suggested that the number of markers required for a genome-wide association study may be very large, with estimates ranging as high as 500,000 SNPs if LD does not extend beyond 3 kb (Kruglyak, 1999). However, the presence of haplotype "blocks" within the genome with sizes in tens of kb has been reported (Daly et al, 2001). These blocks are separated by apparent recombination "hot-spots", and show low allelic diversity within each block. The creation of haplotype "tags", that is, markers whose allelic diversity accurately reflects the haplotypes within each block, could allow the lower estimate of the number of markers required to extract close to full information. The construction of haplotype maps of the genome will allow detection of association between a trait and these ancestral blocks, rather than the current methods which focus on sporadic markers, with uncharacterised interrelationships.

To avoid some of the pitfalls in detecting linkage disequilibrium with a disease locus, it has been recommended that association studies would increase power by examination of variants affecting protein structure or expression (VAPSEs; Jones and Craddock, 2001). These polymorphisms have a higher prior probability of being the variants directly

responsible for an increase in risk. However, it is rarely the case that any polymorphism in or near a gene can be definitively excluded from affecting protein structure or expression, *a priori*.

The advent of rapid sequencing techniques from companies such as Celera Genomics and Solexa may bypass many of the methodological difficulties in the construction of haplotype maps. However, it remains to be seen if the cost of whole-genome sequence analysis or sequence variation analysis can be brought down to a level where studies with samples of hundreds or thousands of individuals are possible.

Statistical Power

The problem of the sample size required to detect genetic effects of small magnitude (see section 7.2.1) will remain relevant to genome-wide association studies. The use of statistical methods which detect association to many loci simultaneously (e.g. Koeleman et al, 2000) may be useful in this context, by allowing the examination of gene-gene interactions. Another way to increase power may be to use samples based on pedigrees, in which the occurrence of BPD is less likely to be sporadic, or 'environmental'. Many samples of this type have already been collected for genetic linkage studies (see section 7.4.3).

In such cases, where pedigrees are genotyped at a very high marker density, the sample would be informative for both linkage and association. The combination of linkage and association into a unified maximum likelihood statistic as devised by Freimer et al (1996), or a related statistic could perhaps be usefully applied in this case. This statistic was previously developed for use in the fine-mapping of a chromosome 18 candidate region in a sample comprising a large pedigree within a population isolate sample.

7.4.3 Linkage Studies

Collection of Large Samples

As mentioned above (section 7.3.2), one of the largest problems facing genetic linkage studies in BPD may be the sample size required to detect genes of small effect. Following from the results of the meta-analysis (chapter 6), researchers may begin to collect larger samples in order to increase statistical power to the level of the schizophrenia meta-analysis (Lewis et al, in press), which was more successful. In this regard, whole-genome studies in large samples are still recommended, due to their ability to be easily integrated into meta-analyses and in order to prevent the introduction of bias in perceptions of the evidence for linkage.

For most individual research groups the collection of samples of the required size may be unfeasible, and the expense may not be justified simply in order to increase the size of an ongoing meta-analysis. Therefore other approaches to increase power are also required.

Use of Special Case Pedigrees or Population Isolates

The most significant linkage results in BPD, indeed the only ones that have been significant by the stringent criteria of Lander and Kruglyak (1995), have been obtained in smaller studies examining one or a few extended pedigrees (Blackwood et al, 1996; McInnes et al, 1996; Adams et al, 1998; Morissette et al, 1999). Therefore, while there is no proof of genetic locus heterogeneity in BPD, this might suggest that several rare families segregate a locus of larger effect than in the general patient population, and the collection of such extended, high-density pedigrees might be more cost and time-efficient. Such pedigrees may be more common, and have greater genetic, environmental and cultural homogeneity, in population isolates (Terwilliger and Göring, 2000). These

populations may then permit a variety of linkage disequilibrium mapping techniques to be successfully applied.

Modelling Epistasis to Improve Statistical Power

One flaw in current linkage statistics is the assumption that there is a single underlying locus; this is true even of non-parametric statistics. The property of robustness in the presence of heterogeneity or polygenic inheritance is unsatisfactory if these models are known to exist, and acknowledged to be reducing the statistical power of linkage studies. The examination of epistatically interacting genes may become more relevant in examination of linkage to BPD in the future. The benefits of stratifying a sample on the basis of allele-sharing at one locus, have been show to be minimal (Leal and Ott, 2000), however, there are statistical methods available which measure linkage simultaneously at more than one locus (e.g. Cordell et al, 2000).

Collaborative Studies and Maximising Existing Samples

Individual linkage studies still have an important role to play, particularly if they permit identification of positional candidate genes linked to sub-phenotypes (section 7.4.1), or using statistical techniques which may more accurately model the underlying genetic architecture (see above). A cumulative meta-analysis, proposed as an extension to the Genome Scan Meta-Analysis (chapter 6) will be an important tool in encouraging integrated methodologies, and in the long term may yield significant results. Alternatively, or in parallel, the re-genotyping of all existing datasets for a unified marker map would assist greatly integration of data in further meta-analysis or collaborative studies, without an immediate need for investment in further sample ascertainment. It would permit the extraction of the maximum information from existing pedigrees. A standardised high-density genome-wide map using markers, which preferably, would have been heavily used

by previous studies, should be developed for this purpose. Adherence to such a standard need not mean the limitation of research to the point of excluding other approaches, but may eventually permit a study to attain the statistical power necessary to identify BPD vulnerability genes.

Conclusion

The way forward for investigations into the aetiology of affective disorders will be difficult but should, gradually, and with intelligent experimental design and some measure of insight, bear fruit. The hope, and expectation, of many researchers in the field is that the achievement of advances which can be of direct benefit to the long-suffering patients of BPD, will occur sooner rather than later.

Bibliography

1. Online Resources

CGAP Gene Library Summarizer: http://cgap.nci.nih.gov/Tissues/LibrarySummarizer

CEDAR Genetic Location Database: http://cedar.genetics.soton.ac.uk/public_html/ldb.html

Genome DataBase: http://www.gdb.org

Human Genome mapping Project resource centre: http://www.hgmp.mrc.ac.uk

Marshfield Clinic Center for Medical Genetics: http://research.marshfieldclinic.org/genetics

UCSC Golden Path Genome Browser: http://genome.cse.ucsc.edu/

2. References

1. Adams LJ, Mitchell PB, Fielder SL, Rosso A, Donald JA, Schofield PR (1998). A susceptibility locus for bipolar affective disorder on chromosome 4q35. *American Journal of Human Genetics* **62**(5):1084-1091

2. Albanese V, Biquet NF, Kiefer H, Bayard E, Mallet J, Meloni R (2001). Quantitative effects on gene silencing by allelic variation at a tetranucleotide microsatellite. *Human Molecular Genetics* **10**(17):1785-1792.

3. Allen MG, Cohen S, Pollin W, Greenspan SI (1974). Affective Illness in Veteran Twins: A Diagnostic Review. *American Journal of Psychiatry* **131**(11):1234-1239.

4. Allen MG (1976). Twin Studies of Affective Illness. *Archives of General Psychiatry* **33**:1476-1478.

5. American Psychiatric Association, "*Diagnostic and Statistical Manual of Mental Disorders*", 4th edition. American Psychiatric Association, Washington D.C., 1996.

6. Andeasen NC, Rice J, Endicott J, Coryell W, Grove WM, Reich T (1987). Familial Rates of Affective Disorder. *Archives of General Psychiatry* **44**:461-469.

7. Angst J, Frey R, Lohmeyer B, Zerbin-Rüdin E (1980). Bipolar Manic-Depressive Psychoses: Results of a Genetic Investigation. *Human Genetics* **55**:237-254.

8. Ashcroft GW, Eccleston D, Murray LG, Glen AIM, Crawford TBB, Pullar IA, Shields PJ, Walter DS, Blackburn IM, Connechan J, Lonergan M (1972). Modified amine hypothesis for the aetiology of affective illness. *The Lancet* **2**:573-577.

9. Asherson P, mant R, Williams N, Cardno A, Jones L, Murphy K, Collier DA, Nanko S, Craddock N, Morris S, Muir W, Blackwood D, McGuffin P, Owen MJ (1998). A study of chromosome 4p markers and dopamine D5 receptor gene in schizophrenia and bipolar disorder. *Molecular Psychiatry* **3**:310-320.

10. Badenhop RF, Moses MJ, Scimone A, Mitchell PB, Ewen KR, Rosso A, Donald JA, Adams LJ, Schofield PR (2001). A genome screen of a large bipolar affective disorder pedigree supports evidence for a susceptibility locus on chromosome 13q. *Molecular Psychiatry* **6**:396-403.

11. Badner JA, Goldin LR (1997). Bipolar disorder and chromosome 18: an analysis of multiple data sets. *Genetic Epidemiology* **14**:569-574.

12. Badner JA, Gershon ES (2002a). Regional meta-analysis of published data supports linkage of autism with markers on chromosome 7. *Molecular Psychiatry* 7(1):56-66.

13. Badner JA, Gershon ES (2002b). Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Molecular Psychiatry* 7(4):405-411.

14. Baron M, Risch N, Hamburger R, Mandel B, Kushner S, Newman M, Drumer D, Belmaker RH (1987). Genetic Linkage between X-Chromosome Markers and Bipolar Affective Illness. *Nature* **326**:289-292.

15. Baron M, Freimer NF, Risch N, Lerer B, Alexander JR, Straub RE, Asokan S, Das K, Peterson A, Amos J, Endicott J, Ott J, Gilliam TC (1993). Diminished

Support for Linkage between Manic Depressive Illness and X-Chromosome Markers in Three Israeli Pedigrees. *Nature Genetics* **3**:49-54.

16. Bellivier F, Schurhoff F, Nosten-Bertrand M, Mallet J, Feingold J, Leboyer M (1998). Methodological problems in meta-analysis of association studies between bipolar affective disorder and the tyrosine hydroxylase gene. *American Journal of Medical Genetics* **81**:349-350.

17. Berrettini WH, Goldin LR, Gelernter J, Gejman PV, Gershon ES, Detera-Wadleigh S (1990). X-Chromosome Markers and Manic-Depressive Illness. *Archives of General Psychiatry* **47**:366-373.

18. Berrettini WH, Ferraro TN, Goldin LR, Weeks DE, Detera-Wadleigh S, Nurnberger JI Jr., Gershon ES (1994). Chromosome 18 DNA markers and manic-depressive illness: evidence for a susceptibility gene. *Proceedings of the National Academy of Science USA* **91**:5918-5921.

19. Berrettini WH, Ferraro TN, Goldin LR, Detera-Wadleigh SD, Choi H, Muniec D, Guroff JJ, Kazuba DM, Nurnberger JI Jr., Hsieh WT, Hoehe MR, Gershon ES (1997). A linkage study of bipolar illness. *Archives of General Psychiatry* **54**(1):27-35.

20. Berrettini WH, Vuoristo J, Ferraro TN, Wildenauer D, Ala-Kokko L (1998). Human G(olf) gene polymorphisms and vulnerability to bipolar disorder. *Psychiatric Genetics* **8**(4):235-238.

21. Berrettini WH (2000). Are schizophrenic and bipolar disorders related? A review of family and molecular studies. *Biological Psychiatry* **48**:531-538.

22. Berridge MJ, Downes CP, Hanley MR (1982). Lithium amplifies agonist-dependent phosphatidylinositol responses in brain and salivary glands. *Biochemical Journal* **206**:587-595.

23. Bertelsen A, Harvald B, Hauge M (1977). A Danish Twin Study of Manic-Depressive Disorders. *British Journal of Psychiatry* **130**:330-351.

24. Bertelsen A, Gottesman II (1995). Schizoaffective Psychoses: Genetical Clues to Classification. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* **60**:7-11.

25. Bishop DT, Williamson JA (1990). The power of identity-by-state methods for linkage analysis. *American Journal of Human Genetics* **46**:254-265.

26. Blacker D, Tsuang MT (1993). Unipolar Relatives in Bipolar Pedigrees: are they Bipolar? *Psychiatric Genetics* **3**:5-16.

27. Blacker D, Faraone SV, Rosen AE, Guroff JJ, Adams P, Weissman MM, Gershon ES (1996). Unipolar Relatives in Bipolar Pedigrees: A Search for Elusive Indicators of Underlying Bipolarity. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* **67**:445-454.

28. Blackwood DHR, He L, Morris SW, McLean A, Whitton C, Thomson M, Walker MT, Woodburn K, Sharp CM, Wright AF, Shibasaki Y, St. Clair DM, Porteous DJ, Muir WJ (1996). A locus for bipolar affective disorder on chromosome 4p. *Nature Genetics* **12**(4):427-430

29. Blangero J, Elston RC (1989). Familial Analysis of Bipolar Affective Disorder using Logistic Models. *Genetic Epidemiology* **6**:221-227.

30. Blehar MC, Weissman MM, Gershon ES, Hirschfeld RMA (1988). Family and genetics studies of affective disorders. *Archives of General Psychiatry* **45**:289-292.

31. 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 SE, Pulver AE (1998). Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. *Nature Genetics* **20**(1):70-73

32. Bocchetta A, Piccardi MP, Del Zompo M (1994). Is Bipolar Disorder Linked to Xq28? *Nature Genetics* **6**:224.

33. Bowen T, Kirov G, Gill M, Spurlock G, Vallada HP, Murray RM, McGuffin P, Collier DA, Owen MJ, Craddock N (1999). Linkage studies of bipolar disorder with chromosome 18 markers. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* **88**:503-509.

34. Brownstein MJ, Carpten JD, Smith JR (1996). Modulation of non-templated nucleotide addition by taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* **20**:1004-1008.

35. Bruun TG, Ewald H (1999). Selection bias of susceptibility genes possible when using parent-offspring trios in genetic association studies. *Molecular Psychiatry* **4**:414-416.

36. Bunney WE Jr, Post RM, Andersen AE, Kopanda RT (1977). A neuronal sesitivity mechanism in affective illness (a review of evidence). *Communications in Psychopharmacology* **1**:393-405.

37. Byerley W, Hoff M, Holik J, Coon H (1994). A linkage study with D5 dopamine and alpha 2C-adrenergic receptor genes in six muliplex pedigrees. *Psychiatric Genetics* **4**(3):121-124.

38. Byerley W, Holik J, Hoff M, Coon H (1995). Search for a Gene Predisposing to Manic-Depression on Chromosome 21. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* **60**:231-233.

39. Cadoret RJ (1978). Evidence for Genetic Inheritance of Primary Affective Disorder in Adoptees. *American Journal of Psychiatry* **135**(4):463-466.

40. Carroll BJ, Martin FIR, Davies B (1968). Resistance to dexamethasone of plasma 11-O.H.C.S levels in severe depressed patients. *British Medical Journal* **3**:285-287.

41. Carroll BJ, Feinberg M, Greden JF, Tarika J, Albala AA, Haskett RF, James NM, Kronfol Z, Lohr N, Steiner M, de Vigne JP, Young E (1981). A specific laboratory test for the diagnosis of melancholia. *Archives of General Psychiatry* **38**:15-22.

42. Christie JE, Whalley LJ, Dick H, Blackwood DHR, Blackburn IM, Fink G (1986). Raised plasma cortisol concentrations a feature of drug-free psychotics and not specific for depression. *British Journal of Psychiatry* **148**:58-65.

43. Cichon S, Schmidt-Wolf G, Schumacher J, Müller DJ, Hürter M, Schulzez TG, Albus M, Borrmann-Hassenbach M, Franzek E, Lanczik M, Fritze J, Kreiner R, Weigelt B, Minges J, Lichtermann D, Lerer B, Kanyas K, Strauch K, Windemuth C, Baur MP, Wienker TF, Maier W, Rietschel M, Propping P, Nöthen MM (2001a). A possible susceptibility locus for bipolar affective disorder in chromosomal region 10q25-q26. *Molecular Psychiatry* **6**:342-349.

44. Cichon S, Schumacher J, Müller DJ, Hürter M, Windemuth C, Strauch K, Hemmer S, Schulze TG, Schmidt-Wolf G, Albus M, Borrmann-Hassenbach M, Franzek E, Lanczik M, Fritze J, Kreiner R, Reuner U, Weigelt B, Minges J, Lichtermann D, Lerer B, Kanyas K, Baur M, Wienker TF, Maier W, Rietschel M, Propping P, Nöthen MM (2001b). A genome screen for genes predisposing to bipolar affective disorder detects a new susceptibility locus on 8q. *Human Molecular Genetics* **10**(25):2933-2944.

45. Claes S, Raeymaekers P, Van den Broeck M, Diependaele S, De Bruyn A, Verheyen G, Wils V, Boogaerts A, Tanghe A, Godderis J, Van Broeckhoven C,

Cassiman J-J (1997). A chromosome 18 genetic linkage study in three large Belgian pedigrees with bipolar disorder. *Journal of Affective Disorders* **43**:195-205.

46. Cohen J (1968). Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* **70**(4):213-220.

47. Coon H, Jensen S, Hoff M, Holik J, Plaetke R, Reimherr F, Wender P, Leppert M, Byerley W (1993). A genome-wide search for genes predisposing to manic-depression, assuming autosomal dominant inheritance. *American Journal of Human Genetics* **52**:1234-1249.

48. Coon H, Hoff M, Holik J, Hadley D, Fang N, Reimherr F, Wender P, Byerley W (1996). Analysis of chromosome 18 DNA markers in multiplex pedigrees with manic depression. *Biological Psychiatry* **39**:689-696.

49. Cordell HJ, Todd JA (1995). Multifactorial inheritance in type 1 diabetes. *Trends in Genetics* **11**(12):499-504.

50. Cordell HJ, Wedig GC, Jacobs KB, Elston RC (2000). Multilocus linkage tests based on affected relative pairs. *American Journal of Human Genetics* **66**(4):12733-1286.

51. Corder EH, Saunders AM, Strittmatter WJ, Schmechel D, Gaskell P, Small GW, Roses AD, Haines JL, Pericak-Vance MA (1993). Apolipoprotein E4 gene dose and the risk fof Alzheimer disease in late onset families. *Science* **261**:921-923.

52. Coryell W, Endicott J, Reich T, Andreasen N, Keller M (1984). A Family Study of Bipolar II Disorder. *British Journal of Psychiatry* **145**:49-54.

53. Craddock N, Dawson E, Burge S, Parfitt L, Mant B, Roberts Q, Daniels J, Gill M, McGuffin P, Powell J, Owen M (1993). The gene for Darier's disease maps to chromosome 12q23-q24.1. *Human Molecular Genetetics* **2**(11):1941-1943

54. Craddock N, Owen M, Burge S, Kurian B, Thomas P, McGuffin P (1994). Familial cosegregation of major affective disorder and Darier's disease (keratosis follicularis). *British Journal of Psychiatry* **164**(3):355-358

55. Craddock N, Khodel V, Van Eerdewegh P, Reich T (1995). Mathematical Limits of Multilocus Models: The Genetic Transmission of Bipolar Disorder. *American Journal of Medical Genetics* **57**:690-702

56. Craddock N, Davé S, Greening J (2001). Association studies of bipolar disorder. *Bipolar Disorders* **3**:284-298.

57. Crow TJ (1986). The continuum of psychosis and its implication for the structure of the gene. *British Journal of Psychiatry* **149**:419-429.

58. Cummings CJ, Zoghbi HY (2000). Fourteen and Counting: Unravelling Trinucleotide Repeat Diseases. *Human Molecular Genetics* **9**(6):909-916.

59. Curtis D, Brynjolfsson J, Petursson H, Holmes S, Sherrington R, Brett P, Rifkin L, Murphy P, Moloney E, Melmer G, Gurling HMD (1993). Segregation and linkage analysis in five manic depression pedigrees excludes the 5HT1a receptor gene (*HTR1A*). *Annals of Human Genetics* **57**:27-39.

60. Da Fonseca A. "*Anâlise Heredo-Clínica das Perturbações Affectivas (Estudio de 60 Pares de Gémeos, e seus Consaguíneos*". Porto, Imprensa Portuguesa, 1959.

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

62. Dawson E, Parfitt E, Roberts Q, Daniels J, Lim L, Sham P, Nöthen M, Propping P, Weissenbach J, Gill M, Powell J, McGuffin P, Owen M, Craddock N (1995). Linkage studies of bipolar disorder in the region of the Darier's disease gene on chromosome 12q23-24.1. *American Journal of Medical Genetics* **60**(2):94-102.

63. De Bruyn A, Souery D, Mendelbaum K, Mendlewicz J, Van Broeckhoven C (1996). Linkage analysis of families with bipolar illness and chromosome 18 markers. *Biological Psychiatry* **39**:679-688.

64. Del Zompo M, Bocchetta A, Goldin LR, Corsini GU (1984). Linkage between X-Chromosome Markers and Manic-Depressive Illness. *Acta Psychiatrica Scandinavica* **70**:282-287.

65. Detera-Wadleigh SD, Berrettini WH, Goldin LR, Boorman D, Anderson S, Gershon ES (1987). Close linkage of c-Harvey-ras-1 and the insulin gene to affective disorder is ruled out in three North American pedigrees. *Nature* **325**(6107):806-808.

66. Detera-Wadleigh SD, Hsieh WT, Berrettini WH, Goldin LR, Rollins DY, Muniec D, Grewal R, Guroff JJ, Turner G, Coffman D, Barrick J, Mills K, Murray J, Donohue SJ, Klein DC, Sanders J, Nurnberger JI Jr., Gershon ES (1994). Genetic Linkage Mapping for a Susceptibility Locus to bipolar illness: Chromosomes 2, 3, 4, 7, 9, 10p, 11p, 22 and Xpter. *American Journal of Medical Genetics* **54**:206-218.

67. Detera-Wadleigh SD, Badner JA, Goldin LR, Berrettini WH, Sanders AR, Rollins DY, Turner G, Moses T, Haerian H, Muniec D, Nurnberger JI Jr, Gershon ES (1996). Affected-Sib-Pair Analyses Reveal Support of Prior Evidence for a Susceptibility Locus for Bipolar Disorder, on 21q. *American Journal of Human Genetics* **58**:1279-1285.

68. Detera-Wadleigh SD, Badner JA, Yoshikawa T, Sanders AR, Goldin LR, Turner G, Rollins DY, Moses T, Guroff JJ, Kazuba D, Maxwell ME, Edenberg HJ, Foroud T, Lahiri D, Nurnberger JI Jr, Stine OC, McMahon F, Meyers DA, MacKinnon D, Simpson S, McInnis M, DePaulo JR, Rice J, Goate A, Gershon ES (1997). Initial genome scan of the NIMH Genetics Initiative bipolar pedigrees: Chromosomes 4, 7, 9, 18, 19, 20, and 21q. *American Journal of Medical Genetics* **74**(3):254-262

69. Detera-Wadleigh, SD, Badner JA, Berrettini WH, Yoshikawa T, Goldin LR, Turner G, Rollins DY, Moses T, Sanders AR, Karkera JD, Esterling LE, Zeng J, Ferraro TN, Guroff JJ, Kazuba D, Maxwell ME, Nurnberger JI Jr., Gershon ES (1999). A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11.2. *Proceedings of the National Academy of Science USA* **96**:5604-5609

70. Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G Morissette J, Weissenbach J et al (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**(6570):152-154.

71. Dixon JF, Hokin LE (1997). The antibipolar drug valproate mimics lithium in stimulating glutamate release and inositol 1,4,5-trisphosphate accumulation in brain cortex slices but not accumulation of inositol monophosphates and bisphosphates. *Proceedings of the National Academy of Sciences USA* **94**:4757-4760.

72. Douglas JA, Boehnke M, Lange K (2000). A multipoint method for detecting genotyping errors and mutations in sibling-pair linkage data. *American Journal of Human Genetics* **66**:1287-1297.

73. Drevets WC, Öngür D, Price JL (1998). Neuroimaging abnormalities in the subgenual prefrontal cortex: implications for the psychophysiology of familial mood disorders. *Molecular Psychiatry* **3**:220-226.

74. Du Fort GG, Kovess V, Boivin J-F (1994). Spouse Similarity for Psychological Distress and Well-Being: A Population Study. *Psychological Medicine* **24**(2):431-447.

75. Durner M, Vieland VJ, Greenberg DA (1999). Further evidence for the increased power of LOD scores compared wth nonparametric methods. *American Journal of Human Genetics* **64**(1):281-289.

76. Ebert MH, Post RM, Goodwin FK (1972). Effect of physical activity on urinary MHPG excretion in depressed patients. *The Lancet* **2**:766-767.

77. Edenberg HJ, Foroud T, Conneally PM, Sorbel JJ, Carr K, Crose C, Willig C, Zhao J, Miller M, Bowman E, Mayeda A, Rau NL, Smiley C, Rice JP, Goate A, Reich T, Stine OC, McMahon F, DePaulo JR, Meyers D, Detera-Wadleigh SD, Goldin LR, Gershon ES, Blehar MC, Nurnberger JI Jr. (1997). Initial genomic scan of the NIMH Genetics Initiative bipolar pedigrees: Chromosomes 3, 5, 15, 16, 27 and 22. *American Journal of Medical Genetics* **74**(3):238-246.

78. Egeland JA, Gerhard DS, Pauls DL, Sussex JN, Kidd KK, Allen CR, Hostetter AM, Housman DE (1987). Bipolar affective disorders linked to DNA markers on chromosome 11. *Nature* **325**(6107):783-7

79. Ehm MG, Wagner M (1998). A test statistic to detect errors in sib-pair relationships. *American Journal of Human Genetics* **62**:181-188.

