MOLECULAR GENETICS OF THE 8p21-22

SCHIZOPHRENIA SUSCEPTIBILITY LOCUS

Ekaterini Blaveri

Molecular Psychiatry Laboratory
Department of Psychiatry and Behavioural Sciences
Royal Free and University College London Medical School
The Windeyer Institute of Medical Sciences
46 Cleveland Street
London W1P 6DB
United Kingdom

This thesis is submitted for the degree of Doctor of Philosophy at

University College London

October 2000
TABLE OF CONTENTS

i  Acknowledgements 16
ii Abbreviations 17
iii Abstract 19

CHAPTER 1: INTRODUCTION

1.1  INTRODUCTION 22
1.2  HISTORICAL BACKGROUND 22
1.3  CLINICAL FEATURES
   1.3.1  Thought content and thought process 23
   1.3.2  Perception 24
   1.3.3  Mood, feelings and affect 25
   1.3.4  Abnormal behaviour 25
1.4  DIAGNOSIS OF SCHIZOPHRENIA 26
   1.4.1  Positive and negative symptoms 27
   1.4.2  Subtypes of schizophrenia 27
   1.4.3  Schizophrenia-like syndromes 28
   1.4.4  Diagnostic categories with respect to molecular genetic studies 29
1.5  EPIDEMIOLOGY 30
   1.5.1  Family studies 31
   1.5.2  Twin studies 33
   1.5.3  Adoption studies 34
1.6  GENETIC TRANSMISSION MODELS 36
   1.6.1  Single Major Locus (SML) model 36
2.2.1 Linkage analysis studies

2.2.2 Allelic association studies

2.3 LABORATORY METHODS

2.3.1 Laboratory reagents

2.3.2 Bacterial strains and growth conditions
   2.3.2.1 Bacterial strains
   2.3.2.2 Propagation of Bacteria
   2.3.2.3 Preparation of electrocompetent cells
   2.3.2.4 Transformation by electroporation

2.3.3 DNA extraction techniques
   2.3.3.1 Isolation of genomic DNA from blood
   2.3.3.2 Small Scale Plasmid DNA extraction Top10F' from cells
   2.3.3.3 Large Scale Plasmid DNA Extraction

2.3.4 Genotyping
   2.3.4.1 PCR reactions
   2.3.4.2 Polyacrylamide Gel Electrophoresis
   2.3.4.3 Reading genotypes and genotype checking
   2.3.4.4 Genotyping of an MboII RFLP in the NAT1 gene

2.3.5 Cloning techniques
   2.3.5.1 Agarose Gel Electrophoresis
   2.3.5.2 Restriction Digests
   2.3.5.3 DNA Ligations
   2.3.5.4 DNA blotting
   2.3.5.5 Hybridization
   2.3.5.6 DNA Sequencing
2.3.5.7 Library screening 88

2.4 STATISTICAL ANALYSIS 92
2.4.1 Parametric linkage analysis 92
2.4.2 “Model-free” linkage analysis 93
2.4.3 Allelic association analysis using CLUMP 94
2.4.4 Pair-wise linkage disequilibrium calculations using EH 95

CHAPTER 3: GENETIC LINKAGE STUDIES ON CHROMOSOME 8p21-22
3.1 INTRODUCTION 97
3.2 RESULTS 100
3.3 DISCUSSION 107

CHAPTER 4: ISOLATION OF A NEW POLYMORPHIC MARKER IN A CANDIDATE GENE AND AN ALLELIC ASSOCIATION STUDY BETWEEN SCHIZOPHRENIA AND TWO CANDIDATE GENES: REPRONOCICEPTIN (PNOC) AND NEURONAL CHOLINERGIC RECEPTOR, NICOTINIC, ALPHA POLYPEPTIDE 2 (CHARNA2)
4.1 INTRODUCTION 109
4.2 ISOLATION OF A (CA)$_n$ REPEAT 112
4.2.1 RPCI1 PAC library screening 112
4.2.2 Isolation of a PAC clone containing the human PNOC gene 113
4.2.3 Identification of a (CA)$_n$ repetitive element in the PAC clone containing the human PNOC gene

4.3 CHARACTERIZATION OF THE NOVEL (CA)$_n$ REPEAT

4.4 CASE-CONTROL ALLELIC ASSOCIATION STUDY OF THE PNOC GENE

4.5 CASE-CONTROL ALLELIC ASSOCIATION STUDY OF THE CHRNA2 GENE

4.6 DISCUSSION

CHAPTER 5: FINE MAPPING OF THE CHROMOSOME 8p21-22 REGION

BY ALLELIC ASSOCIATION STUDIES

5.1 INTRODUCTION

5.2 A PRELIMINARY ALLELIC ASSOCIATION STUDY OF THE 8p21-22 REGION IMPLICATED IN LINKAGE STUDIES OF SCHIZOPHRENIA

5.3 IDENTIFICATION OF NOVEL POLYMORPHISMS IN THE 8p21.3-22 REGION

5.4 ALLELIC ASSOCIATION STUDIES IN THE LONDON (UK) CASE-CONTROL SAMPLE WITH NOVEL POLYMORPHISMS IN THE 8p21.3-22 REGION

5.5 REPLICATION STUDIES IN TWO CASE-CONTROL SAMPLES FROM SCOTLAND

5.6 ASSESMENT OF ALLELIC ASSOCIATION FROM ALL AVAILABLE SAMPLES
CHAPTER 6: A CANDIDATE GENE STUDY OF HUMAN PERICENTRIOLAR MATERIAL 1 GENE (PCM1)

6.1 INTRODUCTION 175
6.2 GENOMIC ORGANIZATION OF PCM1 177
6.3 ANALYSIS OF PCM1 DNA VARIATION BY AUTOMATED BI-DIRECTIONAL DNA SEQUENCING OF AMPLIFIED EXONS 189
   6.3.1 Analysis of exon 4 and its exon/intron boundaries 191
   6.3.2 Analysis of exon 5 and its exon/intron boundaries 201
6.4 DISCUSSION 211

CHAPTER 7: DISCUSSION 214

REFERENCES 223

APPENDIX 1 Pedigree diagrams for the 16 families used in the study 242
INDEX OF FIGURES

CHAPTER 1

Figure 1.1 Lifetime risks of developing schizophrenia for first degree (parents, sibs, children), second degree (uncles, nephews, grand children, half sibs, DZ twins) and third degree (first cousins) relatives of schizophrenics (adapted from Gottesman, 1991) 32

CHAPTER 2

Figure 2.1 PCR using tailed primers 86
Figure 2.2 pCYPAC2 vector used to construct the PAC RPCI1 library 89
Figure 2.3 PCR screening of the RPCI1 library 91

CHAPTER 4

Figure 4.1 GDB Comprehensive map showing the positions of PNOC and CHRNA2 on 8p21 111
Figure 4.2 RPCI1 library screening by PCR 114
Figure 4.3 A. FISH of the PAC 122n11 DNA 115
   B. PCR amplification of the PAC 122n11 DNA (P) using the same primers and conditions used for screening the PAC library 115
Figure 4.4 H.sapiens prepronociceptin exon 3 sequence 117
Figure 4.5 BLAST 2.0 output after searching for a match for the 97 bp of sequence obtained from cycle sequencing PAC 122n11 DNA clone against the 'nr' (non redundant) database 118
Figure 4.6 Southern blot analysis of the PAC 122n11 DNA clone 120
Figure 4.7 The DNA sequence containing a (CA)$_{26}$ repeat obtained from a single subclone of a PstI fragment of PAC 122n11 DNA

Figure 4.8 BLAST 2.0 output after searching for a match for the 589bp of sequence lying next to the novel (CA)$_{n}$ repeat against the 'htgs' database (unfinished Human genome sequences database)

CHAPTER 5

Figure 5.1 Part of the Whitehead contig WC8.1 map and the Marshfield map of the 8p21-22 region

Figure 5.2 Part of the NT_000501 contig (1,210,381 bp) on chromosome 8p21.3-22

CHAPTER 6

Figure 6.1 The longest open reading frame of Homo sapiens pericentriolar material 1 (PCM1) mRNA and the translated product

Figure 6.2 LI-COR AlignIR alignment report for a novel DNA variant identified in intronic sequence 3’ to exon 4 of PCM1 (position 80254 of the AB020866 clone)

Figure 6.3 Chromatograms of the sequence containing a novel DNA variant identified in intronic sequence 3’ to exon 4 of PCM1 (position 80254 of the AB020866 clone)

Figure 6.4 LI-COR AlignIR alignment report for a novel DNA variant identified in exon 4 of PCM1 (position 80123 of the AB020866 clone)
Figure 6.5 Chromatograms of the sequence containing a novel DNA variant identified in exon 4 of PCM1 (position 80123 of the AB020866 clone) 200

Figure 6.6 LI-COR AlignIR alignment report for a novel DNA variant identified in intronic sequence 5' to exon 5 of PCM1 (position 87366 of the AB020866 clone) 203

Figure 6.7 Chromatograms of the sequence containing a novel DNA variant identified in intronic sequence 5' to exon 5 of PCM1 (position 87366 of the AB020866 clone) 206

Figure 6.8 LI-COR AlignIR alignment report for a novel DNA variant identified in intronic sequence 5' to exon 5 of PCM1 (position 87507 of the AB020866 clone) 207

Figure 6.9 Chromatograms of the sequence containing a novel DNA variant identified in intronic sequence 5' to exon 5 of PCM1 (position 87507 of the AB020866 clone) 210
INDEX OF TABLES

CHAPTER 1

Table 1.1 Statistically significant and other supportive linkage data for a schizophrenia susceptibility locus on chromosome 8p 56

Table 1.2 Statistically significant and other supportive linkage data for a schizophrenia susceptibility locus obtained from genome-wide searches 61

CHAPTER 3

Table 3.1 Information on markers used in the linkage study on chromosome 8p21-22 99

Table 3.2 Two-point MLINK lod and lod2 scores between markers at 8p21-22 and schizophrenia under the DOMS affection model and dominant mode of inheritance 101

Table 3.3 Two-point MLINK lod and lod2 scores between markers at 8p21-22 and schizophrenia under the DOMS affection model and recessive mode of inheritance 102

Table 3.4 Two-point MLINK lod and lod2 scores between markers at 8p21-22 and schizophrenia under the DOMSS affection model and dominant mode of inheritance 103

Table 3.5 Two-point MLINK lod and lod2 scores between markers at 8p21-22 and schizophrenia under the DOMSS affection model and recessive mode of inheritance 104
Table 3.6 Two-point MFLINK lod scores between schizophrenia and markers on chromosome 8p21-22 under the DOMS and DOMSS affection models

CHAPTER 4

Table 4.1 Size in base pairs and frequency of the alleles of the D8S2611 microsatellite marker near the human prepronociceptin locus 125

Table 4.2 Allele counts and frequencies (in parentheses) at marker D8S2611 near the PNOC gene 128

Table 4.3 Allele counts and frequencies (in parentheses) at marker D8S131 129

Table 4.4 Allele counts and frequencies (in parentheses) at marker D8S131P 129

CHAPTER 5

Table 5.1 Information on markers used in the preliminary allelic association studies on chromosome 8p21-22 135

Table 5.2 Allele counts and frequencies (in parentheses) at marker AFM333th1 from the London (UK) case-control sample 140

Table 5.3 Allele counts and frequencies (in parentheses) at marker D8S261 from the London (UK) case-control sample 140

Table 5.4 Allele counts and frequencies (in parentheses) at marker D8S136 from the London (UK) case-control sample 141

Table 5.5 Allele counts and frequencies (in parentheses) at marker D8S1771 from the London (UK) case-control sample 141

Table 5.6 Allele counts and frequencies (in parentheses) at marker D8S1145 from the London (UK) case-control sample 142
Table 5.7  Allele counts and frequencies (in parentheses) for the MboII RFLP at the NAT1 gene from the London (UK) case-control sample

Table 5.8  CLUMP test results for preliminary association studies between schizophrenia and alleles at polymorphic loci on chromosome 8p21-22 in the London (UK) case-control sample

Table 5.9  The clones that constitute the NT_000501 contig (121,0381 bases) of the human 8p21.3-p22 sequence

Table 5.10 Primer sequences and allele set of the novel microsatellite repeats identified in the NT_000501 contig on chromosome 8p21.3-22

Table 5.11  Allele counts and frequencies (in parentheses) at marker D8S2612 from the London (UK) case-control sample

Table 5.12  Allele counts and frequencies (in parentheses) at marker D8S2617 from the London (UK) case-control sample

Table 5.13  Allele counts and frequencies (in parentheses) at marker D8S2616 from the London (UK) case-control sample

Table 5.14  Allele counts and frequencies (in parentheses) at marker D8S2615 from the London (UK) case-control sample

Table 5.15  Allele counts and frequencies (in parentheses) at marker D8S2614 from the London (UK) case-control sample

Table 5.16  Allele counts and frequencies (in parentheses) at marker D8S2613 from the London (UK) case-control sample

Table 5.17  Allele counts and frequencies (in parentheses) at marker D8S2616 from the London (UK) case-control sample
Table 5.18  CLUMP test results for association between schizophrenia and alleles at novel polymorphic loci on chromosome 8p21.3-22 in the London (UK) case-control sample

Table 5.19  Allele counts and frequencies (in parentheses) at marker D8S261 from the first case-control sample from Scotland

Table 5.20  Allele counts and frequencies (in parentheses) at marker D8S261 from the second case-control sample from Scotland

Table 5.21  Allele counts and frequencies (in parentheses) at marker D8S261 from both case-control samples from Scotland

Table 5.22  Allele counts and frequencies (in parentheses) at marker D8S261 from all three case-control samples

Table 5.23  Allele counts and frequencies (in parentheses) at marker D8S2616 from the first case-control sample from Scotland

Table 5.24  Allele counts and frequencies (in parentheses) at marker D8S2616 from the second case-control sample from Scotland

Table 5.25  Allele counts and frequencies (in parentheses) at marker D8S2616 from both case-control samples from Scotland

Table 5.26  Allele counts and frequencies (in parentheses) at marker D8S2616 from all three case-control samples

Table 5.27  Allele counts and frequencies (in parentheses) at marker D8S2615 from the first case-control sample from Scotland

Table 5.28  Allele counts and frequencies (in parentheses) at marker D8S2615 from the second case-control sample from Scotland
Table 5.29 Allele counts and frequencies (in parentheses) at marker D8S2615 from both case-control samples from Scotland 165

Table 5.30 Allele counts and frequencies (in parentheses) at marker D8S2615 from all three case-control samples 165

Table 5.31 CLUMP test results for association between schizophrenia and alleles at polymorphic loci on chromosome 8p21.3-22 from the first case-control sample from Scotland 166

Table 5.32 CLUMP test results for association between schizophrenia and alleles at polymorphic loci on chromosome 8p21.3-22 from the second case-control sample from Scotland 166

Table 5.33 CLUMP test results for association between schizophrenia and alleles at polymorphic loci on chromosome 8p21.3-22 for the combined case-control samples from Scotland 166

Table 5.34 CLUMP test results for association between schizophrenia and alleles at polymorphic loci in chromosome 8p21.3-22 from all three case-control samples 167

Table 5.35 Linkage disequilibrium between paired microsatellite markers on chromosome 8p21.3-22 169

Table 5.36 Linkage disequilibrium between paired microsatellite markers on chromosome 8p21.3-22 from the combined Scottish sample 170

CHAPTER 6

Table 6.1 Intron-exon boundary sequences of PCM1 184
Acknowledgements

Firstly I would like to thank my supervisor Professor Hugh Gurling for giving me the opportunity to realise one of my biggest dreams which was to work in the ever so challenging field of Molecular Genetics and offer my small contribution to it. I would also like to thank him for his support, guidance and enthusiasm throughout this project.

I am extremely grateful to Dr Dave Curtis for performing all the statistical analysis for this thesis and for all his help and guidance in the understanding of the statistical issues dealt in this study.

I would like to thank my colleagues, Dr Gursharan Kalsi, Dr Andrew McQuillan and Dr Karl D. Murray for generously offering me their help and sharing their expertise with me. Also, Dr Margaret Fox for performing the FISH analysis.

I am grateful to Dr Douglas Blackwood and Dr Marian Thompson for providing me with the Scottish DNA samples.

I would like to thank my parents and my sister, Mina and brother in law, Vaggelis for their constant love and encouragement when I needed it most. Also, many of my friends for their support and kind words.

My special thanks to Dr Michael Robinson for his friendship, help, support and patience throughout the completion of this thesis.

Finally, I would like to express my extreme gratefulness to the medical doctors who were involved in the collection of the blood samples but most of all to the patients and their families without whom this work would not have been possible.
### ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>APM</td>
<td>affected pedigree member</td>
</tr>
<tr>
<td>ARP</td>
<td>affected relative pair</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>cpm</td>
<td>counts per minute</td>
</tr>
<tr>
<td>dATP</td>
<td>deoxyadenosine triphosphate</td>
</tr>
<tr>
<td>dCTP</td>
<td>deoxycytidine triphosphate</td>
</tr>
<tr>
<td>dGTP</td>
<td>deoxyguanosine triphosphate</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxynucleotide triphosphate</td>
</tr>
<tr>
<td>DSMIII-R</td>
<td>Diagnostic and Statistical Manual of Mental Disorders, third edition, revised</td>
</tr>
<tr>
<td>DSMIV</td>
<td>Diagnostic and Statistical Manual of Mental Disorders, fourth edition</td>
</tr>
<tr>
<td>dTTP</td>
<td>deoxythymidine triphosphate</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediamine tetra acetic acid</td>
</tr>
<tr>
<td>g</td>
<td>gram</td>
</tr>
<tr>
<td>hrs</td>
<td>hours</td>
</tr>
<tr>
<td>IBD</td>
<td>identical by descent</td>
</tr>
<tr>
<td>IBS</td>
<td>identical by state</td>
</tr>
<tr>
<td>lod</td>
<td>logarithm of the odds</td>
</tr>
<tr>
<td>M</td>
<td>molar</td>
</tr>
<tr>
<td>mg</td>
<td>milligram</td>
</tr>
<tr>
<td>μg</td>
<td>microgram</td>
</tr>
<tr>
<td>min</td>
<td>minute</td>
</tr>
<tr>
<td>ml</td>
<td>millilitre</td>
</tr>
<tr>
<td>μl</td>
<td>microlitre</td>
</tr>
<tr>
<td>mM</td>
<td>millimolar</td>
</tr>
<tr>
<td>ng</td>
<td>nanogram</td>
</tr>
<tr>
<td>NPL</td>
<td>nonparametric linkage</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
</tbody>
</table>
pmol: picomole
RDC: Research Diagnostic Criteria
RFLP: restriction fragment length polymorphism
SADS-L: Schedule for Affective Disorder and Schizophrenia, lifetime version
SDS: sodium dodecyl sulphate
SML: single major locus
SSC: standard sodium citrate
STRs: short tandem repeats
TAE: tris-acetate/EDTA
TBE: tris-borate/EDTA
TDT: transmission disequilibrium test
TE: tris/EDTA
Tris: tris(hydroxymethyl)aminomethane
UV: ultraviolet
VNTRs: variable number of tandem repeats
Abstract

Evidence for a susceptibility locus for schizophrenia mapping to chromosome 8p21-p22 has been reported by several investigators. This thesis will describe work carried out in order to narrow down the susceptibility area on this chromosomal region making it amenable to positional cloning and positional candidate studies.

A linkage study was performed in 16 English families containing cases of schizophrenia with five microsatellite polymorphisms on chromosome 8p21-p22. No evidence of linkage was obtained between any of these markers and schizophrenia in this family sample.

Next, case-control allelic association studies were performed in order to assess the candidacy of prepronociceptin (PNOC) and neuronal nicotinic acetylcholine receptor subunit alpha 2 (CHRNA2) genes, in this region. No evidence of allelic association was found between the DNA variations at or near the two genes that were examined and schizophrenia in the London (UK) case-control sample.

Linkage disequilibrium mapping was also performed in order to narrow down the region of chromosome 8p21-22 that is implicated in schizophrenia. An allelic association was obtained between microsatellite marker D8S261 and schizophrenia (CLUMP T3: $\chi^2=9.9929$, $p=0.01$) in the London (UK) case-control sample. Analysis of additional polymorphisms covering a region of ~700 kb around this marker in our population sample also revealed significant evidence for allelic association between schizophrenia and two novel, neighbouring polymorphisms, D8S2616 (CLUMP T1: $\chi^2=19.9236$, $p=0.024$) and D8S2615 (CLUMP T1: $\chi^2=15.1777$, $p=0.004$). The three markers, D8S261, D8S2616 and D8S2615 cover a region of approximately 108 kb on
the 8p21.3-22 region and also showed statistically significant marker-to-marker linkage disequilibrium.

A replication study was attempted in two case-control samples of Scottish ancestry. The first case-control Scottish sample consisting of 100 cases and 100 controls did not show evidence of allelic association between any of the three previously associated markers and schizophrenia. The second case-control Scottish sample consisting again of 100 cases and 100 controls demonstrated evidence of allelic association only between D8S2616 and schizophrenia (CLUMP T1: $\chi^2=16.3893$, $p=0.043$). The two case-control Scottish samples were combined but failed to demonstrate allelic association with any of the three positive markers.

Finally, significant allelic association was obtained by the CLUMP T3 statistic between D8S261 ($\chi^2=7.9706$, $p=0.047$), D8S2616 ($\chi^2=8.1593$, $p=0.033$), D8S2615 ($\chi^2=5.9546$, $p=0.05$) and schizophrenia when all three case-control samples were combined together.

The microsatellite markers, D8S261 and D8S2616 are localised in the intronic region of the pericentriolar material 1 (PCM1) gene on chromosome 8p21.3-22 while D8S2615 lies 75 kb upstream the translation initiation codon of this gene. The study of the DNA variation of this gene is underway in order to assess its possible involvement in the liability to schizophrenia. Some preliminary data are described in this thesis.
CHAPTER 1

INTRODUCTION
1.1 INTRODUCTION

The schizophrenias are a collection of mental disorders consisting of variable clusters of symptoms. Variability of such symptoms is also associated with considerable prognostic variation. The schizophrenias are mainly characterized by a disruption in the processes of thought, perception, behaviour, affect as well as social and occupational deterioration. In about one third of cases, it is a chronic disorder perhaps with periods of relative improvement but evidence of continuous deterioration. In a further one third there is good recovery between episodes with treatment while in another one third there can be spontaneous remission without treatment. The onset can be sudden or insidious and because of the early age of onset, schizophrenia is responsible for very high levels of morbidity. The mean age of onset is about 23 in males and 27 in females. Schizophrenia is characterized not only by psychopathological features but also by physiological and anatomical changes. Some of these changes can be found in the first and second-degree relatives of schizophrenic probands. The precise aetiology for the schizophrenias is unknown but the most consistent and well-documented evidence is a strong influence of familial factors, which can consist of both genetic and environmental effects.

1.2 HISTORICAL BACKGROUND

Schizophrenia was only demarcated as a specific diagnostic entity at the end of the 19th century. Earlier accounts appear in the literature and in ancient writings although they do not correspond precisely to current diagnosis.
The first clinically characterized cases of schizophrenia were described separately in 1809 by two physicians, John Haslam (1764-1844) in England and Phillipe Pinel (1745-1826) in France.

In 1852, Benedict Morel (1809-1873) introduced the term *démence précoce* in order to describe the illness and to emphasis the deterioration in mental abilities of the patients and the early age of onset of the disease.

Emil Kraepelin (1856-1926) used the same name, *dementia praecox*, to describe the chronic type of illness but also included other types of psychosis such as catatonia, paranoia and hebephrenia, setting the foundations for classifying schizophrenia into the different subtypes that are known today.

Finally, Eugen Bleuler (1857-1939) changed the name to *schizophrenia* in order to emphasis the disruption of the thought processes and feelings of patients as well as to the fact that they can recover.

Kraepelin's and Bleuler's nosological theories have been refined and expanded by modern psychiatrists who have used operational criteria to improve the diagnosis of the schizophrenias in reliable and valid diagnostic instruments.

1.3 CLINICAL FEATURES

Schizophrenia is characterized by symptoms that reflect multiple mental processes:

1.3.1 Thought content and thought process

Delusions, or abnormalities in inferential thinking, are a classic symptom in the schizophrenias. These may be persecutory, grandiose, religious or somatic. The schizophrenic can sometimes believe that some outside force is controlling their
thoughts and behaviours or that they are controlling outside events in some extraordinary fashion. Paranoid delusions commonly observed in patients are of being spied upon, talked about or being at risk from harm. A central feature of schizophrenic delusions is the certainty with which a patient holds these beliefs and the fact that they have no insight into the fact they are false. Patients may lack a clear sense of their own body, of their own thoughts and their own abilities. They can believe in effects and influences from animate and inanimate objects. For example, the patient may believe that other people or the television is talking about them or that they have fused with another object or disintegrated completely.

Thought disturbances found in schizophrenia include loosening of associations, incoherence, tangentiality, flight of ideas, neologisms, echolalia, verbigeration, mutism, thought blocking, impaired attention, overinclusion, illogical ideas, vagueness and poverty of content. The patients however do not seem to be aware that their communication is abnormal.

Disorganised speech, or abnormalities in language, is a common feature of schizophrenia. Such schizophrenic speech is called “thought disorder” and has a tendency to be filled with bizarre or symbolic images the meaning of which is difficult to comprehend. Poverty of speech content refers to a lack of speech during an interview even though the patient may appear to be preoccupied with profound and involving thoughts.

1.3.2 Perception

Hallucinations, or abnormalities in perception, are a prominent feature of schizophrenia. Hallucinations can occur in any of the five sensory modalities, however up to 75% are auditory. The auditory hallucinations may consist of voices of
God or the devil; sometimes they are of neighbours, deceased relatives or unrecognised individuals. Voices may make obscene comments about the patient, and the patient may hold audible conversations with the voices. Hallucinations of smell are quite frequent as well as hallucinations of a somatic kind. Patients experience induced sensations of heat, cold, pain, or electric shock. Visual hallucinations are less common in schizophrenia.

1.3.3 Mood, feelings and affect

These can be grossly reduced, extremely exaggerated, or patently bizarre. Many patients show reduced emotional responses and seem to be indifferent with emotional shallowness. Often the emotional responses of a schizophrenic are inappropriate to the situation. For example, they may smile whilst talking about a morbid subject or show unaccountable anger. The patients may also show marked sensitivity to emotional trauma, being easily hurt by very mildly aggressive or rejecting behaviour by others.

1.3.4 Abnormal behaviour

Disorganized behaviour, or abnormalities in behavioural monitoring and control, is common. First impressions of a schizophrenic can sometimes consist of extreme bizarreness, agitation or withdrawal, exhibited as a set facial expression, lack of sustained eye contact and staring at inanimate objects. Their personal appearance tends to deteriorate and they may exhibit idiosyncratic manners or offensive behaviour. Lack of motivation and will is demonstrated by an inability to continue an occupation or being uninterested in future plans. In chronic patients, stereotypic behaviour may sometimes be present as repetitive patterns of moving or walking,
strange gestures, or endless repetitions of the same phrase or question. Social withdrawal is a very common symptom in schizophrenia.

1.4 DIAGNOSIS OF SCHIZOPHRENIA

Modern diagnostic schemes were developed out of the earlier nosological theories of Kraepelin and Bleuler. The major diagnostic systems for schizophrenia include: Schneider's (Schneider, 1959), the various revisions of the International Classification of Disease (ICD-10, WHO 1996), the St Louis Criteria (also called Feighner's criteria, (Feighner et al., 1972), the Research Diagnostic Criteria (Spitzer et al., 1978b), the Present State Examination (PSE/CATEGO), (Wing et al., 1974) and the various revisions of the American Psychiatric Association's, Diagnostic and Statistical Manual of Mental Disorders (DSM). The current version of DSM, namely, DSM-IV (1994) is a refinement of the previous diagnostic systems criteria based on statistical tests of reliability of diagnosis. The St Louis and DSM-IV criteria require a reduced level of functioning for the diagnosis. The RDC, St Louis and DSM-IV systems stipulate a minimum duration of symptoms, and the St Louis system requires onset of symptoms before the age of 45. DSM-IV is the most widely used, although ICD-10 provides a similar alternative. Both diagnostic schemes are criterion or operationally based, taking a purely phenomenological approach to the symptoms and defining schizophrenia based on a clustering of clinical signs, symptoms and prognosis. Particular emphasis is placed on affective and psychotic symptoms as a defining feature. RDC, DSM-III, DSM-IV and ICD-10 criteria have been found to be
as reliable and valid as diagnostic instruments as those used in other branches of medicine.

1.4.1 Positive and negative symptoms

The symptoms of schizophrenia have sometimes been divided into positive (or florid, productive or type I) and negative (or defect, deficit or type II).

Positive symptoms represent a distortion or exaggeration of normal functions. These include: delusions, hallucinations, agitated behaviour and are associated with acute onset.

Negative symptoms represent a diminution or absence of mental functions that are normally present. These include: affective flattening, social withdrawal, avolition, anhedonia, attentional impairment and poverty of content in thought and speech and are usually associated with an insidious onset and a chronic course of the illness.

However, positive and negative symptoms can be mixed together and can both disappear spontaneously or remit with suitable treatment. Thus, this classification of schizophrenia has little use at the present time.

1.4.2 Subtypes of schizophrenia

Other subtypes of schizophrenia have been defined by the predominant symptomatology at the time of diagnosis.

Paranoid

Paranoid schizophrenia can be the least severe of the subtypes with relatively late age of onset and less deterioration of thought, social behaviour and/or emotional response in some cases. It is mainly characterized by delusions of persecution or
grandeur. However many patients with paranoid schizophrenia have a chronic severe type of illness.