80. El-Mallakh RS, Wyatt RJ (1995). The Na,K-ATPase hypothesis for bipolar illness. *Biological Psychiatry* **37**:235-244.

81. Emory AEH. "*Methodology in Medical Genetics: An Introduction to Statistical Methods*", 2nd edition. Churchill Livingstone, New York, 1986.

82. Escamilla MA, McInnes LA, Spesny M, Reus VI, Service SK, Shimayoshi N, Typler DJ, Silva S, Molina J, Gallegos A, Meza L, Cruz ML, Batki S, Vinogradov S, Neylan T, Nguyen JB, Fournier E, Araya C, Barondes SH, Leon P, Sandkuijl LA, Freimer NB (1999). Assessing the feasibility of linkage disequilibrium methods for mapping complex traits: an initial screen for bipolar disorder loci on chromosome 18. *American Journal of Human Genetics* **64**:1670-1678.

83. Ewald H, Mors O, Eiberg H, Flint T, Kruse TA (1995). No evidence of linkage between manic depressive illness and the dopa decarboxylase gene or nearby region on chromosome 7p. *Psychiatric Genetics* **5**:161-169.

84. Ewald H, Mors O, Koed K, Eiberg H, Kruse TA (1997). Susceptibility loci for bipolar affective disorder on chromosome 18? A review and a study of Danish families. *Psychiatric Genetics* **7**(1):1-12.

85. Ewald H, Degn B, Mors O, Kruse TA (1998). Support for the possible locus on chromosome 4p16 for bipolar affective disorder. *Molecular Psychiatry* **3**(5):442-448.

86. Ewald H, Flint T, Kruse TA, Mors O (2002). A genome-wide scan shows significant linkage between bipolar disorder and chromosome 12q24.3 and suggestive linkage to chromosomes 1p22-21, 4p16, 6q14-22, 10q26 and 16p13.3. *Molecular Psychiatry* **7**(7):734-744.

87. Falk CT, Rubinstein P (1987). Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Annals of Human Genetics* **51**: 227-233.

88. Faraone SV, Matise T, Svrakic D, Pepple J, Malaspina D, Suarez B, Hampe C, Zambuto CT, Schmitt K, Meyer J, Markel P, Lee H, Harkavy Friedman J, Kaufmann C, Cloninger CR, Tsuang MT (1998). Genome scan of European-American schizophrenia pedigrees: results of NIMH Genetics Initiative and Millennium Consortium. *American Journal of Medical Genetics* **81**(4):290-195.

89. Faraone SV, Meyer J, Matise T, Svrakic D, Pepple J, Malaspina D, Suarez B, Hampe C, Chan C, Aelony A, Friedman JH, Kaufmann C, Cloninger CR, Tsuang MT (1999). Suggestive linkage of chromosome 10p to schizophrenia is not due to transmission ratio distortion. *American Journal of Medical Genetics* **88**(6):607-608.

90. Farrer LA, Cupples LA. *Determining the Genetic Component of a Disease;* in "Approaches to Gene Mapping in Complex Human Diseases", Haines JL, Pericak-Vance MA, eds.; Wiley-Liss, New York, 1998, pp 93-130.

91. Fenton I, Sandkuijl KA (1992) A genetic database for polycystic kidney disease. *Contributions to Nephrology* **97**:118-127.

92. Fisher RA. "*Statistical Methods for Research Workers*". Oliver & Boyd, London, 1932.

93. Foroud T, Castelluccio PF, Koller DL, Edenberg HJ, Miller M, Bowman E, Rau NL, Smiley C, Rice JP, Goate A, Armstrong C, Bierut LJ, Reich T, Detera-Wadleigh SD, Goldin LR, Badner JA, Guroff JJ, Gershon ES, McMahon FJ, Simpson S, MacKinnon D, McInnis M, Stine OC, DePaulo JR, Blehar MC, Nurnberger JI Jr (2000) Suggestive evidence of a locus on chromosome 10p using the NIMH genetics initiative bipolar affective disorder pedigrees. *American Journal of Medical Genetics* **96**, 18-23.

94. Freimer NB, Reus VI, Escamilla MA, McInnes LA, Spesny M, Leon P, Service SK, Smith LB, Silva S, Rojas E, Gallegos A, Meza L, Fournier E, Baharloo S, Blankenship K, Tyler DJ, Batki S, Vinogradov S, Weissenback J, Barondes SH, Sandkuijl LA (1996). Genetic mapping using haplotype, association and linkage methods suggests a locus for severe bipolar disorder (BPI) at 18q22-q23. *Nature Genetics* **12**:436-441.

95. Friddle C, Koskela R, Ranade K, Hebert J, Cargill M, Clark CD, McInnis M, Simpson S, McMahon F, Stine OC, Meyers D, Xu J, MacKinnon D, Swift-Scanlan T, Jamison K, Folstein S, Daly M, Kruglyak L, Marr T, DePaulo JR, Botstein D (2000). Full-genome scan for linkage in 50 families segregating the bipolar affective disorder phenotype. *American Journal of Human Genetics* **66**:205-215

96. Furlong RA, Ho L, Walsh C, Rubinsztein JS, Paykel ES, Easton DF, Rubinsztein DC (1998). Analysis and meta-analysis of two serotonin transporter gene polymorphisms in bipolar and unipolar affective disorders. *American Journal of Mdeical Genetics* **81**:58-63.

97. Furlong RA, Rubinsztein JS, Ho, L, Walsh C, Coleman TA, Muir WJ, Paykel ES, Blackwood DHR, Rubinsztein DC (1999a). Analysis and meta-analysis of two

polymorphisms within the tyrosine hydroxylase gene in bipolar and unipolar affective disorders. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* **88**: 88-94.

98. Furlong RA, Ho L, Rubinsztein JS, Walsh C, Paykel ES, Rubinsztein DC (1999b). Analysis of the monoamine oxidase A (MAOA) gene in bipolar affective disorder by association studies, meta-analyses, and sequencing of the promoter. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* **88**:398-406.

99. Garner C, McInnes LA, Service SK, Spesny M, Fournier E, Leon P, Freimer NB (2001). Linkage analysis of a complex pedigree with severe bipolar disorder, using a markov chain monte carlo method. *American Journal of Human Genetics* **68**(4):1061-1064

100. Gelder M, Gath D, Mayou R (eds). "Oxford Textbook of Psychiatry", 2nd edition. Oxford Medical Publications, Oxford, 1991.

101. Gerhard DS, LaBuda MC, Bland SD, Allen C, Egeland JA, Pauls DL (1994). Initial report of a genome search for the affective disorder predisposition gene in the Old Order Amish Pedigrees: Chromosomes 1 and 11. *American Journal of Medical Genetics* **54**:398-404

102. Gershon ES, Mark A, Cohen N, Belizon N, Baron M, Knobe KE (1975). Transmitted Factors in the Morbid Risk of Affective Disorders: A Controlled Study. *Journal of Psychiatric Research* **12**:283-299.

103. Gershon ES, Bunney WE Jr, Leckman JF, Van Eerdewegh M, DeBauche BA (1976). The Inheritance of Affective Disorders: A Review of Data and of Hypotheses. *Behavior Genetics* **6**(3):227-261.

104. Gershon ES, Hamovit J, Guroff JJ, Dibble E, Leckman JF, Sceery W, Targum SD, Nurnberger JI Jr, Goldin LR, Bunney WE Jr (1982). A Family Study of

Schizoaffective, Bipolar I, Bipolar II, Unipolar and Normal Control Probands. *Archives of General Psychiatry* **39**:1157-1167.

105. Gershon ES, Badner JA, Detera-Wadleigh SD, Ferraro TN, Berrettini WH (1996). Maternal inheritance and chromosome 18 allele sharing in unilineal bipolar illness pedigrees. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* **67**:202-207.

106. Ghosh S, Karanjawala ZE, Hauser ER, Ally D, Knapp JI, Rayman JB, Musick A, Tannenbaum J, Te C, Shapiro S, Elridge W, Musick T, Martin C, Smith JR, Carpten JD, Brownstein MJ, Powell JI, Whiten R, Chines P, Nylund SJ, Magnuson VL, Boehnke M, Collins FS, FUSION Study Group (1997). Methods for precise sizing, automated binning of alleles, and reduction of error rates in large-scale genotyping using fluorescently labeled dinucleotide markers. *Genome Research* 7:165-178.

107. Gill M, McKeon P, Humphries P (1988). Linkage analysis of manic depression in an Irish family using H-ras 1 and INS DNA markers. *Journal of Medical Genetics* **25**(9):634-635.

108. Gill M, Castle D, Duggan C (1992). Cosegregation of Christmas Disease and Major Affective Disorder in a Pedigree. *British Journal of Psychiatry* **160**:112-114.

109. Ginns EI, Ott J, Egeland JA, Allen CR, Fann CSJ, Pauls DL, Weissenbach J, Carulli JP, Falls KM, Keith TP, Paul SM (1996). A genome-wide search for chromosomal loci linked to bipolar affective disorder in the Old Order Amish. *Nature Genetics* 12:431-435.

110. Ginns EI, St. Jean P, Philibert RA, Galdzicka M, Damschroder-Williams P, Thiel B, Long RT, Ingraham LJ, Dalwaldi H, Murray MA, Ehlert M, Paul S, Remortel BG, Patel AP, Anderson MCH, Shaio C, Lau E, Dymarskaia I, Martin BM, Stubblefield B, Falls KM, Carulli JP, Keith TP, Fann CSJ, Lacy LG, Allen CR,

Hostetter AM, Elston RC, Schork NJ, Egeland JA, Paul SM (1998). A genome-wide search for chromosomal loci linked to mental health wellness in relatives at high risk for bipolar affective disorder among the Old Order Amish. *Proceedings of the National Academy of Science USA* **95**:15531-15536.

111. Göring HH, Ott J (1997). Relationship estimation in affected sib pair analysis of late-onset diseases. *European Journal of Human Genetics* **5**:69-77.

112. Gottschalk A, Bauer MS, Whybrow PC (1995). Evidence of chaotic mood variation in bipolar disorder. *Archives of General Psychiatry* **52**(11):947-959.

113. Grigoroiu-Serbanescu M, Nöthen M, Propping P, Poustka F, Magureanu S, Vasilescu R, Marinescu E, Ardelean V (1995). Clinical evidence for genomic imprinting in bipolar I disorder. *Acta Psychiatrica Scandinavica* **92**(5):365-370.

114. Gudbjartsson DF, Jonasson K, Frigge ML, Kong A (2000). Allegro, a new computer program for multipoint linkage analysis. *Nature Genetics* **25**(1):12-13.

115. Gurling H, Smyth C, Kalsi G, Moloney E, Rifkin L, O'Neill J, Murphy P, Curtis D, Petursson H, Brynjolfsson J (1995). Linkage findings in bipolar disorder. *Nature Genetics* **10**:8-9.

116. Guttiérez B, Arranz MJ, Collier DA, Vallès V, Guillamat R, Bertranpetit J, Murray RM, Fañanás L (1998). Serotonin transporter gene and risk for bipolar affective disorder: an association study in a Spanish population. *Biological Psychiatry* **43**:843-847.

117. Guy CA, Bowen T, Jones I, McCandless F, Owen MJ, Craddock N, O'Donovan MC (1999). CTG18.1 and ERDA-1 CAG/CTG Repeat Size in Bipolar Disorder. *Neurobiology of Disease* **6**:302-307.

118. Hafen T, Wollnik F (1994). Effect of lithium carbonate on activity level and circadian period in different strains of rats. *Pharmacology Biochemistry and Behavior* **4**:975-983.

119. Hamer D, Sirota L (2000). Beware the chopsticks gene. *Molecular Psychiatry* **5**:11-13.

120. Harada H, Nakajima T, Inazawa J, Abe T (1995). Bipolar disorder associated with beta-thallasaemia minor. *Biological Psychiatry* **37**(7):477-479.

121. Harvald B, Hauge M. Heredity factors elucidated by twin studies. In: Neel JV (ed): Genetics and the Epidemiology of Chronic Diseases. *Public Health Service Publication* **1163**:61-76. US Government Printing Office, 1965.

122. Hawi Z, Straub RE, O'Neill A, Kendler KS, Walsh D, Gill M (1998). No linkage or linkage disequilibrium between brain-derived neurotrophic factor (BDNF) dinucleotide repeat polymorphism and schizophrenia in Irish families. *Psychiatry Research* **81**:111-116.

123. Hawi Z, Mynett-Johnson L, Gill M, Murphy V, Straub RE, Kendler KS, Walsh D, Machen F, Connell H, McKeon P, Shields D (1999). Pseudoautosomal Gene: Possible Association with Bipolar Males but not with Schizophrenia. *Psychiatric Genetics* **9**:129-134.

124. Hebebrand J (1992). A Critical Appraisal of X-Linked Bipolar Illness: Evidence for the Assumed Mode of Inheritance id Lacking. *British Journal of Psychiatry* **160**:7-11.

125. Hirschfeld RM (2000). History and Evolution of the monoamine hypothesis of depression. *Journal of Clinical Psychiatry* **61**(Suppl 6):4-6.

126. Ho LW, Furlong RA, Rubinsztein JS, Walsh C, Paykel ES, Rubinsztein DC (2000). Genetic associations with clinical characteristics in bipolar affective disorder and recurrent unipolar depressive disorder. *American Journal of Medical Genetics* **96**(1): 36-42.

127. Hodge SE (1993). Linkage analysis versus association analysis: distinguuishing between two models that explain disease-marker associations. *American Journal of Human Genetics* **53**:367-384.

128. Hodgkinson S, Sherrington R, Gurling H, Marchbanks R, Reeders S, Mallet J, McInnis M, Petursson H, Brynjolfsson J (1987). Molecular genetic evidence for heterogeneity in manic depression. *Nature* **325**:805-806.

129. Holmans P (1993). Asymptotic Properties of Affected-Sib-Pair Linkage Analysis. *American Journal of Human Genetics* **52**:362-374.

130. Holmans P, Clayton D (1995). Efficiency of typing unaffected relatives in an affected sib-pair linkage study with single locus and multiple tightly-linked markers. *American Journal of Human Genetics* **57**:1221-1232.

131. Holmans P, Craddock N (1997). Efficient strategies for genome scanning using maximum-likelihood affected-sib-pair analysis. *American Journal of Human Genetics* **60**:657-666.

132. Holmes D, Brynjolfsson J, Brett P, Curtis D, Petursson H, Sherrington R, Gurling H (1991). No evidence for a susceptibility locus predisposing to manic depression in the region of the dopamine (D2) receptor gene. *British Journal of Psychiatry* **158**:635-641.

133. Horrobin DF, Bennett CN (1999). New gene tragets related to schizophrenia and other psychiatric disorder: enzymes, binding proteins and transport proteins involved in phospholipid and fatty acid metabolism. *Prostaglandins, Leukotrienes and Essential Fatty Acids* **60**(3):141-167.

134. Horton RW, Katona CLE. "*Biological Aspects of Affective Disorders*" Academic Press, London, 1991.

135. Hough C, Lu S-J, Davis CL, Chuang D-W, Post RM (1999). Elevated basal and thiapsigargin-stimulated intracellular calcium of platelets and lymphocytes from

bipolar affective disorder patients measured by a fluorometric microassay. *Biological Psychiatry* **46**:247-255.

136. Inglehearn CF (1997). Intelligent linkage analysis using gene density estimates. *Nature Genetics* **16**:15.

137. Inglehearn CF (1998). Molecular genetics of human retinal dystrophies. *Eye* **12**(3b):571-579.

138. Jacobsen NJ, Franks EK, Owen MJ, Craddock NJ (1999). Mutational analysis of phospholipase A2A: a poitional candidate susceptibility gene for bipolar disorder. *Molecular Psychiatry* **4**:274-279.

139. Jeffries FM, Reiss AL, Brown WT, Meyers DA, Clicksman AC, Bandyopadhyay S (1993). Bipolar Spectrum Disorder and Fragile X Syndrome: A Family Study. *Biological Psychiatry* **33**:213-216.

140. Jeunemaitre X, Soubrier F, Kotelevtsev YV, Lifton RP, Williams CS, Charru A, Hunt SC, Hopkins PN, Williams RR, Lalouel JM (1992). Molecular basis of human hypertension: role of angiotensin. *Cell* **71**:169-180.

141. Joffe RT, Horvath Z, Tarvydas I (1986). Bipolar Disorder and thalassemia minor. *American Journal of Psychiatry* **143**:933.

142. Jones I, Craddock N (2001). Candidate gene studies of bipolar disorder. *Annals of Medicine* **33**:248-256.

143. Jope RS (1999a). A bimodal model of the mechanism of action of lithium. *Molecular Psychiatry* **4**:21-25.

144. Jope RS (1999b). Anti-bipolar therapy: mechanism of action of lithium. *Molecular Psychiatry* **4**:117-128.

145. Kallmann F. "Heredity in Health and Mental Disorder". New York, WW Norton & Co, 1953.

146. Kandel ER, Schwartz JH, Jessell TM, eds. "*Essentials of Neural Science and Behavior*". Appleton & Lange, Stamford, 1995.

147. Kato T, Winokur G, Coryell W, Rice J, Endicott J, Keller MB, Akiskal HS (1998). Failure to demonstrate parent-of-origin effect in transmission of bipolar II disorder. *Journal of Affective Disorders* **50**:135-141.

148. Kelsoe JR, Ginns EI, Egeland JA, Gerhard DS, Goldstein AM, Bale SJ, Pauls DL, Long RT, Kidd KK, Conte G, Housman DE, Paul SM (1989). Re-evaluation of the linkage relationship between chromosome 11p loci and the gene for bipolar affective disorder in the Old Order Amish. *Nature* **342**(6247):238-243.

149. Kelsoe JR, Spence MA, Loetscher E, Foguet M, Sadovnick AD, Remick RA, Flodman P, Khristich J, Mroczkowski-Parker Z, Brown JL, Masser D, Ungerleider S, Rapaport MH, Wishart WL, Luebbert H (2001). A genome survey indicates a possible susceptibility locus for bipolar disorder on chromosome 22. *Proceedings of the National Acadademy of Science USA* **98**(2):585-590.

150. Knapp M (1999). A note on power approximations for the transmission/disequilibrium test. *American Journal of Human Genetics* **64**(4):1177-1185.

151. Knowles JA, Rao PA, Cox-Matise T, Loth JE, de Jesus GM, Levine L, Das K, Penchaszadeh GK, Alexander JR, Lerer B, Endicott J, Ott J, Gilliam TC, Baron M (1998). No evidence for significant linkage between bipolar affective disorder and chromosome 18 pericentromeric markers in a large series of multiplex extended pedigrees. *American Journal of Human Genetics* **62**:916-924.

152. Koeleman BP, Dudbridge F, Cordell HJ, Todd JA (2000). Adaptation of the extended transmission/disequilibrium test to distinguish disease associations of multiple loci: the Conditional Extended Transmission/Disequilibrium Test. *Annals of Human Genetics* **64**(3):207-213.

153. Kong A, Cox NJ (1997). Allele-sharing models: LOD scores and accurate linkage tests. *American Journal of Human Genetics* **61**(5):1179-1188.

154. Kornberg JR, Brown JL, Sadovnick AD, Remick RA, Keck PE Jr., McElroy SL, Rapaport MH, Thompson PM, Kaul JB, Vrabel CM, Schommer SC, Wilson T, Pizzuco D, Jameson S, Schibuk L, Kelsoe JR (2000). Evaluating the parnet-of-origin effect in bipolar affective disorder. Is a more penetrant subtype transmitted paternally? *Journal of Affective Disorders* **59**(3):183-192.

155. Körner J, Rietschel M, Hunt N, Castle D, Gill M, Nöthen MM, Craddock N, Daniels J, Owen M, Fimmers R, Fritze J, Möller H-J, Propping P (1994). Association and haplotype analysis at the tyrosine hydroxylase locus in a combined German-British samples of manic depressive patients and controls. *Psychiatric Genetics* **4**: 167-175.

156. Kosaki K, Suzuki T, Kosaki R, Yoshihashi H, Itoh M, Goto Y, Matsuo N (2001). Human homolog of the mouse imprinted gene *Impact* resides at the pericentric region of chromosome 18 within the critical region for bipolar affective disorder. *Molecular Psychiatry* **6**:87-91.

157. Kraeplin E. "*Manic-depressive insanity and paranoia*". E & S Livingstone, Edinburgh, 1921.

158. Kringlen E. "Heredity and Environment in the Functional Psychoses: An Epidemiological-Clinical Twin Study". Oslo, Universitsforlaget, 1967.

159. Kripke DF, Mullaney DJ, Atkinson M, Wolf S (1978). Circadian rhythm disorders in manic-depressives. *Biological Psychiatry* **13**(3):335-351.

160. Kruglyak L, Lander ES (1995). Complete Multipoint Sib-Pair Analysis of Qualitative and Quantitative Traits. *American Journal of Human Genetics* **57:**439-454.

161. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996). Parametric and nonparametric linkage analysis: a unified multipoint approach. *American Journal of Human Genetics* **58:**1347-1363.

162. Kruglyak L (1999). Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics* **22**:139-144.

163. Kwok JBJ, Adams LJ, Salmon JA, Donald JA, Mitchell PB, Schofield PR (1999). Nonparametric Simulation-Based Statistical Analyses for Bipolar Affective Disorder Locus on Chromosome 21q22.3. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* **88**:99-102.

164. LaBuda MC, Maldonado M, Marshall D, Otten K, Gerhard DS (1996). A follow-up report of a genome search for affective disorder predisposition loci in the Old Order Amish. *American Journal of Human Genetics* **59**:1343-1362.

165. Lachman HM, Morrow B, Shprintzen R, Veit S, Parsia SS, Faedda G, Goldberg R, Kucherlapati R, Papolos DF (1996). Association of codon 108/158 catechol-O-methyltransferase gene polymorphism with the psychiatric manifestations of velo-cardio-facial syndrome. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* **67**:468-472.

101. Lander ES, Schork NJ (1994). Genetic dissection of complex traits. *Science* **265**(5181):2037-2048.

102. Lander E, Kruglyak L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* **11**:241-247.

103. Leal SM, Ott J (2000). Effects of stratification in the analysis of affected-sib-pair data: benefits and costs. *American Journal of Human Genetics* **66**:567-575.

104. Lecrubier Y, Puech AJ, Jouvent R, Simon P, Widlocjer D (1980). A beta adrenergic stimulant (salbutamol) versus clomipramine in depression: a controlled study. *British Journal of Psychiatry* **136**:354-358.

105. Leonhard K. "*The classification of endogenous psychoses*", 5th edition. Irvington Publishers, New York, 1979.

106. Levinson DF, Levinson MD, Segurado R, Lewis CM. Genome scan meta-analysis of schizophrenia and bipolar disorder I: methods and power analysis. *American Journal of Human Genetics*, submitted.

107. Lewis CM, Levinson DL, Wise LH, DeLisi LE, Straub RE, Hovatta I, Williams NM, Schwab SG, Pulver AE, Faraone SV, Brzustowicz LM, Kaufmann CA, Garver DL, Gurling HMD, Lindholm E, Coon H, Moises HWM, Byerley W, Shaw SH, Mesen A, Sherrington R, O'Neill FA, Walsh D, Kendler KS, Ekelund J, Paunio T, Lönnqvist J, Peltonen L, O'Donovan MC, Owen MJ, Wildenauer DB, Nestadt G, Blouin J-L, 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, McQuillan A, Curtis D, Brynjolfson J, Sigmundsson T, Petursson H, Jazin E, Zoëga T, Helgason T. Genome scan meta-analysis of schizophrenia and bipolar disorder II: schizophrenia. *American Journal of Human Genetics*, submitted.

108. Lim LCC, Gurling H, Curtis D, Brynjolfsson J, Petursson H, Gill M (1993). Linkage between tyrosine hydroxylase gene and affective disorder cannot be excluded in two of six pedigrees. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* **48**(4):223-228.

109. Lindblad K, Mylander P-O, Zander C, Yuan Q-P, Ståhle L, Engström C, Balciuniene J, Pettersson U, Breschel T, McInnis M, Ross CA, Adolfsson R, Schalling M (1998). Two Commonly Expanded CAG/CTG Repeat Loci: Involvement in Affective Disorders?. *Molecular Psychiatry* **3**:405-410.

110. Luxenburger H (1928). Vorläufiger bericht über psychiatrische Serienuntersuchungen an Zwillingen. *Z Gesamte Neurol Psychiatr* **116**:297-326

111. Luxenburger H (1930). Psychiatrisch-neurologisch Zwillingspathologie. *Zentralb Gesamte Neurol Psychiatr* **56**:145-180.

112. Mahieu B, Souery D, Lipp O, Mendelbaum K, Verheyen G, De Maertelaer V, Van Broeckhoven C, Mendlewicz J (1997). No association between bipolar affective disorder and a serotonin receptor (5-HT2A) polymorphism. *Psychiatry Research* **70**:65-69.

113. Margolis RL, McInnis MG, Rosenblatt A, Ross CA (1999). Trinucleotide Repeat Expansion and Neuropsychiatric Disease. *Archives of General Psychiatry* **56**:1019-1031.

114. Mayeux R, Ottman R, Maestre G, Ngai C, Tang MX, Ginsberg H, Chun M, Tycko B, Shelanski M (1995). Synergistic effects of traumativ head injury and apolipoprotein-epsilon 4 in patients with Alzheimer's disease. *Neyrology* **45**(3 Pt 1):555-557.