**Catatonic**

Its essential feature is a marked psychomotor disturbance. Patients can be in a state of complete stupor with waxy flexibility and stereotyped behaviour or a decrease in spontaneous movements and activity may occur. However, other patients may switch into an excited catatonic state, which is a state of extreme psychomotor agitation.

**Disorganised (or hebephrenic)**

This can be a severe subtype and usually has an early and insidious onset, a continuous course and may be without significant remissions. It is characterized by disorganised behaviour, thought, speech and flat and highly inappropriate affect.

**Residual**

It is characterized by a history of schizophrenia but the current evaluation of the patients does not present any positive psychotic symptoms. However, there can be evidence of some negative symptoms.

**Undifferentiated**

It presents with multiple psychotic symptoms, which do not fulfil any one of criteria to be categorised as one of the previous subtypes.

**1.4.3 Schizophrenia-like syndromes**

A substantial number of patients do not meet the diagnostic criteria for schizophrenia yet exhibit psychotic symptoms characteristic of that disorder. The most common of these schizophrenia-like phenotypes are:
Schizoaffective disorder

It demonstrates as an admixture of schizophrenic and affective symptoms. Such cases are usually predominantly schizophrenic or bipolar and represent an overlap between the two major psychoses. The existence of a group of individuals suffering from a specific subtype of schizoaffective psychosis has also been confirmed but is probably relatively rare.

Schizophreniform disorder

Usually applies to a “symptomatic” schizophrenia where there is a known organic cause and where the duration is more than one but less than six months.

Brief reactive psychosis

It is diagnosed when schizophrenia-like symptoms have been present for a short period, no more than one month. There is thought to be a clear precipitating factor that remits.

Schizotypal personality disorder

Represents a “forme fruste” of schizophrenia or a partial case, which meets some but not all of the diagnostic criteria for schizophrenia. Affected individuals may be aloof, with eccentric and bizarre behaviour but have only one psychotic symptom. Individuals may be severely or mildly affected. This category is closely related to schizophrenia.

1.4.4 Diagnostic categories with respect to molecular genetic studies

Several similar definitions of the schizophrenia phenotypes are used in different molecular genetic studies. Narrow diagnostic schemes are those that treat only individuals with schizophrenia and schizoaffective disorder as affecteds. Studies including only schizophrenia could be described as using a very narrow spectrum of
classification. Intermediate schemes include the core and spectrum diagnoses such as schizotypal disorder and all other nonaffective psychoses. Broad schemes include the above plus psychotic affective illness and paranoid and schizoid personality disorder. Very broad schemes include any other psychiatric diagnoses found in the relatives of schizophrenics.

1.5 EPIDEMIOLOGY

The lifetime morbid risk (MR) of schizophrenia is approximately 1% (0.7-1.4%) under narrow diagnosis in a wide variety of geographic regions (Jablensky et al., 1992). However, a higher or lower incidence can been found in certain isolated populations (Book et al., 1978), (Eaton, 1991).

The age of onset for the bulk of schizophrenia cases varies between the ages of 15 and 35 although a small proportion of cases are known in children as well as appearing later in life (Gottesman, 1991).

Gender differences have been observed with males having an earlier age of onset by 2-3 years and being more severely affected than females.

Numerous risk factors that may or may not have a genetic basis have been implicated in the aetiology of schizophrenia including a winter-spring season of birth, perinatal or obstetrical complications, urban birth and rearing and/or viral infection (McDonald & Murray, 2000). However, none of these factors shows consistent or large effects on disease risk. It is possible that some of these factors are confounded, or interact with each other or genetic factors.
1.5.1 Family studies

Studies of family members constitute the basic sampling design in determining the genetic aetiology of a trait. It involves collecting information about the risk of developing the disease in specific relative sets, within twins (special case) or in entire pedigrees and compares it with the risk in the general population.

The first systematic family studies of schizophrenia were performed in the early 20th century. These studies demonstrated that different subtypes frequently recur within the same family and regardless of the subtype for the original proband, siblings were at a much higher risk for developing schizophrenia than the risk for the general population. Data from the pooling of information from about 40 studies conducted in Western Europe from 1920 to 1987 was analysed by Gottesman et al. (Figure 1.1) (Gottesman, 1991). These take show that the relatives of schizophrenics have a higher risk of developing schizophrenia than the general population (1%), with the non age corrected risk being highest amongst MZ twins (48%) and offspring of two schizophrenic parents (46%) and lowest amongst third degree relatives (2%). It is also noteworthy that apart from the degree of genetic relationship, the risk varies directly with the severity of the proband’s illness and the number of other relatives affected.

Although the validity of the previous studies has been challenged for not employing explicit operational diagnostic criteria, recent studies using more sophisticated sampling methods and narrowly defined diagnostic criteria provide added support to the earlier data (Kendler et al., 1985), (Kendler & Diehl, 1993).

It is clear that family studies provide strong evidence for the familial nature of schizophrenia. However, this does not necessarily imply that it is due to genetic factors. Further disentanglement of environmental, cultural and/or genetic factors that
Figure 1.1 Lifetime risks of developing schizophrenia for first degree (parents, sibs, children), second degree (uncles, nephews, grand children, half sibs, DZ twins) and third degree (first cousins) relatives of schizophrenics (adapted from Gottesman, 1991).
may contribute to the familiar clustering will be provided by two other types of study, twin and adoption studies.

1.5.2 Twin studies

MZ twins share a common set of genes whereas fraternal or dizygotic twins (DZ) twins share on average 50% of their genes. If one assumes that on average, both types of twins share the same common and specific environment, then a significantly higher concordance for a disorder in MZ than DZ twins would indicate a genetic aetiology. The above concepts form the basis of what is called the classical twin study.

Such studies have been of importance in establishing a genetic contribution to the aetiology of schizophrenia. Summarized data from European register-based studies between 1963 and 1987 produced a non age corrected weighted average probandwise concordance rate of 48% in MZ compared with 17% in DZ twins (Gottesman, 1991).

Four recent European studies and one Japanese study substantiate the earlier findings producing probandwise concordance rates of 41-65% in MZ and 0-28% in DZ twins (Klaning et al., 1996), (Cannon et al., 1998), (Franzek & Beckmann, 1998), (Cardno et al., 1999), (Cardno & Gottesman, 2000).

The difference between the probandwise concordance rates for MZ and DZ twins indicates an important contribution from genetic variation to disease susceptibility. However, the relatively high discordance rate in MZ twins indicates the presence of incomplete penetrance due to chronogenetic and/or environmental effects that are ill understood.
Another way of evaluating the magnitude of the genetic contribution is by studying the offspring of discordant MZ twins. Where this has been done (Gottesman & Bertelsen, 1989), the rate in the offspring of affected and unaffected MZ twins have been the same or over 20%, providing further evidence of incomplete penetrance.

Finally, MZ twins that have been reared apart, although very rare, are extremely valuable because they combine advantages of both the twin and adoption study designs. Results from such studies indicate a higher number of concordant than discordant pairs, which may indicate the non-familiarity of environmental factors in the aetiology of schizophrenia (Kringlen, 1990). It must be stressed however that more studies are necessary in order to validate these results.

1.5.3 Adoption studies

Adoption studies complement the finding of genetic factors in twin studies of the aetiology of schizophrenia because, unlike the twin method, genes and environment are much better separated. There are mainly two types of adoption studies: the high-risk adoptees approach and the adoptees' families approach.

In the high-risk adoptees approach, adopted-away children of schizophrenic patients (usually mothers) are studied for an increased risk of developing schizophrenia. If these children develop the disease more often than control adoptees then their liability can be attributed to genetic factors. The first such study was performed by (Heston, 1966) and the results show a higher risk of schizophrenia in foster home and institution reared children of mothers with schizophrenia than in the case matched controls. A more recent study conducted in Finland substantiates this result (Tienari et al., 1994).
The crossfostering adoption study method takes exactly the opposite approach than the high-risk adoptees method by examining the offspring of normal biological parents reared by adoptive parents with severe psychopathology or psychosis. The only study of this kind was reported by Wender (Wender et al., 1974) and showed that such children did not have an increased risk to develop schizophrenia.

In the adoptees’ families approach, the biological and adoptive relatives of individuals who were adopted at an early age before exhibiting any signs of schizophrenia are studied for an increased risk of developing the disease. The biological and adoptive relatives of carefully matched control adoptees with no history of mental illness are also studied for comparison. If there is a higher incidence of schizophrenia among members of the biological families of adoptees with schizophrenia then this due to shared genetic factors. The most conclusive results come from combined data of two independent studies both of which made use of the total registers of Danish adoptees. These were: a Danish-American collaboration, which used a nationwide sample (Kety et al., 1994), (Kendler et al., 1994) and a Danish study by Kety and his colleagues who used an independent sample from the provinces of Copenhagen (the Provincial sample). The results from all three studies are consistent and demonstrate a highly significant concentration of schizophrenia among the biological relatives of chronic schizophrenic adoptees. The results stand after reanalysis of the data using DSM-III diagnosis (Kendler et al., 1994).
1.6 GENETIC TRANSMISSION MODELS

Family, twin and adoption studies demonstrate a genetic contribution to the aetiology of schizophrenia. Once familial aggregation with a probable genetic aetiology has been established for a trait, one may consider using segregation analysis to evaluate whether major or minor genes contribute to the expression of the phenotype. Segregation analysis is one of the most established methods for this purpose. It aims to determine the transmission pattern of the trait within families and to test this pattern against predictions from specific genetic models. Maximum likelihood statistics are used to compare likelihoods for an observed pattern occurring by chance (null hypothesis) to a maximum likelihood under a particular hypothesis.

Segregation analysis for schizophrenia using the simple model of Mendelian inheritance with high penetrance has been shown to be incompatible with the observed familial recurrence rates (Tsuang \textit{et al.}, 1982), (McGue & Gottesman, 1989) while the purely environmental models cannot adequately explain the results of adoption studies (Gottesman & Shields, 1982), (Kety, 1983).

More plausible models, which might fit the data, are the generalized single major locus model and the multifactorial polygenic model. The relative validity of the different models has been investigated in several studies. The findings of many of these have been reviewed extensively by Baron (Baron, 1986).

1.6.1 Single Major Locus (SML) Model

This model is often based on the assumption that the inheritance of a disorder is a consequence of a single locus with two alleles (Slater & Cowie, 1971), (Morton \textit{et al.}, 1971). The concepts of reduced penetrance and phenocopies are introduced to
account for the deviations from classical Mendelian inheritance. Several analyses demonstrated that the SML model was sufficient to correctly predict the data from twin and family studies (Slater & Cowie, 1971), (Kidd & Cavalli-Sforza, 1973), (Karlsson, 1974) while other analyses found it insufficient (Matthysse & Kidd, 1976), (Baron, 1982a), (O'Rourke et al., 1982), (Risch & Baron, 1984), (McGue et al., 1985).

1.6.2 Polygenic or oligogenic models

A polygenic (or oligogenic) model is one where two or more disease alleles at two or more distinct loci are needed before a disease is expressed. Such models can be used to explain the recurrence of many common diseases. However, methodology in the area is undeveloped. Further understanding of how two or more loci interact will depend on the accurate identification of each locus.

1.6.2.1 Oligogenic model

The oligogenic model assumes that several genes may act additively, interactively or multiplicatively on the aetiology of the illness. Risch (Risch, 1990) assessed the compatibility of multilocus models with the observed recurrence risks in schizophrenia and suggested that there should be multiplicative effects from at least three loci acting on the risk for this illness. However, Risch and others assume a single subtype of schizophrenia with an equal effect size from oligogenes in every individual. The linkage evidence to date is incompatible with this assumption and therefore multiple models of transmission with heterogeneity in the oligogenes increasing susceptibility to schizophrenia are more likely.
1.6.2.2 Multifactorial-Polygenic (MFP) Model

In this model, Falconer's method of partitioning genetic liability has been widely adopted. In Falconer's model, the trait is assumed to be the result of many genes that have additive effects and unspecified environmental factors the sum of which follows a normal distribution in the general population (Falconer, 1965). The disease is manifested when the liability exceeds a certain threshold where all individuals above this point are affected and those below are normal. Relatives have a greater risk for a genetic disorder than that of the general population and their mean liability is higher.

This model has produced a good fit with the observed risk in several family studies of schizophrenia especially when environmental factors are taken into account (Gottesman & Shields, 1967), (Kidd & Cavalli-Sforza, 1973), (Rao et al., 1981), (McGue et al., 1983), (McGue et al., 1985) but it has also been rejected in several other studies (Matthysse & Kidd, 1976), (Baron, 1982a), (Tsuang et al., 1983).

1.6.2.3 Mixed model

While early segregation analyses considered SML or polygenic models, more recent approaches have considered a mixed model in which the phenotype may be the result of a combination of these two (Lalouel et al., 1983).

1.6.2.4 Other Models

Other models include: the two-locus theory (Kidd & Gladstein, 1978), in which the phenotype is the result of the interaction of two separate loci and a polygenic model with graduated gene effects to allow variable contributions to the liability from the different loci (Matthysse et al., 1979).
Segregation analyses have produced controversial results among the different studies. This is mainly due to limitations of the genetic models examined as well as diagnostic uncertainties. Most of the models used did not take into account factors associated with schizophrenia such as genetic heterogeneity, assortative mating, reduced fertility, and social isolation. The limitations of the analytical methodologies are further compounded by reduced penetrance, phenocopies, diagnostic difficulties, sampling bias, ascertainment bias, mortality, variable age of onset and the lack of enough family data to estimate the large number of unknown parameters required to accurately model the complex trait.

However, it is clear that schizophrenia is a complex disorder that does not show a clear pattern of Mendelian inheritance in all families. For example, both dominant and recessive transmission is plausible for different subtypes. Yet, further subtypes may be oligogenic. As for other common disorders, a mixture of different genetic and non-genetic subtypes is highly likely. Some of these subtypes may be influenced by the environment or by multiple genes. Classical segregation analysis does not have the power to identify a definitive mode of transmission for this disorder and will not do so in the future due to the drawbacks mentioned above. However, it has been argued that linkage and/or association (linkage disequilibrium) studies have enough power to detect the susceptibility genes underlying a common disorder (Clerget-Darpoux, 1991), (Risch & Merikangas, 1996).
1.7 GENETIC LINKAGE ANALYSIS

1.7.1 Genetic markers

RFLPs (Restriction Fragment Length Polymorphisms) were introduced in 1978 (Kan & Dozy, 1978), (Botstein et al., 1980) and became the first genetic markers to be used in a successful linkage (Huntington’s disease) (Gusella et al., 1983). They are based on a single base pair change that creates or obliterates a cleavage site for a specific restriction enzyme. The resulting variation between individuals can be detected by digestion of the DNA by the appropriate restriction enzyme. RFLPs are inherited as simple Mendelian codominant markers, which can be readily identified in families. A disadvantage of RFLPs is that because they are biallelic they are not very informative, having low heterozygosity (usually <0.40).

The heterozygosity of the genotyping markers was greatly increased by the identification of VNTRs (Variable Number of Tandem Repeats) in 1985 (Nakamura et al., 1987). This class of marker is made of a specific set of consensus sequences that vary between 14 and 100 base pairs in length. They are remarkably polymorphic, with a high heterozygosity rate (usually > 0.6) in the population. However, there is only a small number of VNTRs available and their distribution is rather limited in the genome, often tending to cluster toward human telomeres. In addition, their high heterozygosity together with their large size make these markers difficult to genotype.

STRs (Short Tandem Repeats) or microsatellites were initially described by Weber and May (Weber & May, 1989) and Litt and Luty (Litt & Luty, 1989). STRs are distributed widely and evenly in the genome. They sometimes have high heterozygosities (usually > 0.7) and are relatively easy to score. The number of the repeated motifs varies, the most common consisting of two (dinucleotide), three
(trinucleotide) or four (tetranucleotide) bases. The dinucleotide \((CA)_n-(GT)_n\) is the most common repeat, with a highly polymorphic form of the repeat occurring approximately every 0.4 cM. Their polymorphic status usually depends on two factors: the size of the repeat and whether a perfect or imperfect repeat is present. A marker having 15, or more, perfect (uninterrupted) repeat motifs is usually, but not always, highly polymorphic (Weber, 1990). Following the success of the dinucleotide repeat markers, other polymorphic microsatellites were isolated and characterized, namely, tri- and tetra-nucleotide repeat markers, which have helped to further saturate, the genome with additional polymorphic markers for use in genetic studies.

Finally, in recent years, attention has been focused on the use of single nucleotide polymorphisms (SNPs) as genetic markers. They are the most common type of human DNA variation and as the name suggests they represent a position at which two alternative bases occur at an appreciable frequency (>1%) in the human population (Wang et al., 1998). On average, they occur 1 per 300-1000 base pairs (Collins et al., 1997). Although individual SNPs are less informative than typical multi-allelic simple sequence length polymorphisms, they are more abundant and their genotyping can be automated with the use of DNA chip-based microarrays (Hacia et al., 1996), which allow the very rapid analysis of very large numbers of SNPs. Large collections of mapped SNPs are being developed and will provide a powerful tool for human genetic studies (Kruglyak, 1997), (Masood, 1999).
1.7.2 Parametric approach: Lod score analysis

The term genetic linkage is used to describe the tendency of alleles from two or more loci, which lie in close physical proximity to each other to segregate together in a family. Therefore, the extent of linkage depends on distance, which in genetic terms can be measured with the frequency of recombination, also known as the recombination fraction, or theta (\( \theta \)) between loci. Hence, the further apart the loci, the greater the recombination fraction and vice versa. Theta ranges from 0 for loci that are completely linked to 0.5 for unlinked loci on the same or different chromosomes.

Lod score analysis is a likelihood-based parametric linkage approach to estimate the recombination fraction and the significance of the evidence for linkage. Two likelihoods \( (L) \) of observing a particular configuration of a disease locus and a genetic marker locus in a pedigree data set are calculated: one under various degrees of linkage over a selected range of recombination fractions (i.e., \( 0 < \theta < 0.5 \)) and the other under no linkage (i.e., \( \theta = 0.5 \)). The log\(_{10}\) of the ratio for these two likelihoods, also known as lod score \( z(x) \) is used as a measure of support for linkage versus non-linkage (Morton, 1955):

\[
z(x) = \log_{10} \left( \frac{L(\text{pedigree given } \theta=x) / L(\text{pedigree given } \theta=0.5)}{1} \right)
\]

where \( x \) is a particular value of theta. The value is termed the two-point lod when it involves linkage between the disease locus and a genetic marker locus and multilocus, or multipoint lod score when it involves linkage between a disease locus and several linked marker loci. Lod scores across families can be summed but only at the same recombination fraction and the overall lod score of the data set is then examined.

Morton suggested that lods of 3 or more are indicative of linkage because theoretical considerations based on Bayesian probability and empirical evidence indicate, for a Mendelian trait, less than 5% of linkages declared at this level are false
positive. A lod score of −2 or less are indicative of non-linkage, while values between 3 and −2 are considered as inconclusive (Morton, 1955).

The lod score (or parametric) method of linkage analysis requires a genetic model for the disease locus as well as specification of certain parameters such as: recombination fraction, disease allele frequencies and penetrance values for each possible disease genotype. Since a model of inheritance is unknown for schizophrenia, researchers have tended to use a range of different models in which the parameters are assumed. However, power to detect linkage is decreased when the wrong assumptions are made. Several studies have shown that the problem with applying a parametric method when the genetic model is unknown is not the false conclusion of linkage but rather the increased probability of missing a linkage when it is present (Clerget-Darpoux et al., 1986).

One way to circumvent mis-specification of penetrance and/or the existence of phenocopies is by performing an affecteds-only analysis in which one records unaffected individuals as "phenotype unknown". In this way maximum likelihood estimates come from the most certain of the trait information, the phenotypes of affected individuals. However, the inheritance pattern remains to be specified.

Another important issue when performing linkage analysis for complex traits is locus heterogeneity. The lod score method provides statistical tests to calculate heterogeneity (Ott, 1999). In this case, the likelihood ratio for linkage given heterogeneity \((h\text{lod})\) may be written as

\[
Z(\theta) = \log_{10} \left[ \frac{L(\theta, \alpha)}{L(\theta = 1/2, \alpha = 0)} \right]
\]

where \(\alpha\) the proportion of families being linked

The lod score method has several advantages when the mode of inheritance is known. It is very powerful method for detecting linkage, it provides an estimate of
recombination fraction and it enables formal tests of genetic heterogeneity. However, for complex genetic traits such as schizophrenia, with an unknown mode of transmission it is necessary either to use arbitrary penetrances to evaluate linkage or to use a “model free” method in which penetrance is varied automatically such that the maximum lod is produced with gene frequency and penetrances that are appropriate. Several investigators have argued for alternative approaches to the detection of linkage described below.

1.7.3 Non-parametric approach

Non-parametric methods of linkage analysis, as opposed to parametric ones, do not require prior assumptions about the mode of inheritance. The basis for such methods is that affected related individuals should share the same alleles for a genetic marker locus or loci that are linked to the susceptibility locus more often than expected by chance. Two commonly used nonparametric methods are the affected sib-pair method (Penrose, 1935), and the affected-pedigree member (APM) (Lange, 1986).

Non-parametric methods of linkage analysis can be divided into identical by state (IBS) analyses or identical by descent (IBD) analyses. Two alleles are IBS in siblings when they have inherited the same variant of some polymorphic system even though the parental genotypes are unknown. Any individuals whether related or not can share 0, 1 or 2 alleles IBS. Two individuals are IBD when they share the same allele or alleles that were known to be present in their parents.

The simplest case is when the numbers of sib pairs sharing 0, 1 or 2 alleles IBD can unambiguously be determined. Data can be then analysed using standard statistical methods such as a simple chi-square test, which compares the observed
portion of sib pairs sharing 0, 1 or 2 alleles IBD with the expected portions of 0.25, 0.5 and 0.25 under the hypothesis of no linkage between the marker and the disease locus.

In the case when IBD status is not known with certainty, a likelihood method can be used to estimate the relative probabilities of sharing an allele or alleles in two affected siblings based on the genotypes of the parents and on the probability that there is no IBD sharing under the null hypothesis. A maximum lod score can be calculated for the observed genotypic data as a function of the proportion of sharing 0, 1, and 2 alleles IBD \((Z_0, Z_1\) and \(Z_2)\) with the corresponding probability under the null hypothesis of no linkage. The \(Z_0, Z_1\) and \(Z_2\) values for which this ratio is the largest are considered the optimal values of the sharing probabilities. To be analogous to parametric linkage analysis, the \(\log_{10}\) of the likelihood ratio is calculated to produce a lod score.

The major disadvantage of the sib pair analysis is its inability to use additional information that can be extracted from other affected relative pairs or distantly related individuals. Methods that take greater advantage of extended families have been developed and are called affected relative pair (ARP) methods.

The Affected Pedigree Member (APM) analysis as proposed by Lange (Lange, 1986) uses data from affected relatives other than just siblings. In contrast to sib pair methods, APM aims to detect increased allele sharing IBS rather than IBD. Multipoint analysis can be performed increasing the amount of information that can be extracted from a pedigree. The major disadvantage of this method is that it relies on IBS statistics that are prone to type I errors. It also considers the data from only one pair of affected individuals at a time. Thus sharing of alleles in affected
individuals from the same family does not weigh more than sharing of alleles in affected individuals in separate families.

Another ARP method is the NPL (nonparametric linkage) analysis, which is part of the computer program GENEHUNTER (Kruglyak et al., 1996). The NPL statistic measures allele sharing among affected individuals within a pedigree. It can be used to analyse data from affected relative pairs or can provide a simultaneous comparison of alleles in all affected individuals in a pedigree. It can use multipoint linkage data. However, its usefulness is limited only to moderate size pedigrees particularly when a large number of markers need to be examined.

Finally, the likelihood-based model-free analysis, implemented by the computer program MFLINK (Curtis & Sham, 1995) has the advantage of retaining much of the power of the traditional lod score method without the need to specify transmission model parameters. In addition, the method can be applied equally to sib pairs and pedigrees and can deal with multipoint data.

The major drawback of the non-parametric approach is that these methods do not utilise all the information from a pedigree data set. In addition, it has been demonstrated through simulated data sets that LOD scores calculated under the correct or even approximately correct genetic model have a greater power to detect linkage than “model-free” methods (Goldin & Weeks, 1993), (Greenberg et al., 1996), (Durner et al., 1999). Finally, one of the major advantages of the parametric linkage analysis over the non-parametric one is the ability to account for heterogeneity. Therefore, the parametric lod method remains so far, comparatively more useful in the search for susceptibility loci for complex traits such as schizophrenia.
1.7.4 Genetic linkage studies of schizophrenia

Previous linkage studies of schizophrenia have been carried out by targeting “favoured” loci, which have been disrupted by cytogenetic abnormalities previously found to be associated with schizophrenia or by targeting the chromosomal loci of candidate genes, or by systematic scans of all the chromosomes. Many other studies have tested limited regions on many chromosomes and these are reviewed below together with the genome scans. Loci in several chromosomes have been implicated for linkage with schizophrenia but only linkage data to chromosomes with multiple overlapping findings will be mentioned below. Particular emphasis will be given on chromosome 8 linkage studies.

Chromosome 1

A balanced t(1;11)(q42.1;q14.3) translocation was found to cosegregate with schizophrenia and other types of psychiatric disorder in a single large Scottish pedigree (St Clair et al., 1990). Linkage analysis between the translocation and disease produced lod scores between 2.2 and 4.3. When the diagnoses were revised in a thirty-year follow-up, leading to the identification of new cases of psychiatric disorder, the lod score increased from 3.1 to 6.0 (Blackwood et al., 1998). Recent studies from the cloning and sequencing of the translocation breakpoint report the identification of two novel genes, DISC1 and DISC2, which are both disrupted by the translocation on chromosome 1. It is proposed that truncation and/or abnormal regulation of expression of one or both these proteins caused by the translocation confers susceptibility towards psychiatric illness in the translocation carriers (Millar et al., 2000).

Hovatta and coworkers (Hovatta et al., 1999) have reported positive linkage to schizophrenia with markers on chromosome 1q32.2-41. In a three-stage genome scan
by linkage of a population isolate from Finland, three markers gave LOD scores above 3.0 with the maximum LOD score being 3.82 with the marker D1S2891 under a dominant affecteds-only model.

Further support for the chromosome 1q32.2 region comes from the genome scan undertaken from our group (Gurling et al., 1999), (Gurling et al., 2000, submitted). A five-point multipoint analysis yielded an admixture lod of 3.2 with the markers D1S194, D1S196, D1S2815 and D1S191 under a recessive, narrowly defined model.

A second genome scan in Finland (Ekelund et al., 1997) using a new national sib pair sample has also implicated this a region on chromosome 1 but with less definitive statistical significance and at a slightly more centromeric position on chromosome 1 than the study of the first genetic isolate.

The most impressive result however comes from a genome scan in a group of 22 multiply affected Canadian families (Brzustowicz et al., 2000). Multipoint analysis produced a maximum lod score of 6.5 between the markers D1S1653 and D1S1679 (1q21-22) under the recessive-narrow model and with an estimated 75% of the families linked to this locus.

Finally, there have also been suggestive linkage reports for chromosome 1q22-23 under a recessive model (Shaw et al., 1998).

**Chromosome 2**

A genome scan using data from 5 families under narrow phenotype definition produced a significance value of $P=0.000001$ at D2S135 (2q12-13) (Moises et al., 1995). However, this result should be interpreted with caution as schizophrenic families with bipolar cases were included and a non-parametric test was used. In addition, the family sample was very small.
The NIMH/Millennium genome screen of 43 European-American families detected a maximum non-parametric lod score (NPLZ) of 2.41 ($P=0.008$) at marker D2S293 (2q12) under a narrow definition (Faraone et al., 1998).

Another genome scan in a mixed ethnicity sample of 43 North American and Australian families reported suggestive linkage with an NPLZ of 2.01 ($P<0.01$) for D2S410 (2q14.2) and intermediate phenotype definition (Levinson et al., 1998).

On the other arm of chromosome 2, data from a genome scan of 70 families gave an NPLZ of 2.13 ($P<0.01$) for D2S1337 (2p13-14) under broad definition of the phenotype (Shaw et al., 1998).

A study of families from Palau in Micronesia, gave a LOD of 2.17 at D2S441 (2p12-13) (Coon et al., 1998) in a large multiplex schizophrenia pedigree under dominant model. However, the LOD decreased to 1.69 when 16 more families were typed.

**Chromosome 3**

Linkage was suggested on chromosome 3p26-p24 by a study of 57 families in Maryland USA (Pulver et al., 1995). A LOD score of 2.34 was obtained for marker D3S1283 under affecteds-only dominant model.

A large multicenter collaborative linkage study failed to replicate this result. The authors concluded that it might have been a false positive (Levinson et al., 1996). So far, no more data supporting linkage to this chromosomal region have been published.

**Chromosome 4**

A cytogenetic abnormality of chromosome 4 [inv 4 (p15.2; q21.3)] has been reported in a man presenting with undifferentiated schizophrenia (DSMII-R) and in his mother who had schizotypal personality disorder (Palmour et al., 1994).
Shaw and coworkers (Shaw et al., 1998), Levinson et al (Levinson et al., 1998) and Kaufman et al (Kaufmann et al., 1998) have all reported positive but non-significant lod scores with schizophrenia in the 4q13-31 chromosomal region.

In the study of the Finnish isolated population, a LOD of 2.74 was obtained for marker D4S1586 (4q31) under dominant model and narrow definition of the phenotype (Hovatta et al., 1999).