115. McGinnis R, Shifman S, Danrasi A (2002). Power and efficiency of the TDT and case control design for association scans. *Behavior Genetics* **32**(2):135-144.

116. McGuffin P, Katz R, Watkins S, Rutherford J (1996). A Hospital-Based Twin Register of the Heritability of DSM-IV Unipolar Depression. *Archives of General Psychiatry* **53**:129-136.

117. McInnes LA, Escamilla MA, Service SK, Reus VI, Leon P, Silva S, Rojas E, Spesny M, Baharloo S, Blankenship K, Peterson A, Tyler D, Shimayoshi N, Tobey C, Batki S, Vinogradov S, Meza L, Gallegos A, Fournier E, Smith LB, Barondes SH, Sandkuijl LA, Freimer NB (1996). A complete genome screen for genes predisposing to severe bipolar disorder in two Costa Rican pedigrees. *Proceedings of the National Academy of Science USA* **93**(23):13060-13065

118. McInnes LA, Service SK, Reus VI, Barnes G, Charlat O, Jawahar S, Lewitzky S, Yang Q, Duong Q, Spesny M, Araya C, Araya X, Gallegos A, Meza L, Molina J,

Ramirez R, Mendez R, Silva S, Fournier E, Batki SL, Mathews CA, Neylan T, Glatt CE, Escamilla MA, Luo D, Galiwala P, Song T, Crook S, Hguyen JB, Roche E, Meyer JM, Leon P, Sandkuijl LA, Freimer NB, Chen H (2001). Fine-scale mapping of a locus for severe bipolar mood disorder on chromosome 18p11.3 in the Costa Rican population. *Proceedings of the National Academy of Sciences USA* **98**(20):11485-11490.

119. McKay AP, Tarbuck AF, Shapleske J, McKenna PJ (1995). Neuropsychological Function in Manic-Depressive Psychosis: Evidence for Persistent Deficits in Patients with Chronis, Severe Illness. *British Journal of Psychiatry* **167**:51-57.

120. McMahon FJ, Stine OC, Meyers DA, Simpson SG, DePaulo JR (1995). Patterns of maternal transmission in bipolar affective disorder. *American Journal of Human Genetics* **56**:1277-1286.

121. McMahon FJ, Hopkins PJ, Xu J, McInnis MG, Shaw S, Cardon L, Simpson SG, MacKinnon DF, Stine OC, Sherrington R, Meyers DA, DePaulo JR (1997). Linkage of bipolar affective disorder to chromosome 18 markers in a new pedigree series. *American Journal of Human Genetics* **61**(6):1397-1404

122. McMahon FJ, Simpson SG, McInnis M, Badner JA, MacKinnon DF, DePaulo JR (2001). Linkage of bipolar disorder to chromosome 18q and the validity of bipolar II disorder. *Archives of Generl Psychiatry* **58**:1025-1031.

123. Mendelbaum K, Sevy S, Souery D, Papadimitriou GN, De Bruyn A, Raeymaekers P, Van Broeckhoven C, Mendlewicz J (1995). *Neuropsychobiology* **31**:58-63.

124. Mendlewicz J, Fleiss JL, Fieve RR (1972). Evidence for X-Linkage in the Transmission of Manic-Depressive Illness. *Journal of the American Medical Association* **222**(13):1624-1627.

125. Mendlewicz J, Fleiss JL (1974). Linkage Studies with X-Chromosome Markers in Bipolar (Manic-Depressive) and Unipolar (Depressive) Illnesses. *Biological Psychiatry* **9**(3):261-294.

126. Mendlewicz J, Rainer JD (1977). Adoption Study supporting Genetic Transmission in Manic-Depressive Illness. *Nature* **268**: 327-329.

127. Mendlewicz J, Linkowski P, Wilmotte J (1980). Linkage between Glucose-6-Phosphate Dehydrogenase Deficiency and Manic-Depressive Psychosis. *British Journal of Psychiatry* **137**:337-342.

128. Mendlewicz J, Leboyer M, De Bruyn A, Malafosse A, Sevy S, Hirsch D, Van Broeckhoven C, Mallet J (1991). Absence of linkage between chromosome 11p15 markers and manic-depressive illness in a Belgian pedigree. *American Journal of Psychiatry* **148**(12):1683:1687.

129. Mendlewicz J. Molecular Genetic Studies in Affective Illness. *in* "Genetic Studies in Affective Disorders", Papolos DF, Lackman HM (eds.), Wiley-Interscience, New York, 1994.

130. Merikangas KR (1982). Assortative Mating for Psychiatric Disorder and Psychological Traits. *Archives of General Psychiatry* **39**:1173-1180.

131. Mirow AL, Kristbjanarson H, Egeland JA, Shilling P, Helgason T, Gillin JC, Hirsch S, Kelsoe JR (1994). A linkage study of distal chromosome 5q and bipolar disorder. *Biological Psychiatry* **36**:223-229.

132. Mitchell P, Selbie L, Waters B, Donald J, Vivero C, Tully M, Shine J (1992). Exclusion of close linkage of bipolar disorder to dopamine D_1 and D_2 receptor gene markers. *Journal of Affective Disorders* **25**:1-12.

133. Mitchell P, MacKinnon A, Waters B (1993). The genetics of bipolar disorder. *Australia and New Zealand Journal of Psychiatry* **27**:560-580

134. Mitterauer B (2000). Clock genes, feedback loops and their possible role in the etiology of bipolar disorders: an intergrative model. *Medical Hypotheses* **55**(2):155-159.

135. Morissette J, Villeneuve A, Bordeleau L, Rochette D, Laberge C, Gagne B, Laprise C, Bouchard G, Plante M, Gobeil L, Shink E, Weissenbach J, Barden N (1999). Genome-wide search for linkage of bipolar affective disorders in a very large pedigree derived from a homogeneous population in Quebec points to a locus of major effect on chromosome 12q23-q24. *American Journal of Medical Genetics* **88**(5):567-587

136. Morton NE (1955). Sequential tests for the detection of linkage. *American Journal of Human Genetics* 7:277-318.

137. Murphy VE, Mynett-Johnson LA, Claffey E, Bergin P, McAuliffe M, Kealey C, McKeon P (2000). Search for bipolar disorder susceptibility loci: the application of a modified genome scan concentrating on gene-rich regions. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* **96**(6):728-732.

138. Myles-Worsley M, Coon H, McDowell J, Brenner C, Hoff M, Lind B, Bennett P, Freedman R, Clementz B, Byerley W (1999). Linkage of a composite inhibitory phenotype to a chromosome 22q locus in eight Utah families. *American Journal of Medical Genetics* **88**(5):544-550.

139. Mynett-Johnson LA, Murphy VE, Claffey E, Shields DC, McKeon P (1998). Preliminary evidence of an association between bipolar disorder in females and the catechol-O-methyltransferase gene. *Psychiatric Genetics* **8**:221-225.

140. Negrão AB (1997). Mood disorders across the continents. *Molecular Psychiatry* **2**(6):439-441.

141. Nelson E, Rice J, Rochberg N, Endicott J, Coryell W, Akiskal HS (1995). Affective Illness in Family Members and Matched Controls. *Acta Psychiatrica Scandinavica* **91**:146-151.

142. Neves-Pereira M, Mundo E, Muglia P, King N, Macciardi F, Kennedy JL (2002). The brain-derived neurotrophic factor gene confers susceptibility to bipolar disorder: evidence from a family-based association study. *American Journal of Human Genetics* **71**(3):651-655.

143. NIMH Genetics Initiative Bipolar Group: Nurnberger JI Jr. (1997). Genomic survey of bipolar illness in the NIMH genetics Initiative pedigrees: A preliminary report. *American Journal of Medical Genetics* **74**(3):227-237

144. Nonaka S, Hough CJ, Chuang D-M (1998). Chronic lithium treatment robustly protects neurons in the central nervous system against excitotoxicity by inhibiting *N*-methyl-D-aspartate receptor-mediated calcium influx. *Proceedings of the National Academy of Science USA* **95**:2642-2647.

145. North BV, Curtis D, Sham PC (2002). A note on the calculation of empirical p-values from Monte Carlo procedures. *American Journal of Human Genetics* **71**:439-441.

146. Nöthen MM, Cichon S, Rohleder H, Hemmer S, Franzek E, Fritze J, Albus M, Borrmann-Hassenbach M, Kreiner R, Weigelt B, Craddock N, Fimmers R, Höller T, Baur MP, Rietschel M, Propping P (1999). Evaluation of linkage of bipolar affective disorder to chromosome 18 in a sample of 57 German families. *Molecular Psychiatry* **4**:76-84.

147. Nyholt DR (2000). All LODs are not created equal. *American Journal of Human Genetics* **67**: 282-288.

148. O'Connell JR, Weeks DE (1998). PedCheck: a program for identifying marker typing incompatibilities in linkage analysis. *American Journal of Human Genetics* **10**, 259-266.

149. O'Mahony E, Corvin A, O'Connell R, Comerford C, Larsen B, Jones IR, McCandless F, Kirov G, Cardno AG, Craddock N, Gill M (2002). Sibling pairs with affective disorders: resemblance of demographic and clinical features. *Psychological Medicine* **32**(1):55-61.

150. O'Rourke DH, McGuffin P, Reich T (1983). Genetic Analysis of Manic-Depressive Illness. *American Journal of Physical Anthropology* **62**:51-59.

151. Ophoff RS, Escamilla MA, Service SK, Spesny M, Meshi DB, Poon W, Molina J, Fournier E, Gallegos A, Mathews C, Neylan T, Batki SL, Roche E, Ramirez M, Silva S, De Mille MC, Dong P, Leon PE, Reus VI, Sandkuijl LA, Freimer NB (2002). Genomewide linkage disequilibrium mapping of severe bipolar disorder in a population isolate. *American Journal of Human Genetics* **71**:565-574.

152. Ott J (1990). Invited Editorial: Cutting a Gordian knot in the linkage analysis of complex human traits. *American Journal of Human Genetics* **46**:219-221.

153. Ott J. "*Analysis of Human Genetic Linkage*", 3rd edition. Johns Hopkins University Press, Baltimore, 1999.

154. Owen MJ, Cardino AG, O'Donovan MC (2000). Psychiatric genetics: back to the future. *Molecular Psychiatry* **5**:22-31.

155. Paterson AD, Petronis A (1999a). Transmission ratio distortion in females on chromosome 10p11-p15. *American Journal of Medical Genetics* **88**:657-661.

156. Paterson AD, Petronis A (1999b). Sex based linkage analysis of alcoholism. *Genetic Epidemiology* **17**(Suppl 1):S289-294.

157. Pauls DL, Gerhard DS, Lacy LG, Hostetter AM, Allen CR, Bland SD, LaBuda MC, Egeland JA (1991). Linkage of bipolar affective disorders to markers on

chromosome 11p is excluded in a second lateral extension of Amish pedigree 110. *Genomics* **11**(3):730-736.

158. Pauls DL, Ott J, Paul SM, Allen CR, Fann CSJ, Carulli JP, Falls KM, Bouthillier CA, Gravius TC, Keith TP, Egeland JA, Ginns EI (1995). Linkage analysis of chromosome 18 markers do not identify a major susceptibility locus for bipolar affective disorder in the Old Order Amish. *American Journal of Human Genetics* **57**:636-643.

159. Pekkarinen P, Terwilliger J, Bredbacka PE, Lonnqvist J, Peltonen L (1995). Evidence of a predisposing locus to bipolar disorder on Xq24-q27.1 in an extended Finnish pedigree. *Genome Research* **5**(2):105-115.

160. Penrose LS (1953). The general purpose sib-pair linkage test. *Annals of Eugenics* **18**:120-124

161. Pericak-Vance MA, Bebout JL, Gaskell PC, Yamaoka LA, Hung W-Y, Alberts MJ, Walker AP, Bartlett RJ, Haynes CS, Welsh KA, Earl NL, Heyman A, Clark CM, Roses AD (1991). Linkage studies in familial Alzheimer's disease: evidence for chromosome 19 linkage. *American Journal of Human Genetics* **48**:10344-1050.

162. Perris C (1966).A stufy of bipolar (manic-depressive) and unipolar recurrent depressive psychoses. *Acta Psychiatrica Scandinavica* **42**(Suppl 194):1-188.

163. Petrie WM, Maffucci RJ, Woosley RL (1982). Propanolol and Depression. *American Journal of Psychiatry* **139**:92-94.

164. Pollin W, Allen M, Hoffer A, Stabenau JR, Hrubec Z (1969). Psychopathology in 15,909 pairs of veteran twins: Evidence for a Genetic Factor in the Pathogenesis of Schizophrenia and it's Relative Absence in Psychoneurosis. *American Journal of Psychiatry* **126**(5):597-610.

165. Polymeropoulos MH, Xiao H, Rath DS, Merril CR (1991). Tetranucleotide repeat polymorphism at the human tyrosine hydroxylase gene (TH). *Nucleic Acids Research* **19**(13):3753.

166. Polymeropoulos MH, Schaffer AA (1996). Scanning the genome with 1772 microsatellite markers in search of a bipolar disorder susceptibility gene. *Molecular Psychiatry* **1**(5):404-407

167. Popoli M, Brunello N, Perez J, Racagni G (2000). Second messenger-regulated protein kinases in the brain: their functional role and the action of antidepressant drugs. *Journal of Neurochemistry* **74**:21-33.

168. Post RM, Roy-Byrne PP, Uhde TW (1988). Graphic representation of the life course of illness in patients with affective disorder. *American Journal of Psychiatry* **145**(7):844-848.

169. Preisig M, Bellivier F, Fenton BT, Baud P, Berney A, Courtet P, Hardy P, Golaz J, Leboyer M, Mallet J, Matthey ML, Mouthon D, Neidhart E, Nosten-Bertrand M, Stadelmann-Dubuis E, Guimon J, Ferrero F, Buresi C, Malafosse A (2000). Association between bipolar disorder and monoamine oxidase A gene polymorphisms: results of a multicenter study. *American Journal of Psychiatry* **157**:948-955.

170. Pulver AE, Karayiorgou M, Wolyniec PS, Lasseter VK, Kasch L, Nestadt G, Antonarakis S, Housman D, Kazazian HH, Meyers D, Nestadt G, Ott J, Lamacz M, Liang K-L, Hanfelt J, Ullrich G, DeMarchi N, Ramu E, McHugh PR, Adler L, Thomas M, Carpender WT, Manschreck T, Gordon CT, Kimberland M, Babb R, Puck J, Childs B (1994). Sequential strategy to identify a susceptibility locus for schizophrenia: report of a potential linkage on chromosome 22q12-q13.1: Part 1. *American Journal of Medical Genetics* **54**(1):36-43.

171. Radhakrishna U, Senol S, Herken H, Gucuyener K, Gehrig C, Blouin JL, Akarsu NA, Antonarakis SE (2001). An apparently dominant bipolar affective disorder (BPAD) locus on chromosome 20p11.2-q11.2 in a large Turkish pedigree. *European Journal of Human Genetics* **9**(1):39-44

172. Reus VI, Freimer NB (1997). Understanding the Genetic Basis of Mood Disorders: Where Do We Stand? *American Journal of Human Genetics* **60**:1283-1288.

173. Rice J, Reich T, Andreasen NC, Endicott J, Van Eerdewegh M, Fishman R, Hirschfeld RMA, Klerman GL (1987). The Familial Transmission of Bipolar Illness. *Archives of General Psychiatry* **44**:441-447.

174. Rice JP, Goate A, Williams JT, Bierut L, Dorr D, Wu W, Shears S, Gopalakrishnan G, Edenberg HJ, Foroud T, Nurnberger J Jr, Gershon ES, Detera-Wadleigh SD, Goldin LR, Guroff JJ, McMahon FJ, Simpson S, MacKinnon D, McInnis M, Stine OC, DePaulo JR, Blehar MC, Reich T (1997). Initial genome scan of the NIMH Genetics Initiative bipolar pedigrees: Chromosomes 1, 6, 8, 10, and 12. *American Journal of Medical Genetics* **74**(3):247-253

175. Risch N, Baron M, Mendlewicz J (1986). Assessing the Role of X-Linked Inheritance in Bipolar-Related Major Affective Disorder. *Journal of Psychiatric Research* **20**(4):275-288.

176. Risch N (1990a). Linkage strategies for genetically complex traits. I Multilocus models. *American Journal of Human Genetics* **46**:222-228

177. Risch N (1990b). Linkage strategies for genetically complex traits. II The power of affected relative pairs. *American Journal of Human Genetics* **46**:229-241

178. Risch N (1990c). Linkage strategies for genetically complex traits. III The effect of marker polymorphism on analysis of affected relative pairs. *American Journal of Human Genetics* **46**:242-253

179. Risch N, Merikangas K (1996). The future of genetic studies of complex human diseases. *Science* **273**:1516-1517.

180. Roberts SB, MacLean CJ, Neale MC, Eaves LJ, Kendler KS (1999). Replication of linkage studies of complex traits: an examination of variation in location estimates. *American Journal of Human Genetics* **65**(3):876-884

181. Robins LN, Helzer JE, Weissman MM, Orvaschel H, Gruenberg E, Burke JD Jr, Regier DA (1984). Lifetime Prevalence of Specific Psychiatric Disorders in Three Sites. *Archives of General Psychiatry* **41**: 949-958.

182. Rogers MP, Whybrow PC (1971). Clinical hypothyroidism occurring during lithium treatment: two case histories and a review of thyroid function in 19 patients. *American Journal of Psychiatry* **128**:158-163.

183. Rojas K, Liang L, Johnson EI, Berrettini WH, Overhauser J (2000). Identification of candidate genes for psychiatric disorders on 18p11. *Molecular Psychiatry* **5**:389-395

184. Rosanoff AJ, Handy L, Plesset IR (1935). The Etiology of Manic-Depressive Syndromes with Special References to their Occurrence in Twins. *American Journal of Psychiatry* **91**:725-762.

185. Ross CA, McInnis MG, Margolis RL, Li S-H (1993). Genes with Triplet Repeats: Candidate Mediators of Neuropsychiatric Disorders. *Trends in Neurosciences* **16**(7):254-260.Sacher EJ, Asnis G, Nathan S, Halbreich U, Tabrizi MA, Halpern FS (1980). Dextroamphetamine and cortisol in depression. *Archives of General Psychiatry* **37**;755-757.

186. Sachar EJ, Asnis G, Nathan RS, Halbreich U, Tabrizi MA, Halpern FS (1980). Dextroamphetamine and cortisol in depression. *Archives of General Psychiatry* **37**:755-757.

187. Saito T, Papolos DF, Chernak D, Rapaport MH, Kelsoe JR, Lachman HM (1999). Analysis of GNAZ gene polymorphism in bipolar affective disorder. *American Journal of medical Genetics (Neuropsychiatric Genetics)* **88**:324-328.

188. Sambrook J, Fritsch EF, Maniatis T "*Molecular Cloning: a laboratory manual*". Cold Spring Harbor Press, New York, 1982.

189. Schalling M, Hudson TJ, Buetow KH, Housman DE (1993). Direct Detection of Novel Expanded Trinucleotide Repeats in the Human Genome. *Nature Genetics* **4**:135-139.

190. Schildkraut JJ (1965). The catecholamine hypothesis of affective disorder: a review of supporting evidence. *American Journal of Psychiatry* **122**: 509-522.

191. Schildkraut JM. Examining Complex Genetic Interactions; in "*Approaches to Gene Mapping in Complex Human Diseases*", Haines JL, Pericak-Vance MA, eds. Wiley-Liss, New York, 1998; pp 379-410.

192. Schwab SG, Hallmayer J, Lerer B, Albus M, Borrmann M, Hönig S, Strauβ M, Segman R, Lichtermann D, Knapp M, Trixler M, Maier W, Wildenauer DB (1998). Support for a chromosome 18p locus conferring susceptibility to functional psychoses in families with schizophrenia, by association and linkage analysis. *American Journal of Human Genetics* **63**(4):1139-1152.

193. Segurado R, Gill M (2001). "Genome-wide genetic linkage studies in bipolar disorder: a review": presentation at the IXth World Congress of Psychaitric Genetics. *American Journal of Medical Genetics* **105**(7):O60.

194. Serretti A, Macciardi F, Cusin C, Verga M, Pedrini S, Smeraldi E (1998a). Tyrosine hydroxylase gene in linkage disequilibrium with mood disorders. *Molecular Psychiatry* **3**(2): 169-74.

195. Serretti A, Macciardi F, Verga M, Cusin C, Pedrini S, Smeraldi E (1998b). Tyrosine hydroxylase gene associated with depressive symptomatology in mood disorder. *American Journal of Medical Genetics* **81**(2): 127-30.

196. Sham PC, Curtis D (1995). An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Annals of Human Genetics* **59**:323-336.

197. Sherman WR, Gish BG, Honchar MP, Munsell LY (1986). Effects of lithium on phosphoinositide metabolism in vivo. *Federation Proceedings* **45**:2639-2646.

198. Shprintzen RJ, Goldberg RB, Golding-Kushner KJ, Marion RW (1992). Late-onset psychosis in the velo-cardio-facial syndrome. *American Journal of Medical Genetics* **42**:141-142.

199. Singer I, Rotenberg D (1973). Mechanisms of lithium action. *The New England Journal of Medicine* **289**:254-260.

200. Sjoholt G, Gulbrandsen AK, Lovlie R, Berle JO, Molven A, Steen VM (2000). A human myo-inositol monophosphatase gene (IMPA2) localized in a putative susceptibility region for bipolar disorder on chromosome 18p11.2: genomic structure and polymorphism screening in manic-depressive patients. *Molecular Psychiatry* **5**(2):172-180.

201. Sklar P (2002). Linkage Analysis in Psychiatric Disorders: the emerging picture. *Annual Review of Genomics and Human Genetics* **3**:371-413.

202. Sklar P, Gabriel SB, McInnis MG, Bennett P, Lim YM, Tsan G, Schaffner S, Kirov G, Jones I, Owen M, Craddock N, DePaulo JR, Lander ES (2002).Family-based association study of 76 candidate genes in bipolar disorder: BDNF is a potential risk locus. *Molecular Psychiatry* **7**(6):579-593.

203. Slater E. Psychotic and Neurotic Illness in Twins. *Medical Research Council Special Report Series* **No.278.** London, Her Majesty's Stationary Office, 1953.

204. Smyth C, Kalsi G, Brynjolfsson J, O'Neill J, Curtis D, Rifkin L, Moloney E, Murphy P, Sherrington R, Petursson H, Gurling H (1996). Further tests for linkage of bipolar affective disorder to the tyrosine hydroxylase gene locus on chromosome 11p15 in a new series of multiplex British affective disorder pedigrees. *American Journal of Psychiatry* **153**(2):271-274.

205. Smyth C, Kalsi G, Brynolfsson J, O'Neill J, Curtis D, Rifkin L, Moloney E, Murphy P, Petursson H, Gurling H (1997). Test of Xq26.3-28 Linkage in Bipolar and Unipolar Affective Disorder in Families Selected for Absence of Male to Male Transmission. *British Journal of Psychiatry* **171**:578-581.

206. Souery D, Lipp O, Mahieu B, Mendelbaum K, De Martelaer V, Van Broeckhoven C, Mendlewicz J (1996). Association study of bipolar disorder with candidate genes involved in catecholamine neurotransmission: DRD2, DRD33, DAT1, and TH genes. *American Journal of Medical Genetics* **67**(6):551-555.

207. Souery D, Lipp O, Rivelli SK, Massat I, Serretti A, Cavallini C, Ackenheil M, Adolfsson R, Aschauer H, Blackwood D, Dam H, Dikeos D, Fuchshuber S, Heiden M, Jakovljevic M, Kaneva R, Kessing L, Lerer B, Lonnqvist J, Mellerup T, Milanova V, Muir W, Nylander PO, Oruc L, Mendlewicz J, et al. (1999). Tyrosine hydroxylase polymorphism and phenotypic heterogeneity in bipolar affective disorder: a multicenter association study. *American Journal of Medical Genetics* **88**(5): 527-32.

208. Spence MA, Flodman PL, Sadovnick AD, Bailey-Wilson JE, Ameli H, Remick RA (1995). Bipolar disorder: Evidence for a Major Locus. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* **60**:370-376.

209. Spielman RS, McGinnis RE, Ewens WJ (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal of Human Genetics* **52**: 506-516.

210. Spielman RS, Ewens WJ (1998). A Sibship Test for Linkage in the Presence of Association: The Sib Transmission/Disequilibrium Test. *American Journal of Human Genetics* **62**:450-458.

211. Stine OC, Xu J, Koskela R, McMahon FJ, Gschwend M, Friddle C, Clark CD, McInnis MG, Simpson SG, Breschel TS, Vishio E, Riskin K, Feilotter H, Chen E, Shen S, Folstein S, Meyers DA, Botstein D, Marr TG, DePaulo JR (1995). Evidence for linkage of bipolar disorder to chromosome 18 with a parent-of-origin effect. *American Journal of Human Genetics* **57**:1384-1394

212. Stine OC, McMahon FJ, Chen L, Xu J, Meyers DA, MacKinnon DF, Simpson S, McInnis MG, Rice JP, Goate A, Reich T, Edenberg HJ, Foroud T, Nurnberger JI Jr, Detera-Wadleigh SD, Goldin LR, Guroff J, Gershon ES, Blehar MC, DePaulo JR (1997). Initial genome scan for bipolar disorder in the NIMH Genetics Initiative pedigrees: Chromosomes 2, 11, 13, 14, and X. *American Journal of Medical Genetics* **74**(3):263-269

213. Straub RE, Lehner T, Luo Y, Loth JE, Shao W, Sharpe L, Alexander JR, Das K, Simon R, Fieve RR, Lerer B, Endicott J, Ott J, Gilliam C, Baron M (1994). A possible vulnerability locus for bipolar affective disorder on chromosome 21q22.3. *Nature Genetics* **8**(3):291-296

214. Straub RE, MacLean CJ, O'Neill FA, Burke J, Murphy B, Duke F, Shinkwin R, Webb BT, Zhang J, Walsh D, Kendler KK (1995). A potential vulnerability locus for schizophrenia on chromosome 6p24-22 and evidence for genetic heterogeneity. *Nature Genetics* **11**:287-293.