**Chromosome 5**

In genetic linkage analyses in 7 Icelandic and UK families a LOD score of 6.49 was obtained with marker D5S76, which was originally thought to be localized at chromosome 5q11-13 (Sherrington et al., 1988). Further analyses in a larger sample of 23 families with newly created genetic linkage markers (Sherrington et al., 1993) showed diminished support for this region (Kalsi et al., 1999). The original study and attempted replications at that time by other research groups used low heterozygosity restriction fragment length polymorphisms (RFLPs) to attempt replication. In addition, these markers were poorly localized. Since then the locus D5S76, which gave the highest lod, has been repositioned on to the short arm of chromosome 5 at 5p14.1-13.1, between D5S419 and D5S395 in the Genome Database (Buetow, 1994), (Gyapay et al., 1994).

Interestingly, Silverman et al. (Silverman et al., 1996) have reported a multipoint lod of 4.37 between schizophrenia and D5S111 in a single large nuclear family assuming autosomal dominant transmission and broad phenotype definition. This marker also maps to chromosome 5p14.1-13.1 and, like D5S76, lies between the markers D5S419 and D5S395.
One further study of the 5p14.1-13.1 region found an NPLZ of 2.49 and 2.55 with the markers D5S111 and D5S426 respectively (Garver et al., 1998). These loci are closely linked to D5S419.

A further study of this region in Canadian schizophrenic families (King et al., 1997) failed to find any evidence for linkage, as have several genome scans of schizophrenia.

A second region on chromosome 5, 5q22-31, has been implicated in the genetics of schizophrenia. The genome scan of 265 Irish high-density schizophrenia families produced a multipoint HLOD of 3.35 ($P=0.0002$) at D5S804 under narrow phenotypic definition and recessive genetic model (Straub et al., 1997).

Results from our own genome scan (Gurling et al., 1999), (Gurling et al., 2000, submitted) provided evidence for linkage between schizophrenia and chromosome 5q33.2 markers. An admixture lod of 3.6 was obtained with markers D5S410, D5S422, D5S400 and D5S2111 assuming a dominant mode of transmission and with the spectrum model.

**Chromosome 6**

The genome scan of the 265 Irish high-density schizophrenia families implicated a region of chromosome 6, 6p22-24, producing an HLOD score of 3.51 with D6S296 under a broad definition of the phenotype (Straub et al., 1995).

A previous study of a subset of the same Irish sample had also produced a positive HLOD score of 3.9 with a slightly different set of markers (F13A and D6S269) (Wang et al., 1995).

A large multicenter collaborative linkage study including the data sets mentioned above produced only suggestive evidence for linkage at this chromosomal region (Levinson et al., 1996).
Attempted replication of linkage to schizophrenia at 6p22-25 in our sample of 23 UK/Icelandic families was negative (Gurling et al., 1995).

Other studies also failed to support the role of this locus (Arolt et al., 1996), (Daniels et al., 1997), (Kaufmann et al., 1998), (Lind et al., 1998), (Riley et al., 1996).

Subsequently a second locus on chromosome 6 at 6q13-26 was suggested (Cao et al., 1997). A genome scan study of 53 US families containing a total of 81 affected sib pairs showed evidence for excess allele sharing among affected siblings at 69% shared IBD ($P=0.00024$) for marker D6S416. Two replication studies that followed from the same group provided further suggestive evidence for linkage at this region. The strongest result was obtained after the combination of the two replication data sets which gave an overall nominal $P=0.000014$ (MLS of 3.82) and IBD sharing of 63.8% for the interval between D6S424 and D6S301 (Martinez et al., 1999).

Further suggestive but non-significant support for this locus was obtained in the NIMH sample of African-American schizophrenia families (Kaufmann et al., 1998).

**Chromosome 7**

A genome-wide scan in a Finnish schizophrenia sample has implicated chromosome 7q22 with LOD of 3.18 at D7S486 under a dominant model and broad phenotype definition (Ekelund et al., 2000).

In addition, the genome scan of the 54 Maryland families produced suggestive evidence for linkage on this chromosome (D7S2212, 7q21.1-21.3, NPLZ of 2.5) (Blouin et al., 1998).
**Chromosome 8**

The 8p21-22 region on chromosome 8 was originally reported as showing possible genetic linkage to schizophrenia by Pulver and coworkers (Pulver et al., 1995). As part of a genome wide search, a maximum LOD of 2.35 for a dominant and 2.20 for a recessive affecteds-only model was obtained at D8S136. In addition, affected sib pair analyses were conducted testing the mean proportion of alleles shared IBD. The results showed significant linkage \( (P<0.001) \) with four markers spanning a distance of 10 cM on the telomeric side of D8S136: D8S261 \( (P=0.0004) \), D8S258 \( (P=0.00004) \), D8S282 \( (P=0.0002) \) and D8S133 \( (P=0.0001) \).

The genome scan of the 265 Irish high-density schizophrenia families (Kendler et al., 1996) tested fifteen markers covering 30 cM on chromosome 8p21-22. According to multipoint heterogeneity lod score analyses lod scores exceeded 1.9 over a 10 cM region, between markers D8S1715 and D8S1739. The maximum HLOD obtained was 2.34 with marker D8S258 under a dominant-broad model.

Preliminary positive lod scores from a sample of 23 families using three markers on chromosome 8p22.1 were also published by our group (Kalsi, 1996). A maximised lod score of 2.17 with the marker D8S136 under a dominant-narrow model was found.

In response to the initial schizophrenia linkage findings on chromosome 8, a multicentre study was conducted. 14 microsatellite markers were genotyped in a collaborative sample of 403-567 pedigrees, which also included our pedigrees and those in the Kendler study, which are mentioned above. Some analyses also included the family sample from the Pulver study, which had originally shown the linkage with chromosome 8p markers. Two-point heterogeneity lod scores of 2.22 for the new
sample \( P = 0.0014 \) and 3.06 \( P = 0.00018 \) for the combined sample (including Pulver) were obtained with D8S261 under recessive-narrow model (Levinson et al., 1996).

The more recent genome scan of the 54 Maryland families (46 of which had been used in the previous report of potential linkage to chromosomes 3 and 8) provided further evidence for a susceptibility locus for schizophrenia on the 8p21-22 region. An HLOD of 4.54 under dominant and 2.01 under recessive-narrow models were obtained while, non-parametric analysis provided further support producing an NPLZ of 3.64 \( P = 0.0001 \) near D8S1771. To add support, they analysed 51 additional families for 17 chromosome 8p21-22 markers (32 cM). However, the NPLZ was reduced to 1.95 \( P = 0.023 \) at D8S1752 (~3 cM telomeric to D8S1771) (Blouin et al., 1998).

Further evidence of linkage for this region comes from a sample of 21 Canadian families, which was genotyped with 21 chromosome 8p21-22 markers. A maximum two-point lod of 3.49 was obtained under a dominant-narrow model with D8S136 and was the same under homogeneity and heterogeneity. A four-point multipoint analysis with closely flanking markers reduced the maximum LOD score in this region to 2.13 (recessive-narrow model) (Brzustowicz et al., 1999).

In our recently performed genome scan of 13 families containing multiple cases with schizophrenia a 3-point multipoint lod of 3.5 was obtained between markers D8S504 and D8S503 under a recessive-narrow model. In addition, a 3-point lod of 2.9 was obtained between markers D8S503 and D8S552. The position of these findings is somewhat more telomeric than the position reported by previous investigators. However, a 3-point lod of 2.8 was obtained between markers D8S1771 and D8S283, which are more centromeric and closer to the Blouin et al, 1998 study. Furthermore, a 5-point lod score of 3.2 was obtained with markers D8S261, D8S258,
D8S1771 and D8S283 under recessive mode of transmission and narrow phenotype definition. Overall high lod scores were obtained over a ~60 cM region on chromosome 8p (Gurling et al., 1999), (Gurling et al., 2000, submitted).

Some other genome scans have produced weak evidence for linkage on chromosome 8p21-22 (Moises et al., 1995), (Kaufmann et al., 1998), (Shaw et al., 1998) while other studies have been negative (Hovatta et al., 1999), (Williams et al., 1999). A summary of all the previous statistically significant and other supportive linkage data for a schizophrenia susceptibility locus on chromosome 8p is provided on table 1.1.
<table>
<thead>
<tr>
<th>REFERENCE</th>
<th>NO. OF FAMILIES</th>
<th>NO. OF AFFECTED</th>
<th>DIAGNOSTIC DEFINITION</th>
<th>HOMOGENEITY LOD</th>
<th>HETEROGENEITY LOD</th>
<th>P-VALUE</th>
<th>PEAK AT MARKER</th>
<th>MARKER POSITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blouin et al., (1998)</td>
<td>54</td>
<td>146</td>
<td>Narrow</td>
<td>HLOD=4.54</td>
<td>HLOD=2.01</td>
<td>3.63</td>
<td>0.0001</td>
<td>D8S1771</td>
</tr>
<tr>
<td></td>
<td>51</td>
<td>146</td>
<td></td>
<td>HLOD=0.68</td>
<td></td>
<td>1.95</td>
<td>0.023</td>
<td>D8S1752</td>
</tr>
<tr>
<td>Brzustowicz et al., (1999)</td>
<td>21</td>
<td>71</td>
<td>Narrow</td>
<td>HLOD=3.49</td>
<td>HLOD=3.11</td>
<td></td>
<td></td>
<td>D8S136</td>
</tr>
<tr>
<td>Gurling et al., 2000</td>
<td>13</td>
<td>68</td>
<td>Narrow</td>
<td>HLOD=3.6</td>
<td></td>
<td></td>
<td></td>
<td>D8S503</td>
</tr>
<tr>
<td>Pulver et al., (1995)</td>
<td>57</td>
<td>161</td>
<td>Narrow</td>
<td>LOD=2.35</td>
<td>LOD=2.20</td>
<td>0.0004</td>
<td>D8S258</td>
<td>41.55</td>
</tr>
<tr>
<td>Kendler et al., (1996)</td>
<td>265</td>
<td>754</td>
<td>Broad</td>
<td>HLOD=2.34</td>
<td></td>
<td>2.51</td>
<td>0.006</td>
<td>D8S298</td>
</tr>
<tr>
<td>Levinson et al., (1996)</td>
<td>656</td>
<td>1745</td>
<td>Narrow</td>
<td>HLOD=0.99</td>
<td>HLOD=2.2</td>
<td>0.005</td>
<td>D8S261</td>
<td>37.04</td>
</tr>
<tr>
<td>Shaw et al., (1998)</td>
<td>70</td>
<td>171</td>
<td>Narrow</td>
<td>LOD=1.99</td>
<td></td>
<td>1.20</td>
<td>0.07</td>
<td>D8S560</td>
</tr>
<tr>
<td>Kaufmann et al., (1998)</td>
<td>30</td>
<td>79</td>
<td>Narrow</td>
<td></td>
<td></td>
<td>2.27</td>
<td>0.013</td>
<td>D8S1819</td>
</tr>
<tr>
<td>Moises et al., (1995)</td>
<td>9</td>
<td>37</td>
<td>Narrow</td>
<td></td>
<td></td>
<td>1.88</td>
<td>0.03</td>
<td>D8S532</td>
</tr>
</tbody>
</table>

LOD: maximum homogeneity lod score; HLOD: maximum heterogeneity lod score
NPLZ: multipoint non-parametric lod scores computed by GENEHUNTER
*: multipoint analysis results
**: in Kosambi centimorgans, derived from the Marshfield Map positions
**Chromosome 9**

Data from the African-American families in the NIMH/Millennium genome scan study yielded suggestive linkage for three consecutive markers spanning 15 cM on chromosome 9q34.3. An NPLZ of 1.91 ($P=0.0306$) with marker D9S1825, NPLZ of 2.17 ($P=0.0167$) with marker D9S1830 and an NPLZ of 2.04 ($P=0.0229$) with marker D9S1818 were obtained (Kaufmann *et al.*, 1998).

Another region of chromosome 9, 9q21, has also produced suggestive evidence for linkage with schizophrenia. The genome scan results from the Finnish isolate produced a LOD of 1.95 under a dominant, high penetrance model (Hovatta *et al.*, 1999).

Less significant results have been obtained from other studies (Moises *et al.*, 1995), (Levinson *et al.*, 1998).

**Chromosome 10**

The NIMH/Millennium genome scan of 43 European-American families has provided evidence for linkage on chromosome 10p. Using non-parametric linkage methods an NPLZ of 3.4 ($P=0.0004$) at D10S1423 and an NPLZ of 3.2 ($P=0.0006$) at D10S582 were obtained under narrow phenotype definition (Faraone *et al.*, 1998).

In a genome scan containing only a subset (88) of the 265 Irish high-density schizophrenia families, the marker D10S674 produced a pairwise HLOD score of 3.2 ($P=0.0004$). A follow up study testing more markers across this region of the chromosome in the entire family data set produced a pairwise HLOD of 1.95 ($P=0.005$) at D10S2443 under recessive model and intermediate phenotype definition. A multipoint HLOD of 1.91 ($P=0.006$) extending over 11 cM from marker D10S674 to marker D10S1426 (10p11-15) was also obtained (Straub *et al.*, 1998).
A German and Israeli sample of 72 families also found evidence for linkage at this region with the marker D10S582 ($P=0.0058$) and D10S1423 ($P=0.029$). The maximum NPLZ was 3.2 ($P=0.0007$) for marker D10S1714 (Schwab et al., 1998), (Wildenauer et al., 1999).

*Chromosome II*

A balanced t(1;11)(q42.1;q14.3) translocation was found to cosegregate with schizophrenia and other types of psychiatric disorder in a single large Scottish pedigree (St Clair et al., 1990). Studies of the translocation breakpoint have reported a low gene density around this area of chromosome 11. Cloning the breakpoint on chromosome 1q has identified two novel genes on chromosome 1, DISC1 and DISC2 that are directly disrupted by this translocation as mentioned above. It is therefore possible to hypothesize that the disrupted expression of genes in the breakpoint region of chromosome 1 is responsible for the schizophrenia or alternatively either the breakpoint on chromosome 1q and chromosome 11q are simply acting as linkage markers to susceptibility loci at a distance from the breakpoints.

The 11q21-22 region, including the DRD2 locus, was studied by linkage analysis in a sample of four large multigenerational pedigrees densely affected by schizophrenia in Eastern Quebec and Northern New Brunswick (Maziade et al., 1995). A maximum lod score of 3.41 was obtained with marker D11S35 in an affecteds-only analysis of one large schizophrenia family.

Subsequently a lod of 3.00 was found in this region of chromosome 11 in our recently performed genome scan with marker D11S934 under a recessive mode of transmission and broad definition of the phenotype (Gurling et al., 1999, Gurling et al., 2000, submitted).
Other systematic genome scans and linkage analyses failed to detect any evidence of linkage on chromosome 11 (Gill et al., 1993), (Moises et al., 1995), (Kalsi et al., 1995b), (Mulcrone et al., 1995), (Faraone et al., 1998), (Kaufmann et al., 1998).

Chromosome 13

The most significant evidence for linkage on chromosome 13q31-32 was found in a three-point analysis in a sample of 21 Canadian families. An HLOD of 4.42 at D13S793 was obtained under a recessive-broad model (Brzustowicz et al., 1999).

Also the genome scan of 54 Maryland families using a narrow disease definition produced an HLOD of 1.84 under a dominant and an HLOD of 3.19 under a recessive model and an NPLZ of 4.18 ($P=0.00002$) at marker D13S174 (Blouin et al., 1998). However, further analysis of 51 additional families produced an NPLZ of only 2.36 ($P=0.007$) at D13S779 (7 cM telomeric to the previous peak at D13S174).

Several other studies had previously reported suggestive positive, but not significant findings of linkage to chromosome 13q (Lin et al., 1995), (Shaw et al., 1998).

Chromosome 15

Evidence for a susceptibility locus for schizophrenia on chromosome 15q13-14 comes from a study in 9 families Utah. Marker D15S1360, which is only 500 kb from the neuronal nicotinic acetylcholine receptor a7 subunit (CHRNA7), produced a LOD of 5.3 when tested against the P50 evoked EEG response but showed only weakly positive and non significant linkage to schizophrenia (LOD of 1.3) (Freedman et al., 1997).
Chromosome 22

In the first 39 families of the Maryland sample, Pulver and colleagues found a positive lod of 2.82 at the IL2RB marker on 22q13 under an affecteds-only dominant model (Pulver et al., 1994b).

In a collaborative analysis in a larger sample containing 256 pedigrees (Pulver et al., 1994a), this linkage could not be confirmed assuming either homogeneity or heterogeneity.

Several other studies have produced suggestive but not significant evidence for linkage on this chromosomal region (Coon et al., 1994), (Vallada et al., 1995), (Gill et al., 1996), while other studies failed to find linkage with this region of chromosome 22 (Polymeropoulos et al., 1994), (Kalsi et al., 1995a), (Schwab et al., 1995).

Independent evidence from research on velo-cardiofacial syndrome, which is associated with schizophrenia (Bassett et al., 1998) and is caused by a micro-deletion of chromosome 22q11 strongly confirms the possibility that there is a schizophrenia susceptibility locus on chromosome 22. Furthermore, two genome scans, (Blouin et al., 1998) and (Shaw et al., 1998) have provided support for this chromosomal region (22q11) with an NPLZ of 2.42 ($P=0.009$) at D22S1265 from the former and an NPLZ of 2.16 ($P=0.01$) for the latter study.

A summary of the most statistically significant and other supportive linkage data for a schizophrenia susceptibility locus obtained from genome-wide searches is given on table 1.2.
<table>
<thead>
<tr>
<th>CHROMOSOME</th>
<th>REFERENCE</th>
<th>PEAK AT MARKER</th>
<th>EVIDENCE</th>
<th>MODEL OF INHERITANCE</th>
<th>PHENOTYPE DEFINITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Brzustowicz et al., (2000)</td>
<td>D1S1679 (1q21-22)</td>
<td>HLOD=6.5 (P&lt;0.0002)</td>
<td>Recessive-narrow</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gurling et al., (1999)</td>
<td>D1S196 (1q32.2)</td>
<td>HLOD=3.2</td>
<td>Recessive-narrow</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hovatta et al., (1999)</td>
<td>D1S2891 (1q32.2-41)</td>
<td>LOD=3.82</td>
<td>Dominant-narrow</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Shaw et al., (1998)</td>
<td>D2S1337 (2p13-14)</td>
<td>NPLZ=2.13 (P&lt;0.01)</td>
<td>Broad</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Faraone et al., (1998)</td>
<td>D2S293 (2q12)</td>
<td>NPLZ=2.4 (P=0.008)</td>
<td>Narrow</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Levinson et al., (1998)</td>
<td>D2S5410 (2q14.2)</td>
<td>NPLZ=2.01 (P&lt;0.01)</td>
<td>Intermediate</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Hovatta et al., (1999)</td>
<td>D4S1586 (4q31)</td>
<td>LOD=2.74</td>
<td>Dominant-narrow</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Straub et al., (1997)</td>
<td>D5S804 (5q22-31)</td>
<td>HLOD=3.35 (P=0.0002)</td>
<td>Recessive-narrow</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gurling et al., (1999)</td>
<td>D5S422 (5q33.2)</td>
<td>HLOD=3.6</td>
<td>Dominant-broad</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Straub et al., (1995)</td>
<td>D6S296 (6p22-24)</td>
<td>HLOD=3.52 (P=0.0002)</td>
<td>Dominant-broad</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cao et al., (1997)</td>
<td>D6S416 (6q13-26)</td>
<td>IBD 69% (P=0.00024)</td>
<td>Narrow</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Ekelund et al., (2000)</td>
<td>D7S486 (7q22)</td>
<td>LOD=3.18</td>
<td>Dominant-broad</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Blouin et al., (1998)</td>
<td>D7S2212 (7q21.1-21.3)</td>
<td>NPLZ=2.5 (P=0.007)</td>
<td>Narrow</td>
<td></td>
</tr>
</tbody>
</table>

LOD: maximum homogeneity lod score; HLOD: maximum heterogeneity lod score; *: multipoint analysis results
NPLZ: multipoint non-parametric lod-scores computed by GENEHUNTER
Table 1.2 (continues) Statistically significant and other supportive linkage data for a schizophrenia susceptibility locus obtained from genome-wide searches

<table>
<thead>
<tr>
<th>CHROMOSOME</th>
<th>REFERENCE</th>
<th>PEAK AT MARKER</th>
<th>EVIDENCE</th>
<th>MODEL OF INHERITANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Blouin et al., (1998)</td>
<td>D8S1771 (8p21)</td>
<td>HLOD=4.54</td>
<td>Dominant-narrow</td>
</tr>
<tr>
<td></td>
<td>Gurling et al., (2000)</td>
<td>D8S503 (8p23)</td>
<td>HLOD=3.5</td>
<td>Recessive-narrow</td>
</tr>
<tr>
<td>9</td>
<td>Kaufman et al., (1998)</td>
<td>D9S1830 (9q33-34)</td>
<td>NPLZ=2.17 (P=0.02)</td>
<td>Narrow</td>
</tr>
<tr>
<td></td>
<td>Hovatta et al., (1999)</td>
<td>D9S922 (9q21)</td>
<td>LOD=1.95</td>
<td>Dominant-narrow</td>
</tr>
<tr>
<td></td>
<td>Faraone et al., (1998)</td>
<td>D10S1423 (10p13)</td>
<td>NPLZ=3.4 (P=0.0004)</td>
<td>Narrow</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D10S582 (10p13)</td>
<td>NPLZ=3.2 (P=0.0006)</td>
<td>Narrow</td>
</tr>
<tr>
<td>10</td>
<td>Straub et al., (1998)</td>
<td>D10S2443 (10p11-15)</td>
<td>HLOD=1.95 (P=0.005)</td>
<td>Recessive-intermediate</td>
</tr>
<tr>
<td></td>
<td>Schwab et al., (1998)</td>
<td>D10S1714 (10p11-14)</td>
<td>NPLZ=2.8 (P=0.003)</td>
<td>Narrow</td>
</tr>
<tr>
<td></td>
<td>Faraone et al., (1998)</td>
<td>D10S604 (10q11)</td>
<td>NPLZ=3.2 (P=0.0007)</td>
<td>Narrow</td>
</tr>
<tr>
<td>13</td>
<td>Blouin et al., (1998)</td>
<td>D13S174 (13q32)</td>
<td>HLOD=3.19 NPLZ=4.2 (P=0.0002)</td>
<td>Recessive-narrow</td>
</tr>
<tr>
<td>22</td>
<td>Blouin et al., (1998)</td>
<td>D22S1265 (22q11)</td>
<td>NPLZ=2.4 (P=0.009)</td>
<td>Narrow</td>
</tr>
<tr>
<td></td>
<td>Shaw et al., (1998)</td>
<td>D22S446 (22q11)</td>
<td>NPLZ=2.16 (P=0.01)</td>
<td>Narrow</td>
</tr>
</tbody>
</table>

LOD: maximum homogeneity lod score; HLOD: maximum heterogeneity lod score; *: multipoint analysis results
NPLZ: multipoint non-parametric lod-scores computed by GENEHUNTER
1.8 LINKAGE DISEQUILIBRIUM AND ALLELIC ASSOCIATION

Allelic association studies can be a useful tool for the analysis of complex traits, in the evaluation of candidate gene loci, in the fine mapping of a region that has been implicated by linkage studies and finally for genome-wide searches for susceptibility genes.

In contrast to linkage analysis, allelic association studies do not concern familial inheritance patterns. Instead, they test the increased or decreased frequency of a marker allele in combination with a disease trait between affected and unaffected individuals from a population. Thus, allelic association analysis involves correlated occurrence in population as opposed to linkage studies which show correlated transmission within a family (Lander & Schork, 1994).

Such an association can arise if the polymorphism being used is the pathogenic mutation itself. This can occur if the polymorphism is in the coding region of the gene and alters the amino acid sequence of the protein being coded or if it lies in the non-coding sequence of the gene and then it affects expression pattern, mRNA stability or RNA splicing.

Allelic association can also occur if the marker allele is in linkage disequilibrium with the disease allele. Linkage disequilibrium is a term used to describe the non-random occurrence of certain allele combinations on a haplotype due to tight linkage of the loci that constitute this haplotype. Conceptually it is the same as standard linkage analysis except that the recombination distances being measured are very small (<1 cM). In this way repeated recombination is restricted and therefore certain marker alleles tend to occur on the same haplotype(s) with the disease allele. The rate of decay of linkage disequilibrium is a function of time, recombination
fraction between the two loci, and whether the forces that established the
disequilibrium continue.

There are also a number of possible causes of association at the population level. Spurious evidence for association can occur due to population stratification. If there are subgroups living in one geographical location but random mating between them is not achieved then allele frequencies at two or more loci may differ between the two subgroups. Recent admixture of two populations with different allele frequencies can be the cause of false allelic association. Another form of population stratification is the inappropriate matching of patients and controls. If individuals in the sample are derived from different geographical locations or from different ethnic groups, the resulting sample will be a mixed population and can produce a false association.

1.8.1 Tests for allelic association/linkage disequilibrium

1.8.1.1 Case-control studies

The selection of unrelated cases and controls from a single population is the simplest protocol that can be used when testing the hypothesis of linkage disequilibrium using allelic association analysis. The method involves comparing allele frequencies in a sample of unrelated affected individuals and a sample of controls. The frequencies of alleles in cases and controls can be compared by using a chi-square test and the significance of any allelic differences can be assessed using suitable corrections or by using empirical methods with Monte Carlo simulations. As it was mentioned before, population stratification can confound allelic association studies therefore the two samples should be carefully matched for ethnicity and geographical origin.
1.8.1.2 Family based studies

A method to overcome spurious associations due to population stratification is by obtaining control alleles from relatives of cases. Various tests have been developed for this but the one most commonly used is the transmission disequilibrium test (TDT) (Spielman et al., 1993). It uses simplex families, which include only one affected individual and his or her parents per family. The test has the premise that a parent heterozygous for an associated allele and a non-associated allele should more often transmit the former than the latter to an affected child.

In instances where it may be difficult to obtain parents for a TDT analysis, unaffected siblings can be used as internal controls (sib-TDT) (Curtis, 1997). In order to enhance the power, it is advisable to use a sibling who is as genetically distinct from the case as possible.

The original test proposed by Spielman et al (1989) was limited to a biallelic marker system. Since then several tests have been developed which allow for multiple allelic markers like the extended-TDT (E-TDT) test (Sham & Curtis, 1995a).

1.8.2 Applications of association studies

To date association studies have been restricted either to studies of candidate genes or to attempts to refine the map positions of susceptibility genes that have been localized by linkage studies. Only recently have attempts been made to localize genes increasing susceptibility to schizophrenia at regions previously implicated by linkage analysis as reported in this thesis.
1.8.2.1 Candidate genes

This term is used to describe any gene which based on prior information has a plausible claim to involvement in the determining variation of the disease phenotype. Prior information supporting candidacy can come from a variety of sources such as biological plausibility from the pathophysiology of the disease, from experiments in animal models and/or from genomic position (positional candidates).

Once a candidate gene has been proposed, evaluation proceeds by identifying genetic variation in the locus and subsequently accumulating evidence for (or against) a correlation between the candidate polymorphisms and phenotypic variation relative to the disease. Candidate polymorphisms that lie within the coding or regulatory regions of the gene and are likely to result in altered structure or expression of the protein should be studied, as they are more likely to be the causative mutation. Anonymous polymorphisms can also be used but they should be in very close proximity to the candidate gene in order to maintain strong LD with functional polymorphisms within the gene. However, it should be stressed that even associations between functional variants may still reflect LD rather than involvement per se in the disease pathogenesis (‘hitch hiking’ polymorphisms) and in practice it may be hard to discriminate the two especially if they lie in non-coding sequences. A way to circumvent this problem is by comparing the results in a number of distinct ethnic groups or by using a biological test to assess its effect on gene expression. Variants showing association with the disease in diverse ethnic groups are likely to be directly pathogenic rather than reflecting ethnic stratification.

As more genes are being identified with the completion of the Human Genome Project, it is likely that more attention will focus on candidate genes and the study of their common variants.
1.8.2.2 Fine mapping

Due to the fact that LD is maintained over small distances (usually <1 cM) allelic association studies offer the opportunity to map genes for complex diseases more precisely than linkage studies. By focusing upon specific regions implicated by linkage as harbouring a susceptibility gene for a disease and saturating the candidate region with markers will allow the characterization of the pattern of the LD in the region and the finer localization of the disease gene.

1.8.2.3 Genome-wide screens

Systematic genome-wide association searches with many thousands of markers are theoretically possible (Risch & Merikangas, 1996) and are increasingly technically feasible with the advent of SNPs and their potential high-throughput genotyping on DNA microarrays.

Two different ways to perform whole-genome association studies have been recommended (Collins et al., 1997). One is a direct method, which involves the systematic cataloguing and direct testing of all the common functional variants of every single gene. The other one is an indirect method, which will make use of a very dense map of SNPs across the whole genome and will systematically screen for evidence of linkage disequilibrium between the marker alleles and disease variants.

However, there are still some hurdles that need to be overcome such as understanding the patterns of LD across the genome in various population, deciding on the optimum spacing of markers for whole genome LD studies and optimising further the DNA chip technology as a cheap, accurate and accessible method for genotyping, before genome wide association studies become a reality. Until then, the direct and indirect approaches can be used on a smaller scale, by using cSNPs to
assess a wide range of functional and positional candidate genes. Furthermore, by saturating with markers the regions implicated by linkage studies it should be possible to map the disease susceptibility gene(s) to within regions of less than 1 cM.