215. Straub RE, MacLean CJ, Martin RB, Ma Y, Myakishev MV, Harris-Kerr C, Webb BT, O'Neill FA, Walsh D, Kendler KS (1998). A schizophrenia locus may be located in region 10p15-p11. *American Journal of Medical Genetics* **81**:296-301.

216. Suzuki K, Kusumi I, Sasaki Y, Koyama T (2001). Serotonin-induced platelet intracellular calcium mobilization in various psychiatric disorders: is it specific to bipolar disorder? *Journal of Affective Disorders* **64**:291-296.

217. Terwilliger JD, Ott J. "*Handbook of Human Genetic Linkage*". Johns Hopkins University Press, Maryland, 1994.

218. Terwilliger JD, Weiss KM (1998). Linkage disequilibrium mapping of complex disease: fantasy or reality? *Current Opinion in Biotechnology* **9**:578-594.

219. Terwilliger JD, Göring HH (2000). A review of gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. *Human Biology* **72**(1):63-132.

220. Torgersen S (1986). Genetic Factors in Moderately Severe and Mild Affective Disorders. *Archives of General Psychiatry* **43**:222-226.

221. Tsiouris SJ, Breschel TS, Xu J, McInnis MG, McMahon FJ (1996). Linkage disequilibrium analysis of G-olf alpha (GNAL) in bipolar affective disorder. *American Journal of Medical Genetics* **67**(5):491-494.

222. Turecki G, Rouleau GA, Mari J, Joober R, Morgan K (1997). Lack of association between bipolar disorder and tyrosine hydroxylase: a meta-analysis. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* **74**:348-352.

223. Turecki G, Grof P, Cavazzoni P, Duffy A, Grof E, Martin R, Joober R, Rouleau GA, Alda M (1999). Lithium responsive bipolar disorder, unilineality, and chromosome 18: a linkage study. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* **88**:411-415.

224. Turecki G, Grof P, Grof E, D'Souza V, Lebuis L, Marineau C, Cavazzoni P, Duffy A, Bétard C, Xvolský P, Robertson C, Brewer C, Hudson TJ, Rouleau GA, Alda M (2001). Mapping susceptibility genes for bipolar disorder: a pharmacogenetic approach based on excellent response to lithium. *Molecular Psychiatry* **6**:570-578.

225. Vallada H, Craddock N, Vasques L, Curtis D, Kirov G, Lauriano V, Gentil V, Passos-Bueno R, Murray RM, Zatz M, McGuffin P, Powell JF, Gill M, Owen M, Collier DA (1996). Linkage Studies in Bipolar Affective Disorder with Markers on Chromosome 21. *Journal of Affective Disorders* **41**:217-221.

226. Vallada HP, Vasques L, Curtis D, Zatz M, Kirov G, Lauriano V, Gentil V, Murray RM, McGuffin P, Owen M, Gill M, Craddock N, Collier DA (1998). Linkage Analysis between Bipolar Affective Disorder and Markers on Chromosome X. *Psychiatric Genetics* **8**:183-186.

227. Van Praag HM. Therapy-Resistant Depression: Biochemical and Pharmacological Considerations. *in "Psychosomatic Medicine"*, Lipowski ZJ, Lipsitt DR, Whybrow PC, eds. Oxford University Press, New York, 1977.

228. Vance JM, Othmane KB. Methods of Genotyping; *in "Approaches to Gene Mapping in Complex Human Diseases"*; Haines JL, Pericak-Vance MA, eds, Wiley-Liss, New York, 1998; pp 213-228.

229. Vawter MP, Freed WJ, Kleinman JE (2000). Neuropathology of bipolar disorder. *Biological Psychiatry* **48**(6):486-504.

230. Verheyen GR, Del-Favera J, Mendlewicz J, Lindblad K, Van Zand K, Aalbregtse M, Schalling M, Souery D, Van Broeckhoven C (1999). Molecular Interpretation of Expanded RED Products in Bipolar Disorder by CAG/CTG Repeats Located at Chromosomes 17q and 18q. *Neurobiology of Disease* **6**:424-432.

231. Vincent JB, Klempan T, Parikh SS, Sasaki T, Meltzer HY, Sirugo G, Cola P, Petronis A, Kennedy JL (1996). Frequency analysis of large CAG/CTG trinucleotide repears in schizophrenia and bipolar affective disorder. *Molecular Psychiatry* **1**:141-148.

232. Vincent JB, Petronis A, Strong E, Parikh SV, Meltzer HY, Lieberman J, Kennedy JL (1999). Analysis of genome-wide CAG/CTG repeats, and at *SEF2-1B*

and *ERDA1* in schiophrenia and bipolar affective disorder. *Molecular Psychiatry* **4**:229-234.

233. Von Knorring AL, Cloninger CR, Bohman M, Sigvardsson S (1983). An Adoption Study of Depressive Disorders and Substance Abuse. *Archives of General Psychiatry* **40**:943-950.

234. Vuoristo JT, Berrettini WH, Overhauser J, Prockop DJ, Ferraro TN, Ala-Kokko L (2000). Sequence and genomic organization of the human G-protein Golfalpha gene (GNAL) on chromosome 18p11, a susceptibility region for bipolar disorder and schizophrenia. *Molecular Psychiatry* **5**(5):495-501.

235. Weber JL, May PE (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* **44**: 388-396.

236. Weissman MM, Gershon ES, Kidd KK, Prusoff BA, Leckman JF, Dibble E, Hamovit J, Thompson WD, Pauls DL, Guroff JJ (1984). Psychiatric Disorders in the Relatives of Probands with Affective Disorders. *Archives of General Psychiatry* **41**:13-21.

237. Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu H-G, Joyce PR, Karam EG, Lee C-K, Lellouch J, Lépine J-P, Newman SC, Rubio-Stipec M, Wells E, Wickramaratne PJ, Wittchen H-U, Yeh E-K (1996). Cross-National Epidemiology of Major Depression and Bipolar disorder. *Journal of the American Medical Association* **276**(4):293-299.

238. Wender PH, Kety SS, Rosenthal D, Schulsinger F, Ortmann J, Lunde I (1986). Psychiatric Disorders in the Biological and Adoptive Families of Adopted Individuals with Affective Disorders. *Archives of General Psychiatry* **43**:923-929.

239. Whittemore AS, Halpern J (1994). A class of tests for linkage using affected pedigree members. *Biometrics* **50**:118-127.

240. Whybrow PC, Akiskal HS, McKinney WT, eds. "*Mood Disorders: Toward a New Psychobiology*"; Plenum Press, New York, 1984.

241. Wildenauer DB, Schwab SG (1999). Chromosomes 8 and 10 workshop. *American Journal of Medical Genetics* **88**(3):239-243.

242. Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, Jablenski A, Regier D, Sartorius N (1990). SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Archives of General Psychiatry* **47**(6):589-593.

243. Winokur G, Coryell W, Endicott J, Akiskal H (1993). Further Distinctions between Manic-Depressive Illness (Bipolar disorder) and Primary Depressive Disorder (Unipolar Depression). *American Journal of Psychiatry* **150**(8):1176-1181.

244. Wise LH, Lanchbury JS, Lewis CM (1999). Meta-analysis of genome searches. *Annals of Human Genetics* **63**:263-272

245. Woolf B (1955). On estimating the relation between blood group and disease. *Annals of Human Genetics* **19**: 251-253.

246. World Health Organisation. "The International Statistical Classification of Diseases and Related Health Problems: Tenth Revision". Geneva, Switzerland, 1992.

247. Woyshville MJ, Lackamp JM, Eisengart JA, Gilliland JAM (1999). On the meaning and measurement of affective instability: clues from chaos theory. *Biological Psychiatry* **45**(3):261-269.

248. Xu J, Meyers DA, Pericak-Vance MA. Lod Score Analysis; *in* "*Approaches to Gene Mapping in Complex Human Diseases*", Haines JL, Pericak-Vance MA, eds.; Wiley-Liss, New York, 1998, pp 253-272.

249.Xue F, Kidd JR, Pakstis AJ, Castiglione CM, Mallet J, Kidd KK (1988). *Genomics* **2**(4):288-293.

Appendix A. Genome Scan Comma Separated Value File

Example of comma-separated value (CSV) file for importation of genotype data into the MegaBase database. Each of the three section of the file shown represents a genotyping gel, headed by its CON value (see text, chapter 2 section 2.5.6). Each gel then contains a line indicating the marker name, the individual identification number in the form FamilyID.PersonalID, and two allele codes in "ABI units", respectively (see chapter 1, section 2.5).

```
CON=80.12
D2S112,50002.01,82.24,85.3
D2S112,50028.01,87.91,87.91
D2S112,50016.02,80,85.72
D2S112,50002.02,82.02,85.72
D2S112,50028.02,82.24,87.77
D2S112,50016.03,76.19,80
D2S112,50002.03,85.72,85.72
D2S112,50007.01,85.88,89.77
D2S112,50016.04,80.24,87.88
D2S112,50002.04,82.24,85.3
D2S112,50007.02,80.24,85.88
D2S112,50018.01,76.19,87.74
D2S112,50002.05,82.12,85.18
D2S112,50007.03,85.72,85.72
D2S112,50018.02,76.07,85.83
D2S112,50002.06,85.72,85.72
D2S112,50007.04,80.12,89.76
D2S112,50018.03,76.35,85.88
D2S112,50003.01,85.77,87.77
D2S112,50007.05,80.12,85.83
D2S112,50003.02,80.12,87.74
```

D2S112,50010.01,80.12,82.02

D2S112,50004.01,83.81,87.74

D2S112,50010.02,80.12,82.02

D2S112,50004.02,76.31,87.74

D2S112,50011.01,80,89.64

D2S112,50011.02,76.19,89.76

D2S112,50054.01,80.24,85.77

D2S112,50013.01,85.72,87.62

D2S112,50054.02,80.12,85.77

D2S112,50013.02,80.24,85.77

D2S112,50022.01,76.35,80.24

D2S112,50013.03,85.88,85.88

D2S112,50022.02,76.24,89.65

D2S112,50540.02,80.24,87.77

D2S112,50013.04,80,85.72

D2S112,50022.03,80.12,85.77

D2S112,50540.03,87.68,87.68

D2S112,50016.01,76.51,87.79

D2S112,ceph,80.12,85.77

CON=80.24

D2S112,50022.04,80,89.64

D2S112,50031.03,85.88,87.77

D2S112,50038.02,80,89.64

D2S112,50022.05,76.19,89.76

D2S112,50031.04,85.72,85.72

D2S112,50039.01,85.72,85.72

D2S112,50022.06,80.12,85.72

D2S112,50032.01,85.95,89.88

D2S112,50039.02,85.88,87.88

D2S112,50024.01,76.43,85.83

D2S112,50032.02,85.83,89.76

D2S112,50039.03,85.3,87.77

D2S112,50024.02,76.31,85.83

D2S112,50032.03,87.74,89.64

D2S112,50039.04,85.83,85.83

D2S112,50025.01,80.12,85.72

D2S112,50034.01,80,85.72

D2S112,50054.03,83.77,83.77

D2S112,50025.02,80.12,85.83

D2S112,50034.02,80.24,85.88

D2S112,50054.04,80.12,83.77

D2S112,50025.03,80.12,80.12

D2S112,50036.01,82.12,85.88

D2S112,50054.05,83.95,83.95

D2S112,50026.01,85.77,85.77

D2S112,50036.02,87.91,87.91

D2S112,50042.01,80.12,87.77

D2S112,50026.02,85.72,87.74

D2S112,50036.03,82.33,88.02

D2S112,50042.02,76.24,85.88

D2S112,50036.04,87.77,87.77

D2S112,50042.03,85.77,87.77

D2S112,50029.02,85.72,85.72

D2S112,50037.01,85.88,87.77

D2S112,50042.04,76.24,80.12

D2S112,50030.01,80.24,85.77

D2S112,50037.02,85.72,87.62

D2S112,50043.01,83.88,85.88

D2S112,50030.02,76.35,85.77

D2S112,50037.03,85.77,87.77

D2S112,50043.02,76.24,83.77

D2S112,50031.01,85.72,87.62

D2S112,50031.02,85.72,85.72

D2S112,50038.01,76.07,87.62

CON=80.32

D2S112,50049.03,82.09,85.61

D2S112,50504.02,76.52,83.81

D2S112,50045.02,76.11,80

D2S112,50050.01,76.48,76.48

D2S112,50504.03,85.65,85.65

D2S112,50050.02,76.48,76.48

D2S112,50504.04,83.81,85.76

D2S112,50045.04,76.52,85.87

D2S112,50050.03,76.37,87.8

D2S112,50505.01,76.41,87.72

D2S112,50046.01,77.11,86.67

D2S112,50051.01,85.87,85.87

D2S112,50505.02,76.41,87.72

D2S112,50046.02,76.37,85.83

D2S112,50051.02,77.39,86.74

D2S112,50505.03,76.41,80.22

D2S112,50046.03,78.06,86.84

D2S112,50500.01,76.37,80.11

D2S112,50507.01,83.91,85.87

D2S112,50500.02,80.33,85.87

D2S112,50507.02,85.91,87.85

D2S112,50047.01,76.48,86.37

D2S112,50500.03,76.41,85.87

D2S112,50507.03,83.98,85.91

D2S112,50047.02,80.11,85.72

D2S112,50500.04,80.22,85.87

D2S112,50507.04,83.98,85.91

D2S112,50047.03,85.72,85.72

D2S112,50500.05,80.22,85.87

D2S112,50507.05,85.81,87.85

D2S112,50048.01,80.33,80.33

D2S112,50500.06,80.33,80.33

D2S112,50507.06,85.76,85.76

D2S112,50048.02,80.33,80.33

D2S112,50502.01,80.22,87.83

D2S112,50508.01,76.2,85.65

D2S112,50048.03,80.33,80.33

D2S112,50502.02,80.22,87.72

D2S112,50508.02,76.56,85.91

D2S112,50049.01,82.17,85.87

D2S112,50502.03,85.76,87.72

D2S112,50508.03,76.31,85.65

D2S112,50049.02,80.33,82.17

D2S112,50504.01,76.31,85.76

Appendix B. Genome Scan Linkage Analysis Batch Files

1. SPLINK batch files

Example of SPLINK batch files for analysis of chromosome 9 under Model 1 (see chapter 2, section 2.6.6); files were converted to UNIX executable files using the *chmod* command, and run at the command line prompt. The first file, 9m1.batch, creates SPLINK input files for each marker from the full-chromosome LINKAGE format pedigree file.

```
ped2spl n sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn <9m1.ped >288.spl
ped2spl sn n sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn <9m1.ped >286.spl
ped2spl sn sn n sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn <9m1.ped >285.spl
ped2spl sn sn sn n sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn <9m1.ped >157.spl
ped2spl sn sn sn sn n sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn <9m1.ped >171.spl
ped2spl sn sn sn sn sn n sn sn sn sn sn sn sn sn sn sn sn sn sn sn <9m1.ped >161.spl
ped2spl sn sn sn sn sn sn n sn sn sn sn sn sn sn sn sn sn sn sn sn <9m1.ped >1817.spl
ped2spl sn sn sn sn sn sn sn n sn sn sn sn sn sn sn sn sn sn sn sn <9m1.ped >273.spl
ped2spl sn sn sn sn sn sn sn sn n sn sn sn sn sn sn sn sn sn sn sn <9m1.ped >175.spl
ped2spl sn sn sn sn sn sn sn sn sn n sn sn sn sn sn sn sn sn sn sn <9m1.ped >167.spl
ped2spl sn sn sn sn sn sn sn sn sn sn n sn sn sn sn sn sn sn sn sn <9m1.ped >283.spl
ped2spl sn sn sn sn sn sn sn sn sn sn sn n sn sn sn sn sn sn sn sn <9m1.ped >287.spl
ped2spl sn sn sn sn sn sn sn sn sn sn sn sn n sn sn sn sn sn sn sn <9m1.ped >1690.spl
ped2spl sn sn sn sn sn sn sn sn sn sn sn sn sn n sn sn sn sn sn sn <9m1.ped >1677.spl
ped2spl sn sn sn sn sn sn sn sn sn sn sn sn sn sn n sn sn sn sn sn <9m1.ped >1776.spl
ped2spl sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn n sn sn sn sn <9m1.ped >1682.spl
ped2spl sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn n sn sn sn <9m1.ped >290.spl
ped2spl sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn n sn sn <9m1.ped >164.spl
ped2spl sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn n sn <9m1.ped >1818.spl
ped2spl sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn n <9m1.ped >1826.spl
ped2spl sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn sn n <9m1.ped >158.spl
```

The second file, s9m1.batch, inputs each marker file into SPLINK, using the –d flag to output a file containing the allele frequencies (xxxx.frq), and outputting the results of the single-point linkage analysis (c9m1_x.txt).

```
splink -d288.frq <288.spl >c9m1_1.txt
splink -d286.frq <286.spl >c9m1_2.txt
splink -d285.frq <285.spl >c9m1_3.txt
splink -d157.frq <157.spl >c9m1_4.txt
splink -d171.frq <171.spl >c9m1_5.txt
splink -d161.frq <161.spl >c9m1_6.txt
splink -d1817.frq <1817.spl >c9m1_7.txt
splink -d273.frq <273.spl >c9m1_8.txt
splink -d175.frq <175.spl >c9m1_9.txt
splink -d167.frq <167.spl >c9m1_10.txt
splink -d283.frq <283.spl >c9m1_11.txt
splink -d287.frq <287.spl >c9m1_12.txt
splink -d1690.frq <1690.spl >c9m1_13.txt
splink -d1677.frq <1677.spl >c9m1_14.txt
splink -d1776.frq <1776.spl >c9m1_15.txt
splink -d1682.frq <1682.spl >c9m1_16.txt
splink -d290.frq <290.spl >c9m1_17.txt
splink -d164.frq <164.spl >c9m1_18.txt
splink -d1818.frq <1818.spl >c9m1_19.txt
splink -d1826.frq <1826.spl >c9m1_20.txt
splink -d158.frq <158.spl >c9m1_21.txt
```

2. GENEHUNTER batch file

Example of Genehunter batch file for analysis of chromosome 9 under all five disease models. The file was run from the Genehunter prompt using the *run* command. It consists of a preparation section to *load* the LINKAGE format data file (9.dat) and request analysis 10cM past the edges of the genetic map with *off end*. For each model it then inputs the LINKAGE format pedigree file (9mX.ped) with *scan*, turns on the postscript graph output with *ps*, and text output with *photo* directed to an output file (9mX.gho) and initiates analysis with the *total stat het* command. It specifies the postscript output graph files for the NPL score (9mXnpl.ps), the lod score (9mXlod.ps) and information content (9mXinfo.ps). Finally it terminates output before commencing input of the pedigree file for the next disease model.

File 9gh.batch

```
load 9.dat
off end 10
scan 9m1.ped
ps on
photo 9m1.gho
total stat het
9m1npl.ps
9m1lod.ps
9m1info.ps
photo off
ps off
scan 9m2.ped
ps on
photo 9m2.gho
total stat het
9m2npl.ps
9m2lod.ps
9m2info.ps
photo off
ps off
scan 9m3.ped
ps on
photo 9m3.gho
total stat het
9m3npl.ps
9m3lod.ps
9m3info.ps
photo off
ps off
scan 9m4.ped
ps on
photo 9m4.gho
total stat het
9m4npl.ps
9m4lod.ps
9m4info.ps
```

```
photo off
ps off
scan 9m5.ped
ps on
photo 9m5.gho
total stat het
9m5npl.ps
9m5lod.ps
9m5info.ps
photo off
ps off
quit
```

Appendix C. Linkage Mapping Set v.2

Perkin Elmer Applied Biosciences International Linkage Mapping Set, version 2 (LMSv2), now renamed LMS v2.5 Medium Density map (MD-10). Shown are the panels for the 12 Dublin-genotyped chromosomes:

2 (panels 3 and 4); 5 and 6 (panels 8, 9 and 10);

9, 10 and 11 (panels 13, 14, 15 and 16);

17 and 18 (panels 23 and 24);

19, 20 21 and 22 (panels 25, 26 and 17).

The labelling dye (FAM, HEX or NED), the allele size range (ASR) and the genotype of CEPH individual 1347.02 (CEPH gt) are indicated for each microsatellite marker. Markers genotyped by the author are indicated by one (first stage genome scan) or two (second stage) asterisks. The full LMSv2 mapping set can be obtained from http://www.appliedbiosystems.com/.

Panel 3				
Locus	Dye Label	Het	ASR	CEPH gt
D2S286	FAM	0.66	85-107	87,97
D2S265	FAM	0.85	145-177	156,162
D2S261	FAM	0.78	211-227	216,218
D2S2211	FAM	0.74	241-261	248,250
D2S367	FAM	0.86	309-341	320,314
D2S125	HEX	0.82	92-112	99,101
D2S206	HEX	0.80	130-166	153,155
D2S117	HEX	0.82	193-221	194,212
D2S142	HEX	0.76	239-259	242,246
D2S2333	NED	0.82	84-106	98,104
D2S126	NED	0.82	119-149	138,142
D2S325	NED	0.82	158-186	161,171
D2S364	NED	0.80	234-260	138,242
D2S337	NED	0.88	295-319	309,311

Panel 4				
Locus	Dye Label	Het	ASR	CEPH gt
D2S112	FAM	0.71	80-94	81,87
D2S162	FAM	0.75	124-152	125,137
D2S2330	FAM	0.81	171-189	172,180
D2S2216	FAM	0.76	213-227	213,223
D2S347	FAM	0.80	271-301	292,292
D2S2259	FAM	0.79	322-342	322,340
D2S319	HEX	0.73	132-144	134,136
D2S168	HEX	0.82	159-183	165,177
D2S151	HEX	0.82	228-256	246,250
D2S2382	HEX	0.81	299-337	310,318
D2S2368	NED	0.83	97-121	108,110
D2S391	NED	0.79	147-161	154,156
D2S335	NED	0.79	189-209	198,202
D2S396	NED	0.83	236-254	241,245
D2S338	NED	0.81	270-294	280,280
D2S305	NED	0.72	316-338	324,338

Panel 8					
Locus	Dye Label	Het	ASR	CEPH gt	
D5S407	FAM	0.86	87-115	105,107	*
D6S289	FAM	0.79	163-185	173,175	*
D6S1610	FAM	0.84	204-218	208,212	*
D6S1581	FAM	0.72	261-275	265,265	*
D6S422	FAM	0.77	300-322	306,308	*
D5S644	HEX	0.85	86-116	88,100	*
D6S281	HEX	0.68	136-156	140,140	*
D6S262	HEX	0.82	171-191	173,175	*
D5S424	HEX	0.76	215-237	219,221	*
D5S419	HEX	0.81	258-290	274,280	*
D5S433	NED	0.86	69-99	83,91	*
D5S422	NED	0.84	118-140	120,136	*
D5S406	NED	0.79	168-196	180,192	*
D5S400	NED	0.82	221-243	227,231	*
D6S309	NED	0.83	307-333	313,321	*

Panel 9

Locus	Dye Label	Het	ASR	CEPH gt	
D6S264	FAM	0.70	113-135	117,199	*
D6S1574	FAM	0.84	150-176	156,158	*
D6S276	FAM	0.83	205-237	211,227	*
D5S408	FAM	0.73	253-289	255,261	*
D6S308	FAM	0.75	328-356	338,344	*
D6S287	HEX	0.85	110-144	136,140	*
D6S292	HEX	0.83	158-180	162,162	*
D6S434	HEX	0.86	206-250	210,212	*
D5S426	HEX	0.80	279-303	293,297	*
D5S1981	NED	0.73	121-131	125,129	*
D6S257	NED	0.87	170-198	184,186	*
D6S446	NED	0.62	222-234	222,228	*
D5S641	NED	0.77	301-341	315,317	*

Panel 10

Locus	Dye Label	Het	ASR	CEPH gt	
D5S2027	FAM	0.78	184-206	196,198	*
D5S436	FAM	0.83	242-262	244,250	*
D6S460	FAM	0.81	283-307	293,303	*
D5S410	FAM	0.79	332-354	334,344	*
D6S462	HEX	0.68	108-124	114,116	*
D5S2115	HEX	0.76	146-172	172,172	*
D5S418	HEX	0.80	212-232	214,216	*
D5S428	HEX	0.76	248-266	252,252	*
D5S630	HEX	0.89	287-395	289,320	*
D6S470	NED	0.80	125-145	133,137	*
D6S441	NED	0.86	166-200	180,186	*
D5S471	NED	0.76	241-261	251,255	*
D5S416	NED	0.77	289-301	293,295	*
D5S647	NED	0.82	328-368	342,348	*