1.8.3 Association studies in schizophrenia

Association studies in schizophrenia have concentrated on candidate genes, which derived their status mainly from their neurobiological function such as genes belonging to the dopaminergic and serotonergic systems and are involved in monoamine transmission.

Unfortunately, because we understand very little about the psychopathology underlying most psychiatric disorders, it is rarely possible to specify highly plausible candidates and the prior odds against true association are therefore substantial (Crowe, 1993), (Kidd, 1993). This together with the effects of multiple testing and population stratification is likely to result in high type I error rate of such studies and therefore positive findings should be interpreted cautiously until replicated. A better strategy, which will increase the prior odds in favour of finding associations, would be to identify candidate genes in a region that has already been implicated by linkage to contain a susceptibility locus.

Overall, the results of candidate gene studies in schizophrenia have been disappointing. However, it should be noted that the sample sizes used in many of the older studies would generally be regarded as inadequate, because the polymorphic markers in question did not in themselves represent functional variants and because few genes have been systematically screened even for common variants. The most promising reports of candidate gene association studies so far are those of the serotonin 5HT2a receptor gene and the dopamine D3 receptor gene.
The European multicentre association study of schizophrenia involved seven European centres, 571 patients and 639 controls demonstrated an allelic association between schizophrenia and allele 2 of the T102C polymorphism at the 5HT2a receptor gene (13q14-21), suggesting that variation at or near this locus contributes to susceptibility to schizophrenia (Williams et al., 1996). This finding was replicated in a family based association study (Spurlock et al., 1998). Several negative studies have also been reported (Hawi et al., 1997), (Shinkai et al., 1998). Results of a recent meta-analysis of all published data, which included over 3000 individuals, support the original finding (Williams et al., 1997). The T102C polymorphism occurs at a degenerate base and therefore does not alter the amino acid sequence of the receptor, making it unlikely to be of some functional significance. So far, two other polymorphisms have been identified in the promoter region, one of which is in complete linkage disequilibrium with the T102C polymorphism and therefore a candidate pathogenic variant for schizophrenia itself (Spurlock et al., 1998).

The DRD3 receptor gene (3q13.3) has a polymorphism in its first exon that creates a Ball restriction site and gives rise to a glycine to serine substitution in the N-terminal extracellular domain of the receptor. Genotypic association studies demonstrated an increased homozygosity for this polymorphism in schizophrenic patients when compared to healthy controls (Crocq et al., 1992). Replication studies in independent samples followed providing further support of this finding (Mant et al., 1994), (Asherson et al., 1996), while negative studies have also been reported (Nanko et al., 1993), (Nothen et al., 1993). A meta-analysis of all the available results, which included 5351 subjects, demonstrated a significant association between homozygosity for the DRD3 polymorphism and schizophrenia ($P=0.0002$) (Williams
et al., 1998). So far, the functional significance of this polymorphism remains unknown while no other mutations have been found in the coding region of this gene.

1.9 AIMS OF THIS THESIS

To carry out a linkage study of schizophrenia on chromosome 8p21-22 in order to provide a further test of linkage with schizophrenia.

To identify new polymorphic markers and carry out case-control allelic association studies of candidate genes in the 8p21-22 chromosomal region.

To carry out linkage disequilibrium studies with closely spaced polymorphic markers in order to narrow down the susceptibility region implicated by the linkage studies.

To optimize and carry out the identification of DNA variation in a positional candidate gene pinpointed by the allelic association/linkage disequilibrium studies in this region with the ultimate aim of identifying the causative mutation(s) in the susceptibility to schizophrenia.
CHAPTER 2

MATERIALS AND METHODS
2.1 ASCERTAINMENT OF PEDIGREES

Sixteen pedigrees of English ancestry containing cases of schizophrenia were studied. The pedigree structures are given in Appendix 1. These pedigrees were predominantly medium to small size.

The evidence from twin and family studies suggest that manic depression and schizophrenia have predominantly different genetic aetiologies (Baron, 1982b), (Gurling, 1996). The aim was to obtain a sample of families as homogeneous as possible, therefore all families containing cases of bipolar illness were excluded from the study. The reason for not excluding just those individuals who were bipolar was the possibility that there might be two genetically distinct diseases segregating within the pedigree and as a result, there might be some cases (eg schizoaffective disorder or atypical cases) that could not be reliably distinguished as belonging to the schizophrenia or the bipolar phenotypes.

Equal care was taken to ensure that there was only one unilineal source for a schizophrenia susceptibility allele segregating into a sampled pedigree. Such an approach requires that one of the spouses comes from a family that is free from major psychiatric conditions.

2.2 DIAGNOSTIC CRITERIA

2.2.1 Linkage analysis studies

Diagnoses were assigned using Research Diagnostic Criteria (Spitzer et al., 1978a). Subjects gave informed consent and were interviewed by a psychiatrist using
the Lifetime Version of the Schizophrenia and Affective Disorders Schedule (SADS-L) (Spitzer & Endicott, 1977). This information was supplemented by material from case-notes. One psychiatrist blind to genotyping assigned diagnoses according to Research Diagnostic Criteria (Spitzer et al., 1978a) based on information from the SADS-L interviews, hospital case-notes and informants. Two affection classes were used for the linkage analyses: "core schizophrenia" denoted DOMS (for DOMinant Schizophrenia), consisting of schizophrenia and unspecified functional psychosis; "schizophrenia spectrum" denoted DOMSS, consisting additionally of schizoid and schizotypal personality disorder. The remainder, consisting of any psychiatric diagnosis, were classed into a fringe category but are not included in the analyses presented here.

In the 16 families 146 people were assessed but only 97 had blood taken from them and these all had SADS-L interviews. Amongst the 97 that were interviewed 26 were schizophrenic and 8 had schizotypal disorder.

2.2.2 Allelic association studies

134 cases of schizophrenia have been studied together with 316-screened normal controls. The cases and controls were all asked if both their parents and grandparents were of Irish, Welsh, Scottish or English ancestry (the London (UK) case-control sample). Those subjects with two or more grandparents who had any other ancestry were excluded but subjects with a single grandparent with a European ancestry were included. After complete description of the study to the subjects, written informed consent was obtained. All subjects were interviewed by a psychiatrist using the Lifetime Version of the Schizophrenia and Affective Disorders
Schedule (SADS-L) (Spitzer et al., 1978a). The cases were rated using the 90-item OPCRIT checklist, (McGuffin et al., 1991) the family history was recorded and pedigree diagram was drawn. Diagnoses were assigned using Research Diagnostic Criteria (Spitzer & Endicott, 1977). 10ml or more of blood sample was taken from each individual for DNA analysis. Genomic DNA was extracted using the phenol-chloroform method.

The 200 schizophrenia cases and 200 normal controls from Scotland were interviewed by a psychiatrist using the Lifetime Version of the Schizophrenia and Affective Disorders Schedule (SADS-L) (Spitzer et al., 1978a). Cases and controls were of British or Irish ancestry.

2.3 LABORATORY METHODS

2.3.1 Laboratory reagents

General laboratory chemicals were of analytical grade and purchased from either Merck Ltd, Poole, Dorset, U.K.; Boehringer Mannheim, Lewes, East Sussex, U.K.; Difco Laboratories, Detroit Laboratories, Detroit, U.S.A.-, Sigma Chemical Company Ltd, Poole, Dorset, U. Life Technologies, Paisley, Renfrewshire, U.K.- or Biometra Ltd, U.K. dNTPs were obtained from Pharmacia Biotechnology Ltd, St. Albans UK. Radiochemicals and Hybond N+ nylon membranes were obtained from Amersham International Plc. Little Chalfont, Bucks., UK. Kodak X-OMAT imaging photographic film was purchased from Sigma Chemical Co. Ltd. Kodak Professional 64T colour film was also purchased from Sigma chemical company.
2.3.2 Bacterial strains and growth conditions

2.3.2.1 Bacterial strains

The One Shot® Top10F' E.coli cells (Invitrogen BV) were used in this thesis for all plasmid cloning. The genotype of these cells is the following:

F'\{lacT^\text{R} \text{Tn10} (Tet^\text{R})\} mcrA \Delta(mrr-hsdRMS-mcrBC) \Phi80lacZ\Delta M15 \Delta lacX74 deoR
recA1 araD139 \Delta(ara-leu)7697 galU galK rpsL endA1 nupG.

The genotype of the E.coli DH10B cells has as follows:

F' mcrA \Delta(mrr-hsdRMS-mcrBC) \Phi80dlacZ\Delta M15 \Delta lacX74 deoR recA1 endA1
araD139 \Delta(ara, leu)7697 galU galK \lambda rpsL nupG/bMON14272 / pMON7124

2.3.2.2 Propagation of Bacteria

2.3.2.2.1 Growth of E.coli DH10B cells containing P1-derived constructs

This method was used to isolate the pCYPAC cloning vector containing an insert of approximately 110kb of human genomic DNA. The recombinant construct was transformed into DH10B electrocompetent cells (BRL Life technologies). 3ml of 2x YT (1.6% (w/v) Bacto®-tryptone, 0.5 % (w/v) NaCl, 1% (w/v) Bacto®-yeast extract) containing 25μl/ml kanamycin was inoculated with a single colony from a selective plate and was allowed to grow overnight, shaking at 37°C. The next day 500ml of 2x YT containing 25μl/ml kanamycin was inoculated with 0.5ml of the overnight culture. The culture was incubated at 37°C in an orbital shaker at 2000rpm until an OD_{550} of 0.15 was reached. 5ml of freshly made, filter-sterilized 0.1M IPTG was added to the 500ml culture and the culture was allowed to grow at 37°C with vigorous shaking until an OD_{550} of 1.3-1.5 was reached (IPTG is added to induce P1 replication). The cells were harvested by centrifugation at 5500rpm for 10min at 4°C.
2.3.2.2 Growth of E.coli Top10F’ cells

Top10F’ cells were grown in 10ml of Luria Bertani (LB) media containing either no antibiotic or 10µg/ml tetracyclin (Sigma Chemical Company, Poole, Dorset, U.K.) overnight in a Gallenkamp orbital shaker at 200rpm. LB media was autoclaved at 120°C for 20 mins at 10lb square inch\(^{-1}\). Colonies of Top10F’ were isolated by growing on LB plates containing 2% Bacto®-agar.

Luria Bertani (LB Media) 1% (w/v) Bacto®-tryptone
1% (w/v) NaCl
0.5% (w/v) Bacto®-yeast extract

2.3.2.3 Preparation of electrocompetent cells

The One Shot® Top10F’ cells cannot be used for electroporation. In order to make them electrocompetent an overnight culture was set up from a single colony of Top10F’ cells in LB. Next day a 10µl of the overnight culture was added to 200ml of LB. The culture was incubated at 37°C with vigorous shaking until the cells reached the mid-log phase of growth (OD\(_{600}\) of 0.5-0.6). The cells were chilled on ice and then centrifuged at 2K for 10min at 4°C. The pelleted cells were resuspended in an equal volume (200ml) of ice-cold sterile dH\(_2\)O. The cells were washed in the same manner with 100ml and then with 15ml ice-cold sterile dH\(_2\)O. Finally, cells were resuspended in 1ml of 10% glycerol in sterile dH\(_2\)O. Aliquots of 50µl were made which could be stored at -70°C.

2.3.2.4 Transformation by electroporation

During electroporation, the membrane of a cell is exposed to high intensity electric field pulses that can temporarily cause its destabilization in specific regions. During the destabilization period, the cell membrane is highly permeable to...
exogenous molecules present in the surrounding media. 10μl of DNA in low ionic
strength buffer, preferably water, is mixed with 50ml of electrocompetent Top10F’
cells and transferred to a to a pre-chilled electroporation cuvette (BioRad). The
cuvette was dried carefully and placed in a BioRad electroporator. Each sample was
pulsed at 0.25kV and 25μF (average time constant should be between 4.7 and 4.8).
The sample was then carefully removed from the cuvette and added to 1ml of LB
(prewarmed to 37°C). Incubation in an orbital shaker for 1 hour at 37°C/200rpm
followed. The cells were plated onto LB agar plates containing 100μg/ml of
ampicillin. Stocks of ampicillin (Sigma Chemical Company, Poole, Dorset, U.K.)
were made at 100mg/ml (1000x) and stored at -20°C. 40μl of 100mM IPTG was
added to each selection plate in order to induce expression from the lac promoter. If
blue/white screening was required to select for transformants, plates containing 50μl
of 20mg/ml stock of 4-cloro, 5-bromo, 3-indoly1-β-galactosidase (X-Gal, Insight
Biotechnology Ltd.) dissolved in dimethyl formamide, were used.

2.3.3 DNA extraction techniques

2.3.3.1 Isolation of genomic DNA from blood

The peripheral lymphocytes from 10ml of blood were lysed in an equal volume of
lysis solution (10mM NaCl, 10mM EDTA, pH7) and pelleted at 3,000 rpm for 15
minutes at 4°C. This step was repeated once more. The clean nuclei were
resuspended in 10ml of protease-k buffer (100mM NaCl, 10mM Tris-Cl pH8.5,
25mM EDTA) in the presence of SDS. Protein digestion was carried out with
protease-k (10mg/ml) in a shaking water-bath at 55°C overnight. The DNA was
extracted with an equal volume of (25:24:1) phenol/chloroform/isoamyl alcohol, and
precipitated with 1/10 the volume of 3M sodium acetate (pH5.2) and 2.5 volumes of
100% ethanol. The DNA was winded onto a glass rod, rinsed with 70% ethanol and allowed to dry. Finally, it was resuspended in 0.5ml of TE and left to dissolve at 4°C. The DNA concentration was calculated by measuring the absorbance at 260nm (1OD=50µg/ml). This procedure produces high molecular weight, pure DNA that has not been extensively sheared during isolation and can be used for further studies such as genotyping and cloning.

### 2.3.3.2 Small Scale Plasmid DNA extraction Top10F' from cells

Single colonies of Top10F' cells were used to inoculate 5mls of LB containing 100µg/ml ampicillin, and were incubated in an orbital shaker overnight, as before (see section 2.3.2.2.2). The cells from 1.5ml of culture were then pelleted by centrifugation at 14000 rpm for 1min and the pellet was resuspended in 100µl of sterile dH₂O. Cells were lysed by addition of 200µl of freshly made 0.2M NaOH, 1% (w/v) SDS and then neutralized by the addition of 150µl of ice-cold 3M potassium acetate pH 4.8. The cell lysate was then centrifuged for 5min at 4°C and 14000rpm. 400µl of the supernatant was placed in a new eppendorf tube and an equal volume of phenol/CHCl₃/IAA (25:24:1) was added. Centrifugation followed for 5min at 4°C and 14000 rpm. The aqueous phase was removed to a new eppendorf tube and 2.5 volumes of 100% ethanol (EtOH) were added. The solution was left for 10min at room temperature and then was centrifuged for 5min at 4°C and 14000 rpm. The supernatant was discarded and the DNA pellet was washed with 70% ethanol (EtOH). The DNA pellet was air dried and then resuspended in 100µl of double-distilled water (ddH₂O). Plasmid DNA was stored at −20°C.
2.3.3.3 Large Scale Plasmid DNA Extraction

The Qiagen Midi-Prep Kit (Qiagen, Chatsworth, USA) was used for both the P1-vector DNA extraction and plasmid (PUC18) DNA extraction according to the manufacturer's instructions.

2.3.4 Genotyping

2.3.4.1 PCR reactions

100ng of genomic DNA was amplified in a total volume of 12.5μl containing 2.0pmol of each primer, 200μM each of dCTP, dTTP, dGTP and dATP, 0.2 units of Taq polymerase (Qiagen) and 1.25μl of manufacturer's buffer. One of the primers, usually the forward primer, was end-labelled with γ32P-ATP using T4 polynucleotide kinase (Promega, UK). The end-labeling was usually performed in a bulk mix of 20μl volume containing 200pmol of PCR primer and 6pmol of γ32P-ATP, which was enough for 100 PCR reactions. The amplification cycling conditions were as follows: an initial denaturing step at 95°C for 5min followed by 30-35 cycles at 95°C, 30-45 sec at the appropriate annealing temperature calculated according to the base constitution of the primers (between 52°C to 60°C) and finally an extension step at 72°C for 30-45 sec. A final cycle at 72°C for 7min was also included. The touchdown PCR that uses a cycling program with varying annealing temperatures was used to amplify some of the polymorphic markers. It is a useful approach, which increases the specificity of the initial primer-template duplex formation and hence the specificity of the final PCR product. The annealing temperature in the initial cycle was 65°C and was subsequently decreased in steps of 1°C/cycle until it reached 55°C. This annealing temperature was then maintained for the following 25 cycles. Denaturation temperature was 95°C for an initial 4min and then 1 min for each cycle.
Extension was performed at 72°C for 2 min for each cycle, with a final extension step of 7 min at 72°C at the end of the PCR.

The PCR products, which correspond to the various alleles being amplified, were mixed with an equal volume of loading buffer (10 ml formamide, 200 μl of 0.5 M EDTA, 10 mg of xylene-cyanol and 10 mg of bromophenol blue), heated to 95°C for 3 min and then loaded on a pre-run denaturing gel at 1.8 kvolts for 3 to 4 hours depending on the size of the alleles.

2.3.4.2 Polyacrylamide Gel Electrophoresis

The alleles were separated on 6% denaturing polyacrylamide gels (National Diagnostics Ltd) comprised of 5.7% (w/v) acrylamide, 0.3% (w/v) methylene bisacrylamide and 7 M urea in 0.1 M Tris-borate-2 M EDTA buffer pH 8.3 (1 x TBE).

On completion of electrophoresis, the gel was blotted onto 3MM Whatman paper, covered with Saran wrap and transferred to an autoradiography cassette. The alleles were visualised by autoradiography using Fuji AR X-ray film with intensifying screen for 24 to 72 hours at -70°C.

2.3.4.3 Reading genotypes and genotype checking

The autoradiograms were read independently of each other and blind to clinical diagnostic status by a colleague and me. Ambiguous genotypes were repeated and in situations where consensus could not be reached, the genotypes were either repeated or omitted from the final analysis.
2.3.4.4 Genotyping of an \textit{MboII} RFLP in the NAT1 gene

There is a frequent sequence variant at position 1095 of the NAT1 gene, which is a C→A substitution. The method used to genotype this variant is a modification of the method described by Bell et al (Bell \textit{et al.}, 1995b) and Payton et al (Payton & Sim, 1998). Primers N769 (F): 5' ACT CTG AGT GAG GTA GAA ATA 3' and N1113 (R): 5' ACA GGC CAT CTT TAG AA3' were used to PCR amplify the RFLP. The N1113 primer contains a partial \textit{MboII} restriction site, which forms a complete site when a C is at position 1095 instead of an A. PCR reactions were performed in a volume of 12.5μl containing: 100 ng of genomic DNA, 20 pmoles of each primer, 200 μM dNTPs, 10mM Tris-Cl (pH 8.4), 50 mM KCl, 2.0 mM MgCl2 and 0.5 units of Taq DNA polymerase (Qiagen). Reactions were incubated at 95°C for 5 minutes followed by 30 cycles of denaturing for 30 seconds at 95°C, annealing for 30 seconds at 52°C followed by extension at 68°C for 30 seconds. Amplification was completed by incubation at 68°C for 7 minutes. The PCR products (344 bp) were digested overnight at 37°C with 3 units of \textit{MboII} (Promega, UK) and separated by electrophoresis through 1.0% (w/v) Metaphor agarose (FMC Bioproducts) gels. Products were detected by ethidium bromide staining and visualized under ultra-violet light. Fragments of 164 bp and 124 bp were representing allele 1 while fragments of 164 and 144 bp were representing allele 2. When fragments of 164, 144 and 124 bp were present in one lane then this individual was considered heterozygous. Genotypes were determined by two independent raters blind to each other's assessment of genotypes. Where different genotypes were observed, gels were re-examined and if agreement could not be reached, the genotypes were repeated or discarded.
2.3.5 Cloning techniques

2.3.5.1 Agarose Gel Electrophoresis

1% (w/v) agarose gels were made using 1 x TAE (0.4M Tris, 0.2M sodium acetate, 20mM EDTA, pH8.3). Ethidium bromide was added to a final concentration of 0.5μg/ml. Loading buffer (1 x TAE, 50% v/v glycerol, 0.025% bromophenol blue) was added to DNA samples prior to loading.

2.3.5.2 Restriction Digests

Restriction digests of PAC or plasmid DNA fragments were carried out in a total volume of 50μl containing 5μg of midi-prep DNA and the appropriate buffer. Digest were incubated overnight and then run on a 1% agarose gel. DNA bands were visualized on a UV transilluminator. When required, the desired bands were carefully excised using a scalpel. DNA was then extracted from the agarose using the QIAquick Gel Extraction kit protocol (QIAGEN) as per manufacturer’s instructions. The DNA was then eluted in a final volume of 20μl.

2.3.5.3 DNA Ligations

Ligations were performed in 0.5ml reaction tubes in a total volume of 20μl. Reactions contained 1-2μl of gel-purified vector, 7-10μl of gel-purified insert, 1x ligase buffer, 1-3 units of T4 DNA ligase (Promega, Southampton, UK) in ddH2O. Reactions were placed in a thermocycler (PE Thermocycler) and incubated at 16°C overnight. The ligation reaction was then ethanol precipitated and resuspended in 10μl of sterile dH2O that was subsequently transformed into 50μl of competent Top10F' cells by electroporation, as described earlier.
2.3.5.4 DNA blotting

2.3.5.4.1 Southern blotting

A Southern blot was performed on a PAC clone containing an insert of human genomic DNA in order to determine the presence of CA-repeat. 5μg of DNA prep was digested overnight with the appropriate enzymes and buffers in a total volume of 50μl. The same amount of a control DNA clone, known to contain a CA-repeat, was digested in a total volume of 50μl. The digest reactions were then run on a 1% TAE gel as described above. The DNA was visualised on UV transilluminator and photographed. The gel was left on the transilluminator for a further 2min. The gel was then placed in denaturing solution (1.5M NaCl, 0.5M NaOH) for 45min and was then transferred to neutralising solution (2M NaCl, 1M Tris pH5.5) for a further 45min. It was then placed upside down on a plastic support which was covered in a triple layer of 3mm Whatman paper which was used as a wick placed in a resevoir of 20x SSC (150mM NaCl, 15mM sodium citrate pH8.0). A piece of Hybond N+ nylon membrane cut to the same size as the gel was pre-soaked in the neutralising solution and then carefully placed on the gel ensuring there were no air bubbles present. 10 pieces of 3mm Whatman paper, pre-soaked in 20 x SSC, where placed on top of the nylon membrane and a suitable stack of dry paper towels was placed on top with a weight. The DNA was then transferred by capillary action to the nylon membrane overnight. The membrane was removed and washed in 3x SSC and then crosslinked (UV Stratalinker 2400). The membrane was then air dried for 30min. The membrane if necessary could be stored at 4°C prior to hybridization.
2.3.5.4.2 Colony blotting

Hybond N+ nylon membrane was placed carefully on to the agar surface and was marked using a sterile needle to ensure correct orientation of colonies. The membrane was removed after 1 min and placed colony side up on a pad of absorbent filter paper soaked in denaturing solution for 7 min. It was then transferred in an absorbent filter paper soaked in neutralizing solution and left for 3 min. This step was repeated with a fresh pad soaked in neutralizing solution. Filters were then washed in 2X SSC, transferred to dry filter paper and left to air dry. The DNA was fixed by UV crosslinking (UV Stratalinker 2400).

2.3.5.5 Hybridization

The nylon membranes were prehybridized for between 2-5 hours at 65°C with 30 ml of prehybridization solution (0.005M phosphate buffer pH 7, 7% SDS, 1% BSA). The volume of the prehybridization solution was reduced to 5 ml prior to adding the oligonucleotide probe. The probe used was 100 pmol of (CA)$_{15}$ oligonucleotide that was end-labeled in a bulk mix of 30 µl volume with 3 pmol of $\gamma^{32}$P-ATP using T4 polynucleotide kinase (Promega, UK). The probe was hybridized to the membrane overnight at 65°C. The membrane was then washed twice for 10 min in 2X SSC/0.1% w/v SDS at room temperature. The filters were then wrapped in cling film and exposed to X-ray film at -70°C.
2.3.5.6 DNA Sequencing

2.3.5.6.1 Cycle sequencing of PCR products

The fmol® DNA Sequencing System from Promega was used that includes a Sequencing Grade Taq DNA polymerase. The sequencing primer was end-labeled in a bulk mix of 10μl volume by incubating 10pmol of the primer with 10pmol of γ32P-ATP using T4 polynucleotide kinase (Promega, UK). The template DNA for each set of sequencing reactions consisted of 4μl of PCR product that had been previously purified using the Wizard™ DNA Clean-up resin system (Promega). The sequencing reactions were performed as per manufacturer’s instructions. The cycling profile had an initial denaturation step at 95°C for 2min followed by 30 cycles at 95°C for 30sec and 70°C for another 30sec. After cycling had finished sequencing reactions were run on a 6% denaturing polyacrylamide gels (National Diagnostics Ltd) (see Polyacrylamide Gel Electrophoresis).

2.3.5.6.2 Automated bi-directional sequencing of PCR products using tailed primers

In this method, the primers used essentially contain two primer sequences on one oligonucleotide (Figure 2.1). One part of the primer sequence is the specific to the sequence of the PCR product that is being amplified (PCR primer) and the other part is the tail sequence. The tail sequence usually corresponds to a readily available standard sequencing primer such as an M13 forward or reverse primer. The tailed sequence is incorporated into the PCR product during the amplification reaction. For this project the primers used were as follows:
Figure 2.1 PCR using tailed primers

Both PCR primers (forward and reverse) have added tails (tail 2 and tail 1 respectively). Tails can be any standard primer sequence such as M13 forward (-29), T3 promoter or T7 promoter. The tails are incorporated into the PCR product during the amplification reaction. These PCR products can then be sequenced in both directions by using standard primers labeled with an appropriate dye. This method is particularly appealing because the same standard primers can be used to sequence any PCR product.
The PCR product is then tested on a 1% (w/v) agarose gel. Single band PCR products without artifacts are required for optimal sequencing results.

The PCR product pre-sequencing kit (Amersham) is used according to the manufacturer's instructions to prepare PCR products for sequencing. It contains two enzymes, exonuclease I, which removes residual single-stranded primers and any extraneous ssDNA produced by the PCR and shrimp alkaline phosphatase, which removes the excess of deoxynucleoside triphosphates (dNTPs) from the PCR mixture.

Following purification, the 6μl of this mix were added directly to a standard 20μl sequencing reaction with 1pmol of each of the appropriate end-labeled sequencing primer. In bi-directional sequencing, both forward and reverse sequencing primers are labeled on their 5' with different near-infrared (NIR) fluorescent dyes (IRD700 and IRD800) and combined in one reaction. The sequencing primers used had as follows:

5'-IRD700-GGATAACAATTTTCACACAGG-3’
   M13 reverse primer (-21)
5'-IRD800-CACGACGTTGATAAACGAC- 3’
   M13 forward primer (-29)

4.5μl of this DNA/primer mix was added to 1.5μl of the G, A, T, C termination mix of the Thermo Sequenase fluorescent labeled primer cycle sequencing kit with 7-deaza-dGTP (Amersham). Thermal cycling conditions for the elongation/termination reactions were as follows: an initial denaturing step at 95°C for 5min followed by 30 cycles at 95°C, 60°C 72°C for 30-45 sec each step. The
sequencing products were mixed with an equal volume of stop solution (0.1% of fuchsin in formamide), denatured to 95°C for 3 min and then 1 μl was loaded on a pre-run denaturing gel.

A 6% denaturing polyacrylamide gel was prepared as follows: 7.5 ml of RapidGel™-XL (40%) solution (Amersham), 5 ml 10X TBE Long Run buffer, 4 ml formamide (Sigma), 21 gr Urea (Sigma) and ddH2O was added to reach a volume of 50 ml. 350 μl of 10% ammonium persulphate and 75 μl of TEMED were added to induce polymerization.

10X Long run TBE Tris Base (1340 mM),
Boric acid (450 mM) and
EDTA (Na2, 2XH2O) (25 mM)

41 cm glass plates were used and RapidGel™ XL is especially suited for the very long reads obtained with these gels.

Electrophoresis conditions were 45 W constant power and the temperature was maintained at 45°C until completion of the run.

LI-COR’s dual-dye automated DNA sequencer (Model 4200 IR² System) was used to electrophorese and analyze the forward and reverse sequences of the bi-directional sequencing reactions in parallel. Sequence data were interpreted using the automated base-calling algorithms of LI-COR’s Base ImagIR™ software.

2.3.5.7 Library screening

The PAC library RPCII was constructed by Pieter de Jong and his group at the Roswell Park Cancer Institute, Buffalo (Ioannou & Jong, 1996). The vector is pCYPAC2 (Figure 2.2) and can accommodate large DNA fragments. The DNA source is a normal male blood donor, and the average insert size 110 Kb. The library
Figure 2.2 pCYPAC2 vector used to construct the PAC RPC1 library.

PUC-LINK is stuffer fragment, being replaced by insert DNA during cloning.

consists of approx. 120,000 clones. 25% of the clones lack insert. The library is available from the HGMP-Resource Centre either as pools for screening by PCR, or as high-density gridded filters for screening by hybridisation.

Pools of library clones were sent out as live cells on dry ice. A three stage PCR screening system was used in order to isolate a genomic clone containing the human prepronociceptin gene (Figure 2.3). The primer sequences used for the PCR were the following: forward primer: 5'GCA GGA AGA GCC CGA GCC3' and reverse primer: 5'GGG GCG GGG AGA TGA ATG3', located respectively in the coding and 3'-uncoding regions of the human PNOC gene (Mollereau et al., 1996). The reaction amplified a fragment of 356bp. PCR was performed in a volume of 12.5μl, containing, 10mM Tris-Cl (pH 8.4), 50mM KCl, 1.0mM MgCl2, 200μM dNTPs, 2.0pmol of forward primer and an equal amount of reverse primer, and 0.2 units of Taq polymerase (QIAGEN). Cycle conditions were, 94°C for 5mins then 30 cycles of 94°C 1min, 57°C for 2min, 72°C for 3min, with a final extension step of 7min at 72°C, in a Gene Amp 9700 System (Perkin Elmer Cetus, Palo Alto, CA, USA).

In the first stage of the screening, 3μl of each one of 21 primary pools (Master plate PAC from A to U), each consisting of cells from 15 384-well microtitre plates, was used for PCR.

In the second stage, the 15 secondary pools for each positive primary pool were obtained from the HGMP-Resource Centre. 3μl of each one of the secondary pools were used for the PCR reaction.

In the final stage of the screening, for each positive secondary pool one 384-well plate was obtained containing cells with individual clones in each well. As the Resource Centre does not have the robotics necessary for pooling rows and columns,
Figure 2.3 PCR screening of the RPCI1 library

1st STAGE: 21 primary pools each consisting of clones from 15 384-well plates.

2nd STAGE: for each primary pool, there are 15 secondary pools, each consisting of clones from one 384-well plate.

3rd STAGE: 384-well microtitre plate containing individual clones in each well.
the following pooling scheme was followed: 5µl of liquid was removed from each of
the 15 wells of each row (letters A-P) with a multichannel pipette. The liquid was
pooled together in a Petri dish and transferred to a tube. After identification of the
positive row, a PCR was performed in order to identify which one of the 15 wells in
this row contained the clone of interest. Cells from this well were streaked out on
plates with kanamycin selection to obtain single colonies (see section 2.3.2.2.1).

2.4 STATISTICAL ANALYSIS

2.4.1 Parametric linkage analysis

The LINKAGE package (Lathrop & Lalouel, 1984, Lathrop et al., 1984)
includes computer programs which enable the calculation of lod scores at various
positions and which are conditional of transmission parameters of the disease locus,
the relevant allele frequencies and the genetic distance between the disease locus and
the genetic markers. For this study, the linkage analysis was carried out using the
computer program MLINK. For the analyses carried out assuming dominant
transmission the gene frequency of the abnormal allele was set to 0.0085 and the
heterozygote penetrance was set to the same value as for the abnormal homozygote.
For analyses carried out assuming recessive transmission the gene frequency of the
abnormal allele was set to 0.13 and the heterozygote penetrance was set to the same
value as for the normal homozygote. The penetrance values used were 0.4 and 0.6 for
the DOMS and DOMSS models respectively. Sporadic cases were accounted for by
assuming a penetrance of 0.005 and 0.01 for the normal homozygote for the DOMS
and DOMSS affection classes respectively.
In order to investigate the possibility that only a subset of pedigrees might have a susceptibility locus in the region studied, admixture test was also carried out (Ott, 1999). This is reported as the lod2 statistic in the presence of an estimated portion (alpha) of families linked (Risch, 1989).

2.4.2 “Model-free” linkage analysis

“Model-free” linkage analysis was carried out using the computer program MFLINK (Curtis & Sham, 1995). This analysis is model-free in the sense that the transmission model parameters are not fixed in advance (as in classical lod score method) but are freed in both the numerator and denominator of the likelihood ratio. MFLINK calculates the likelihood of the data with the disease locus at a given map position using a range of different dominant and recessive transmission models, all yielding the same disease prevalence ($K_p$) and parameterized using a single variable, the heterozygote penetrance ($f_h$, which is varied from 0 to 1).

The likelihood for the observed data under the hypothesis that a locus at a particular test position influences susceptibility in a proportion of families and under the hypothesis that it has no effect are compared. Both likelihoods are maximized independently over the heterozygote penetrance, $f_h$, and the likelihood under linkage is additionally maximized over alpha (the proportion of families linked). The difference between the two maximized log likelihoods provides the “model-free” lod score (MFLOD) for the position tested.

MFLINK also reports the maximum lod scores under the assumptions of homogeneity (MLOD) and admixture (MALOD), which are obtained over disease-model parameters subject to constraints that are the same as those for MFLOD.
In this study, MFLINK analyses were carried out using both the core and spectrum affection models, with population prevalence being set to 0.011 or 0.019 respectively, to match the prevalences used in the lod score analyses. Two-point analyses were carried out with each marker using a test position at a recombination fraction of 0.05 with the marker.

2.4.3 Allelic association analysis using CLUMP

The data obtained from the case-control allelic association studies was analysed using the CLUMP program (Sham & Curtis, 1995b). It is a program designed to assess the significance of the departure of observed values from the expected values under the null hypothesis for a contingency table conditional on marginal totals.

The chi-squared statistic ($\chi^2$) associated with a contingency table is defined in the usual way as the sum over all cells of the squared difference between observed and expected value divided by the expected value. The expected value for a cell is the total for its row multiplied by the total for its column divided by the overall total number of observations.

The four $\chi^2$ statistics produced by CLUMP are:

(i) T1, a straightforward Pearson’s $\chi^2$ statistic of the ‘raw’ contingency table,

(ii) T2, the $\chi^2$ statistic of a table with rare alleles grouped together to prevent small expected cell counts,

(iii) T3, the largest of the $\chi^2$ statistics of 2x2 tables each of which compares one allele against the rest grouped together and finally

(iv) T4, the largest of the $\chi^2$ statistics of all possible 2x2 tables comparing any combination of alleles against the rest.
The significance of each of the \( \chi^2 \) values is evaluated using a Monte Carlo approach, by performing repeated simulations to generate tables having the same marginal totals as the one under consideration. The proportion of times the \( \chi^2 \) value produced by the real data is reached yields an estimate of the significance of the departure of the observed data from the expectation under the null hypothesis. In this way an empirical \( p \) value is produced.

This method obviates problems related to dealing with sparse contingency tables or to correcting for testing of multiple alleles.

One thousand simulations were performed for each analysis.

2.4.4 Pair-wise linkage disequilibrium calculations using EH

Pair-wise linkage disequilibrium was calculated between each marker locus using the computer program EH (Terwilliger & Ott, 1994). Maximum likelihood values for the haplotype frequencies for each pairwise combination of alleles are estimated under \( H_0 \) (no allelic association) and \( H_1 \) (allelic association). A \( \chi^2 \) statistic is produced, which is the difference in 2 \( \ln(\text{likelihood}) \) and therefore pairwise \( p \)-values can be obtained for each test.
CHAPTER 3

GENETIC LINKAGE STUDIES

ON CHROMOSOME 8p21-22
3.1 INTRODUCTION

There are several linkage studies that provide supportive evidence for a susceptibility locus for schizophrenia on chromosome 8p21-22 (Blouin et al., 1998, Brzustowicz et al., 1999, Kendler et al., 1996). These studies have been reviewed extensively in chapter 1 of this thesis (see section 1.7.4).

So far, three linkage studies for schizophrenia have been performed by our group on chromosome 8p21-22.

The first study was performed with 23 families (11 Icelandic and 12 British pedigrees) containing multiple cases of schizophrenia, containing 95 individuals with "core" (DOMS) and 17 individuals with "spectrum" schizophrenia (DOMSS). Four microsatellite markers, D8S261, D8S136, D8S133 and D8S283, that span the chromosome 8p21-22 region were chosen for the linkage analysis. Two-point lod analyses using MLINK assuming homogeneity and lack of homogeneity (lod2) were carried out. In addition, "model-free" analyses were carried out with MFLINK to ensure that all compatible recessive and dominant models were tested for, again assuming homogeneity and lack of homogeneity. In this study a two-point admixture lod score of 2.006 at a recombination fraction of 0.05 was obtained with marker D8S136 under a narrowly defined "core" model of affection (DOMS) and dominant mode of inheritance. The "model-free" analysis provided a maximized admixture lod score of 2.17 again with the DOMS model of affection (Kalsi, 1996).

The second linkage study was performed with 13 pedigrees (8 Icelandic and 5 British), which comprised the most densely affected and largest from the 23 pedigrees used in the previous linkage study and contained 56 individuals with core (DOMS) and 12 with spectrum schizophrenia (DOMSS). These pedigrees contributed to a
large multi-centre collaboration organized by the European Science Foundation (ESF), who arranged for the genotyping to be performed. In this study, a 3-point multipoint lod of 3.5 was obtained between markers D8S504 and D8S503 under a narrow model of affection and recessive mode of inheritance. In addition, a 3-point lod of 2.9 was obtained between markers D8S503 and D8S552 and a 3-point lod of 2.8 between markers D8S1771 and D8S283, which are more centromeric and closer to the Blouin et al, 1998 study. Furthermore, a 5-point lod score of 3.2 was obtained with markers D8S261, D8S258, D8S1771 and D8S283 under recessive mode of transmission and narrow phenotype definition and it was the same under homogeneity and heterogeneity (Gurling et al., 1999), (Gurling et al., 2000, submitted).

The present study evaluated the evidence for linkage of schizophrenia to chromosome 8p21-22 in a sample of 16 English families. A more detailed description of these families is given in chapter 2 (see section 2.1 and 2.2).

Two affection classes were used for the linkage analyses as before: "core schizophrenia" model denoted DOMS, consisting of schizophrenia and unspecified functional psychosis and "schizophrenia spectrum" denoted DOMSS, consisting additionally of schizoid and schizotypal personality disorder. In this study, there were 26 individuals with DOMS and 8 with DOMSS definition of schizophrenia.

The five markers chosen for this study, D8S261, D8S258, D8S136, D8S133 and D8S283 cover the region of 8p21-22 and four of them are the same as in our first linkage study. Information was obtained on these markers from The Genome Database (http://www.gdb.org) and is provided on Table 3.1.
Table 3.1 Information on markers used in the linkage study on chromosome 8p21-22

<table>
<thead>
<tr>
<th>MARKER</th>
<th>PRIMERS</th>
<th>TYPE OF REPEAT</th>
<th>No. ALLELES</th>
<th>ALLELE SIZES (bp)</th>
<th>HET.</th>
<th>POSITION^1</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S261</td>
<td>5'TGC CAC TGT CTT GAA AAT CC3' 5'TAT GGC CCA GCA ATG TGT AT3'</td>
<td>(CA)_n</td>
<td>8</td>
<td>144-128</td>
<td>0.7760</td>
<td>37</td>
</tr>
<tr>
<td>D8S258</td>
<td>5'CTGCCAGGAATCAACTGAG3' 5'TTGACAGGGACCACG3'</td>
<td>(CA)_n</td>
<td>5</td>
<td>154-144</td>
<td>0.7139</td>
<td>41</td>
</tr>
<tr>
<td>D8S136</td>
<td>5'GCCCAAAGAGGAGAATAAA3' 5'CTGTTCACACCGAAGC3'</td>
<td>(CA)_n</td>
<td>9</td>
<td>89-73</td>
<td>0.8810</td>
<td>43</td>
</tr>
<tr>
<td>D8S133</td>
<td>5'CAGGTGGGAAAACGTGAGGGA3' 5'AGCAACTGTCAACATATTGCTC3'</td>
<td>(CA)_n</td>
<td>10</td>
<td>112-94</td>
<td>0.7750</td>
<td>45</td>
</tr>
<tr>
<td>D8S283</td>
<td>5'ATTCATGTCTAGGCCATTGC3' 5'AGATACAGATGTAGATCTCTCCG3'</td>
<td>(CA)_n</td>
<td>10</td>
<td>129-103</td>
<td>0.7996</td>
<td>60</td>
</tr>
</tbody>
</table>

1: in Kosambi centimorgans, derived from the Marshfield Map positions apart from D8S133 which was derived from the NIH/CEPH linkage map (1992)
Two-point lod analyses using MLINK assuming homogeneity and lack of homogeneity (lod2) were carried out at recombination fraction (\(\theta\)) 0,.01,.05,.1,.2,.3 and .4. Four genetic models were selected to be analysed, one dominant and one recessive for each of the DOMS and DOMSS definitions of affection.

In addition, "model-free" analyses were undertaken with MFLINK to ensure that all compatible recessive and dominant models were tested for, again assuming homogeneity and lack of homogeneity.

3.2 RESULTS

The total two-point lod scores and the admixture lod scores (lod2) obtained with the "core" (DOMS) definition of the phenotype are presented in table 3.2 and 3.3 under dominant and recessive mode of inheritance respectively. Furthermore, the total two-point lod scores and the admixture lod scores (lod2) obtained with the "spectrum" (DOMSS) definition of the phenotype are presented in table 3.4 and 3.5 under dominant and recessive modes of inheritance respectively. It can be seen that the results are very similar for each affection model and do not indicate any evidence for linkage on chromosome 8p21-22 for schizophrenia in this set of families. The highest total two-point lod score was 0.384 for marker D8S258 under DOMS affection model and dominant mode of inheritance. In addition, this marker produced the best results for all four genetic models tested in this study.

Examination of the lod scores by family did not reveal any suggestion of subgroup of families having strongly positive scores, and the admixture test did not provide any significant evidence to reject homogeneity.
TABLE 3.2 Two-point MLINK lod and lod2 scores between markers at 8p21-22 and schizophrenia under the DOMS affection model and dominant mode of inheritance

<table>
<thead>
<tr>
<th>MARKER</th>
<th>theta</th>
<th>0.000</th>
<th>0.010</th>
<th>0.050</th>
<th>0.100</th>
<th>0.200</th>
<th>0.300</th>
<th>0.400</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S261</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>-2.415</td>
<td>-2.218</td>
<td>-1.651</td>
<td>-1.191</td>
<td>-0.629</td>
<td>-0.300</td>
<td>-0.104</td>
<td></td>
</tr>
<tr>
<td>lod2</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>D8S258</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.384</td>
<td>0.371</td>
<td>0.321</td>
<td>0.264</td>
<td>0.170</td>
<td>0.098</td>
<td>0.043</td>
<td></td>
</tr>
<tr>
<td>lod2</td>
<td>0.384</td>
<td>0.371</td>
<td>0.321</td>
<td>0.264</td>
<td>0.170</td>
<td>0.098</td>
<td>0.043</td>
<td></td>
</tr>
<tr>
<td>D8S136</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>-1.177</td>
<td>-0.979</td>
<td>-0.531</td>
<td>-0.268</td>
<td>-0.060</td>
<td>-0.005</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>lod2</td>
<td>0.052</td>
<td>0.051</td>
<td>0.044</td>
<td>0.036</td>
<td>0.023</td>
<td>0.011</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>D8S133</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>-1.072</td>
<td>-0.983</td>
<td>-0.740</td>
<td>-0.567</td>
<td>-0.388</td>
<td>-0.274</td>
<td>-0.154</td>
<td></td>
</tr>
<tr>
<td>lod2</td>
<td>0.038</td>
<td>0.032</td>
<td>0.013</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>D8S283</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>-1.933</td>
<td>-1.675</td>
<td>-1.057</td>
<td>-0.660</td>
<td>-0.287</td>
<td>-0.128</td>
<td>-0.050</td>
<td></td>
</tr>
<tr>
<td>lod2</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>
TABLE 3.3 Two-point MLINK lod and lod2 scores between markers at 8p21-22 and schizophrenia under the DOMS affection model and recessive mode of inheritance

<table>
<thead>
<tr>
<th>MARKER</th>
<th>theta</th>
<th>0.000</th>
<th>0.010</th>
<th>0.050</th>
<th>0.100</th>
<th>0.200</th>
<th>0.300</th>
<th>0.400</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S261</td>
<td>Total</td>
<td>-2.585</td>
<td>-2.414</td>
<td>-1.854</td>
<td>-1.339</td>
<td>-0.681</td>
<td>-0.308</td>
<td>-0.099</td>
</tr>
<tr>
<td></td>
<td>lod2</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>D8S258</td>
<td>Total</td>
<td>0.252</td>
<td>0.248</td>
<td>0.228</td>
<td>0.198</td>
<td>0.132</td>
<td>0.071</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>lod2</td>
<td>0.252</td>
<td>0.248</td>
<td>0.228</td>
<td>0.198</td>
<td>0.132</td>
<td>0.071</td>
<td>0.026</td>
</tr>
<tr>
<td>D8S136</td>
<td>Total</td>
<td>-1.115</td>
<td>-1.001</td>
<td>-0.674</td>
<td>-0.419</td>
<td>-0.162</td>
<td>-0.064</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td>lod2</td>
<td>0.055</td>
<td>0.051</td>
<td>0.036</td>
<td>0.021</td>
<td>0.003</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>D8S133</td>
<td>Total</td>
<td>-0.084</td>
<td>-0.019</td>
<td>0.105</td>
<td>0.133</td>
<td>0.071</td>
<td>-0.009</td>
<td>-0.043</td>
</tr>
<tr>
<td></td>
<td>lod2</td>
<td>0.182</td>
<td>0.183</td>
<td>0.179</td>
<td>0.157</td>
<td>0.076</td>
<td>0.005</td>
<td>0.000</td>
</tr>
<tr>
<td>D8S283</td>
<td>Total</td>
<td>-1.939</td>
<td>-1.859</td>
<td>-1.517</td>
<td>-1.117</td>
<td>-0.552</td>
<td>-0.240</td>
<td>-0.077</td>
</tr>
<tr>
<td></td>
<td>lod2</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>MARKER</td>
<td>theta</td>
<td>0.000</td>
<td>0.010</td>
<td>0.050</td>
<td>0.100</td>
<td>0.200</td>
<td>0.300</td>
<td>0.400</td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lod2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D8S261</td>
<td>Total</td>
<td>-3.461</td>
<td>-3.093</td>
<td>-2.211</td>
<td>-1.578</td>
<td>-0.840</td>
<td>-0.413</td>
<td>-0.149</td>
</tr>
<tr>
<td></td>
<td>lod2</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>D8S258</td>
<td>Total</td>
<td>-0.338</td>
<td>-0.334</td>
<td>-0.319</td>
<td>-0.292</td>
<td>-0.200</td>
<td>-0.093</td>
<td>-0.019</td>
</tr>
<tr>
<td></td>
<td>lod2</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>D8S136</td>
<td>Total</td>
<td>-0.820</td>
<td>-0.653</td>
<td>-0.226</td>
<td>0.032</td>
<td>0.170</td>
<td>0.115</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>lod2</td>
<td>0.127</td>
<td>0.132</td>
<td>0.149</td>
<td>0.167</td>
<td>0.172</td>
<td>0.115</td>
<td>0.027</td>
</tr>
<tr>
<td>D8S133</td>
<td>Total</td>
<td>-1.316</td>
<td>-1.065</td>
<td>-0.524</td>
<td>-0.218</td>
<td>0.001</td>
<td>0.032</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>lod2</td>
<td>0.027</td>
<td>0.030</td>
<td>0.038</td>
<td>0.044</td>
<td>0.044</td>
<td>0.033</td>
<td>0.012</td>
</tr>
<tr>
<td>D8S283</td>
<td>Total</td>
<td>-1.881</td>
<td>-1.549</td>
<td>-0.856</td>
<td>-0.473</td>
<td>-0.180</td>
<td>-0.094</td>
<td>-0.051</td>
</tr>
<tr>
<td></td>
<td>lod2</td>
<td>0.111</td>
<td>0.104</td>
<td>0.077</td>
<td>0.049</td>
<td>0.009</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
TABLE 3.5 Two-point MLINK lod and lod2 scores between markers at 8p21-22 and schizophrenia under the DOMSS affection model and recessive mode of inheritance

<table>
<thead>
<tr>
<th>MARKER</th>
<th>theta</th>
<th>0.000</th>
<th>0.010</th>
<th>0.050</th>
<th>0.100</th>
<th>0.200</th>
<th>0.300</th>
<th>0.400</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S261</td>
<td>Total</td>
<td>-2.418</td>
<td>-2.267</td>
<td>-1.778</td>
<td>-1.309</td>
<td>-0.673</td>
<td>-0.305</td>
<td>-0.098</td>
</tr>
<tr>
<td></td>
<td>lod2</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>D8S261</td>
<td>Total</td>
<td>-0.606</td>
<td>-0.536</td>
<td>-0.332</td>
<td>-0.180</td>
<td>-0.044</td>
<td>-0.003</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>lod2</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>D8S156</td>
<td>Total</td>
<td>-0.820</td>
<td>-0.653</td>
<td>-0.226</td>
<td>0.032</td>
<td>0.170</td>
<td>0.115</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>lod2</td>
<td>0.127</td>
<td>0.132</td>
<td>0.149</td>
<td>0.167</td>
<td>0.172</td>
<td>0.115</td>
<td>0.027</td>
</tr>
<tr>
<td>D8S133</td>
<td>Total</td>
<td>-1.127</td>
<td>-1.034</td>
<td>-0.725</td>
<td>-0.457</td>
<td>-0.191</td>
<td>-0.107</td>
<td>-0.070</td>
</tr>
<tr>
<td></td>
<td>lod2</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>D8S283</td>
<td>Total</td>
<td>-2.305</td>
<td>-2.016</td>
<td>-1.295</td>
<td>-0.806</td>
<td>-0.336</td>
<td>-0.143</td>
<td>-0.053</td>
</tr>
<tr>
<td></td>
<td>lod2</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
The “model-free” analysis did not provide any further support for linkage (Table 3.6). The best result was a maximized admixture lod score (MALOD) and a “model-free” lod score (MFLOD) of 0.325 with marker D8S136 and DOMS affection model, while a maximized lod score (MLOD) and an MALOD of 0.310 was obtained for marker D8S258 and DOMS affection model.
TABLE 3.6 Two-point MFLINK lod scores between schizophrenia and markers on chromosome 8p21-22 under the DOMS and DOMSS affection models

<table>
<thead>
<tr>
<th>MARKER</th>
<th>AFFECTION MODEL</th>
<th>MLOD</th>
<th>MALOD</th>
<th>MFLOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S261</td>
<td>DOMS</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>DOMSS</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>D8S258</td>
<td>DOMS</td>
<td>0.310</td>
<td>0.310</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>DOMSS</td>
<td>0.019</td>
<td>0.022</td>
<td>0.000</td>
</tr>
<tr>
<td>D8S136</td>
<td>DOMS</td>
<td>0.000</td>
<td>0.325</td>
<td>0.325</td>
</tr>
<tr>
<td></td>
<td>DOMSS</td>
<td>0.153</td>
<td>0.210</td>
<td>0.210</td>
</tr>
<tr>
<td>D8S133</td>
<td>DOMS</td>
<td>0.132</td>
<td>0.139</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>DOMSS</td>
<td>0.000</td>
<td>0.072</td>
<td>0.000</td>
</tr>
<tr>
<td>D8S283</td>
<td>DOMS</td>
<td>0.000</td>
<td>0.008</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>DOMSS</td>
<td>0.000</td>
<td>0.155</td>
<td>0.000</td>
</tr>
</tbody>
</table>

MLOD: maximized lod score
MALOD: maximized admixture lod score
MFLOD: "model-free" lod score

All the analyses were performed at theta 0.05
3.3 DISCUSSION

This study was performed as an attempt to provide further support for linkage between schizophrenia and the chromosome 8p21-22 susceptibility locus.

A parametric method of linkage analysis was applied using four different genetic models, consisting of either dominant or recessive transmission and a "core" or a "spectrum" definition of the phenotype. Parametric linkage analysis was conducted because it has been suggested to be more powerful than non-parametric methods (Dumer et al., 1999) and is a robust method for detecting linkage despite errors or simplifications in the analysing model, as long as both a dominant and a recessive model are used (Abreu et al., 1999, Dumer et al., 1999, Greenberg et al., 1998).

In addition "model-free" likelihood based analysis was conducted as a further validation of the robustness of the linkage findings.

The results obtained in this study provide no support for the hypothesis that a locus for schizophrenia exists in this region in the set of families examined assuming that the transmission models are approximately correct.

Additionally, the results of the "model-free" analyses provide some reassurance that negative results have not been obtained due to an artefact of an erroneous transmission model.

The differences from the previous linkage findings in this region obtained by our group may be due to random variation in the proportion of linked families at this schizophrenia susceptibility locus. Furthermore, the families in this study were considerably smaller and had less affected members than the previously examined pedigrees.
CHAPTER 4

ISOLATION OF A NEW POLYMORPHIC MARKER IN A CANDIDATE GENE AND AN ALLELIC ASSOCIATION STUDY BETWEEN SCHIZOPHRENIA AND TWO CANDIDATE GENES: PREPRONOCICEPTIN (PNOC) AND NEURONAL CHOLINERGIC RECEPTOR, NICOTINIC, ALPHA POLYPEPTIDE 2 (CHRNA2)
4.1 INTRODUCTION

A positional candidate approach has been employed to elucidate the role of two potential candidate genes that map to the region of chromosome 8 implicated in schizophrenia. The genes are: prepronociceptin (PNOC), neuronal nicotinic acetylcholine receptor subunit alpha 2 (CHRNA2).

The gene structure and organization of PNOC is very similar to that of the precursors of the endogenous opioid peptides; preproenkephalin, preprodynorphin, and prepro-opiomelanocortin, suggesting that these four genes have evolved from a common ancestor. The gene is also highly conserved between murine and human species (Mollereau et al., 1996). The PNOC gene is predominantly expressed in the brain and spinal cord and at lower levels in the ovary, which is the sole peripheral organ expressing the gene. Proteolytic processing of the prepronociceptin precursor generates at least three biologically active peptides: nocistatin, nociceptin and nocII/nocIII. Nociceptin (orphanin FQ), a heptadecapeptide, is an endogenous ligand of the orphan opioid receptor-like receptor 1 (ORL1) (Meunier et al., 1995). Nocistatin lies at the N-terminal end of nociceptin. Its length varies in human, rat and mouse with 30, 35 and 41 aminoacid residues respectively. NocII a heptadecapeptide resides at the C-terminal end of nociceptin in the PNOC precursor. Further processing of nocII results in the nocIII, which is only 3 arginine residues shorter.

The ORL1 receptor has a strong affinity for nociceptin, but not the two other bioactive peptides derived from its precursor. Specific receptors for nocII and nocistatin have not yet been identified.

The various roles of these peptides in the control of brain functions are not yet fully understood. Nociceptin and its receptor are expressed in a wide range of CNS
regions and several studies demonstrate their involvement in a great number of
diverse brain functions, such as: pain, learning, memory, fear and anxiety responses,
auditory processing, neuroendocrine control, sleep and neuronal development
(Darland et al., 1998). Nocistatin, although produced by the same precursor as
nociceptin seems to play opposite roles in pain transmission as well as in learning and
memory (Okuda-Ashitaka et al., 1998). Finally, nocII has been implicated in
affecting locomotion in a manner similar to that of nociceptin (Florin et al., 1997).

The action of these peptides in many parts of the nervous system makes them
potential targets for new drugs and candidates in many neurological disorders.

The PNOC gene has been localized to the 8p21 region, between sequence-
tagged site markers WI-5833 and WI-1172, by using a radiation hybrid panel (Figure
4.1). The expression pattern and the wide variety of functions as well as its position
made the PNOC gene an attractive candidate for schizophrenia. In order to assess the
candidacy of this gene in schizophrenia a highly polymorphic microsatellite repeat
was isolated from a genomic clone containing the human PNOC gene. This marker
was characterized and used in a case-control allelic association study.

The next gene studied was the neuronal cholinergic receptor, nicotinic, alpha
polypeptide 2 (CHRNA 2). It consists part of the neuronal acetylcholine receptors,
which are members of a family of multisubunit, neurotransmitter-gated ion channels
that are responsible for rapid intercellular communication (McGehee, 1999). These
receptors are composed of pentamers of two acetylcholine binding α subunits and
three structural β subunits.

The alpha 7 subunit of the neuronal nicotinic receptor (CHRNA7) is
considered a candidate gene for the P50 deficit, a neurophysiological deficit
associated with schizophrenia (Freedman et al., 1997). This and the fact that
Figure 4.1 GDB Comprehensive map showing the positions of PNOC and CHRNA2 on 8p21. The broader chromosomal area that is implicated in linkage with schizophrenia lies approximately between microsatellite markers D8S261 and D8S1771. According to Gene Map 98 PNOC maps to the interval between microsatellite markers D8S1734 and D8S1829 and their respective position is shown on the map. Finally, microsatellite marker D8S131 is localized within the CHRNA gene. Obtained from The Genome Database, http://www.gdb.org.
CHRNA2 maps to 8p21 (Wood et al., 1995), a region implicated by several linkage studies to contain a locus for schizophrenia, made it a strong candidate.

The human CHRNA2 gene has been positioned on 8p21 (Figure 4.1). Two previously characterized microsatellites markers D8S131 and D8S131P, both upstream of exon 2 of the CHRNA2 gene, were used to look for evidence of allelic association with schizophrenia (Wood et al., 1995).

4.2 ISOLATION OF A (CA)n REPEAT

4.2.1 RPCI1 PAC library screening

The PNOC gene was identified from the PAC human genomic library-RPCI1 (Ioannou & Jong, 1996). Pools of library clones, which were sent out as live cells on dry ice, were screened by PCR. The primer sequences used were the following: forward primer: 5'GCA GGA AGA GCC CGA GCC3' and reverse primer: 5'GGG GCG GGG AGA TGA ATG3', located respectively in the coding and 3'-noncoding regions of the human PNOC gene (Mollereau et al., 1996). The reaction amplified a fragment of 356bp.