Panel 13					
Locus	Dye Label	Het	ASR	CEPH gt	
D11S987	FAM	0.82	100-138	121,125	
D11S937	FAM	0.88	149-183	164,166	
D11S935	FAM	0.73	200-222	204,214	
D9S1677	FAM	0.81	234-258	235,257	
D11S1314	HEX	0.78	97-123	98,104	
D11S902	HEX	0.80	152-172	155,161	
D11S904	HEX	0.83	187-215	188,200	
D10S547	HEX	0.74	238-260	242,252	**
D11S905	HEX	0.75	272-298	277,277	
D9S285	NED	0.78	85-113	109,113	
D10S249	NED	0.74	122-144	136,136	
D9S171	NED	0.79	164-188	165,165	
D9S273	NED	0.74	206-226	210,212	
D10S192	NED	0.77	242-268	244,252	
D11S4175	NED	0.89	290-342	324,334	

Panel 14					
Locus	Dye Label	Het	ASR	CEPH gt	
D9S161	FAM	0.78	126-144	126,128	
D10S197	FAM	0.75	169-183	173,177	
D10S185	FAM	0.77	206-222	211,213	
D9S175	FAM	0.85	259-291	270,264	
D11S901	FAM	0.82	314-330	318,322	
D10S1653	HEX	0.77	122-136	128,128	**
D10S212	HEX	0.71	193-209	202,202	
D10S1686	HEX	0.86	247-285	259,263	
D9S287	HEX	0.67	299-317	302,304	
D11S4046	NED	0.86	105-129	133,123	
D9S288	NED	0.84	138-158	141,149	
D10S208	NED	0.79	177-197	189,191	**
D9S157	NED	0.84	229-253	229,235	
D9S167	NED	0.87	307-341	319,335	

Panel 15				
Locus	Dye Label	Het	ASR	CEPH gt
D9S164	FAM	0.80	90-106	97,99
D9S286	FAM	0.88	139-173	151,165
D9S1690	FAM	0.78	230-244	240,242
D11S1320	FAM	0.68	264-280	267,273
D11S4151	FAM	0.79	335-347	336,338
D11S4191	HEX	0.87	100-124	96,100
D11S968	HEX	0.81	144-164	153,153
D9S1776	HEX	0.84	176-214	180,182
D11S1338	HEX	0.74	259-277	271,271
D10S591	HEX	0.71	312-340	320,332
D10S587	NED	0.80	96-118	100,104
D10S537	NED	0.83	143-169	153,157
D10S189	NED	0.72	184-200	189,195
D11S925	NED	0.84	265-293	266,266

Panel 16					
Locus	Dye Label	Het	ASR	CEPH gt	
D10S217	FAM	0.81	100-124	106,110	*
D11S898	FAM	0.85	144-166	151,158	*
D10S548	FAM	0.70	186-202	190,194	*
D9S1826	FAM	0.69	219-233	222,222	*
D9S290	FAM	0.83	246-268	252,254	*
D9S1817	FAM	0.88	283-317	300,306	*
D9S158	FAM	0.69	333-357	340,342	*
D10S196	HEX	0.77	109-119	114,114	*
D9S1682	HEX	0.68	152-162	152,154	*
D11S908	HEX	0.76	175-191	182,184	*
D10S1693	HEX	0.80	217-231	221,223	*
D10S597	HEX	0.64	278-298	293,293	*
D9S283	NED	0.80	94-120	94,96	*
D10S1651	NED	0.80	210-232	211,227	*
D10S1652	NED	0.78	273-299	291,293	*

Panel 23					
Locus	Dye Label	Het	ASR	CEPH gt	
D18S70	FAM	0.83	115-135	117,177	
D17S949	FAM	0.80	214-232	220,220	
D18S478	FAM	0.64	247-261	251,257	
D17S1852	FAM	0.87	283-311	299,311	*
D17S831	HEX	0.82	112-134	116,120	
D17S1857	HEX	0.64	167-177	169,171	
D17S799	HEX	0.68	189-211	193,199	
D17S1868	HEX	0.73	258-272	264,264	
D17S798	HEX	0.80	300-324	310,316	
D18S1102	NED	0.79	96-108	100,100	
D17S787	NED	0.81	142-178	148,158	
D18S61	NED	0.87	213-243	229,231	*
D17S849	NED	0.67	257-271	261,265	*
D18S462	NED	0.70	300-322	306,306	

Panel 24					
Locus	Dye Label	Het	ASR	CEPH gt	
D18S474	FAM	0.82	126-148	130,142	
D18S53	FAM	0.79	156-186	1,668,170	
D17S938	FAM	0.76	242-262	252,254	
D18S464	FAM	0.65	303-319	313,313	
D18S63	HEX	0.79	80-106	98,102	*
D18S59	HEX	0.81	154-176	166,170	
D17S921	HEX	0.72	196-214	202,210	
D17S784	HEX	0.77	233-247	239,241	
D18S64	HEX	0.74	319-345	323,327	*
D17S928	NED	0.76	74-108	88,92	
D18S452	NED	0.83	131-149	133,143	
D17S785	NED	0.83	169-197	175,177	
D18S1161	NED	0.82	215-241	231,233	
D18S68	NED	0.68	273-299	291,293	
D17S944	NED	0.75	333-347	321,331	

Panel 25					
Locus	Dye Label	Het	ASR	CEPH gt	
D20S889	FAM	0.83	93-129	107,117	*
D20S117	FAM	0.84	154-190	160,178	*
D20S112	FAM	0.81	217-241	227,231	*
D19S220	FAM	0.84	271-295	281,283	*
D19S221	HEX	0.86	91-115	101,109	*
D20S171	HEX	0.78	130-158	140,144	*
D19S210	HEX	0.74	175-195	183,189	*
D20S100	HEX	0.76	213-239	225,227	*
D19S420	NED	0.79	101-121	110,112	*
D19S414	NED	0.78	168-198	170,188	*
D20S115	NED	0.66	238-250	242,242	*
D20S196	NED	0.81	263-299	265,287	*

Panel 26					
Locus	Dye Label	Het	ASR	CEPH gt	
D20S119	FAM	0.82	109-129	117,123	*
D21S266	FAM	0.59	159-181	163,175	*
D20S107	FAM	0.80	200-224	208,216	*
D19S902	FAM	0.79	241-277	245,257	*
D20S186	FAM	0.86	121-143	119,139	*
D22S420	HEX	0.77	156-172	160,160	*
D22S280	HEX	0.82	216-228	222,222	*
D19S216	HEX	0.76	258-276	266,266	*
D22S423	HEX	0.82	290-312	308,308	*
D19S884	NED	0.86	97-117	101,107	*
D21S1252	NED	0.80	156-180	150,166	*
D22S539	NED	0.58	203-221	205,205	*
D22S274	NED	0.77	280-302	288,292	*

Panel 27					
Locus	Dye Label	Het	ASR	CEPH gt	
D20S195	FAM	0.81	133-159	145,145	*
D22S315	FAM	0.78	184-214	198,200	*
D19S209	FAM	0.77	243-259	245,257	*
D19S418	HEX	0.66	90-110	94,94	*
D20S173	HEX	0.67	131-185	177,177	*
D21S263	HEX	0.75	198-232	200,204	*
D21S1914	HEX	0.86	262-284	266,276	*
D21S1256	NED	0.65	101-121	115,115	*
D22S283	NED	0.89	132-160	136,156	*
D20S178	NED	0.83	182-198	190,190	*
D19S226	NED	0.85	241-273	245,251	*
D19S571	NED	0.83	290-322	310,316	*

Appendix D. Genome Scan Excel Macros

Excel macros for the recoding of raw MegaBase output files into five LINKAGE format files appropriate for use in linkage programs, corresponding to the five disease classification models (see chapter 4, section 4.2.4).

1. Macro 'ReClass'

The first macro removed unnecessary columns, and recoded the disease affection status column from the 17-part liability class system output by the MegaBase database to a 6-part system reflecting the five diagnoses used in our sample and the unaffected/unknown classification.

```
Sub ReClass()
'
' ReClass Macro
' Macro recorded 20/10/2000 by Ricardo Segurado
'
' Keyboard Shortcut: Ctrl+Shift+R
'
    Range("E:G,I:J").Select
    Range("I1").Activate
    Selection.Delete Shift:=xlToLeft
    Columns("F:F").Select
    Selection.Replace What:="17", Replacement:="0", LookAt:=xlPart, _
        SearchOrder:=xlByRows, MatchCase:=False
    Selection.Replace What:="16", Replacement:="0", LookAt:=xlPart, _
        SearchOrder:=xlByRows, MatchCase:=False
    Selection.Replace What:="15", Replacement:="0", LookAt:=xlPart, _
        SearchOrder:=xlByRows, MatchCase:=False
    Selection.Replace What:="12", Replacement:="0", LookAt:=xlPart, _
```

```
        SearchOrder:=xlByRows, MatchCase:=False
    Selection.Replace What:="10", Replacement:="0", LookAt:=xlPart, _
        SearchOrder:=xlByRows, MatchCase:=False
    Selection.Replace What:="9", Replacement:="0", LookAt:=xlPart, _
        SearchOrder:=xlByRows, MatchCase:=False
    Selection.Replace What:="13", Replacement:="9", LookAt:=xlPart, _
        SearchOrder:=xlByRows, MatchCase:=False
    Selection.Replace What:="4", Replacement:="3", LookAt:=xlPart, _
        SearchOrder:=xlByRows, MatchCase:=False
    Selection.Replace What:="6", Replacement:="4", LookAt:=xlPart, _
        SearchOrder:=xlByRows, MatchCase:=False
    Selection.Replace What:="8", Replacement:="6", LookAt:=xlPart, _
        SearchOrder:=xlByRows, MatchCase:=False
    Selection.Replace What:="9", Replacement:="5", LookAt:=xlPart, _
        SearchOrder:=xlByRows, MatchCase:=False
End Sub
```

2. Macro 'M3Class'

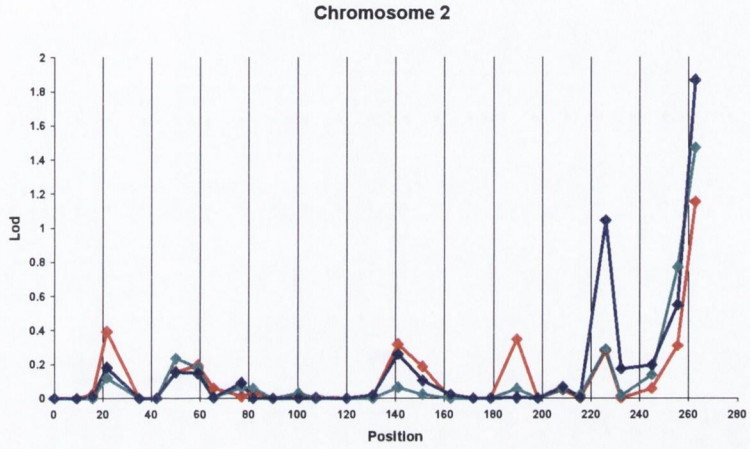
The second file shows the example macro which converts the 6-liability class file to a single affection status file for the intermediate disease model 3. The two broadest disease classifications: 6 (recurrent unipolar) and 5 (bipolar disorder – not otherwise specified), are converted to 0 (unknown phenotype). The other three classes: 4 (bipolar II disorder), 3 (schizoaffective disorder – bipolar subtype) and 2 (bipolar I disorder) are converted to 2 (affected). The macros for the four broader and narrower models are identical, except the 6 disease classes are converted to match the final disease classification model.

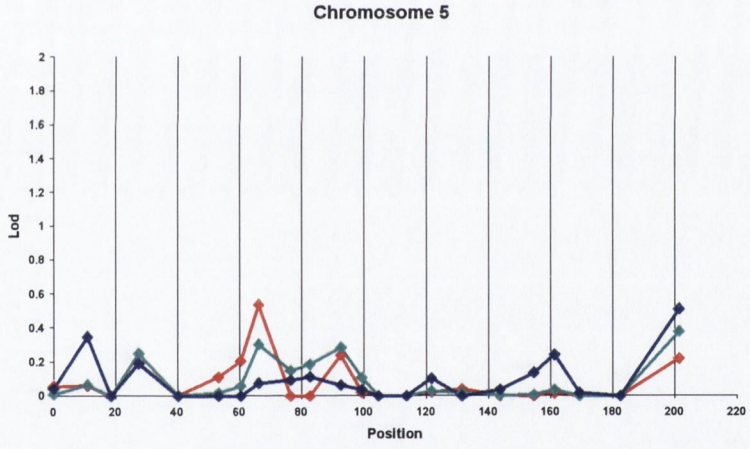
```
Sub M3Class()
'
' M3Class Macro
' Macro recorded 20/10/2000 by Ricardo Segurado
'
' Keyboard Shortcut: Ctrl+Shift+I
'
    Columns("F:F").Select
    Selection.Replace What:="6", Replacement:="0", LookAt:=xlPart, _
        SearchOrder:=xlByRows, MatchCase:=False
    Selection.Replace What:="5", Replacement:="0", LookAt:=xlPart, _
        SearchOrder:=xlByRows, MatchCase:=False
    Selection.Replace What:="4", Replacement:="2", LookAt:=xlPart, _
        SearchOrder:=xlByRows, MatchCase:=False
    Selection.Replace What:="3", Replacement:="2", LookAt:=xlPart, _
        SearchOrder:=xlByRows, MatchCase:=False
End Sub
```

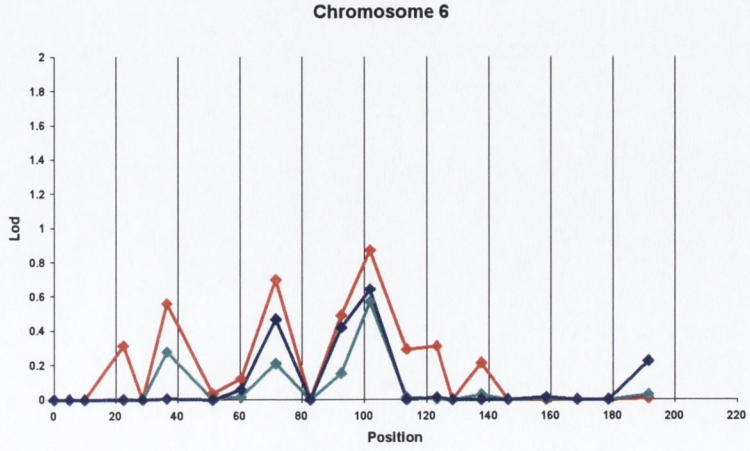
Appendix E. Genome Scan Stage I Results

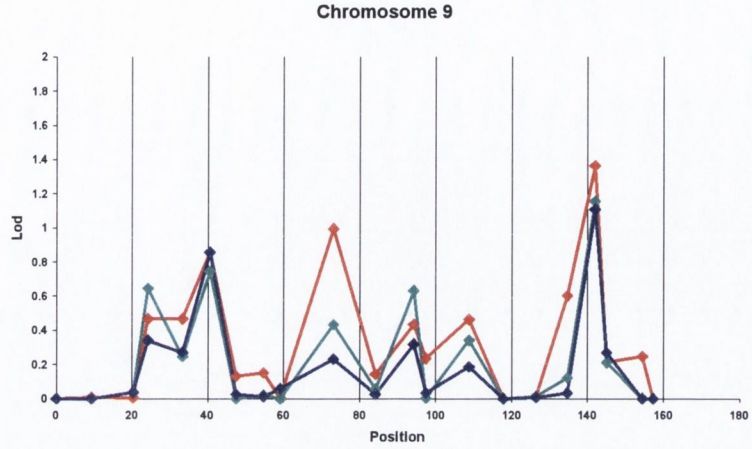
1. Single-Point MLS Results for Dublin Chromosomes

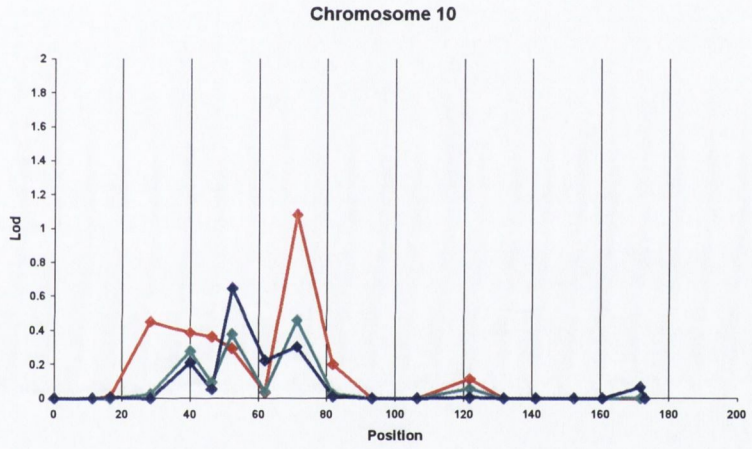
Single-point MLS plots for chromosomes 2, 5, 6, 9, 10, 11, 17, 18, 19, 20, 21 and 22, from SPLINK program. Shown are results under Models 1 (red trace), 3 (green trace) and 5 (blue trace); genetic marker positions are indicated by the trace markers.

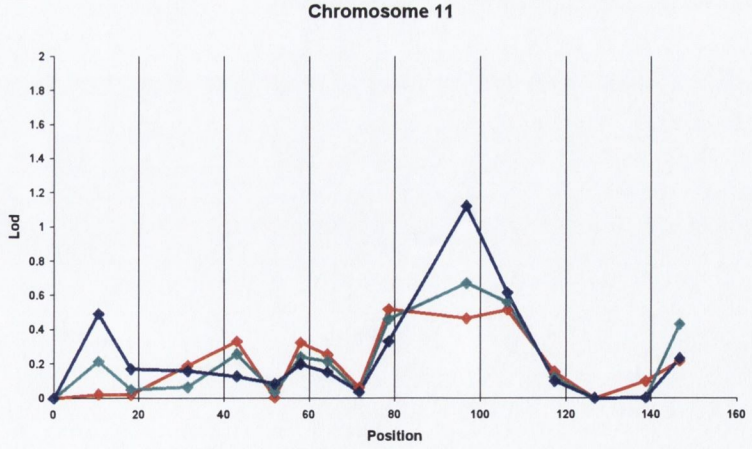


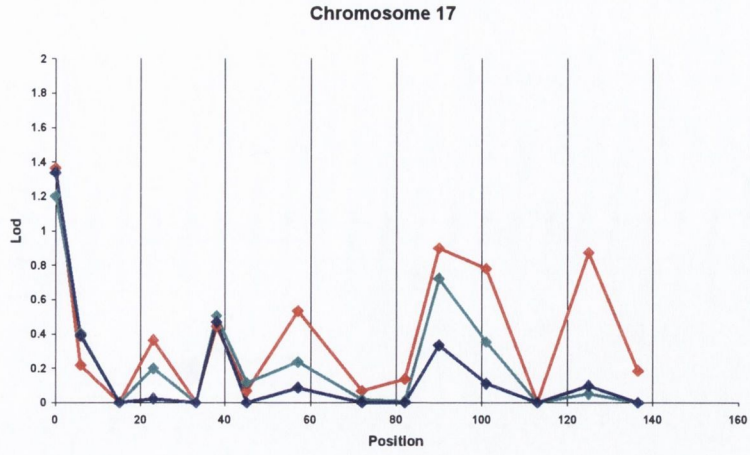


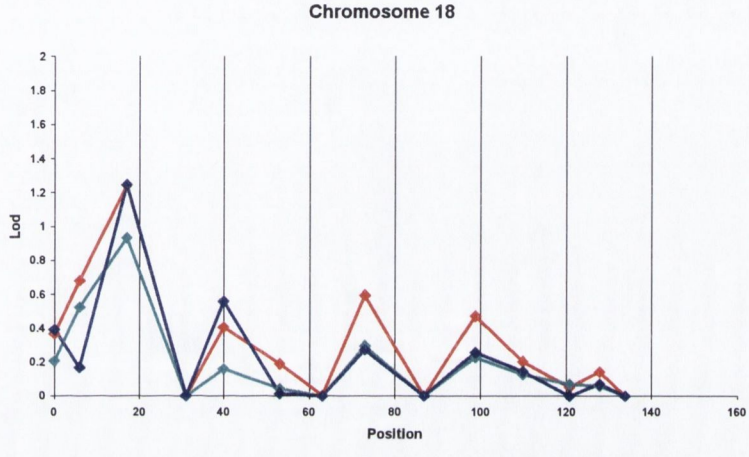


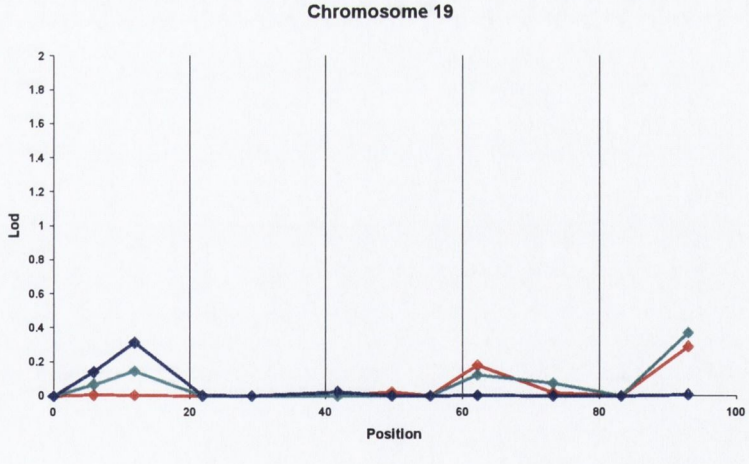

Chromosome 9
Lod
Position
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0
20
40
60
80
100
120
140
160
180

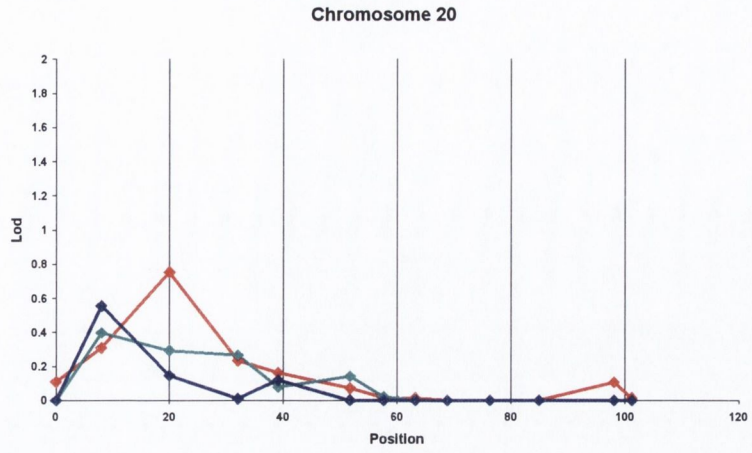

Chromosome 10
Lod
Position
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0
20
40
60
80
100
120
140
160
180
200

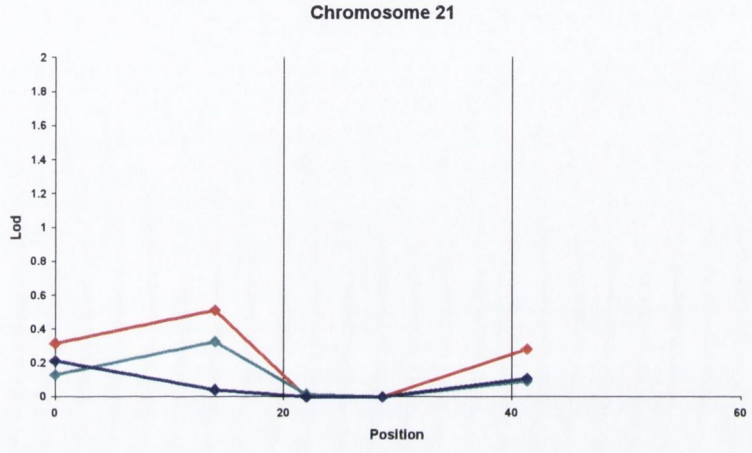

Chromosome 11
Lod
Position
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0
20
40
60
80
100
120
140
160

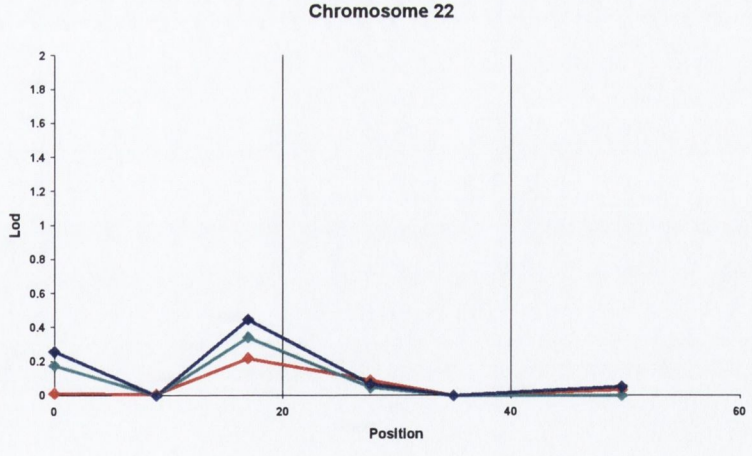

Chromosome 17
Lod
Position
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0
20
40
60
80
100
120
140
160


Chromosome 18
Lod
Position
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0
20
40
60
80
100
120
140
160