A three stage screening system was used as described in materials and methods (see section 2.3.5.7). In the first stage, 21 primary pools coded from A to U, were screened. 3µl of each pool and 2ng of human genomic DNA (positive control) were used for PCR amplification. 3 pools (Master plate D, H and I) were identified as positive (Figure 4.2A). The secondary pools corresponding to each of the positive primary pools were obtained from the HGMP-Resource Centre. Upon PCR
amplification of the secondary pools from Master plate PAC D and Master plate PAC H non-specific products were obtained, indicating that the first screening stage produced false positives. However, PCR amplification of the secondary pools from Master plate PAC I produced a further positive pool, I2 (Master plate I, well #2) (Figure 4.2A). In the final, tertiary stage, a 384-well plate (Plate 122) containing one individual clone/well was supplied from the Resource Centre. The pooling scheme described in materials and methods was used. Finally, the human genomic clone in row N and column 11 of this plate was the only positive clone obtained by PCR amplification (Figure 4.2B and C).

4.2.2 Isolation of a PAC clone containing the human PNOC gene

Cells from the positive clone were streaked out on plates with kanamycin selection (25mg/ml) to obtain single colonies. Single colonies were grown in sufficient quantities to be used for DNA extraction.

The PNOC gene has been localized to the 8p21 region, between sequence –tagged site markers WI-5833 and WI-1172, by using a radiation hybrid panel (Mollereau et al., 1996). 10μg of PAC 122n11 DNA, isolated as mentioned above, containing the PNOC gene were used for FISH, which was performed in collaboration with Dr Margaret Fox at the MRC Human Biochemical Genetics Unit, of UCL. The result from the FISH indicated a strong signal on chromosome 8p12-21 and a weak signal on chromosome 17q25, which may be an artefact or some weak sequence homology (Figure 4.3A). This clone will be described as PAC 122n11 clone from the RPCI11 library.
Figure 4.2 RPCII library screening by PCR.

A. The 1st and 2nd stage of screening. The +ve clones from the 1st stage were D, H and I. When these primary pools were amplified again, pool H gave a very weak band. In the 2nd stage only pool I2 (Master plate I, well #2) produced a band of the expected size. (See Figure 2.3 for a diagrammatic representation of the screening procedure followed).

B. The 3rd stage of screening. The wells of each of the 16 rows (from A to P) of the 384-well plate (Plate 122) were pooled together. Upon PCR amplification the pooled row N gave the correct size band indicating that one of the wells in this row contained the correct clone.

C. All 24 PAC clones contained in each well of row N were PCR amplified. Bands in lanes 2, 3 and 4 are artefacts. The clone in row N and column 11 produced a single band of the correct size. This clone will be described as PAC 122n11 clone from the RPCII library.

(**): indicate the bands from the positive clones.

c, c1, c2 and c3: indicate the positive controls, which were genomic DNA from different individuals.

m: indicates DNA molecular weight markers.
Figure 4.3

A. FISH of the PAC 122n11 DNA performed by Dr Margaret Fox (MRC Human Biochemical Genetics Unit, UCL). The result indicated a strong signal on chromosome 8p12-21 and a weaker signal on chromosome 17q25.

B. PCR amplification of the PAC 122n11 DNA (P) using the same primers and conditions used for screening the PAC library. As +ve control (c) was used human genomic DNA. A single band of the correct size (356bp) was obtained from both samples on an agarose gel. The positions of the DNA molecular weight marker (m) are indicated.
Finally, cycle sequencing was performed using as a sequencing primer the reverse primer (5'GGG GCG GGG AGA TGA ATG3') that was used for screening the PAC library. Part of the sequence that was obtained was the following:

Fasta format of the sequence produced by cycle sequencing.
>cutpaste-12069 [Unknown form], 97 bases, 891DC4CE checksum.
GAGCAAGGCTAAGACAGTTGGGTGTGAGGTCTCTCTCTTGCCTGCCCAGCTT
ATTGGTGCCCAACTCACTCATGGGTGGAGTTGACGCGGTACAGCTGGG

This sequence was shown to correspond to part of exon 3 of the human prepronociceptin gene (Figure 4.4). BLAST 2.0 homology search (Altschul et al., 1997) at the NCBI (National Centre for Biotechnology Information, http://www.ncbi.nlm.nih.gov) also revealed the match to exon 3. The output obtained from the BLAST 2.0 search is shown in Figure 4.5.
Figure 4.4 H.sapiens prepronociceptin exon 3 sequence (in bold).

The primers used for PCR screening the RPC11 library are highlighted in yellow. In addition, the sequence obtained after cycle sequencing the PAC 122n11 DNA clone is highlighted in grey. The sequencing primer used was the reverse primer used in the PCR library screening.
Figure 4.5 BLAST 2.0 output after searching for a match for the 97 bp of sequence obtained from cycle sequencing PAC 122n11 DNA clone against the 'nr' (non redundant) database. The most significant match is with exon 3 of the human prepronociceptin gene. Obtained from http://www.ncbi.nlm.nih.gov.
**Query:** (97 letters)

**Database:** nt 617,642 sequences; 1,858,941,277 total letters

### Distribution of 31 Blast Hits on the Query Sequence

![Distribution of 31 Blast Hits on the Query Sequence](image)

<table>
<thead>
<tr>
<th>Score</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>192</td>
<td>1e-47</td>
</tr>
<tr>
<td>90</td>
<td>1e-16</td>
</tr>
<tr>
<td>62</td>
<td>3e-08</td>
</tr>
<tr>
<td>54</td>
<td>8e-06</td>
</tr>
<tr>
<td>54</td>
<td>8e-06</td>
</tr>
<tr>
<td>54</td>
<td>8e-06</td>
</tr>
<tr>
<td>54</td>
<td>8e-06</td>
</tr>
<tr>
<td>54</td>
<td>8e-06</td>
</tr>
<tr>
<td>54</td>
<td>8e-06</td>
</tr>
<tr>
<td>54</td>
<td>8e-06</td>
</tr>
<tr>
<td>52</td>
<td>3e-05</td>
</tr>
<tr>
<td>44</td>
<td>0.008</td>
</tr>
<tr>
<td>38</td>
<td>0.47</td>
</tr>
<tr>
<td>38</td>
<td>0.47</td>
</tr>
</tbody>
</table>

**Sequences producing significant alignments:**

<table>
<thead>
<tr>
<th>Score (bits)</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>192</td>
<td>1e-47</td>
</tr>
<tr>
<td>90</td>
<td>1e-16</td>
</tr>
<tr>
<td>62</td>
<td>3e-08</td>
</tr>
<tr>
<td>54</td>
<td>8e-06</td>
</tr>
<tr>
<td>54</td>
<td>8e-06</td>
</tr>
<tr>
<td>54</td>
<td>8e-06</td>
</tr>
<tr>
<td>54</td>
<td>8e-06</td>
</tr>
<tr>
<td>54</td>
<td>8e-06</td>
</tr>
<tr>
<td>54</td>
<td>8e-06</td>
</tr>
<tr>
<td>52</td>
<td>3e-05</td>
</tr>
<tr>
<td>44</td>
<td>0.008</td>
</tr>
<tr>
<td>38</td>
<td>0.47</td>
</tr>
<tr>
<td>38</td>
<td>0.47</td>
</tr>
</tbody>
</table>

**emb|X97368.1|HSPFNNX3 H.sapiens gene encoding prepronociceptin...**

**Length = 878**

**Score = 192 bits (97), Expect = 1e-47**

**Identities = 97/97 (100%)**

**Strand = Plus / Minus**

**Query:**
```
1 gagcaagctagacatggtggcagtgtgctcttggtcccccaagctttattgtgcccc 60
```

**Sbjct:**
```
59 gagcaagctagacatggtggcagtgtgctcttggtcccccaagctttattgtgcccc 532
```

**Query:**
```
61 aacctcactcactggtggcagtggccggtatacgctggg 97
```

**Sbjct:**
```
531 aacctcactcactggtggcagtggccggtatacgctggg 495
```
4.2.3 Identification of a (CA)n repetitive element in the PAC clone containing the human PNOC gene

The PAC 122n11 DNA was digested with several different restriction endonucleases with 4- and 6-bp recognition sites, Alul, HaeIII, Sau3A and PstI. A Southern blot was performed using a (CA)15 probe and a DNA fragment produced by the PstI digestion appeared to contain a CA repeat (Figure 4.6).

The positive fragment was subcloned in pUC18 plasmid vector and transformed into E.coli competent cells by electroporation. Colony hybridization was then performed in order to identify single colonies containing a DNA fragment with a CA repeat. DNA was extracted from several positive colonies and was sequenced by MWG-Biotech. A DNA fragment was identified containing a long, uninterrupted (CA)26 repeat (Figure 4.7). PCR primers were designed to flank the repeat sequence using the Primer 3 software program (Whitehead Institute/MIT Center for Genome Research). The primers were as follows:

N13 Forward 5'-AGA CAT GGC CAG AAC AGC TT-3' and
N13 Reverse 5'-CCA GCT CAT GTG GAT GCT TA-3'.

Using these primers 100ng of PAC 122n11 DNA was PCR amplified in a volume of 12.5µl, containing, 10mM Tris-Cl (pH 8.4), 50mM KCl, 1.0mM MgCl2, 200µM dNTPs, 2.0pmol of forward primer and an equal amount of reverse primer, and 0.2 units of Taq polymerase (QIAGEN). Cycle conditions were, 94°C for 5mins then 30 cycles of 94°C, 56°C, 72°C each step for 30sec, with a final extension step of 7min at 72°C, in a Gene Amp 9700 System (Perkin Elmer Cetus, Palo Alto, CA, USA). The PCR product was visualized as a single band on an agarose gel verifying that the repeat element is contained in the original PAC clone that contains the prepronociceptin gene.
Figure 4.6 Southern blot analysis of the PAC 122n11 DNA clone.

The PAC 122n11 DNA clone was digested with *Alul* (lane A), *HaeIII* (lane H), *PstI* (lane P) and *Sau3A* (lane S); as +ve control was used a clone previously known to contain a (CA)n repeat. The probe used was a (CA)$_{15}$ oligonucleotide. Lane P shows a +ve signal in a fragment of approximately 2kb.
Figure 4.7 The DNA sequence containing a (CA)$_{26}$ repeat obtained from a single subclone of a PstI fragment of PAC 122n11 DNA. The primer sequences that are used to PCR amplify the repeat are shown in orange.
Part of the sequence near the repeat (underlined in Figure 4.7) was used to perform a Blast 2.0 search against the unfinished Human genome sequences database (htgs). This search revealed matches with the chromosome 8 clones (Figure 4.8). In addition, a less significant match was shown with chromosome 17, which may explain the results obtained by FISH were due to this sequence homology.

Finally, no match was revealed with the human prepronociceptin gene. The (CA)_n repeat therefore may lie in one of the introns of the gene or some other non-coding sequence of this gene in close proximity. Otherwise, as the average insert size of the PAC clones is 110 kb and the PNOC mRNA is 1015 bp, the repeat should be within ~109 kb from the gene.
Figure 4.8 BLAST 2.0 output after searching for a match for the 589bp of sequence lying next to the novel (CA)n repeat against the ‘htgs’ database (unfinished Human genome sequences database). The most significant match is with human chromosome 8 clones. Obtained from http://www.ncbi.nlm.nih.gov.
Query= (589 letters)

Database: htg  72,116 sequences; 4,095,156,099 total letters

Distribution of 2040 Blast Hits on the Query Sequence

Color Key for Alignment Scores

Scores producing significant alignments:

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Score (bits)</th>
<th>E Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>gb</td>
<td>AF276757.1</td>
<td>AF276757 Homo sapiens chromosome 8 clone RP3-...</td>
</tr>
<tr>
<td>gb</td>
<td>AC021678.5</td>
<td>AC021678 Homo sapiens chromosome 8 clone RP11...</td>
</tr>
<tr>
<td>gb</td>
<td>AC015919.1</td>
<td>AC015919 Homo sapiens chromosome 17 clone -20...</td>
</tr>
</tbody>
</table>

>gb|AF276757.1|AF276757 Homo sapiens chromosome 8 clone RP3-335L1, *** SEQUENCING IN PROGRESS ***, 9 unordered pieces Length = 157523
Score = 747 bits (377), Expect = 0.0
Identities = 377/377 (100%)
Strand = Plus / Plus

Query: 24 ctgcagcctcaaaagccagctagaaatagacatacaaatggtccaacaactgcactgaaa
Sbjct: 16240 ctgcagcctcaaaagccagctagaaatagacatacaaatggtccaacaactgcactgaaa 16299

Query: 84 gctttctattttttcggcaaatagtgtcacgtgaaaatctcctaatccctagatgtttcc
Sbjct: 16300 gctttctattttttcggcaaatagtgtcacgtgaaaatctcctaatccctagatgtttcc 16359

Query: 144 tgtttctagatcagaaaatctcaaatctcaaatctttccttcactaccaccatcagaggc
Sbjct: 16360 tgtttctagatcagaaaatctcaaatctcaaatctttccttcactaccaccatcagaggc 16419

Query: 204 gtgacagcttgccacggtctgaatccgtgcttaaatggaagagagtcccttctttctgct
Sbjct: 16420 gtgacagcttgccacggtctgaatccgtgcttaaatggaagagagtcccttctttctgct 16479

Query: 264
tctcccctgttcctcatcagagtgttggtgtaaagctggtgaaatgccatttcacaaaat
Sbjct: 16480 tctcccctgttcctcatcagagtgttggtgtaaagctggtgaaatgccatttcacaaaat 16539

Query: 324 tgtgtttttgttcatgcaagatgactatcttcccaaagctattttgtggttctggaaaag
Sbjct: 16540 tgtgtttttgttcatgcaagatgactatcttcccaaagctattttgtggttctggaaaag 16599

Query: 384 gaagatatgagacagaa
Sbjct: 16600 gaagatatgagacagaa 16616

>gb|AC021678.5|AC021678 Homo sapiens chromosome 8 clone RP11-380I10
map 8, WORKING DRAFT SEQUENCE, 28 unordered pieces Length = 184626
Score = 728 bits (367), Expect = 0.0
Identities = 373/375 (99%)
Strand = Plus / Plus

>gb|AC015919.1|AC015919 Homo sapiens chromosome 17 clone -2047D24 map 17, LOW-PASS SEQUENCE SAMPLING Length = 152113
Score = 61.9 bits (31), Expect = 4e-07
Identities = 31/31 (100%)
Strand = Plus / Minus
4.3 CHARACTERIZATION OF THE NOVEL (CA)n REPEAT

This novel repeat was further characterized to assess its suitability as a marker for genetic studies. After PCR amplification and denaturing polyacrylamide gel electrophoresis fifteen alleles were detected in eight hundred and seventy-four chromosomes of unrelated European Caucasian individuals. Each allele produced distinct bands, which could be accurately genotyped.

The following equation was used to calculate the heterozygosity of this repeat:

\[ H = 1 - \sum p_i^2 \]

where \( H \) is the probability of an individual being heterozygous and \( p \) the population frequency of the \( 'i' \) alleles. The observed heterozygosity frequency for this marker was calculated to be 0.898. The size and the frequency of the alleles are summarised on table 4.1. The repeat was found to be codominantly inherited in two three-generation families. All the information was submitted to The Genome Database (GDB, http://www.gdb.org) and a D number was obtained for this repeat D8S2611.

This highly polymorphic microsatellite marker was used in a case-control allelic association study in order to assess the candidacy of the human PNOC gene for susceptibility to schizophrenia.
Table 4.1 Size in base pairs and frequency of the alleles of the D8S2611 microsatellite marker near the human prepronociceptin locus.

<table>
<thead>
<tr>
<th>Allele</th>
<th>Size (bp)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>233</td>
<td>0.0046</td>
</tr>
<tr>
<td>2</td>
<td>231</td>
<td>0.0183</td>
</tr>
<tr>
<td>3</td>
<td>229</td>
<td>0.0766</td>
</tr>
<tr>
<td>4</td>
<td>227</td>
<td>0.1247</td>
</tr>
<tr>
<td>5</td>
<td>225</td>
<td>0.1133</td>
</tr>
<tr>
<td>6</td>
<td>223</td>
<td>0.1510</td>
</tr>
<tr>
<td>7</td>
<td>221</td>
<td>0.0870</td>
</tr>
<tr>
<td>8</td>
<td>219</td>
<td>0.0766</td>
</tr>
<tr>
<td>9</td>
<td>217</td>
<td>0.1236</td>
</tr>
<tr>
<td>10</td>
<td>215</td>
<td>0.0229</td>
</tr>
<tr>
<td>11</td>
<td>213</td>
<td>0.0206</td>
</tr>
<tr>
<td>12</td>
<td>211</td>
<td>0.0492</td>
</tr>
<tr>
<td>13</td>
<td>209</td>
<td>0.1053</td>
</tr>
<tr>
<td>14</td>
<td>207</td>
<td>0.0011</td>
</tr>
<tr>
<td>15</td>
<td>205</td>
<td>0.0252</td>
</tr>
</tbody>
</table>
4.4 CASE-CONTROL ALLELIC ASSOCIATION STUDY OF THE PNOC GENE

A case-control allelic association study was performed with marker D8S2611 near the PNOC gene. Genotypes were obtained for 129 schizophrenia patients and 308 ethnically matched normal controls of Irish, Welsh, Scottish or English ancestry (the London (UK) case-control sample) (see section 2.2.2).

Allele counts and frequencies for cases and controls are shown in Table 4.2. The CLUMP program was used to assess the evidence of allelic association between the marker and the disease locus. No significant statistical difference in allele frequency was found for marker D8S2611 between schizophrenia patients and control individuals (T1: \( \chi^2 \)=15.3929, \( p=0.3540 \)).

4.5 CASE-CONTROL ALLELIC ASSOCIATION STUDY OF THE CHRNA2 GENE

For the CHRNA2 gene, subjects were genotyped for microsatellite markers D8S131 (cos132C2) (GDB Amplimer: 215525) and D8S131P (GDB Amplimer: 215528) by PCR. Amplification of D8S131 (cos132C2) marker was achieved using primers 5' AGG GGA TCA GGG CTC AGC A 3' and 5' ATA GGC TGG AGA GTC ACA GG 3'. The amplification conditions were 95°C for 5min, then 30 cycles of 95°C for 40s, 60°C for 40s, and 72°C for 40s, with a final extension step of 7min at 72°C. D8S131P was amplified with oligonucleotide primers 5' GCC CCC CAA GTC ACA GG 3'. The cycling
conditions were 95°C for 5min, then 30 cycles of 95° C for 1min, 58°C for 30s, and 72° C for 2min, with a final extension step of 7min at 72°C. The PCR reactions were set up as described in chapter 2 (section 2.3.4.1).

A case-control allelic association study was also performed with microsatellite markers D8S131 and D8S131P in order to assess the candidacy of CHRNA2 gene in schizophrenia. The London (UK) case-control association sample was used (see section 2.2.2).

Allele counts and frequencies for cases and controls are shown in Table 4.3 for D8S131 and Table 4.4 for D8S131P. The CLUMP program was used to assess the evidence of allelic association between the marker loci and the disease locus. None of the markers showed statistically significant evidence of allelic association with the disease (T1: $\chi^2 = 8.27, p = 0.542$ for D8S131 and T1: $\chi^2 = 13.59, p = 0.404$ for D8S131P).

Pair-wise linkage disequilibrium was calculated between D8S131 and D8S131P using the EH program. Both markers are upstream of exon 2 of the CHRNA2 gene and are themselves 9 to 20 kb apart. The two markers are in very strong linkage disequilibrium with each other as expected due to their close proximity.
Table 4.2 Allele counts and frequencies (in parentheses) at marker D8S2611 near the PNOC gene

<table>
<thead>
<tr>
<th>Alleles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>2 (0.008)</td>
<td>2 (0.008)</td>
<td>17 (0.066)</td>
<td>24 (0.093)</td>
<td>28 (0.109)</td>
<td>40 (0.155)</td>
<td>31 (0.120)</td>
<td>18 (0.07)</td>
<td>37 (0.143)</td>
<td>5 (0.019)</td>
<td>6 (0.023)</td>
</tr>
<tr>
<td>Control individuals</td>
<td>2 (0.003)</td>
<td>14 (0.023)</td>
<td>50 (0.081)</td>
<td>85 (0.138)</td>
<td>71 (0.115)</td>
<td>92 (0.149)</td>
<td>45 (0.073)</td>
<td>49 (0.079)</td>
<td>71 (0.115)</td>
<td>15 (0.024)</td>
<td>12 (0.019)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>30</td>
<td>0</td>
<td>4</td>
<td>258</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>62</td>
<td>1</td>
<td>18</td>
<td>616</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.3 Allele counts and frequencies (in parentheses) at marker D8S131

<table>
<thead>
<tr>
<th>Alleles*</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>a</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia patients</td>
<td>1 (0.004)</td>
<td>0 (0.000)</td>
<td>6 (0.027)</td>
<td>178 (0.795)</td>
<td>1 (0.004)</td>
<td>2 (0.009)</td>
<td>36 (0.161)</td>
<td>0 (0.000)</td>
<td>0 (0.000)</td>
<td>0 (0.000)</td>
</tr>
<tr>
<td>Control individuals</td>
<td>0 (0.000)</td>
<td>1 (0.002)</td>
<td>20 (0.032)</td>
<td>489 (0.774)</td>
<td>0 (0.000)</td>
<td>5 (0.008)</td>
<td>112 (0.177)</td>
<td>3 (0.005)</td>
<td>1 (0.002)</td>
<td>1 (0.002)</td>
</tr>
<tr>
<td><strong>Row Total</strong></td>
<td>224</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*alleles according to GDB

Table 4.4 Allele counts and frequencies (in parentheses) at marker D8S131P

<table>
<thead>
<tr>
<th>Alleles**</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia patients</td>
<td>5 (0.023)</td>
<td>11 (0.050)</td>
<td>7 (0.032)</td>
<td>24 (0.109)</td>
<td>31 (0.141)</td>
<td>95 (0.432)</td>
<td>14 (0.064)</td>
<td>9 (0.041)</td>
<td>0 (0.000)</td>
<td>1 (0.0045)</td>
<td>0 (0.000)</td>
<td>2 (0.009)</td>
<td>19 (0.086)</td>
<td>2 (0.009)</td>
</tr>
<tr>
<td>Control individuals</td>
<td>18 (0.029)</td>
<td>29 (0.046)</td>
<td>19 (0.030)</td>
<td>46 (0.073)</td>
<td>85 (0.136)</td>
<td>292 (0.466)</td>
<td>45 (0.072)</td>
<td>25 (0.040)</td>
<td>1 (0.002)</td>
<td>1 (0.002)</td>
<td>2 (0.003)</td>
<td>1 (0.002)</td>
<td>62 (0.0999)</td>
<td>0 (0.000)</td>
</tr>
<tr>
<td><strong>Row Total</strong></td>
<td>220</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**alleles according to S.Wood et al.
4.6 DISCUSSION

Association studies are gaining popularity because of their power to detect genes of moderate or small effect such as those that could be increasing liability to a proportion of cases of schizophrenia (Long & Langley, 1999).

When choosing a candidate gene for a complex disorder such as schizophrenia that has an unknown pathophysiology, there is a low prior probability of it being involved (Owen et al., 1997), (Crowe, 1993). One way of increasing the prior probability of finding association is by examining the candidacy of genes that map in an interval that has been implicated by previous linkage studies (Berrettini, 1997). In this study, all the candidate genes map to chromosome 8p21, a region implicated by linkage to schizophrenia by several investigators.

As the DNA polymorphisms studied are anonymous polymorphisms, the approach taken was the detection of linkage disequilibrium with the actual disease causing variant(s). So far, studies using RFLPs and repeat markers suggest that allelic association can be detectable up to distances of over 2-3 cM but are more likely to be detected over a distance of 500kb (Jorde et al., 1994), (Petrukhin et al., 1993), (Peterson et al., 1995). The distance of the polymorphisms from the candidate genes in this study does not exceed the limit for detectable LD mentioned above. They are therefore capable of detecting strong LD with a polymorphism at the disease-causing site or the disease-causing variant itself.

However, no significant differences in the allele frequencies of any of the markers studied near PNOC and CHRNA2 were observed in our population of schizophrenic patients in comparison with ethnically matched control individuals. The failure of demonstrating association is unlikely to be due to population
stratification, since both groups of individuals were carefully matched for ethnicity and drawn from as ethnically a homogenous population as possible.

Especially for the CHRNA2 gene, the result appears to be genuine as strong linkage disequilibrium was obtained between D8S131 and D8S131P but neither of the markers showed statistically significant linkage disequilibrium with the disease.

Finally, the absence of association can be due to the existence of multiple variants in the candidate gene that predispose to the disorder or to functional mutations that have an opposite effect on the phenotype. This will render detection of association very difficult due to power considerations.

Examining the candidacy of these two genes was not pursued any further as fine mapping of the 8p21-22 region undertaken simultaneously with the candidate gene studies indicated a very interesting allelic association between schizophrenia and marker D8S261 (see chapter 5).

As more genes are being identified, more attention will focus on candidate gene association studies. In future studies, the overall sequence variability of candidate genes will be determined and their role in the aetiology of a complex disorder investigated in a number of association studies.
CHAPTER 5

FINE MAPPING OF THE CHROMOSOME 8p21-22 REGION

BY ALLELIC ASSOCIATION STUDIES
5.1 INTRODUCTION

Linkage analysis, as a first step in positional cloning of a susceptibility locus for a complex disease, is not expected to narrow down a candidate region to less than 1 Mb (Kruglyak & Lander, 1995). In contrast, association analysis or linkage disequilibrium (LD) mapping can be expected to refine the location of a disease, since it exploits historic recombination events by measuring the LD between disease mutation and marker (Jorde, 1995). Generally, LD becomes stronger as physical distance and recombination rate between two loci decrease. However, there are a number of factors that can affect the assumed linear relationship between physical distance and LD, such as the time and number of generations that have elapsed since the mutation was introduced in the population, population admixture, mutation, and selection. Another factor that may disrupt this relationship is the varying recombination frequencies in different chromosomal regions also known as recombination “hotspots”, which are distributed non-uniformly in the genome.

In the previous chapters, the linkage studies on chromosome 8p were reported. The candidate region implicated by these studies spans an area of approximately 18 cM, which is too broad for positional cloning efforts.

In this study, we have employed LD mapping in an attempt to narrow down the region of chromosome 8p21-22, which is involved in schizophrenia. Marker-to-marker linkage disequilibrium analysis was also performed in order to give us an insight in the pattern of LD in this region and a meaningful interpretation of the association studies.

The results presented here provide further evidence for a susceptibility locus for schizophrenia on chromosome 8p21-22 and demonstrate the great potential of
linkage disequilibrium mapping in refining susceptibility regions in complex disorders using microsatellite markers in relatively heterogeneous populations.

5.2 A PRELIMINARY ALLELIC ASSOCIATION STUDY OF THE 8p21-22 REGION IMPLICATED IN LINKAGE STUDIES OF SCHIZOPHRENIA

A case-control allelic association study was performed with markers D8S261, D8S136, D8S1771. These markers were chosen because they are at positions where the lod scores have peaked in several linkage analyses studies. Genotyping was performed for 134 schizophrenia patients and 316 ethnically matched normal controls of Irish, Welsh, Scottish or English ancestry (the London (UK) case-control sample) (see section 2.2.2). The sequence of the primers and more information about the markers used is provided on table 5.1.

Ten alleles were found for marker D8S261, which was an increase in the number of alleles listed in the Genome Database (GDB, http://www.gdb.org), where eight alleles were reported. Alleles one to seven correspond to the alleles listed in GDB with sizes of the reaction products from 0.128-0.140 kb. Allele eight corresponds to the allele with size of the reaction product 0.142 kb that was missed by GDB. Allele nine corresponds to what is allele eight in GDB with a size of reaction product of 0.144 kb. Finally, allele ten has a size of reaction product of 0.126 kb, it is a rare allele and it is not reported in GDB.
<table>
<thead>
<tr>
<th>MARKER</th>
<th>PRIMERS</th>
<th>TYPE OF REPEAT</th>
<th>No. ALLELES</th>
<th>ALLELE SIZES (bp)</th>
<th>HET.</th>
<th>POSITION¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFM333th1</td>
<td>5'ATC TGT CAG TGA GTG CCC3' 5'ATT ACT TCG GCT CTG CTA TGT AT3'</td>
<td>(CA)&lt;sub&gt;n&lt;/sub&gt;</td>
<td>7</td>
<td>124-112</td>
<td>0.4925</td>
<td>37</td>
</tr>
<tr>
<td>D8S261</td>
<td>5'TGC CAC TGT CTT GAA AAT CC3' 5'TAT GGC CCA GCA ATG TGT AT3'</td>
<td>(CA)&lt;sub&gt;n&lt;/sub&gt;</td>
<td>8</td>
<td>144-128</td>
<td>0.7760</td>
<td>37</td>
</tr>
<tr>
<td>D8S1145</td>
<td>5'TGCTAACTGGCAGGGTCAC3' 5'CAATCCACAGTAATCTATAACTTCAC3'</td>
<td>(TATC)&lt;sub&gt;n&lt;/sub&gt;</td>
<td>7</td>
<td>289-261</td>
<td>0.6250</td>
<td>37</td>
</tr>
<tr>
<td>NAT1 (1095)</td>
<td>5'ACT CTG AGT GAG GTA GAA ATA3' 5'ACA GGC CAT CTT TAG AA3'</td>
<td>RFLP</td>
<td>2</td>
<td>164/124 164/144</td>
<td>0.3606</td>
<td>38</td>
</tr>
<tr>
<td>D8S136</td>
<td>5'GCCAAAAGGAAGGATAAAA3' 5'CTGTTTCCACACCGAAGC3'</td>
<td>(CA)&lt;sub&gt;n&lt;/sub&gt;</td>
<td>9</td>
<td>89-73</td>
<td>0.8810</td>
<td>43</td>
</tr>
<tr>
<td>D8S1771</td>
<td>5'TTT ACA AGA ACC ACC TGC C3' 5'GAT ATA AAA CAT GAC TTT GCT ACC C3'</td>
<td>(CA)&lt;sub&gt;n&lt;/sub&gt;</td>
<td>8</td>
<td>238-226</td>
<td>0.7587</td>
<td>50</td>
</tr>
</tbody>
</table>

1: in Kosambi centimorgans, derived from the Marshfield Map positions
Figure 5.1 Part of the Whitehead contig WC8.1 map and the Marshfield map of the 8p21-22 region. The microsatellite markers AFM333th1 and D8S1145 are placed distal and proximal respectively to D8S261 on the Whitehead contig WC8.1 map. The relative positions of microsatellite markers D8S261, D8S1145, D8S136 and D8S1771 can also be seen on the Marshfield, sex-averaged linkage map. Distances are in Kosambi centimorgans. Obtained from the Genome Database (GDB, http://www.gdb.org)
In this study, fourteen alleles were found for marker D8S136, which is an increase in the number of alleles listed in GDB by five alleles. Alleles from one to nine correspond to the ones reported in GDB with a size of the reaction product from 0.089 to 0.073 kb. The five newly discovered fragments are alleles ten to fourteen and have sizes of reaction products from 0.071 to 0.063.