Chromosome 19
Lod
Position
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0
20
40
60
80
100

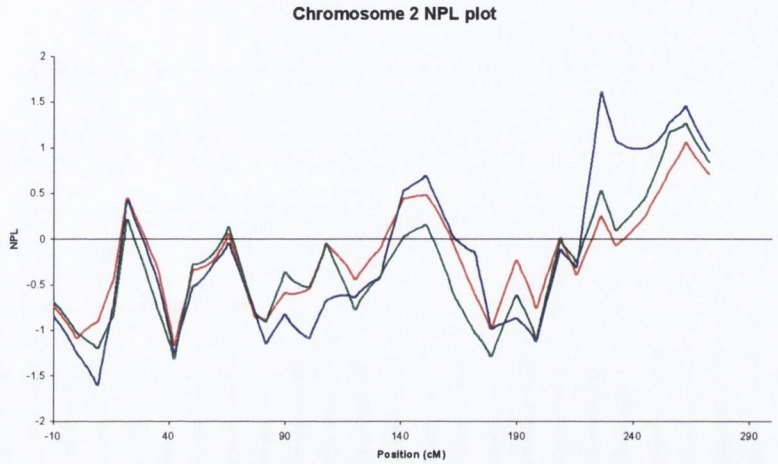

Chromosome 20
Lod
Position
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0
20
40
60
80
100
120

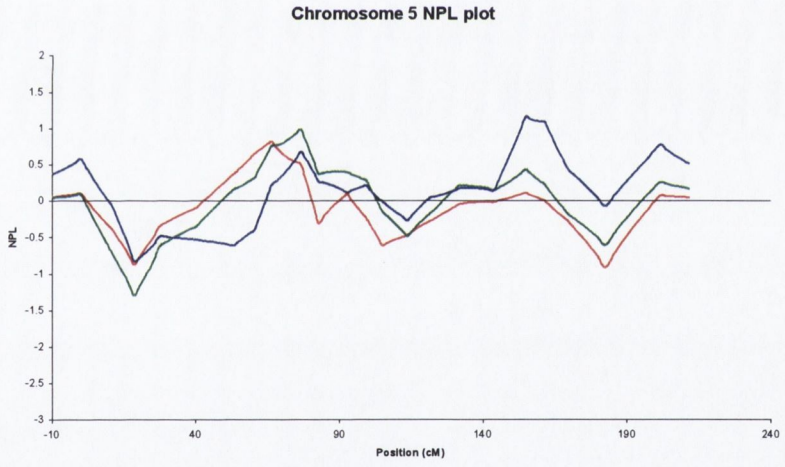

Chromosome 21
Lod
Position
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0
20
40
60

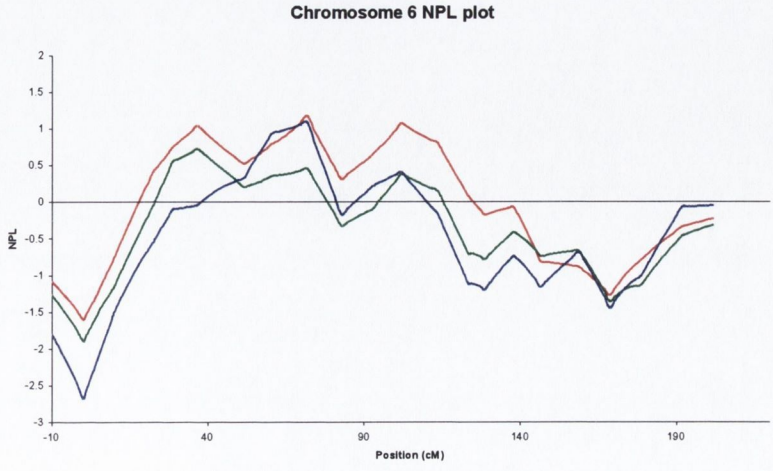

Chromosome 22
Lod
Position
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0
20
40
60

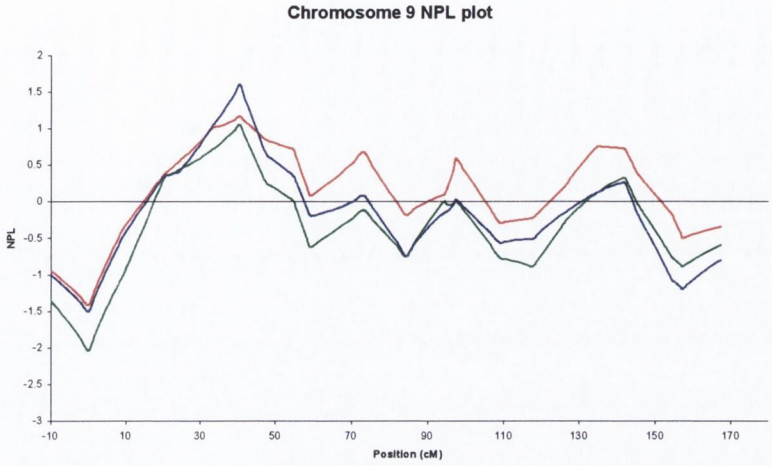
2. Multipoint NPL Results for Dublin Chromosomes

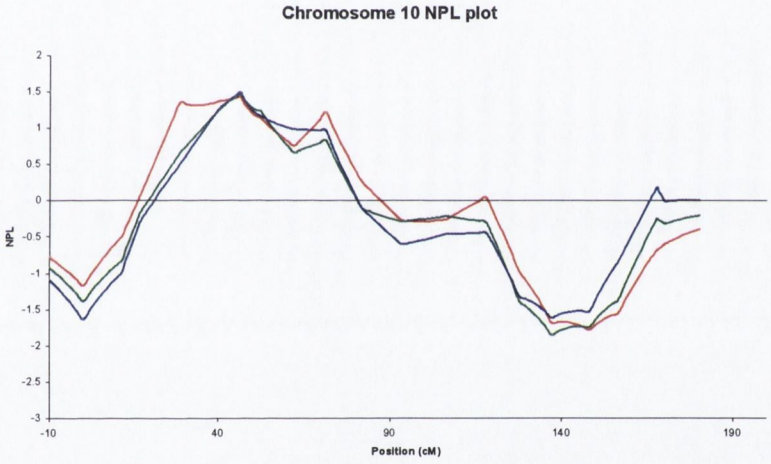
Multipoint NPL plots for chromosomes 2, 5, 6, 9, 10, 11, 17, 18, 19, 20, 21 and 22, from Genehunter. Shown are results under Models 1 (red trace), 3 (green trace) and 5 (blue trace).

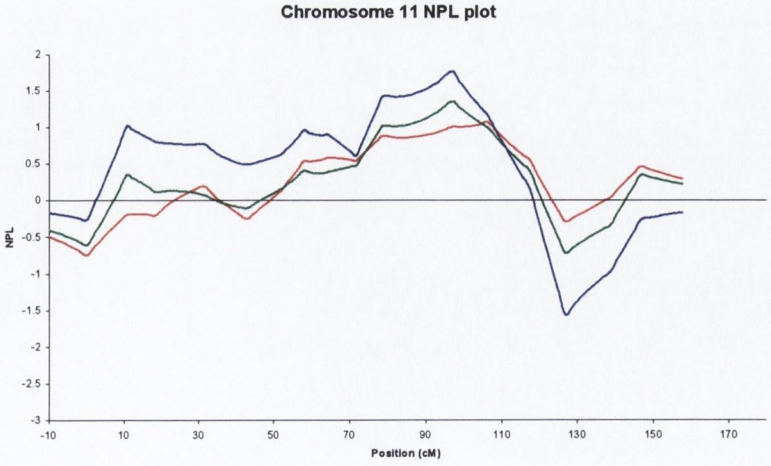




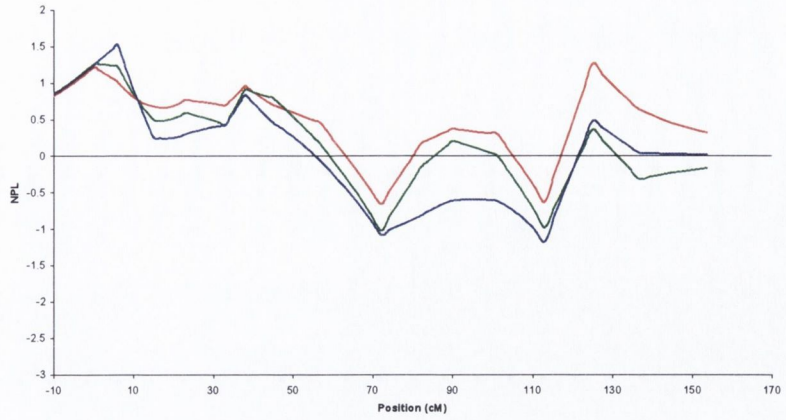



Chromosome 9 NPL plot
NPL
Position (cM)

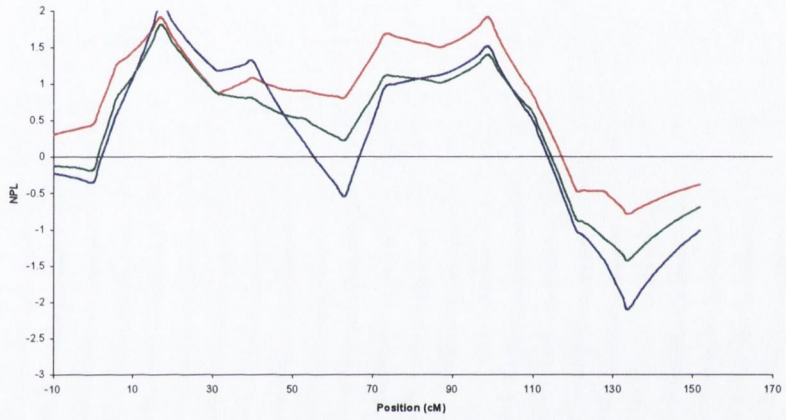

Chromosome 10 NPL plot
NPL
Position (cM)


Chromosome 11 NPL plot
NPL
Position (cM)

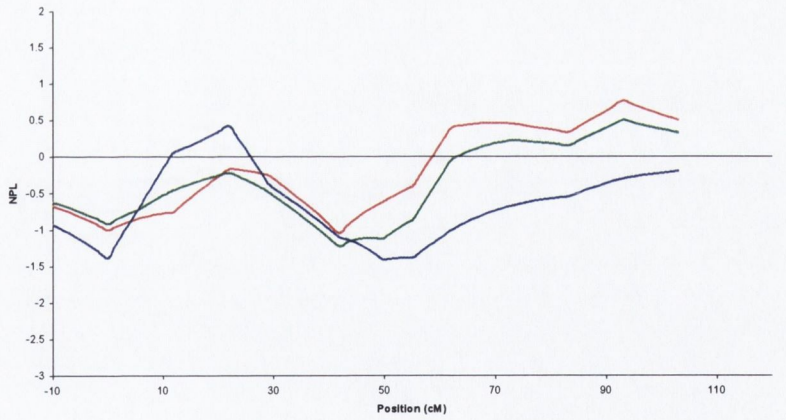
Chromosome 17 NPL plot



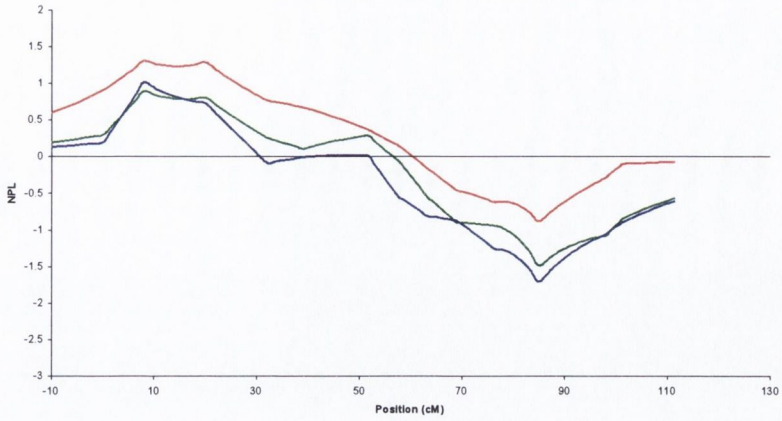
Chromosome 18 NPL plot



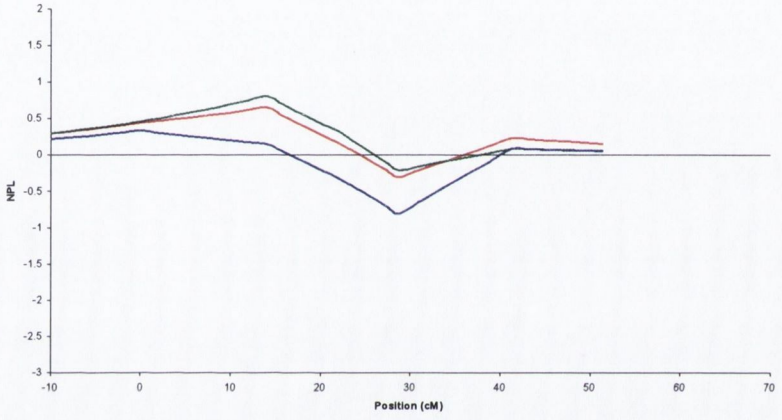
Chromosome 19 NPL plot



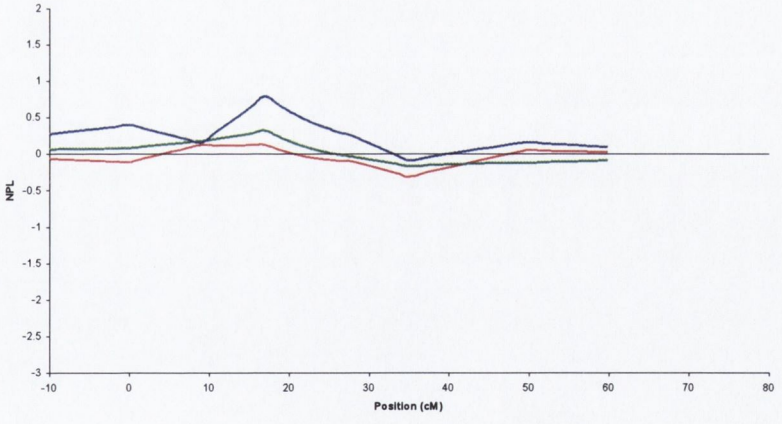
Chromosome 20 NPL plot



Chromosome 21 NPL plot



Chromosome 22 NPL plot



Appendix F. Meta-Analysis Bin Descriptions

Bin	Genetic position (Marsh cM)		Marker boundaries		Width (Marsh cM)	Cytogenetic band (Apr 2002 Genome Browser, UCSC)
	Start	End	Start	End		
1.1	0	20.61	pter	D1S503	20.61	1pter-p36.23
1.2	20.61	54.30	D1S503	D1S2674	33.69	1p36.23-1p35.3
1.3	54.30	83.07	D1S2674	D1S2742	28.77	1p35.3-p32.2
1.4	83.07	113.69	D1S2742	D1S226	30.62	1p32.2-p31.1
1.5	113.69	142.24	D1S226	D1S221	28.55	1p31.1-p13.3
1.6	142.24	170.84	D1S221	D1S2705	28.60	1p13.3-q23.3
1.7	170.84	201.58	D1S2705	D1S202	30.74	1q23.3-q31.1
1.8	201.58	231.11	D1S202	D1S425	29.53	1q31.1-q32
1.9	231.11	256.26	D1S425	D1S2680	25.15	1q32 -q42.3
1.10	256.26	286.66	D1S2680	qter	30.40	1q42.3-qter
2.1	0	26.52	pter	D2S297	26.52	2pter-p25.1
2.2	26.52	47.97	D2S297	D2S2312	21.45	2p25.1-p23.2
2.3	47.97	76.34	D2S2312	D2S2251	28.37	2p23.2-p16.2
2.4	76.34	101.56	D2S2251	D2S139	25.22	2p16.2-p12
2.5	101.56	128.41	D2S139	D2S2254	26.85	2p12-q22.1
2.6	128.41	154.48	D2S2254	D2S2275	26.07	2q22.1-q23.3
2.7	154.48	177.53	D2S2275	D2S326	23.05	2q23.3-q31.1
2.8	177.53	206.74	D2S326	D2S371	29.21	2q31.1-q34
2.9	206.74	233.62	D2S371	D2S2317	26.88	2q34-q35
2.10	233.62	269.07	D2S2317	qter	35.45	2q35-qter
3.1	0	32.36	pter	D3S3589	32.36	3pter-p25.3
3.2	32.36	63.12	D3S3589	D3S3521	30.76	3p25.3-p22.1
3.3	63.12	88.60	D3S3521	D3S1287	25.48	3p22.1-p14.1
3.4	88.60	117.76	D3S1287	D3S3655	29.16	3p14.1-q12.3
3.5	117.76	146.60	D3S3655	D3S1290	28.84	3q12.3-q22.1
3.6	146.60	173.34	D3S1290	D3S1553	26.74	3q22.1-q25.31
3.7	173.34	201.14	D3S1553	D3S3570	27.80	3q25.31-q27.3
3.8	201.14	228.14	D3S3570	qter	27.00	3q27.3-3qter
4.1	0	25.90	pter	D4S403	25.90	4pter-p15.33
4.2	25.90	51.60	D4S403	D4S3040	25.70	4p15.33-p14
4.3	51.60	78.97	D4S3040	D4S1543	27.37	4p14-q13.3
4.4	78.97	109.02	D4S1543	D4S3026	30.05	4q13.3-q24
4.5	109.02	134.74	D4S3026	D4S422	25.72	4q24-q28.3
4.6	134.74	159.30	D4S422	D4S2980	24.56	4q28.3-q32.1
4.7	159.30	181.93	D4S2980	D4S2967	22.63	4q32.1-q35.1
4.8	181.93	211.65	D4S2967	qter	29.72	4q35.1-qter
5.1	0	31.78	pter	D5S486	31.78	5pter-p15.1
5.2	31.78	64.14	D5S486	D5S633	32.36	5p15.1-q11.2
5.3	64.14	97.82	D5S633	D5S2103	33.68	5q11.2-q14.3
5.4	97.82	131.48	D5S2103	D5S2098	33.66	q5q14.3-q23.2
5.5	131.48	164.19	D5S2098	D5S422	32.71	5q23.2-q34
5.6	164.19	197.54	D5S422	qter	33.35	q34-qter

Bin	Genetic position (Marsh cM)		Marker boundaries		Width (Marsh cM)	Cytogenetic band (Apr 2002 Genome Browser, UCSC)
	Start	End	Start	End		
6.1	0	32.62	pter	D6S274	32.62	6pter-p22.3
6.2	32.62	65.14	D6S274	D6S1582	32.52	6p22.3-p21.1
6.3	65.14	99.01	D6S1582	D6S462	33.87	6p21.1-q15
6.4	99.01	131.07	D6S462	D6S472	32.06	6q15-q23.2
6.5	131.07	164.78	D6S472	D6S1581	33.71	6q23.2-q25.3
6.6	164.78	193.14	D6S1581	qter	28.36	6q25.3-qter
7.1	0	29.28	pter	D7S503	29.28	7pter-p21.1
7.2	29.28	59.93	D7S503	D7S2524	30.65	7p21.1-p14.1
7.3	59.93	91.67	D7S2524	D7S2499	31.74	7p14.1-q21.11
7.4	91.67	122.48	D7S2499	D7S2418	30.81	7q12.11-q31.1
7.5	122.48	148.11	D7S2418	D7S2468	25.63	7q31.1-q34
7.6	148.11	181.97	D7S2468	qter	33.86	7q34-qter
8.1	0	27.40	pter	D8S1790	27.40	8pter-p22
8.2	27.40	54.98	D8S1790	D8S1809	27.58	8p22-p21.1
8.3	54.98	82.84	D8S1809	D8S1785	27.86	8p21.1-q13.1
8.4	82.84	110.20	D8S1785	D8S1778	27.36	8q13.1-q22
8.5	110.20	137.92	D8S1778	D8S508	27.72	8q22-q24.21
8.6	137.92	167.90	D8S508	qter	29.98	8q24.21-qter
9.1	0	27.32	pter	D9S1869	27.32	9pter-p22.3
9.2	27.32	53.60	D9S1869	D9S1853	26.28	9p22.3-p21.1
9.3	53.60	84.90	D9S1853	D9S152	31.30	9p21.1-q21.32
9.4	84.90	109.85	D9S152	D9S277	24.95	9q21.32-q31.1
9.5	109.85	136.47	D9S277	D9S1798	26.62	9q31.1-q33.3
9.6	136.47	168.98	D9S1798	qter	32.51	9q33.3-qter
10.1	0	29.15	pter	D10S547	29.15	10pter-p14
10.2	29.15	62.23	D10S547	D10S1768	33.08	10p14-q11.21
10.3	62.23	91.13	D10S1768	D10S1648	28.90	10q11.21-q22.1
10.4	91.13	117.42	D10S1648	D10S1690	26.29	10q22.1-q23.33
10.5	117.42	142.78	D10S1690	D10S1483	25.36	10q23.33-q26.13
10.6	142.78	173.13	D10S1483	qter	30.35	10q26.13-10qter
11.1	0	21.47	pter	D11S4096	21.47	11pter-p15.1
11.2	21.47	47.06	D11S4096	D11S4083	25.59	11q15.1-p12
11.3	47.06	72.82	D11S4083	D11S4196	25.76	11p12-q13.3
11.4	72.82	98.98	D11S4196	D11S4161	26.16	11q13.3-22.3
11.5	98.98	123.00	D11S4161	D11S1316	24.02	11q22.3-q24.1
11.6	123.00	147.77	D11S1316	qter	24.77	11q24.1-qter
12.1	0	24.45	pter	D12S1697	24.45	12pter-p12.1
12.2	24.45	53.28	D12S1697	D12S1692	28.83	12p12.1-p11.21
12.3	53.28	82.12	D12S1692	D12S1722	28.84	12p11.21-q15
12.4	82.12	109.47	D12S1722	D12S318	27.35	12q15-q23.2
12.5	109.47	139.61	D12S318	D12S1349	30.14	12q23.2-q24.31
12.6	139.61	170.60	D12S1349	qter	30.99	12q24.31-qter

Bin	Genetic position (Marsh cM)		Marker boundaries		Width (Marsh cM)	Cytogenetic band (Apr 2002 Genome Browser, UCSC)
	Start	End	Start	End		
13.1	0	26.87	pter	D13S1293	26.87	13pter-q13.2
13.2	26.87	58.54	D13S1293	D13S162	31.67	13q13.2-q22.2
13.3	58.54	85.41	D13S162	D13S280	26.87	13q22.2-q33.1
13.4	85.41	114.98	D13S280	qter	29.57	13q33.1-qter
14.1	0	40.11	pter	D14S70	40.11	14pter-q13.1
14.2	40.11	74.96	D14S70	D14S1011	34.85	14q13.1-q24.1
14.3	74.96	105.00	D14S1011	D14S995	30.04	14q24.1-q32.12
14.4	105.00	138.18	D14S995	qter	33.18	14q32.12-qter
15.1	0	25.86	pter	D15S1007	25.86	15pter-q14
15.2	25.86	52.33	D15S1007	D15S1033	26.47	15q14-q21.3
15.3	52.33	85.64	D15S1033	D15S116	33.31	15q21.3-q26.1
15.4	85.64	122.14	D15S116	qter	36.50	15q26.1-qter
16.1	0	32.07	pter	D16S3103	32.07	16pter-p13
16.2	32.07	67.62	D16S3103	D16S415	35.55	16p13-q12.2
16.3	67.62	100.39	D16S415	D16S516	32.77	16q12.2-q23.1
16.4	100.39	134.12	D16S516	qter	33.73	16q23.1-qter
17.1	0	25.14	pter	D17S1875	25.14	17pter-p12
17.2	25.14	63.62	D17S1875	D17S1860	38.48	17p12-q21.33
17.3	63.62	93.98	D17S1860	D17S1350	30.36	17q21.33-q24.3
17.4	93.98	126.46	D17S1350	qter	32.48	17q24.3-qter
18.1	0	24.08	pter	D18S1163	24.08	18pter-p11
18.2	24.08	62.84	D18S1163	D18S1128	38.76	18p11-q12.3
18.3	62.84	96.48	D18S1128	D18S55	33.64	18q12.3-q22.1
18.4	96.48	126.00	D18S55	qter	29.52	18q22.1-qter
19.1	0	26.37	pter	D19S884	26.37	19pter-p13.3
19.2	26.37	52.59	D19S884	D19S928	26.22	19p13.3-q12
19.3	52.59	75.41	D19S928	D19S604	22.82	19q12-q13.33
19.4	75.41	105.02	D19S604	qter	29.61	19q13.33-qter
20.1	0	21.15	pter	D20S115	21.15	20pter-p12.3
20.2	21.15	47.52	D20S115	D20S190	26.37	20p12.3-p11
20.3	47.52	72.27	D20S190	D20S176	24.75	20p11-q13.13
20.4	72.27	101.22	D20S176	qter	28.95	20q13.13-qter
21.1	0	25.26	pter	D21S1916	25.26	21pter-q21.3
21.2	25.26	57.77	D21S1916	qter	32.51	21q21.3-qter
22.1	0	33.76	pter	D22S424	33.76	22pter-q12.3
22.2	33.76	62.31	D22S424	qter	28.55	22q12.3-pter