Allele counts and frequencies for cases and controls are shown in Table 5.3 for D8S261, Table 5.4 for D8S136 and Table 5.5 for D8S1771.

The CLUMP program was used to assess the evidence of allelic association between the marker loci and the disease locus (see section 2.4.3). For each marker both the $\chi^2$ and $p$-value are shown for each of the four statistics that are produced by CLUMP (T1, T2, T3 and T4). The $p$-value shown is an empirical $p$-value obtained after performing Monte Carlo evaluations of significance. In this way, correction for multiple alleles is not necessary. The results obtained from the CLUMP output are presented on Table 5.8. Evidence for allelic association was obtained between D8S261 and schizophrenia. The T1 statistic (a straightforward Pearson’s $\chi^2$ statistic of the ‘raw’ contingency table) produced an empirical $p$-value of 0.055 while the T3 statistic (the largest of the $\chi^2$ statistics of 2x2 tables each of which compares one allele against the rest grouped together) produced an empirical $p$-value of 0.01. This T3 result was due to an excess of allele 8 of D8S261 in the cases ($p=0.0016$, uncorrected for multiple alleles). No difference in allele frequency was found for D8S136 (T1: $\chi^2=14.4263$, $p=0.350$) and D8S1771 (T1: $\chi^2=15.9475$, $p=0.186$) between schizophrenia patients and control individuals.

After this allelic association result between D8S261 and schizophrenia, the region around D8S261 was searched for more polymorphic markers. At that time, a tetranucleotide repeat marker, D8S1145, was found to be in close proximity to
D8S261. According to the Marshfield chromosome 8 sex averaged linkage map, the two markers had the same coordinates while the Whitehead contig WC8.1 map placed D8S1145 proximal to D8S261 (Figure 5.1).

In addition, a microsatellite marker, AFM333th1, is located distal to D8S261 according to the Whitehead contig WC8.1 map, and the two markers have the same coordinates on the linkage map (Figure 5.1).

Studies by Bookstein and colleagues to establish a YAC contig across the 8p22 region, where D8S261 resides, suggested that the CEPH YAC clone 753_e_4 contains DNA markers D8S261 and D8S21 (Bookstein et al., 1994). Further work on this region revealed that the D8S21 marker lies in the arylamine N-acetyltransferase 2 (NAT2) gene while the arylamine N-acetyltransferase 1 (NAT1) gene lies in between the NAT2 gene and D8S261 (Matas et al., 1997).

NAT1 and NAT2 are two genes encoding functional NAT in humans on 8p22 within 400kb from each other. Their corresponding proteins are enzymes with distinct but overlapping substrate specificities, which are both involved in the metabolism and detoxification of drugs and other foreign chemicals. More specifically, they catalyse the acetyl CoA dependent N-acetylation of aromatic amines resulting in their detoxification. Their biological function did not suggest any direct relevance to the pathophysiology of schizophrenia. Moreover, several studies have implicated these genes in the development of bladder cancer as well as other types of cancer (Bell et al., 1995a), (Bouchardy et al., 1998).

The NAT1 gene has a single nucleotide polymorphism at position 1095. This biallelic polymorphism is a C->A substitution in the 3' UTR of the gene and was detected as a PCR formatted MboII RFLP (see section 2.3.4.4). The maximum distance between this polymorphism and D8S261 is ~1 cM (Matas et al., 1997). This
RFLP was used as a marker to test for allelic association with schizophrenia due to its close proximity to D8S261.

The sequence of the primers used for AFM333ht1, D8S1145 and the MboII RFLP in the NAT1 gene as well as more information about these markers is provided on Table 5.1. Genotyping was performed on the same case-control sample as before of 134 schizophrenia patients and 316 ethnically matched normal controls of Irish, Welsh, Scottish or English ancestry (the London (UK) case-control sample).

There is no information on the allele set of marker AFM333th1 reported in GDB. In this study, seven alleles were found, with allele one to allele seven corresponding to a size of the reaction product from 0.124 to 0.112 kb.

Information on the MboII RFLP genotyping is provided in the materials and methods (see section 2.3.4.4).

Allele counts and frequencies for cases and controls are shown in Table 5.1 for AFM333th1, Table 5.6 for D8S1145 and Table 5.5 for the MboII RFLP.

The results from the CLUMP program did not provide evidence for association between AFM333th1 (T1: \( \chi^2 = 5.3018, p = 0.509 \)), D8S1145 (T1: \( \chi^2 = 6.9705, p = 0.569 \)) or MboII RFLP in the NAT1 gene (T1: \( \chi^2 = 0.2954, p = 0.610 \)) and the disease locus (Table 5.8).
Table 5.2 Allele counts and frequencies (in parentheses) at marker AFM333th1 from the London (UK) case-control sample

<table>
<thead>
<tr>
<th>Alleles*</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>8 (0.031)</td>
<td>5 (0.019)</td>
<td>13 (0.05)</td>
<td>50 (0.192)</td>
<td>173 (0.665)</td>
<td>11 (0.042)</td>
<td>0 (0.000)</td>
<td>260</td>
</tr>
<tr>
<td>Control individuals</td>
<td>10 (0.016)</td>
<td>11 (0.017)</td>
<td>20 (0.032)</td>
<td>120 (0.190)</td>
<td>435 (0.690)</td>
<td>31 (0.049)</td>
<td>3 (0.005)</td>
<td>630</td>
</tr>
</tbody>
</table>

*alleles according to E.Blaveri

Table 5.3 Allele counts and frequencies (in parentheses) at marker D8S261 from the London (UK) case-control sample

<table>
<thead>
<tr>
<th>Alleles*</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>13 (0.053)</td>
<td>76 (0.309)</td>
<td>6 (0.024)</td>
<td>14 (0.057)</td>
<td>75 (0.305)</td>
<td>30 (0.122)</td>
<td>18 (0.073)</td>
<td>12 (0.049)</td>
<td>0 (0.000)</td>
<td>2 (0.008)</td>
<td>246</td>
</tr>
<tr>
<td>Control individuals</td>
<td>28 (0.045)</td>
<td>196 (0.317)</td>
<td>5 (0.008)</td>
<td>36 (0.058)</td>
<td>207 (0.335)</td>
<td>75 (0.121)</td>
<td>49 (0.079)</td>
<td>8 (0.013)</td>
<td>4 (0.006)</td>
<td>10 (0.016)</td>
<td>618</td>
</tr>
</tbody>
</table>

*alleles according to GDB
Table 5.4 Allele counts and frequencies (in parentheses) at marker D8S136 from the London (UK) case-control sample

<table>
<thead>
<tr>
<th>Alleles*</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>1</td>
<td>3</td>
<td>11</td>
<td>7</td>
<td>58</td>
<td>21</td>
<td>7</td>
<td>41</td>
<td>47</td>
<td>54</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>252</td>
</tr>
<tr>
<td>Control individuals</td>
<td>5</td>
<td>2</td>
<td>44</td>
<td>12</td>
<td>139</td>
<td>46</td>
<td>20</td>
<td>127</td>
<td>110</td>
<td>106</td>
<td>10</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>632</td>
</tr>
</tbody>
</table>

*alleles according to GDB

Table 5.5 Allele counts and frequencies (in parentheses) at marker D8S1771 from the London (UK) case-control sample

<table>
<thead>
<tr>
<th>Alleles*</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>101</td>
<td>4</td>
<td>71</td>
<td>19</td>
<td>22</td>
<td>17</td>
<td>8</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>262</td>
</tr>
<tr>
<td>Control individuals</td>
<td>208</td>
<td>12</td>
<td>203</td>
<td>59</td>
<td>65</td>
<td>17</td>
<td>15</td>
<td>20</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>24</td>
<td>632</td>
</tr>
</tbody>
</table>

*alleles according to GDB
Table 5.6 Allele counts and frequencies (in parentheses) at marker D8S1145 from the London (UK) case-control sample

<table>
<thead>
<tr>
<th>Alleles*</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>3 (0.012)</td>
<td>28 (0.109)</td>
<td>23 (0.089)</td>
<td>119 (0.461)</td>
<td>57 (0.221)</td>
<td>18 (0.070)</td>
<td>1 (0.004)</td>
<td>1 (0.004)</td>
<td>8 (0.031)</td>
<td>258</td>
</tr>
<tr>
<td>Control individuals</td>
<td>14 (0.022)</td>
<td>70 (0.112)</td>
<td>78 (0.125)</td>
<td>283 (0.452)</td>
<td>128 (0.204)</td>
<td>32 (0.051)</td>
<td>2 (0.003)</td>
<td>0 (0.000)</td>
<td>19 (0.030)</td>
<td>626</td>
</tr>
</tbody>
</table>

*alleles according to GDB

Table 5.7 Allele counts and frequencies (in parentheses) for the MboII RFLP at the NAT1 gene from the London (UK) case-control sample

<table>
<thead>
<tr>
<th>Alleles*</th>
<th>1</th>
<th>2</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>194 (0.752)</td>
<td>64 (0.248)</td>
<td>258</td>
</tr>
<tr>
<td>Control individuals</td>
<td>486 (0.769)</td>
<td>146 (0.231)</td>
<td>632</td>
</tr>
</tbody>
</table>

*alleles according to E.Blaveri
Table 5.8 CLUMP test results for preliminary association studies between schizophrenia and alleles at polymorphic loci on chromosome 8p21-22 in the London (UK) case-control sample

<table>
<thead>
<tr>
<th>Marker Locus (tel→cen)</th>
<th>T1 $\chi^2$</th>
<th>p-value</th>
<th>T2 $\chi^2$</th>
<th>p-value</th>
<th>T3 $\chi^2$</th>
<th>p-value</th>
<th>T4 $\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFM3337h1</td>
<td>5.3018</td>
<td>0.509</td>
<td>4.1088</td>
<td>0.544</td>
<td>2.0610</td>
<td>0.512</td>
<td>3.7441</td>
<td>0.426</td>
</tr>
<tr>
<td>D8S261</td>
<td>16.6737</td>
<td><strong>0.055</strong></td>
<td>10.6062</td>
<td>0.150</td>
<td>9.9929</td>
<td><strong>0.010</strong></td>
<td>13.8258</td>
<td><strong>0.017</strong></td>
</tr>
<tr>
<td>NAT1 (1095)</td>
<td>0.2954</td>
<td>0.610</td>
<td>0.2954</td>
<td>0.630</td>
<td>0.2954</td>
<td>0.610</td>
<td>0.2954</td>
<td>0.610</td>
</tr>
<tr>
<td>D8S1145</td>
<td>6.9705</td>
<td>0.569</td>
<td>3.7177</td>
<td>0.710</td>
<td>2.2692</td>
<td>0.561</td>
<td>3.3699</td>
<td>0.641</td>
</tr>
<tr>
<td>D8S136</td>
<td>14.4263</td>
<td>0.360</td>
<td>6.5643</td>
<td>0.630</td>
<td>2.6352</td>
<td>0.640</td>
<td>7.9002</td>
<td>0.380</td>
</tr>
<tr>
<td>D8S1771</td>
<td>15.9475</td>
<td>0.186</td>
<td>14.5552</td>
<td>0.068</td>
<td>7.305</td>
<td>0.036</td>
<td>9.5211</td>
<td>0.132</td>
</tr>
</tbody>
</table>

Where,

T1: a straightforward Pearson’s $\chi^2$ statistic of the ‘raw’ contingency table

T2: the $\chi^2$ statistic of a table with rare alleles grouped together to prevent small expected cell counts

T3: the largest of the $\chi^2$ statistics of 2x2 tables each of which compares one allele against the rest grouped together

T4: the largest of the $\chi^2$ statistics of all possible 2x2 tables comparing any combination of alleles against the rest
5.3 IDENTIFICATION OF NOVEL POLYMORPHISMS IN THE 8p21.3-22 REGION

Chromosome 8p is implicated as the site of one or more genes associated with the development of carcinomas in a large spectrum of tissues (Pineau et al., 1999). In particular, the 8p21.3-22 region is commonly deleted in hepatocellular, colorectal and non-small cell lung carcinomas and therefore is considered as a candidate region for a tumour suppressor gene (Chinen et al., 1996). This led to the sequencing of approximately 1Mb of sequence within this region (contig NT_000501, 1,210,381 bp) by the Japanese Foundation for Cancer Research (JST sequencing team, http://www-alis.tokyo.jst.go.jp). Contig NT_000501 is subdivided in 10 clones, AB020858 to AB020868, (Table 5.9) and contains D8S261. The publication of this sequence was very fortuitous for the thesis in aiding the identification of more polymorphic markers in an exact region. The sequence from clones AB020863, AB020864, AB020865, AB020866, AB020867 and AB020868 were searched for di-, tri-, tetra- repeat motifs. Novel polymorphisms were identified which were close to the marker D8S261. Primers flanking the repeat region were designed using the Primer 3 software program (Whitehead Institute/MIT Center for Genome Research) and were used to amplify the repeat regions by PCR. Several different repeat sequences around the area were assessed for their suitability as highly polymorphic markers. Seven (CA)n repeats were observed to be polymorphic. Allele sizes were determined by running next to DNA size standards (Microzone Ltd.). Heterozygosity values were calculated by the formula $H=1-\Sigma p_i^2$, where $p_i$ is the population frequency of the $i$th allele. D numbers for the novel markers were obtained from The Genome Database (GDB, http://www.gdb.org). The relative positions of the markers with respect to the clones
of the NT_000501 contig are shown in Table 5.9 and Figure 5.2 while primer sequences and alleles are shown in Table 5.10.
Table 5.9 The clones that constitute the NT_000501 contig (121,0381 bases) of the human 8p21.3-p22 sequence. The position of seven novel microsatellite markers which developed from this sequence together with two already known markers (D8S261 and AFM333th1) are shown as well as their respective clones and the intermarker distances between the nine markers.

<table>
<thead>
<tr>
<th>CLONES</th>
<th>SIZE (bp)</th>
<th>MARKERS</th>
<th>POSITION IN THE CLONE (bp)</th>
<th>INTERMARKER DISTANCES (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB020858.1</td>
<td>100,000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AB020859.1</td>
<td>100,000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AB020860.1</td>
<td>100,000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AB020861.1</td>
<td>100,000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AB020862.1</td>
<td>100,000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AB020863.1</td>
<td>156,909</td>
<td>D8S2618</td>
<td>14,588</td>
<td>178,562</td>
</tr>
<tr>
<td>AB020864.1</td>
<td>100,000</td>
<td>D8S2613</td>
<td>36,241</td>
<td>90,549</td>
</tr>
<tr>
<td>AB020865.1</td>
<td>100,000</td>
<td>D8S2614</td>
<td>26,790</td>
<td>73,402</td>
</tr>
<tr>
<td>AB020866.1</td>
<td>100,000</td>
<td>D8S2615</td>
<td>192</td>
<td>86,911</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D8S2616</td>
<td>87,103</td>
<td>21,981</td>
</tr>
<tr>
<td>AB020867.1</td>
<td>100,000</td>
<td>D8S261</td>
<td>9,084</td>
<td>6,150</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AFM333th1</td>
<td>15,234</td>
<td>62,533</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D8S2617</td>
<td>77,767</td>
<td></td>
</tr>
<tr>
<td>AB020868.1</td>
<td>153,472</td>
<td>D8S2612</td>
<td>6,167</td>
<td>28,400</td>
</tr>
</tbody>
</table>
Figure 5.2 Part of the NT_000501 contig (1,210,381 bp) on chromosome 8p21.3-22

The NT_000501 contig is subdivided into 10 cosmid clones (AB020858.1 to AB020868). Novel polymorphisms were identified around D8S261 from clones AB020863.1, AB020864.1, AB020865.1, AB020866.1, AB020867.1 and AB020868.1. Their position in respect to the contig clones and the already known microsatellite markers D8S261 and AFM333th1 can be seen.
Table 5.10 Primer sequences and allele set of the novel microsatellite repeats identified in the NT_000501 contig on chromosome 8p21.3-22

D8S2612

Primers: 5' AAT TCC CCA AAC AAA ACA ACA 3'
          5' AGG CTA TCC TTT CCT CAG CA 3'

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Length(bp)</th>
<th>(CA)n</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>161</td>
<td>18</td>
<td>0.0112</td>
</tr>
<tr>
<td>2</td>
<td>159</td>
<td>17</td>
<td>0.0594</td>
</tr>
<tr>
<td>3</td>
<td>157</td>
<td>16</td>
<td>0.1614</td>
</tr>
<tr>
<td>4</td>
<td>155</td>
<td>15</td>
<td>0.2578</td>
</tr>
<tr>
<td>5</td>
<td>153</td>
<td>14</td>
<td>0.4283</td>
</tr>
<tr>
<td>6</td>
<td>151</td>
<td>13</td>
<td>0.0751</td>
</tr>
<tr>
<td>7</td>
<td>149</td>
<td>12</td>
<td>0.0000</td>
</tr>
<tr>
<td>8</td>
<td>147</td>
<td>11</td>
<td>0.0000</td>
</tr>
<tr>
<td>9</td>
<td>145</td>
<td>10</td>
<td>0.0067</td>
</tr>
</tbody>
</table>

Heterozygosity: 0.7147
No. of chromosomes: 892

D8S2617

Primers: 5' ATG TTC AGC CAC CAT CGT CT 3'
          5' CAG TGT CGC TGG AAA GTT GA 3'

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Length(bp)</th>
<th>(CA)n</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>203</td>
<td>25</td>
<td>0.0146</td>
</tr>
<tr>
<td>2</td>
<td>201</td>
<td>24</td>
<td>0.0822</td>
</tr>
<tr>
<td>3</td>
<td>199</td>
<td>23</td>
<td>0.1216</td>
</tr>
<tr>
<td>4</td>
<td>197</td>
<td>22</td>
<td>0.0619</td>
</tr>
<tr>
<td>5</td>
<td>195</td>
<td>21</td>
<td>0.2759</td>
</tr>
<tr>
<td>6</td>
<td>193</td>
<td>20</td>
<td>0.4358</td>
</tr>
<tr>
<td>7</td>
<td>191</td>
<td>19</td>
<td>0.0068</td>
</tr>
<tr>
<td>8</td>
<td>189</td>
<td>18</td>
<td>0.0000</td>
</tr>
<tr>
<td>9</td>
<td>187</td>
<td>17</td>
<td>0.0000</td>
</tr>
<tr>
<td>10</td>
<td>185</td>
<td>16</td>
<td>0.0011</td>
</tr>
</tbody>
</table>

Heterozygosity: 0.7083
No. of chromosomes: 888
Table 5.10 (continues)

**D8S2616**

Primers: 5'TCC CGA AGT GCT AGG ATT ACA3'
5'GCT CAG CAG GAA GAG GAA TG3'

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Length(bp)</th>
<th>(CA)n</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>213</td>
<td>24</td>
<td>0.0011</td>
</tr>
<tr>
<td>4</td>
<td>211</td>
<td>23</td>
<td>0.0113</td>
</tr>
<tr>
<td>5</td>
<td>209</td>
<td>22</td>
<td>0.0317</td>
</tr>
<tr>
<td>6</td>
<td>207</td>
<td>21</td>
<td>0.0588</td>
</tr>
<tr>
<td>7</td>
<td>205</td>
<td>20</td>
<td>0.1301</td>
</tr>
<tr>
<td>8</td>
<td>203</td>
<td>19</td>
<td>0.5848</td>
</tr>
<tr>
<td>9</td>
<td>201</td>
<td>18</td>
<td>0.0532</td>
</tr>
<tr>
<td>10</td>
<td>199</td>
<td>17</td>
<td>0.1199</td>
</tr>
<tr>
<td>11</td>
<td>197</td>
<td>16</td>
<td>0.0011</td>
</tr>
<tr>
<td>12</td>
<td>195</td>
<td>15</td>
<td>0.0057</td>
</tr>
<tr>
<td>13</td>
<td>193</td>
<td>14</td>
<td>0.0000</td>
</tr>
<tr>
<td>14</td>
<td>191</td>
<td>13</td>
<td>0.0011</td>
</tr>
<tr>
<td>15</td>
<td>189</td>
<td>12</td>
<td>0.0011</td>
</tr>
</tbody>
</table>

Heterozygosity: 0.6192
No. of chromosomes: 884

**D8S2615**

Primers: 5'AGA GGC CAG GCA CAA AAG TA3'
5'AAC ATT CCA GCA TCC CAA AG3'

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Length(bp)</th>
<th>(CA)n</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>205</td>
<td>24</td>
<td>0.8807</td>
</tr>
<tr>
<td>2</td>
<td>203</td>
<td>23</td>
<td>0.0148</td>
</tr>
<tr>
<td>3</td>
<td>201</td>
<td>22</td>
<td>0.0875</td>
</tr>
<tr>
<td>4</td>
<td>119</td>
<td>21</td>
<td>0.0000</td>
</tr>
<tr>
<td>5</td>
<td>117</td>
<td>19</td>
<td>0.0000</td>
</tr>
<tr>
<td>6</td>
<td>115</td>
<td>18</td>
<td>0.0136</td>
</tr>
<tr>
<td>7</td>
<td>113</td>
<td>17</td>
<td>0.0011</td>
</tr>
<tr>
<td>8</td>
<td>111</td>
<td>16</td>
<td>0.0000</td>
</tr>
<tr>
<td>9</td>
<td>99</td>
<td>15</td>
<td>0.0000</td>
</tr>
<tr>
<td>10</td>
<td>97</td>
<td>14</td>
<td>0.0023</td>
</tr>
</tbody>
</table>

Heterozygosity: 0.2163
No. of chromosomes: 880
Table 5.10 (continues)

**D8S2614**

Primers: 5'GAC CCA CTG CCA CAC TCT TT3'
5'GGA GTG CGG CAT GAA ATT AT3'

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Length (bp)</th>
<th>(CA)n</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>194</td>
<td>22</td>
<td>0.0057</td>
</tr>
<tr>
<td>2</td>
<td>192</td>
<td>21</td>
<td>0.0588</td>
</tr>
<tr>
<td>3</td>
<td>190</td>
<td>20</td>
<td>0.4570</td>
</tr>
<tr>
<td>4</td>
<td>188</td>
<td>19</td>
<td>0.0916</td>
</tr>
<tr>
<td>5</td>
<td>186</td>
<td>18</td>
<td>0.1437</td>
</tr>
<tr>
<td>6</td>
<td>184</td>
<td>17</td>
<td>0.0045</td>
</tr>
<tr>
<td>7</td>
<td>182</td>
<td>16</td>
<td>0.2285</td>
</tr>
<tr>
<td>8</td>
<td>180</td>
<td>15</td>
<td>0.0000</td>
</tr>
<tr>
<td>9</td>
<td>178</td>
<td>14</td>
<td>0.0011</td>
</tr>
<tr>
<td>10</td>
<td>176</td>
<td>13</td>
<td>0.0079</td>
</tr>
<tr>
<td>11</td>
<td>174</td>
<td>12</td>
<td>0.0011</td>
</tr>
</tbody>
</table>

Heterozygosity: 0.7063
No. of chromosomes: 884

**D8S2613**

Primers: 5'ATA TGT ATA CAA TGT GTA TCT GTA TC3'
5'CCT TTT AGT TCC CAT TCC CAT T3'

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Length (bp)</th>
<th>(CA)n</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>133</td>
<td>16</td>
<td>0.0275</td>
</tr>
<tr>
<td>2</td>
<td>131</td>
<td>15</td>
<td>0.2002</td>
</tr>
<tr>
<td>3</td>
<td>129</td>
<td>14</td>
<td>0.1167</td>
</tr>
<tr>
<td>4</td>
<td>127</td>
<td>13</td>
<td>0.0000</td>
</tr>
<tr>
<td>5</td>
<td>125</td>
<td>12</td>
<td>0.6362</td>
</tr>
<tr>
<td>6</td>
<td>123</td>
<td>11</td>
<td>0.0000</td>
</tr>
<tr>
<td>7</td>
<td>121</td>
<td>10</td>
<td>0.0046</td>
</tr>
<tr>
<td>8</td>
<td>119</td>
<td>9</td>
<td>0.0000</td>
</tr>
<tr>
<td>9</td>
<td>117</td>
<td>8</td>
<td>0.0137</td>
</tr>
<tr>
<td>10</td>
<td>115</td>
<td>7</td>
<td>0.0011</td>
</tr>
</tbody>
</table>

Heterozygosity: 0.5406
No. of chromosomes: 874
Table 5.10 (continues)

D8S2618

Primers: 5'TGA TGC AGG AGA ATT GCT TG3'
       5'CCT ACT TGG CTG GGA TTC TG3'

<table>
<thead>
<tr>
<th>Alleles</th>
<th>Length (bp)</th>
<th>(CA)n</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>182</td>
<td>24</td>
<td>0.00457</td>
</tr>
<tr>
<td>2</td>
<td>180</td>
<td>23</td>
<td>0.00686</td>
</tr>
<tr>
<td>3</td>
<td>178</td>
<td>22</td>
<td>0.05148</td>
</tr>
<tr>
<td>4</td>
<td>176</td>
<td>21</td>
<td>0.01487</td>
</tr>
<tr>
<td>5</td>
<td>174</td>
<td>20</td>
<td>0.07208</td>
</tr>
<tr>
<td>6</td>
<td>172</td>
<td>19</td>
<td>0.04462</td>
</tr>
<tr>
<td>7</td>
<td>170</td>
<td>18</td>
<td>0.02745</td>
</tr>
<tr>
<td>8</td>
<td>168</td>
<td>17</td>
<td>0.36041</td>
</tr>
<tr>
<td>9</td>
<td>166</td>
<td>16</td>
<td>0.02288</td>
</tr>
<tr>
<td>10</td>
<td>164</td>
<td>15</td>
<td>0.00228</td>
</tr>
<tr>
<td>11</td>
<td>162</td>
<td>14</td>
<td>0.01029</td>
</tr>
<tr>
<td>12</td>
<td>160</td>
<td>13</td>
<td>0.37185</td>
</tr>
<tr>
<td>13</td>
<td>158</td>
<td>12</td>
<td>0.00000</td>
</tr>
<tr>
<td>14</td>
<td>156</td>
<td>11</td>
<td>0.00000</td>
</tr>
<tr>
<td>15</td>
<td>154</td>
<td>10</td>
<td>0.00915</td>
</tr>
<tr>
<td>16</td>
<td>152</td>
<td>9</td>
<td>0.00114</td>
</tr>
</tbody>
</table>

Heterozygosity: 0.72023
No. of chromosomes: 874
5.4 ALLELIC ASSOCIATION STUDIES IN THE LONDON (UK) CASE-CONTROL SAMPLE WITH NOVEL POLYMORPHISMS IN THE 8p21.3-22 REGION

Seven novel microsatellite markers were identified near D8S261 as described in the previous section. In addition, with the aid of the published sequence, microsatellite marker AFM333th1 could be accurately placed as being 6 kb telomeric to D8S261. These nine markers cover an area of approximately 700 kb on chromosome 8p21.3-22 and were used to perform case-control allelic association studies in this region. The seven novel polymorphisms were genotyped in the London (UK) sample of 134 schizophrenia patients and 316 ethnically matched normal controls as was done previously with D8S261 and AFM333th1. Allele counts and frequencies for cases and controls are shown in Table 5.11 for D8S2612, Table 5.12 for D8S2617, Table 5.13 for D8S2616, Table 5.14 for D8S2615, Table 5.15 for D8S2614, Table 5.16 for D8S2613 and Table 5.17 for D8S2618.