Appendix G. Meta-Analysis Excel Macros

Excel macros written to aid in organisation and analysis of the simulation studies to determine empirical significance levels in the meta-analysis study (see chapter 6, section 6.2.4)

1. Macro 'RandCreate'

The first macro operates on an Excel file containing the 120 ranked bins of each of the studies for one of the disease classification models. It performs a randomisation of the order of the ranks for each study separately using a constantly renewed random number between 1 and 1,000,000, and pastes the resulting summed rank into a sheet in the same file, named "18 Data".

```
Sub RandCreate()
'
' RandCreate Macro
' Macro recorded 27-04-2002 by Ricardo Segurado
'
' Keyboard Shortcut: Ctrl+Shift+Z
'
    Sheets("18 Data").Select
    Selection.PasteSpecial Paste:=xlValues, Operation:=xlNone,
SkipBlanks:= _False, Transpose:=False
    Sheets("Random Numbers").Select
    ActiveWindow.ScrollColumn = 1
    Columns("A:D").Select
    Range("C1").Activate
    Application.CutCopyMode = False
```

```
    Selection.Sort Key1:=Range("C1"), Order1:=xlAscending,
Header:=xlGuess, _OrderCustom:=1, MatchCase:=False,
Orientation:=xlTopToBottom
    Columns("E:H").Select
    Range("G1").Activate
    Selection.Sort Key1:=Range("G1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    Columns("I:L").Select
    Range("K1").Activate
    Selection.Sort Key1:=Range("K1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    Columns("M:P").Select
    Range("O1").Activate
    Selection.Sort Key1:=Range("O1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    Columns("Q:T").Select
    Range("S1").Activate
    Selection.Sort Key1:=Range("S1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    ActiveWindow.SmallScroll ToRight:=9
    Columns("U:X").Select
    Range("W1").Activate
    Selection.Sort Key1:=Range("W1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    Columns("Y:AB").Select
    Range("AA1").Activate
```

```
    Selection.Sort Key1:=Range("AA1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    ActiveWindow.SmallScroll ToRight:=8
    Columns("AC:AF").Select
    Range("AE1").Activate
    Selection.Sort Key1:=Range("AE1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    Columns("AG:AJ").Select
    Range("AI1").Activate
    Selection.Sort Key1:=Range("AI1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    ActiveWindow.SmallScroll ToRight:=8
    Columns("AK:AN").Select
    Range("AM1").Activate
    Selection.Sort Key1:=Range("AM1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    Columns("AO:AR").Select
    Range("AQ1").Activate
    Selection.Sort Key1:=Range("AQ1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    ActiveWindow.SmallScroll ToRight:=7
    Columns("AS:AV").Select
    Range("AU1").Activate
    Selection.Sort Key1:=Range("AU1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    ActiveWindow.SmallScroll ToRight:=6
```

```
    Columns("AW:AZ").Select
    Range("AY1").Activate
    Selection.Sort Key1:=Range("AY1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    Columns("BA:BD").Select
    Range("BC1").Activate
    Selection.Sort Key1:=Range("BC1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    ActiveWindow.SmallScroll ToRight:=6
    Columns("BE:BH").Select
    Range("BG1").Activate
    Selection.Sort Key1:=Range("BG1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    Columns("BI:BL").Select
    Range("BK1").Activate
    Selection.Sort Key1:=Range("BK1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    ActiveWindow.SmallScroll ToRight:=7
    Columns("BM:BP").Select
    Range("BO1").Activate
    Selection.Sort Key1:=Range("BO1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    ActiveWindow.SmallScroll ToRight:=6
    Columns("BQ:BT").Select
    Range("BS1").Activate
    Selection.Sort Key1:=Range("BS1"), Order1:=xlAscending,
Header:=xlGuess, _
```

```
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    ActiveWindow.SmallScroll ToRight:=6
    Columns("BU:BX").Select
    Range("BW1").Activate
    Selection.Sort Key1:=Range("BW1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    ActiveWindow.SmallScroll ToRight:=6
    Columns("BY:CB").Select
    Range("CA1").Activate
    Selection.Sort Key1:=Range("CA1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    ActiveWindow.SmallScroll ToRight:=6
    Columns("CC:CF").Select
    Range("CE1").Activate
    Selection.Sort Key1:=Range("CE1"), Order1:=xlAscending,
Header:=xlGuess, _
        OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom
    Columns("CM:CM").Select
    Selection.Copy
    Sheets("18 Data").Select
End Sub
```

2. Macro 'FullSheetRC'

The second macro repeatedly calls the first macro, pasting the new summed ranks into each of the 256 columns of the sheet "18 Data".

```
Sub FullSheetRC()
'
' FullSheetRC Macro
' Macro recorded 27-04-2002 by Ricardo Segurado
'
' Keyboard Shortcut: Ctrl+Shift+X
'
    Application.Run "RandCreate"
    Columns("B:B").Select
    Application.Run "RandCreate"
    Columns("C:C").Select
    Application.Run "RandCreate"
    Columns("D:D").Select
    Application.Run "RandCreate"
    Columns("E:E").Select
    Application.Run "RandCreate"
    Columns("F:F").Select
    Application.Run "RandCreate"
    Columns("G:G").Select
    Application.Run "RandCreate"
    Columns("H:H").Select
    Application.Run "RandCreate"
    Columns("I:I").Select
    Application.Run "RandCreate"
    Columns("J:J").Select
    Application.Run "RandCreate"
    Columns("K:K").Select
```

```
Application.Run "RandCreate"
Columns("L:L").Select
Application.Run "RandCreate"
Columns("M:M").Select
Application.Run "RandCreate"
Columns("N:N").Select
Application.Run "RandCreate"
Columns("O:O").Select
Application.Run "RandCreate"
Columns("P:P").Select
Application.Run "RandCreate"
Columns("Q:Q").Select
Application.Run "RandCreate"
Columns("R:R").Select
Application.Run "RandCreate"
Columns("S:S").Select
Application.Run "RandCreate"
Columns("T:T").Select
Application.Run "RandCreate"
Columns("U:U").Select
Application.Run "RandCreate"
Columns("V:V").Select
Application.Run "RandCreate"
Columns("W:W").Select
Application.Run "RandCreate"
Columns("X:X").Select
Application.Run "RandCreate"
Columns("Y:Y").Select
Application.Run "RandCreate"
Columns("Z:Z").Select
Application.Run "RandCreate"
Columns("AA:AA").Select
```

```
Application.Run "RandCreate"
Columns("AB:AB").Select
Application.Run "RandCreate"
Columns("AC:AC").Select
Application.Run "RandCreate"
Columns("AD:AD").Select
Application.Run "RandCreate"
Columns("AE:AE").Select
Application.Run "RandCreate"
Columns("AF:AF").Select
Application.Run "RandCreate"
Columns("AG:AG").Select
Application.Run "RandCreate"
Columns("AH:AH").Select
Application.Run "RandCreate"
Columns("AI:AI").Select
Application.Run "RandCreate"
Columns("AJ:AJ").Select
Application.Run "RandCreate"
Columns("AK:AK").Select
Application.Run "RandCreate"
Columns("AL:AL").Select
Application.Run "RandCreate"
Columns("AM:AM").Select
Application.Run "RandCreate"
Columns("AN:AN").Select
Application.Run "RandCreate"
Columns("AO:AO").Select
Application.Run "RandCreate"
Selection.ColumnWidth = 8.57
Columns("AP:AP").Select
Application.Run "RandCreate"
```

```
Columns("AQ:AQ").Select
Application.Run "RandCreate"
Columns("AR:AR").Select
Application.Run "RandCreate"
Columns("AS:AS").Select
Application.Run "RandCreate"
Application.Run "RandCreate"
Columns("AT:AT").Select
Application.Run "RandCreate"
Columns("AU:AU").Select
Application.Run "RandCreate"
Columns("AV:AV").Select
Application.Run "RandCreate"
Columns("AW:AW").Select
Application.Run "RandCreate"
Columns("AX:AX").Select
Application.Run "RandCreate"
Columns("AY:AY").Select
Application.Run "RandCreate"
Columns("AZ:AZ").Select
Application.Run "RandCreate"
Columns("BA:BA").Select
Application.Run "RandCreate"
Columns("BB:BB").Select
Application.Run "RandCreate"
Range("BC1").Select
Application.Run "RandCreate"
Range("BD1").Select
Application.Run "RandCreate"
Range("BE1").Select
Application.Run "RandCreate"
Range("BF1").Select
```

```
Application.Run "RandCreate"
Range("BG1").Select
Application.Run "RandCreate"
Range("BH1").Select
Application.Run "RandCreate"
Range("BI1").Select
Application.Run "RandCreate"
Range("BJ1").Select
Application.Run "RandCreate"
Range("BK1").Select
Application.Run "RandCreate"
Range("BL1").Select
Application.Run "RandCreate"
Range("BM1").Select
Application.Run "RandCreate"
Range("BN1").Select
Application.Run "RandCreate"
Range("BO1").Select
Application.Run "RandCreate"
Range("BP1").Select
Application.Run "RandCreate"
Range("BQ1").Select
Application.Run "RandCreate"
Range("BR1").Select
Application.Run "RandCreate"
Range("BS1").Select
Application.Run "RandCreate"
Range("BT1").Select
Application.Run "RandCreate"
Range("BU1").Select
Application.Run "RandCreate"
Range("BV1").Select
```

```
Application.Run "RandCreate"
Range("BW1").Select
Application.Run "RandCreate"
Range("BX1").Select
Application.Run "RandCreate"
Range("BY1").Select
Application.Run "RandCreate"
Range("BZ1").Select
Application.Run "RandCreate"
Range("CA1").Select
Application.Run "RandCreate"
Range("CB1").Select
Application.Run "RandCreate"
Range("CC1").Select
Application.Run "RandCreate"
Range("CD1").Select
Application.Run "RandCreate"
Range("CE1").Select
Application.Run "RandCreate"
Range("CF1").Select
Application.Run "RandCreate"
Range("CG1").Select
Application.Run "RandCreate"
Range("CH1").Select
Application.Run "RandCreate"
Range("CI1").Select
Application.Run "RandCreate"
Range("CJ1").Select
Application.Run "RandCreate"
Range("CK1").Select
Application.Run "RandCreate"
Range("CL1").Select
```

```
Application.Run "RandCreate"
Range("CM1").Select
Application.Run "RandCreate"
Range("CN1").Select
Application.Run "RandCreate"
Range("CO1").Select
Application.Run "RandCreate"
Range("CP1").Select
Application.Run "RandCreate"
Range("CQ1").Select
Application.Run "RandCreate"
Range("CR1").Select
Application.Run "RandCreate"
Range("CS1").Select
Application.Run "RandCreate"
Range("CT1").Select
Application.Run "RandCreate"
Range("CU1").Select
Application.Run "RandCreate"
Range("CV1").Select
Application.Run "RandCreate"
Range("CW1").Select
Application.Run "RandCreate"
Range("CX1").Select
Application.Run "RandCreate"
Range("CY1").Select
Application.Run "RandCreate"
Range("CZ1").Select
Application.Run "RandCreate"
Range("DA1").Select
Application.Run "RandCreate"
Range("DB1").Select
```

```
Application.Run "RandCreate"
Range("DC1").Select
Application.Run "RandCreate"
Range("DD1").Select
Application.Run "RandCreate"
Range("DE1").Select
Application.Run "RandCreate"
Range("DF1").Select
Application.Run "RandCreate"
Range("DG1").Select
Application.Run "RandCreate"
Range("DH1").Select
Application.Run "RandCreate"
Range("DI1").Select
Application.Run "RandCreate"
Range("DJ1").Select
Application.Run "RandCreate"
Range("DK1").Select
Application.Run "RandCreate"
Range("DL1").Select
Application.Run "RandCreate"
Range("DM1").Select
Application.Run "RandCreate"
Range("DN1").Select
Application.Run "RandCreate"
Range("DO1").Select
Application.Run "RandCreate"
Range("DP1").Select
Application.Run "RandCreate"
Range("DQ1").Select
Application.Run "RandCreate"
Range("DR1").Select
```

```
Application.Run "RandCreate"
Range("DS1").Select
Application.Run "RandCreate"
Range("DT1").Select
Application.Run "RandCreate"
Range("DU1").Select
Application.Run "RandCreate"
Range("DV1").Select
Application.Run "RandCreate"
Range("DW1").Select
Application.Run "RandCreate"
Range("DX1").Select
Application.Run "RandCreate"
Range("DY1").Select
Application.Run "RandCreate"
Range("DZ1").Select
Application.Run "RandCreate"
Range("EA1").Select
Application.Run "RandCreate"
Range("EB1").Select
Application.Run "RandCreate"
Range("EC1").Select
Application.Run "RandCreate"
Range("ED1").Select
Application.Run "RandCreate"
Range("EE1").Select
Application.Run "RandCreate"
Range("EF1").Select
Application.Run "RandCreate"
Range("EG1").Select
Application.Run "RandCreate"
Range("EH1").Select
```

```
Application.Run "RandCreate"
Range("EI1").Select
Application.Run "RandCreate"
Range("EJ1").Select
Application.Run "RandCreate"
Range("EK1").Select
Application.Run "RandCreate"
Range("EL1").Select
Application.Run "RandCreate"
Range("EM1").Select
Application.Run "RandCreate"
Range("EN1").Select
Application.Run "RandCreate"
Range("EO1").Select
Application.Run "RandCreate"
Range("EP1").Select
Application.Run "RandCreate"
Range("EQ1").Select
Application.Run "RandCreate"
Range("ER1").Select
Application.Run "RandCreate"
Range("ES1").Select
Application.Run "RandCreate"
Range("ET1").Select
Application.Run "RandCreate"
Range("EU1").Select
Application.Run "RandCreate"
Range("EV1").Select
Application.Run "RandCreate"
Range("EW1").Select
Application.Run "RandCreate"
Range("EX1").Select
```

```
Application.Run "RandCreate"
Range("EY1").Select
Application.Run "RandCreate"
Range("EZ1").Select
Application.Run "RandCreate"
Range("FA1").Select
Application.Run "RandCreate"
Range("FB1").Select
Application.Run "RandCreate"
Range("FC1").Select
Application.Run "RandCreate"
Range("FD1").Select
Application.Run "RandCreate"
Range("FE1").Select
Application.Run "RandCreate"
Range("FF1").Select
Application.Run "RandCreate"
Range("FG1").Select
Application.Run "RandCreate"
Range("FH1").Select
Application.Run "RandCreate"
Range("FI1").Select
Application.Run "RandCreate"
Range("FJ1").Select
Application.Run "RandCreate"
Range("FK1").Select
Application.Run "RandCreate"
Range("FL1").Select
Application.Run "RandCreate"
Range("FM1").Select
Application.Run "RandCreate"
Range("FN1").Select
```

```
Application.Run "RandCreate"
Range("FO1").Select
Application.Run "RandCreate"
Range("FP1").Select
Application.Run "RandCreate"
Range("FQ1").Select
Application.Run "RandCreate"
Range("FR1").Select
Application.Run "RandCreate"
Range("FS1").Select
Application.Run "RandCreate"
Range("FT1").Select
Application.Run "RandCreate"
Range("FU1").Select
Application.Run "RandCreate"
Range("FV1").Select
Application.Run "RandCreate"
Range("FW1").Select
Application.Run "RandCreate"
Range("FX1").Select
Application.Run "RandCreate"
Range("FY1").Select
Application.Run "RandCreate"
Range("FZ1").Select
Application.Run "RandCreate"
Range("GA1").Select
Application.Run "RandCreate"
Range("GB1").Select
Application.Run "RandCreate"
Range("GC1").Select
Application.Run "RandCreate"
Range("GD1").Select
```

```
Application.Run "RandCreate"
Range("GE1").Select
Application.Run "RandCreate"
Range("GF1").Select
Application.Run "RandCreate"
Range("GG1").Select
Application.Run "RandCreate"
Range("GH1").Select
Application.Run "RandCreate"
Range("GI1").Select
Application.Run "RandCreate"
Range("GJ1").Select
Application.Run "RandCreate"
Range("GK1").Select
Application.Run "RandCreate"
Range("GL1").Select
Application.Run "RandCreate"
Range("GM1").Select
Application.Run "RandCreate"
Range("GN1").Select
Application.Run "RandCreate"
Range("GO1").Select
Application.Run "RandCreate"
Range("GP1").Select
Application.Run "RandCreate"
Range("GQ1").Select
Application.Run "RandCreate"
Range("GR1").Select
Application.Run "RandCreate"
Range("GS1").Select
Application.Run "RandCreate"
Range("GT1").Select
```

```
Application.Run "RandCreate"
Range("GU1").Select
Application.Run "RandCreate"
Range("GV1").Select
Application.Run "RandCreate"
Range("GW1").Select
Application.Run "RandCreate"
Range("GX1").Select
Application.Run "RandCreate"
Range("GY1").Select
Application.Run "RandCreate"
Range("GZ1").Select
Application.Run "RandCreate"
Range("HA1").Select
Application.Run "RandCreate"
Range("HB1").Select
Application.Run "RandCreate"
Range("HC1").Select
Application.Run "RandCreate"
Range("HD1").Select
Application.Run "RandCreate"
Range("HE1").Select
Application.Run "RandCreate"
Range("HF1").Select
Application.Run "RandCreate"
Range("HG1").Select
Application.Run "RandCreate"
Range("HH1").Select
Application.Run "RandCreate"
Range("HI1").Select
Application.Run "RandCreate"
Range("HJ1").Select
```

```
Application.Run "RandCreate"
Range("HK1").Select
Application.Run "RandCreate"
Range("HL1").Select
Application.Run "RandCreate"
Range("HM1").Select
Application.Run "RandCreate"
Range("HN1").Select
Application.Run "RandCreate"
Range("HO1").Select
Application.Run "RandCreate"
Range("HP1").Select
Application.Run "RandCreate"
Range("HQ1").Select
Application.Run "RandCreate"
Range("HR1").Select
Application.Run "RandCreate"
Range("HS1").Select
Application.Run "RandCreate"
Range("HT1").Select
Application.Run "RandCreate"
Range("HU1").Select
Application.Run "RandCreate"
Range("HV1").Select
Application.Run "RandCreate"
Range("HW1").Select
Application.Run "RandCreate"
Range("HX1").Select
Application.Run "RandCreate"
Range("HY1").Select
Application.Run "RandCreate"
Range("HZ1").Select
```

```
Application.Run "RandCreate"
Range("IA1").Select
Application.Run "RandCreate"
Range("IB1").Select
Application.Run "RandCreate"
Range("IC1").Select
Application.Run "RandCreate"
Range("ID1").Select
Application.Run "RandCreate"
Range("IE1").Select
Application.Run "RandCreate"
Range("IF1").Select
Application.Run "RandCreate"
Range("IG1").Select
Application.Run "RandCreate"
Range("IH1").Select
Application.Run "RandCreate"
Range("II1").Select
Application.Run "RandCreate"
Range("IJ1").Select
Application.Run "RandCreate"
Range("IK1").Select
Application.Run "RandCreate"
Range("IL1").Select
Application.Run "RandCreate"
Range("IM1").Select
Application.Run "RandCreate"
Range("IN1").Select
Application.Run "RandCreate"
Range("IO1").Select
Application.Run "RandCreate"
Range("IP1").Select
```

```
    Application.Run "RandCreate"
    Range("IQ1").Select
    Application.Run "RandCreate"
    Range("IR1").Select
    Application.Run "RandCreate"
    Range("IS1").Select
    Application.Run "RandCreate"
    Range("IT1").Select
    Application.Run "RandCreate"
    Range("IU1").Select
    Application.Run "RandCreate"
    Range("IV1").Select
    Application.Run "RandCreate"
End Sub
```

3. Macro 'TotalAuto'

The third macro repeatedly calls the FullSheetRC macro, applying it to each of 20 sheets, renamed "1" to "20", thus generating 256 columns x 20 sheets = 5120 randomised replicates of the summed rank.

```
Sub TotalAuto()
'
' TotalAuto Macro
' Macro recorded 31-05-2002 by Ricardo Segurado
'
' Keyboard Shortcut: Ctrl+Shift+T
'
    Sheets("1").Select
    Sheets("1").Name = "18 Data"
    Sheets("Random Numbers").Select
    Columns("CM:CM").Select
    Selection.Copy
    Sheets("18 Data").Select
    Application.Run "FullSheetRC"
    Range("A1").Select
    Application.CutCopyMode = False
    ActiveWorkbook.Save
    Sheets("18 Data").Select
    Sheets("18 Data").Name = "1"
        Sheets("2").Select
    Sheets("2").Name = "18 Data"
    Sheets("Random Numbers").Select
    Columns("CM:CM").Select
    Selection.Copy
    Sheets("18 Data").Select
    Application.Run "FullSheetRC"
```

```
Range("A1").Select
Application.CutCopyMode = False
ActiveWorkbook.Save
Sheets("18 Data").Select
Sheets("18 Data").Name = "2"
    Sheets("3").Name = "18 Data"
Sheets("Random Numbers").Select
Columns("CM:CM").Select
Selection.Copy
Sheets("18 Data").Select
Application.Run "FullSheetRC"
Range("A1").Select
Application.CutCopyMode = False
ActiveWorkbook.Save
Sheets("18 Data").Select
Sheets("18 Data").Name = "3"
    Sheets("4").Name = "18 Data"
Sheets("Random Numbers").Select
Columns("CM:CM").Select
Selection.Copy
Sheets("18 Data").Select
Application.Run "FullSheetRC"
Range("A1").Select
Application.CutCopyMode = False
ActiveWorkbook.Save
Sheets("18 Data").Select
Sheets("18 Data").Name = "4"
    Sheets("5").Name = "18 Data"
Sheets("Random Numbers").Select
Columns("CM:CM").Select
Selection.Copy
Sheets("18 Data").Select
```

```
Application.Run "FullSheetRC"
Range("A1").Select
Application.CutCopyMode = False
ActiveWorkbook.Save
Sheets("18 Data").Select
Sheets("18 Data").Name = "5"
    Sheets("6").Name = "18 Data"
Sheets("Random Numbers").Select
Columns("CM:CM").Select
Selection.Copy
Sheets("18 Data").Select
Application.Run "FullSheetRC"
Range("A1").Select
Application.CutCopyMode = False
ActiveWorkbook.Save
Sheets("18 Data").Select
Sheets("18 Data").Name = "6"
    Sheets("7").Name = "18 Data"
Sheets("Random Numbers").Select
Columns("CM:CM").Select
Selection.Copy
Sheets("18 Data").Select
Application.Run "FullSheetRC"
Range("A1").Select
Application.CutCopyMode = False
ActiveWorkbook.Save
Sheets("18 Data").Select
Sheets("18 Data").Name = "7"
    Sheets("8").Name = "18 Data"
Sheets("Random Numbers").Select
Columns("CM:CM").Select
Selection.Copy
```

```
Sheets("18 Data").Select
Application.Run "FullSheetRC"
Range("A1").Select
Application.CutCopyMode = False
ActiveWorkbook.Save
Sheets("18 Data").Select
Sheets("18 Data").Name = "8"
    Sheets("9").Name = "18 Data"
Sheets("Random Numbers").Select
Columns("CM:CM").Select
Selection.Copy
Sheets("18 Data").Select
Application.Run "FullSheetRC"
Range("A1").Select
Application.CutCopyMode = False
ActiveWorkbook.Save
Sheets("18 Data").Select
Sheets("18 Data").Name = "9"
    Sheets("10").Name = "18 Data"
Sheets("Random Numbers").Select
Columns("CM:CM").Select
Selection.Copy
Sheets("18 Data").Select
Application.Run "FullSheetRC"
Range("A1").Select
Application.CutCopyMode = False
ActiveWorkbook.Save
Sheets("18 Data").Select
Sheets("18 Data").Name = "10"
    Sheets("11").Name = "18 Data"
Sheets("Random Numbers").Select
Columns("CM:CM").Select
```

```
Selection.Copy
Sheets("18 Data").Select
Application.Run "FullSheetRC"
Range("A1").Select
Application.CutCopyMode = False
ActiveWorkbook.Save
Sheets("18 Data").Select
Sheets("18 Data").Name = "11"
    Sheets("12").Name = "18 Data"
Sheets("Random Numbers").Select
Columns("CM:CM").Select
Selection.Copy
Sheets("18 Data").Select
Application.Run "FullSheetRC"
Range("A1").Select
Application.CutCopyMode = False
ActiveWorkbook.Save
Sheets("18 Data").Select
Sheets("18 Data").Name = "12"
    Sheets("13").Name = "18 Data"
Sheets("Random Numbers").Select
Columns("CM:CM").Select
Selection.Copy
Sheets("18 Data").Select
Application.Run "FullSheetRC"
Range("A1").Select
Application.CutCopyMode = False
ActiveWorkbook.Save
Sheets("18 Data").Select
Sheets("18 Data").Name = "13"
    Sheets("14").Name = "18 Data"
Sheets("Random Numbers").Select
```

```
Columns("CM:CM").Select
Selection.Copy
Sheets("18 Data").Select
Application.Run "FullSheetRC"
Range("A1").Select
Application.CutCopyMode = False
ActiveWorkbook.Save
Sheets("18 Data").Select
Sheets("18 Data").Name = "14"
    Sheets("15").Name = "18 Data"
Sheets("Random Numbers").Select
Columns("CM:CM").Select
Selection.Copy
Sheets("18 Data").Select
Application.Run "FullSheetRC"
Range("A1").Select
Application.CutCopyMode = False
ActiveWorkbook.Save
Sheets("18 Data").Select
Sheets("18 Data").Name = "15"
    Sheets("16").Name = "18 Data"
Sheets("Random Numbers").Select
Columns("CM:CM").Select
Selection.Copy
Sheets("18 Data").Select
Application.Run "FullSheetRC"
Range("A1").Select
Application.CutCopyMode = False
ActiveWorkbook.Save
Sheets("18 Data").Select
Sheets("18 Data").Name = "16"
    Sheets("17").Name = "18 Data"
```

```
Sheets("Random Numbers").Select
Columns("CM:CM").Select
Selection.Copy
Sheets("18 Data").Select
Application.Run "FullSheetRC"
Range("A1").Select
Application.CutCopyMode = False
ActiveWorkbook.Save
Sheets("18 Data").Select
Sheets("18 Data").Name = "17"
    Sheets("18").Name = "18 Data"
Sheets("Random Numbers").Select
Columns("CM:CM").Select
Selection.Copy
Sheets("18 Data").Select
Application.Run "FullSheetRC"
Range("A1").Select
Application.CutCopyMode = False
ActiveWorkbook.Save
Sheets("18 Data").Select
Sheets("18 Data").Name = "18"
    Sheets("19").Name = "18 Data"
Sheets("Random Numbers").Select
Columns("CM:CM").Select
Selection.Copy
Sheets("18 Data").Select
Application.Run "FullSheetRC"
Range("A1").Select
Application.CutCopyMode = False
ActiveWorkbook.Save
Sheets("18 Data").Select
Sheets("18 Data").Name = "19"
```

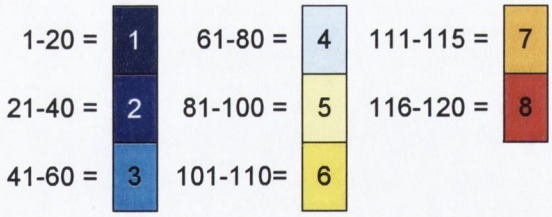
```
        Sheets("20").Name = "18 Data"
    Sheets("Random Numbers").Select
    Columns("CM:CM").Select
    Selection.Copy
    Sheets("18 Data").Select
    Application.Run "FullSheetRC"
    Range("A1").Select
    Application.CutCopyMode = False
    Sheets("18 Data").Select
    Sheets("18 Data").Name = "20"
    ActiveWorkbook.Save
End Sub
```

Appendix H. Meta-Analysis Ranking Results

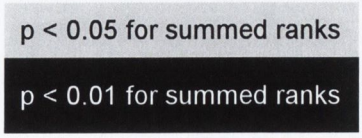
Detailed rank contributions of nine genome scan studies to the meta-analysis under disease Models 1 (A), 2 (B) and 3 (C). The ranks are divided into eight categories for ease of viewing, representing the highest five (8) and next five (7) ranks, followed by the next ten highest ranks (6), and then groups of twenty ranks (5 to 1). The studies are listed on the left, with the name of the first author or (for unpublished studies) the principal investigator (see chapter 6, table 6.1), by reverse order of the number of affected individuals and pedigrees contributed by each under each disease model. The bins are presented in genomic order, with the weighted summed rank (see text, chapter 6, section 6.2.4) , and the order of the summed rank indicated above. Bins achieving bin-wise (BW; see chapter 6, section 6.2.4) significance are shaded (see legend). Indicated below are the distal marker of each bin, and it's position in cM according to the Marshfield genetic map (Mfd cM).

Legend:

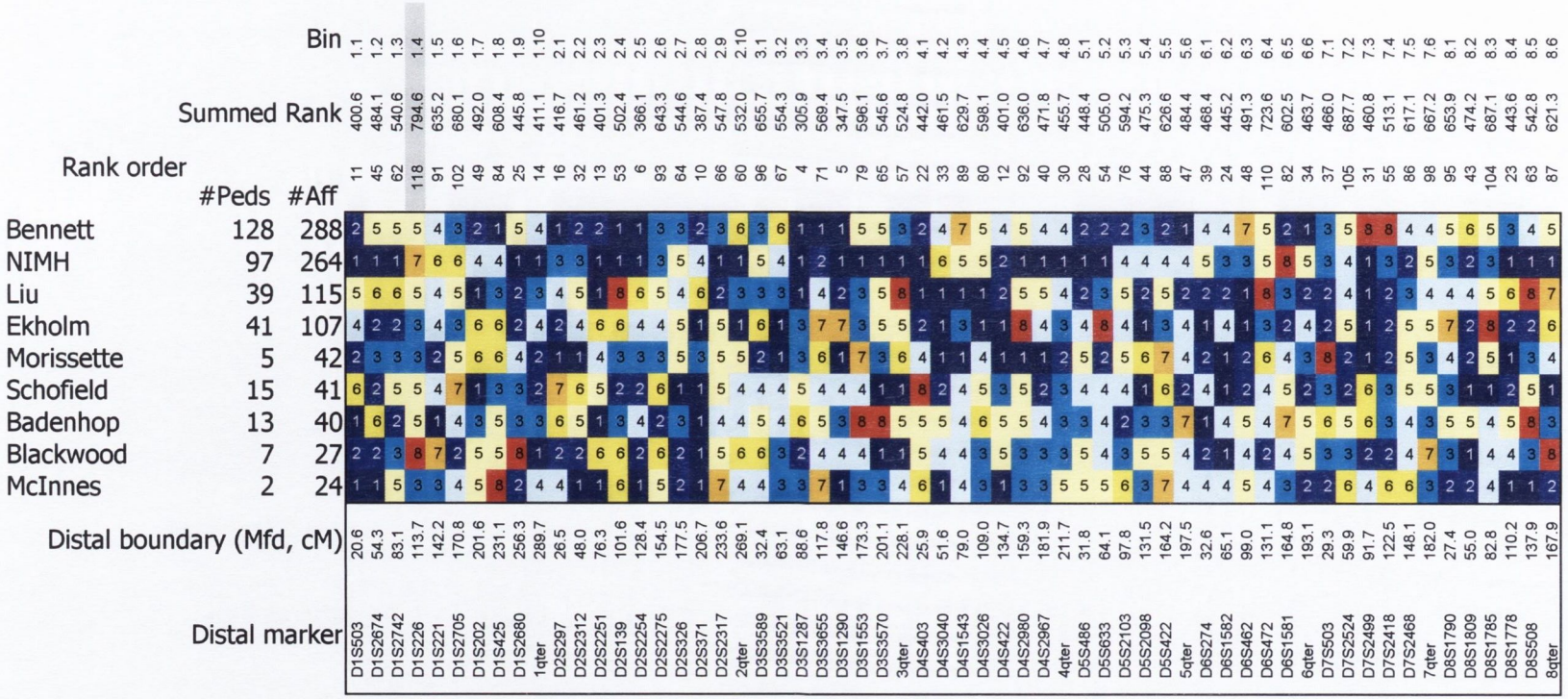
Ranks within studies:



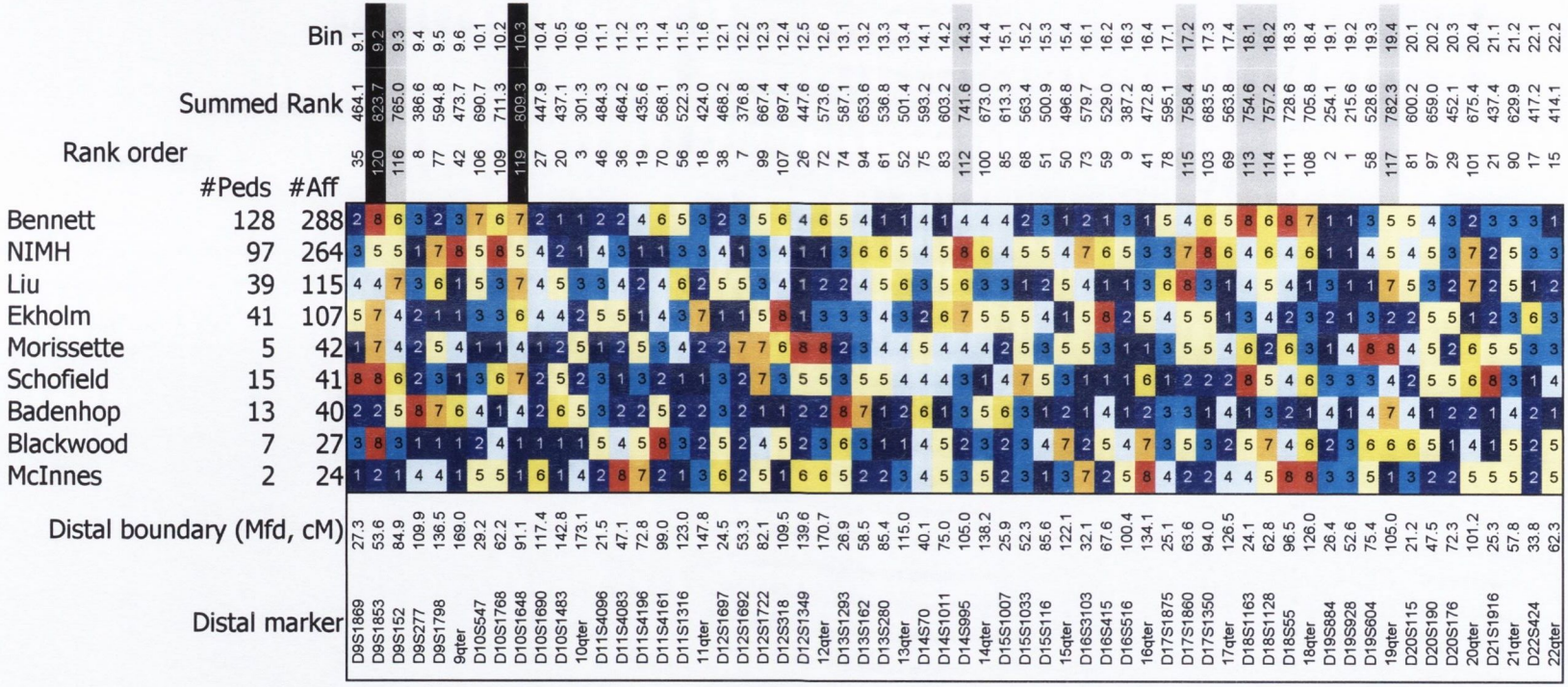
Summed rank significance (weighted analysis):



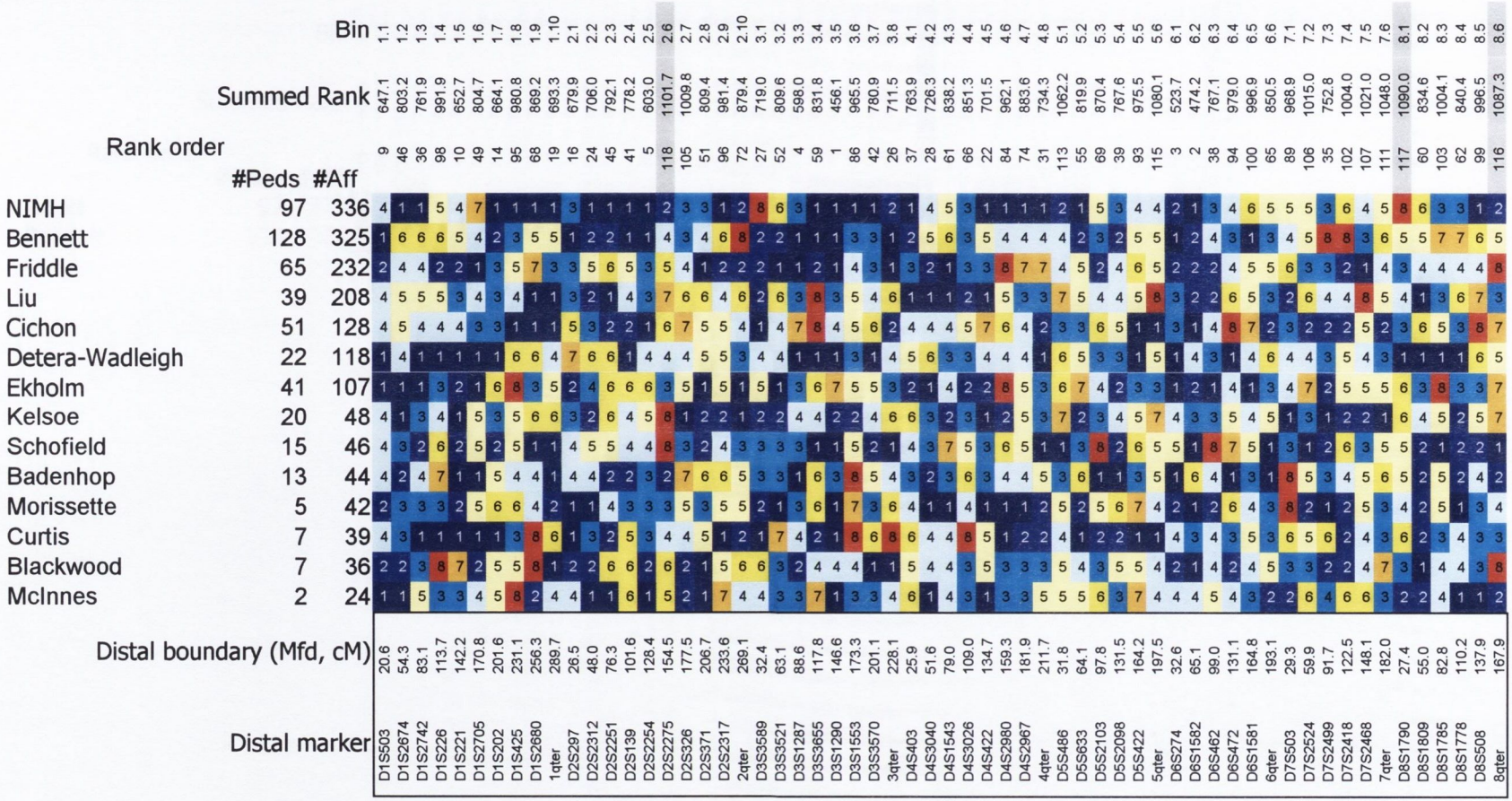
A) Model 1


Bin
Summed Rank
Rank order
#Peds
#Aff
Bennett 128 288
NIMH 97 264
Liu 39 115
Ekholm 41 107
Morissette 5 42
Schofield 15 41
Badenhop 13 40
Blackwood 7 27
McInnes 2 24
Distal boundary (Mfd, cM)
Distal marker
D1S503
D1S2674
D1S2742
D1S226
D1S221
D1S2705
D1S202
D1S425
D1S2680
1qter
D2S297
D2S2312
D2S2251
D2S139
D2S2254
D2S2275
D2S326
D2S371
D2S2317
2qter
D3S3589
D3S3521
D3S1287
D3S3655
D3S1290
D3S1553
D3S3570
3qter
D4S403
D4S3040
D4S1543
D4S3026
D4S422
D4S2980
D4S2967
4qter
D5S486
D5S633
D5S2103
D5S2098
D5S422
5qter
D6S274
D6S1582
D6S462
D6S472
D6S1581
6qter
D7S503
D7S2524
D7S2499
D7S2418
D7S2468
7qter
D8S1790
D8S1809
D8S1785
D8S1778
D8S508
8qter

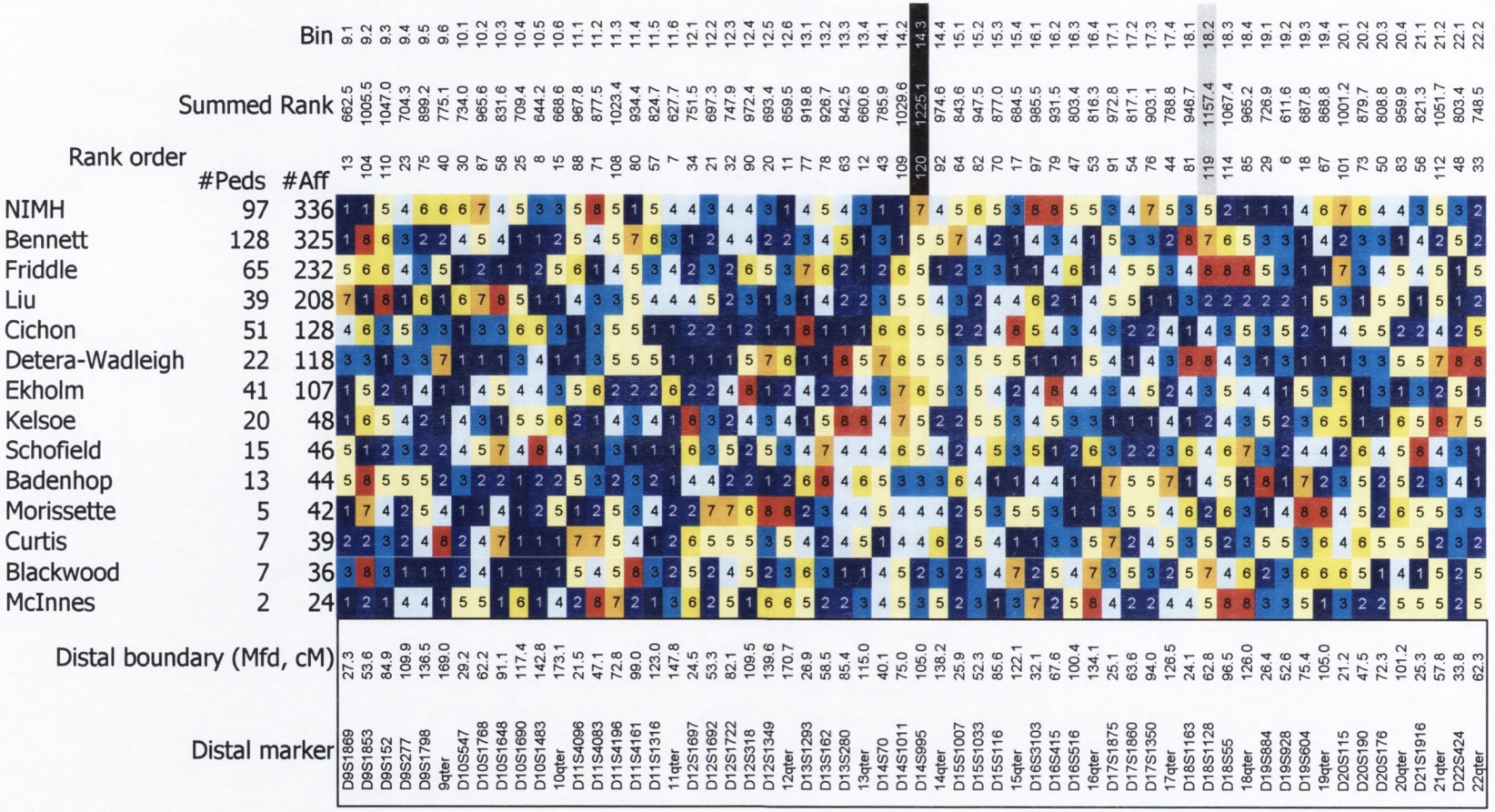
A) Model 1 (cont.)



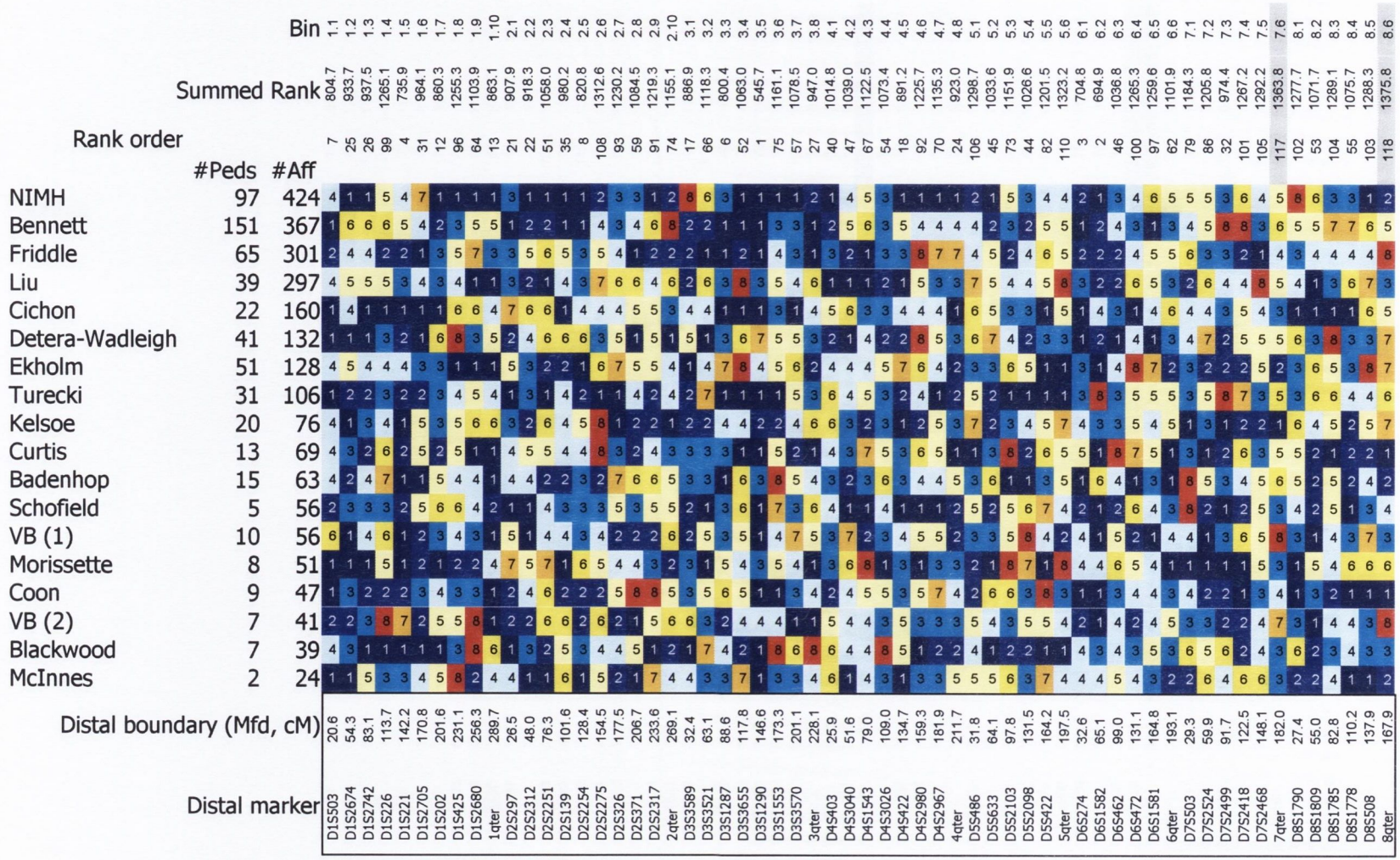
B) Model 2


Bin
Summed Rank
Rank order
#Peds
#Aff
NIMH 97 336
Bennett 128 325
Friddle 65 232
Liu 39 208
Cichon 51 128
Detera-Wadleigh 22 118
Ekholm 41 107
Kelsoe 20 48
Schofield 15 46
Badenhop 13 44
Morissette 5 42
Curtis 7 39
Blackwood 7 36
McInnes 2 24
Distal boundary (Mfd, cM)
Distal marker
D1S503
D1S2674
D1S2742
D1S226
D1S221
D1S2705
D1S202
D1S425
D1S2680
1qter
D2S297
D2S2312
D2S2251
D2S139
D2S2254
D2S2275
D2S326
D2S371
D2S2317
2qter
D3S3589
D3S3521
D3S1287
D3S3655
D3S1290
D3S1553
D3S3570
3qter
D4S403
D4S3040
D4S1543
D4S3026
D4S422
D4S2980
D4S2967
4qter
D5S486
D5S633
D5S2103
D5S2098
D5S422
5qter
D6S274
D6S1582
D6S462
D6S472
D6S1581
6qter
D7S503
D7S2524
D7S2499
D7S2418
D7S2468
7qter
D8S1790
D8S1809
D8S1785
D8S1778
D8S508
8qter

B) Model 2 (cont.)



C) Model 3


Bin
Summed Rank
Rank order
#Peds
#Aff
NIMH 97 424
Bennett 151 367
Friddle 65 301
Liu 39 297
Cichon 22 160
Detera-Wadleigh 41 132
Ekholm 51 128
Turecki 31 106
Kelsoe 20 76
Curtis 13 69
Badenhop 15 63
Schofield 5 56
VB (1) 10 56
Morissette 8 51
Coon 9 47
VB (2) 7 41
Blackwood 7 39
McInnes 2 24
Distal boundary (Mfd, cM)
Distal marker
D1S503 D1S2674 D1S2742 D1S226 D1S221 D1S2705 D1S202 D1S425 D1S2680 1qter D2S297 D2S2312 D2S2251 D2S139 D2S2254 D2S2275 D2S326 D2S371 D2S2317 2qter D3S3589 D3S3521 D3S1287 D3S3655 D3S1290 D3S1553 D3S3570 3qter D4S403 D4S3040 D4S1543 D4S3026 D4S422 D4S2980 D4S2967 4qter D5S486 D5S633 D5S2103 D5S2098 D5S422 5qter D6S274 D6S1582 D6S462 D6S472 D6S1581 6qter D7S503 D7S2524 D7S2499 D7S2418 D7S2468 7qter D8S1790 D8S1809 D8S1785 D8S1778 D8S508 8qter

C) Model 3 (cont.)