The CLUMP program was used to assess the evidence of allelic association between the marker loci and the disease locus as before (Table 5.18). Allelic association was demonstrated between schizophrenia and D8S2615 (T1: $\chi^2=15.1777$, $p=0.004$) as well as schizophrenia and D8S2616 (T1: $\chi^2=19.9236$, $p=0.024$). Therefore, there are three polymorphisms, including D8S261 (T1: $\chi^2=16.6737$, $p=0.055$, T3: $\chi^2=9.9929$, $p=0.01$), associated with schizophrenia, which lie next to each other and cover an area of 108 kb. The order of these markers from the centromere to the telomere is cen-D8S2615-86 kb-D8S2616-22 kb-D8S261-tel. The other markers, distributed on either side of the three markers showing positive allelic
association, did not show any significant linkage disequilibrium with the disease. This is unexpected for marker AFM333th1, which is only 6 kb telomeric to D8S261.
Table 5.11 Allele counts and frequencies (in parentheses) at marker D8S2612 from the London (UK) case-control sample

<table>
<thead>
<tr>
<th>Alleles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>1 (0.004)</td>
<td>18 (0.069)</td>
<td>40 (0.154)</td>
<td>64 (0.246)</td>
<td>111 (0.427)</td>
<td>26 (0.100)</td>
<td>0 (0.000)</td>
<td>260</td>
</tr>
<tr>
<td>Control individuals</td>
<td>9 (0.014)</td>
<td>35 (0.055)</td>
<td>104 (0.165)</td>
<td>166 (0.263)</td>
<td>271 (0.429)</td>
<td>41 (0.065)</td>
<td>6 (0.009)</td>
<td>632</td>
</tr>
</tbody>
</table>

Table 5.12 Allele counts and frequencies (in parentheses) at marker D8S2617 from the London (UK) case-control sample

<table>
<thead>
<tr>
<th>Alleles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>10</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>4 (0.016)</td>
<td>16 (0.062)</td>
<td>36 (0.139)</td>
<td>16 (0.062)</td>
<td>73 (0.283)</td>
<td>109 (0.422)</td>
<td>3 (0.012)</td>
<td>1 (0.004)</td>
<td>258</td>
</tr>
<tr>
<td>Control individuals</td>
<td>9 (0.035)</td>
<td>57 (0.090)</td>
<td>72 (0.114)</td>
<td>39 (0.062)</td>
<td>172 (0.273)</td>
<td>278 (0.441)</td>
<td>3 (0.005)</td>
<td>0 (0.000)</td>
<td>630</td>
</tr>
</tbody>
</table>
Table 5.13 Allele counts and frequencies (in parentheses) at marker D8S2616 from the London (UK) case-control sample

<table>
<thead>
<tr>
<th>Alleles</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>0</td>
<td>6</td>
<td>9</td>
<td>9</td>
<td>44</td>
<td>139</td>
<td>18</td>
<td>31</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(0.000)</td>
<td>(0.023)</td>
<td>(0.035)</td>
<td>(0.035)</td>
<td>(0.171)</td>
<td>(0.539)</td>
<td>(0.070)</td>
<td>(0.120)</td>
<td>(0.000)</td>
<td>(0.004)</td>
<td>(0.000)</td>
<td>(0.004)</td>
<td>258</td>
</tr>
<tr>
<td>Control individuals</td>
<td>1</td>
<td>4</td>
<td>19</td>
<td>43</td>
<td>71</td>
<td>378</td>
<td>29</td>
<td>75</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(0.002)</td>
<td>(0.006)</td>
<td>(0.030)</td>
<td>(0.069)</td>
<td>(0.113)</td>
<td>(0.604)</td>
<td>(0.046)</td>
<td>(0.120)</td>
<td>(0.002)</td>
<td>(0.006)</td>
<td>(0.002)</td>
<td>(0.000)</td>
<td>626</td>
</tr>
</tbody>
</table>

Table 5.14 Allele counts and frequencies (in parentheses) at marker D8S2615 from the London (UK) case-control sample

<table>
<thead>
<tr>
<th>Alleles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>6</th>
<th>7</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>213</td>
<td>7</td>
<td>33</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(0.826)</td>
<td>(0.027)</td>
<td>(0.128)</td>
<td>(0.016)</td>
<td>(0.004)</td>
<td>(0.000)</td>
<td>258</td>
</tr>
<tr>
<td>Control individuals</td>
<td>562</td>
<td>6</td>
<td>44</td>
<td>8</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>(0.984)</td>
<td>(0.010)</td>
<td>(0.071)</td>
<td>(0.013)</td>
<td>(0.000)</td>
<td>(0.003)</td>
<td>622</td>
</tr>
</tbody>
</table>
Table 5.15 Allele counts and frequencies (in parentheses) at marker D8S2614 from the London (UK) case-control sample

<table>
<thead>
<tr>
<th>Alleles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>1 (0.004)</td>
<td>11 (0.043)</td>
<td>115 (0.446)</td>
<td>24 (0.093)</td>
<td>46 (0.178)</td>
<td>1 (0.004)</td>
<td>57 (0.221)</td>
<td>0 (0.000)</td>
<td>3 (0.012)</td>
<td>0 (0.000)</td>
<td>258</td>
</tr>
<tr>
<td>Control individuals</td>
<td>4 (0.006)</td>
<td>41 (0.065)</td>
<td>289 (0.462)</td>
<td>57 (0.091)</td>
<td>81 (0.129)</td>
<td>3 (0.005)</td>
<td>145 (0.232)</td>
<td>1 (0.002)</td>
<td>4 (0.006)</td>
<td>1 (0.002)</td>
<td>626</td>
</tr>
</tbody>
</table>

Table 5.16 Allele counts and frequencies (in parentheses) at marker D8S2613 from the London (UK) case-control sample

<table>
<thead>
<tr>
<th>Alleles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>9</th>
<th>10</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>7 (0.027)</td>
<td>54 (0.208)</td>
<td>29 (0.112)</td>
<td>162 (0.623)</td>
<td>3 (0.050)</td>
<td>5 (0.019)</td>
<td>0 (0.000)</td>
<td>260</td>
</tr>
<tr>
<td>Control individuals</td>
<td>17 (0.028)</td>
<td>121 (0.197)</td>
<td>73 (0.119)</td>
<td>394 (0.642)</td>
<td>1 (0.002)</td>
<td>7 (0.011)</td>
<td>1 (0.002)</td>
<td>614</td>
</tr>
</tbody>
</table>
Table 5.17 Allele counts and frequencies (in parentheses) at marker D8S2618 from the London (UK) case-control sample

<table>
<thead>
<tr>
<th>Alleles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>1 (0.004)</td>
<td>3 (0.011)</td>
<td>11 (0.042)</td>
<td>1 (0.004)</td>
<td>22 (0.084)</td>
<td>13 (0.047)</td>
<td>8 (0.031)</td>
<td>97 (0.370)</td>
<td>6 (0.023)</td>
<td>0 (0.000)</td>
<td>1 (0.004)</td>
<td>99 (0.378)</td>
</tr>
<tr>
<td>Control individuals</td>
<td>3 (0.005)</td>
<td>3 (0.005)</td>
<td>34 (0.056)</td>
<td>12 (0.020)</td>
<td>41 (0.067)</td>
<td>26 (0.042)</td>
<td>16 (0.026)</td>
<td>218 (0.356)</td>
<td>14 (0.023)</td>
<td>2 (0.003)</td>
<td>8 (0.013)</td>
<td>226 (0.369)</td>
</tr>
</tbody>
</table>

Table 5.17 (continues)

<table>
<thead>
<tr>
<th>Alleles</th>
<th>15</th>
<th>16</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>0 (0.000)</td>
<td>0 (0.000)</td>
<td>262</td>
</tr>
<tr>
<td>Control individuals</td>
<td>8 (0.013)</td>
<td>1 (0.002)</td>
<td>612</td>
</tr>
</tbody>
</table>
Table 5.18 CLUMP test results for association between schizophrenia and alleles at novel polymorphic loci on chromosome 8p21.3-22 in the London (UK) case-control sample

<table>
<thead>
<tr>
<th>Marker Locus (tel→cen)</th>
<th>T1 ( \chi^2 )</th>
<th>T1 p-value</th>
<th>T2 ( \chi^2 )</th>
<th>T2 p-value</th>
<th>T3 ( \chi^2 )</th>
<th>T3 p-value</th>
<th>T4 ( \chi^2 )</th>
<th>T4 p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S2612</td>
<td>8.1917</td>
<td>0.215</td>
<td>3.4390</td>
<td>0.497</td>
<td>3.2719</td>
<td>0.287</td>
<td>3.7954</td>
<td>0.414</td>
</tr>
<tr>
<td>D8S2617</td>
<td>6.7142</td>
<td>0.465</td>
<td>4.1393</td>
<td>0.510</td>
<td>1.9651</td>
<td>0.570</td>
<td>1.9651</td>
<td>0.797</td>
</tr>
<tr>
<td>AFM333th1</td>
<td>5.3018</td>
<td>0.509</td>
<td>4.1088</td>
<td>0.544</td>
<td>2.0610</td>
<td>0.512</td>
<td>3.7441</td>
<td>0.426</td>
</tr>
<tr>
<td>D8S261</td>
<td>16.6737</td>
<td>0.055</td>
<td>10.6062</td>
<td>0.150</td>
<td>9.9929</td>
<td>0.010</td>
<td>13.8258</td>
<td>0.017</td>
</tr>
<tr>
<td>D8S2616</td>
<td>19.9236</td>
<td>0.024</td>
<td>12.9969</td>
<td>0.048</td>
<td>5.2682</td>
<td>0.118</td>
<td>11.9128</td>
<td>0.031</td>
</tr>
<tr>
<td>D8S2615</td>
<td>15.1777</td>
<td>0.004</td>
<td>10.5462</td>
<td>0.009</td>
<td>10.5462</td>
<td>0.005</td>
<td>12.1292</td>
<td>0.004</td>
</tr>
<tr>
<td>D8S2614</td>
<td>6.5602</td>
<td>0.720</td>
<td>4.8821</td>
<td>0.445</td>
<td>3.5514</td>
<td>0.260</td>
<td>4.1641</td>
<td>0.474</td>
</tr>
<tr>
<td>D8S2613</td>
<td>5.4494</td>
<td>0.496</td>
<td>2.7285</td>
<td>0.608</td>
<td>0.2735</td>
<td>0.956</td>
<td>3.9373</td>
<td>0.319</td>
</tr>
<tr>
<td>D8S2618</td>
<td>12.3682</td>
<td>0.521</td>
<td>7.9964</td>
<td>0.423</td>
<td>0.7905</td>
<td>0.947</td>
<td>9.3453</td>
<td>0.210</td>
</tr>
</tbody>
</table>

Where,

T1: a straightforward Pearson's \( \chi^2 \) statistic of the ‘raw’ contingency table

T2: the \( \chi^2 \) statistic of a table with rare alleles grouped together to prevent small expected cell counts

T3: the largest of the \( \chi^2 \) statistics of 2x2 tables each of which compares one allele against the rest grouped together

T4: the largest of the \( \chi^2 \) statistics of all possible 2x2 tables comparing any combination of alleles against the rest
5.5 REPLICATION STUDIES IN TWO CASE-CONTROL SAMPLES FROM SCOTLAND

A follow up study was performed in two Scottish independent case-control samples in an attempt to replicate the previous finding. Only the three microsatellite markers that had demonstrated allelic association in the London (UK) case-control sample were genotyped in the two Scottish samples.

In the first replication study, 100 schizophrenia patients and 100 ethnically matched normal controls from Scotland were used. Allele counts and frequencies for cases and controls are shown in Table 5.19 for D8S261, Table 5.23 for D8S2616 and Table 5.27 for D8S2615. CLUMP analysis was performed but failed to demonstrate significant allelic association between any of the marker loci and the disease (Table 5.31).

The replication sample was increased by obtaining another 100 schizophrenia and 100 ethnically matched controls from Scotland. Allele counts and frequencies for cases and controls for the second replication study are shown in Table 5.20 for D8S261, Table 5.24 for D8S2616 and Table 5.28 for D8S2615. CLUMP analysis was performed but failed to demonstrate significant allelic association between D8S261 and D8S2615 and the disease (Table 5.32). However, D8S2616 was significant at 0.043 (T1: $\chi^2=16.3893, p=0.043$).

The data from two case-control allelic association studies from the Scottish samples were combined. Allele counts and frequencies for cases and controls for both case-control samples of Scottish ancestry are shown in Table 5.21 for D8S261, Table 5.25 for D8S2616 and Table 5.29 for D8S2615. CLUMP analysis was performed but
failed to demonstrate significant allelic association between any of the marker loci and the disease (Table 5.33).

5.6 ASSESSMENT OF ALLELIC ASSOCIATION FROM ALL AVAILABLE SAMPLES

The data from the London (UK) case-control sample was combined with the data of the two case-control samples from Scotland. Allele counts and frequencies for this combined case-control sample are shown in Table 5.22 for D8S261, Table 5.26 for D8S2616 and Table 5.30 for D8S2615. CLUMP analysis was performed and demonstrated that significant allelic association was maintained only between D8S2616 and schizophrenia ($\chi^2=20.0613$, $p=0.05$) according to the $T_1$ statistic produced by CLUMP (Table 5.34).

However, the empirical $p$-values of the $T_3$ statistic (the largest of the $\chi^2$ statistics of 2x2 tables each of which compares one allele against the rest grouped together) produced by CLUMP were significant for all three markers, D8S261 ($p=0.047$), D8S2616 ($p=0.033$) and D8S2615 ($p=0.05$). These results were due to an excess of the following alleles (in cases): allele 8 of D8S261 ($p=0.005$, uncorrected for multiple alleles), allele 7 of D8S2616 ($p=0.004$, uncorrected for multiple alleles) and allele 1 of D8S2615 ($p=0.015$, uncorrected for multiple alleles).
Table 5.19 Allele counts and frequencies (in parentheses) at marker D8S261
from the first case-control sample from Scotland

<table>
<thead>
<tr>
<th>Alleles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>6 (0.031)</td>
<td>59 (0.301)</td>
<td>0 (0.000)</td>
<td>8 (0.041)</td>
<td>75 (0.383)</td>
<td>23 (0.117)</td>
<td>21 (0.107)</td>
<td>2 (0.010)</td>
<td>1 (0.005)</td>
<td>1 (0.005)</td>
<td>196</td>
</tr>
<tr>
<td>Control individuals</td>
<td>5 (0.025)</td>
<td>54 (0.273)</td>
<td>2 (0.010)</td>
<td>14 (0.071)</td>
<td>83 (0.419)</td>
<td>16 (0.081)</td>
<td>12 (0.061)</td>
<td>3 (0.015)</td>
<td>1 (0.005)</td>
<td>8 (0.010)</td>
<td>198</td>
</tr>
</tbody>
</table>

Table 5.20 Allele counts and frequencies (in parentheses) at marker D8S261
from the second case-control sample from Scotland

<table>
<thead>
<tr>
<th>Alleles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>11 (0.057)</td>
<td>52 (0.268)</td>
<td>3 (0.015)</td>
<td>10 (0.052)</td>
<td>62 (0.319)</td>
<td>20 (0.103)</td>
<td>22 (0.113)</td>
<td>9 (0.046)</td>
<td>0 (0.000)</td>
<td>4 (0.021)</td>
<td>1 (0.005)</td>
<td>194</td>
</tr>
<tr>
<td>Control individuals</td>
<td>6 (0.030)</td>
<td>66 (0.33)</td>
<td>2 (0.010)</td>
<td>12 (0.060)</td>
<td>59 (0.295)</td>
<td>30 (0.15)</td>
<td>18 (0.09)</td>
<td>4 (0.02)</td>
<td>0 (0.000)</td>
<td>2 (0.010)</td>
<td>1 (0.005)</td>
<td>200</td>
</tr>
</tbody>
</table>
Table 5.21 Allele counts and frequencies (in parentheses) at marker D8S261 from both case-control samples from Scotland

<table>
<thead>
<tr>
<th>Alleles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>17 (0.044)</td>
<td>111 (0.285)</td>
<td>3 (0.008)</td>
<td>18 (0.046)</td>
<td>137 (0.351)</td>
<td>43 (0.110)</td>
<td>43 (0.110)</td>
<td>11 (0.028)</td>
<td>1 (0.003)</td>
<td>5 (0.013)</td>
<td>1 (0.003)</td>
<td>390</td>
</tr>
<tr>
<td>Control individuals</td>
<td>11 (0.028)</td>
<td>120 (0.302)</td>
<td>4 (0.010)</td>
<td>26 (0.065)</td>
<td>142 (0.357)</td>
<td>46 (0.116)</td>
<td>30 (0.075)</td>
<td>7 (0.018)</td>
<td>1 (0.003)</td>
<td>10 (0.025)</td>
<td>1 (0.003)</td>
<td>398</td>
</tr>
</tbody>
</table>

Table 5.22 Allele counts and frequencies (in parentheses) at marker D8S261 from all three case-control samples

<table>
<thead>
<tr>
<th>Alleles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>30 (0.047)</td>
<td>187 (0.294)</td>
<td>9 (0.014)</td>
<td>32 (0.050)</td>
<td>212 (0.333)</td>
<td>73 (0.115)</td>
<td>61 (0.096)</td>
<td>23 (0.036)</td>
<td>1 (0.002)</td>
<td>7 (0.011)</td>
<td>1 (0.002)</td>
<td>636</td>
</tr>
<tr>
<td>Control individuals</td>
<td>39 (0.038)</td>
<td>316 (0.311)</td>
<td>9 (0.009)</td>
<td>62 (0.061)</td>
<td>349 (0.344)</td>
<td>121 (0.139)</td>
<td>79 (0.078)</td>
<td>15 (0.015)</td>
<td>5 (0.005)</td>
<td>20 (0.020)</td>
<td>1 (0.001)</td>
<td>1016</td>
</tr>
</tbody>
</table>
Table 5.23 Allele counts and frequencies (in parentheses) at marker D8S2616 from the first case-control sample from Scotland

<table>
<thead>
<tr>
<th>Alleles</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>2 (0.011)</td>
<td>8 (0.042)</td>
<td>11 (0.058)</td>
<td>30 (0.158)</td>
<td>104 (0.547)</td>
<td>13 (0.069)</td>
<td>16 (0.084)</td>
<td>1 (0.005)</td>
<td>5 (0.026)</td>
<td>190</td>
</tr>
<tr>
<td>Control individuals</td>
<td>2 (0.010)</td>
<td>6 (0.031)</td>
<td>12 (0.062)</td>
<td>22 (0.113)</td>
<td>120 (0.619)</td>
<td>8 (0.041)</td>
<td>22 (0.114)</td>
<td>2 (0.010)</td>
<td>0 (0.000)</td>
<td>194</td>
</tr>
</tbody>
</table>

Table 5.24 Allele counts and frequencies (in parentheses) at marker D8S2616 from the second case-control sample from Scotland

<table>
<thead>
<tr>
<th>Alleles</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>0 (0.000)</td>
<td>0 (0.000)</td>
<td>15 (0.079)</td>
<td>27 (0.143)</td>
<td>112 (0.589)</td>
<td>15 (0.079)</td>
<td>19 (0.100)</td>
<td>1 (0.005)</td>
<td>0 (0.000)</td>
<td>1 (0.005)</td>
<td>190</td>
</tr>
<tr>
<td>Control individuals</td>
<td>2 (0.010)</td>
<td>6 (0.030)</td>
<td>15 (0.076)</td>
<td>19 (0.095)</td>
<td>113 (0.572)</td>
<td>12 (0.061)</td>
<td>25 (0.126)</td>
<td>1 (0.005)</td>
<td>5 (0.025)</td>
<td>0 (0.000)</td>
<td>198</td>
</tr>
</tbody>
</table>
Table 5.25 Allele counts and frequencies (in parentheses) at marker D8S2616 from both case-control samples from Scotland

<table>
<thead>
<tr>
<th>Alleles</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>2 (0.005)</td>
<td>8 (0.021)</td>
<td>26 (0.068)</td>
<td>57 (0.150)</td>
<td>216 (0.568)</td>
<td>28 (0.074)</td>
<td>35 (0.093)</td>
<td>2 (0.005)</td>
<td>5 (0.013)</td>
<td>1 (0.003)</td>
</tr>
<tr>
<td>Control individuals</td>
<td>4 (0.010)</td>
<td>12 (0.031)</td>
<td>27 (0.069)</td>
<td>41 (0.105)</td>
<td>233 (0.593)</td>
<td>20 (0.051)</td>
<td>47 (0.120)</td>
<td>3 (0.008)</td>
<td>5 (0.013)</td>
<td>0 (0.000)</td>
</tr>
</tbody>
</table>

Row Total | 380

Table 5.26 Allele counts and frequencies (in parentheses) at marker D8S2616 from all three case-control samples

<table>
<thead>
<tr>
<th>Alleles</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>0 (0.000)</td>
<td>8 (0.013)</td>
<td>17 (0.027)</td>
<td>35 (0.055)</td>
<td>101 (0.158)</td>
<td>355 (0.556)</td>
<td>46 (0.072)</td>
<td>66 (0.103)</td>
<td>2 (0.003)</td>
<td>6 (0.009)</td>
<td>1 (0.002)</td>
<td>0 (0.000)</td>
<td>1 (0.002)</td>
</tr>
<tr>
<td>Control individuals</td>
<td>1 (0.001)</td>
<td>8 (0.008)</td>
<td>31 (0.030)</td>
<td>70 (0.068)</td>
<td>112 (0.110)</td>
<td>611 (0.600)</td>
<td>49 (0.048)</td>
<td>122 (0.119)</td>
<td>4 (0.004)</td>
<td>9 (0.009)</td>
<td>0 (0.000)</td>
<td>1 (0.001)</td>
<td>0 (0.000)</td>
</tr>
</tbody>
</table>

Row Total | 638

Row Total | 1018
### Table 5.27 Allele counts and frequencies (in parentheses) at marker D8S2615 from the first case-control sample from Scotland

<table>
<thead>
<tr>
<th>Alleles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>6</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>180 (0.910)</td>
<td>3 (0.015)</td>
<td>11 (0.055)</td>
<td>0 (0.000)</td>
<td>4 (0.020)</td>
<td>198</td>
</tr>
<tr>
<td>Control individuals</td>
<td>175 (0.884)</td>
<td>4 (0.020)</td>
<td>13 (0.066)</td>
<td>0 (0.000)</td>
<td>6 (0.030)</td>
<td>198</td>
</tr>
</tbody>
</table>

### Table 5.28 Allele counts and frequencies (in parentheses) at marker D8S2615 from the second case-control sample from Scotland

<table>
<thead>
<tr>
<th>Alleles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>6</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>163 (0.832)</td>
<td>5 (0.026)</td>
<td>20 (0.102)</td>
<td>1 (0.005)</td>
<td>7 (0.035)</td>
<td>196</td>
</tr>
<tr>
<td>Control individuals</td>
<td>172 (0.869)</td>
<td>6 (0.030)</td>
<td>16 (0.081)</td>
<td>0 (0.000)</td>
<td>4 (0.020)</td>
<td>198</td>
</tr>
</tbody>
</table>

### Table 5.29 Allele counts and frequencies (in parentheses) at marker D8S2615 from both case-control samples from Scotland

<table>
<thead>
<tr>
<th>Alleles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>6</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>343 (0.870)</td>
<td>8 (0.020)</td>
<td>31 (0.079)</td>
<td>1 (0.003)</td>
<td>11 (0.028)</td>
<td>394</td>
</tr>
<tr>
<td>Control individuals</td>
<td>347 (0.876)</td>
<td>10 (0.025)</td>
<td>29 (0.074)</td>
<td>0 (0.000)</td>
<td>10 (0.025)</td>
<td>396</td>
</tr>
</tbody>
</table>

### Table 5.30 Allele counts and frequencies (in parentheses) at marker D8S2615 from all three case-control samples

<table>
<thead>
<tr>
<th>Alleles</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>10</th>
<th>Row Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>556 (0.853)</td>
<td>15 (0.023)</td>
<td>64 (0.098)</td>
<td>1 (0.002)</td>
<td>15 (0.023)</td>
<td>1 (0.002)</td>
<td>0 (0.000)</td>
<td>652</td>
</tr>
<tr>
<td>Control individuals</td>
<td>909 (0.893)</td>
<td>16 (0.016)</td>
<td>73 (0.072)</td>
<td>0 (0.000)</td>
<td>18 (0.017)</td>
<td>0 (0.000)</td>
<td>2 (0.002)</td>
<td>1018</td>
</tr>
</tbody>
</table>
Table 5.31 CLUMP test results for association between schizophrenia and alleles at polymorphic loci on chromosome 8p21.3-22 from the first case-control sample from Scotland

<table>
<thead>
<tr>
<th>Marker Locus (tel—cen)</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2$</td>
<td>p-value</td>
<td>$\chi^2$</td>
<td>p-value</td>
</tr>
<tr>
<td>D8S261</td>
<td>13.6992</td>
<td>0.110</td>
<td>11.6102</td>
<td>0.081</td>
</tr>
<tr>
<td>D8S2616</td>
<td>10.1334</td>
<td>0.256</td>
<td>6.1330</td>
<td>0.395</td>
</tr>
<tr>
<td>D8S2615</td>
<td>0.7799</td>
<td>0.854</td>
<td>0.7665</td>
<td>0.657</td>
</tr>
</tbody>
</table>

Table 5.32 CLUMP test results for association between schizophrenia and alleles at polymorphic loci on chromosome 8p21.3-22 from the second case-control sample from Scotland

<table>
<thead>
<tr>
<th>Marker Locus (tel—cen)</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2$</td>
<td>p-value</td>
<td>$\chi^2$</td>
<td>p-value</td>
</tr>
<tr>
<td>D8S261</td>
<td>8.4881</td>
<td>0.488</td>
<td>8.3137</td>
<td>0.317</td>
</tr>
<tr>
<td>D8S2616</td>
<td>16.3893</td>
<td>0.043</td>
<td>11.3872</td>
<td>0.041</td>
</tr>
<tr>
<td>D8S2615</td>
<td>2.5852</td>
<td>0.654</td>
<td>1.4943</td>
<td>0.670</td>
</tr>
</tbody>
</table>

Table 5.33 CLUMP test results for association between schizophrenia and alleles at polymorphic loci on chromosome 8p21.3-22 from both case-control samples from Scotland

<table>
<thead>
<tr>
<th>Marker Locus (tel—cen)</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2$</td>
<td>p-value</td>
<td>$\chi^2$</td>
<td>p-value</td>
</tr>
<tr>
<td>D8S261</td>
<td>8.2148</td>
<td>0.605</td>
<td>8.1628</td>
<td>0.438</td>
</tr>
<tr>
<td>D8S2616</td>
<td>8.8465</td>
<td>0.477</td>
<td>7.1612</td>
<td>0.293</td>
</tr>
<tr>
<td>D8S2615</td>
<td>1.3546</td>
<td>0.933</td>
<td>0.1850</td>
<td>0.974</td>
</tr>
</tbody>
</table>

Where,
T1: a straightforward Pearson’s $\chi^2$ statistic of the ‘raw’ contingency table
T2: the $\chi^2$ statistic of a table with rare alleles grouped together to prevent small expected cell counts
T3: the largest of the $\chi^2$ statistics of 2x2 tables each of which compares one allele against the rest grouped together
T4: the largest of the $\chi^2$ statistics of all possible 2x2 tables comparing any combination of alleles against the rest
Table 5.34 CLUMP test results for association between schizophrenia and alleles at polymorphic loci in chromosome 8p21.3-22 from all three case-control samples

<table>
<thead>
<tr>
<th>Marker Locus (tel→cen)</th>
<th>T1</th>
<th></th>
<th></th>
<th>T2</th>
<th></th>
<th></th>
<th>T3</th>
<th></th>
<th></th>
<th>T4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S261</td>
<td>15.4998</td>
<td>0.112</td>
<td></td>
<td>13.3339</td>
<td>0.098</td>
<td></td>
<td>7.9706</td>
<td>0.047</td>
<td></td>
<td>8.4865</td>
<td>0.137</td>
</tr>
<tr>
<td>D8S2616</td>
<td>20.0613</td>
<td>0.050</td>
<td></td>
<td>15.5574</td>
<td>0.022</td>
<td></td>
<td>8.1593</td>
<td>0.033</td>
<td></td>
<td>15.9036</td>
<td>0.008</td>
</tr>
<tr>
<td>D8S2615</td>
<td>10.2318</td>
<td>0.072</td>
<td></td>
<td>6.0261</td>
<td>0.107</td>
<td></td>
<td>5.9546</td>
<td>0.050</td>
<td></td>
<td>6.6068</td>
<td>0.062</td>
</tr>
</tbody>
</table>

Where,

**T1**: a straightforward Pearson’s $\chi^2$ statistic of the ‘raw’ contingency table

**T2**: the $\chi^2$ statistic of a table with rare alleles grouped together to prevent small expected cell counts

**T3**: the largest of the $\chi^2$ statistics of 2x2 tables each of which compares one allele against the rest grouped together

**T4**: the largest of the $\chi^2$ statistics of all possible 2x2 tables comparing any combination of alleles against the rest
5.7 LINKAGE DISEQUILIBRIUM ANALYSIS

In order to investigate the robustness of the allelic association findings, pairwise linkage disequilibrium was calculated between each pair of marker loci using the EH program (Terwilliger & Ott, 1994). All marker loci demonstrated linkage disequilibrium levels, which corresponded well with their intermarker physical distances; the marker locus, which was furthest away (D8S2618) showing no significant linkage disequilibrium with any of the other markers. The three marker loci, D8S2616, D8S2615, D8S261, which showed significant allelic association with schizophrenia show also highly significant pair-wise linkage disequilibrium with each other as well as with their adjacent marker loci (Table 5.35).

In addition, pairwise linkage disequilibrium was examined in the data from the two case-control Scottish samples combined together. The markers examined in this sample (D8S2615, D8S2616 and D8S261) demonstrate significant pair-wise linkage disequilibrium (Table 5.36).
Table 5.35 Linkage disequilibrium between paired microsatellite markers on chromosome 8p21.3-22

<table>
<thead>
<tr>
<th>Intermarker distances (kb)</th>
<th>markers</th>
<th>D8S2613</th>
<th>D8S2614</th>
<th>D8S2615</th>
<th>D8S2616</th>
<th>D8S2617</th>
<th>Afm333th1</th>
<th>D8S2612</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S2618</td>
<td>0.1920</td>
<td>0.3104</td>
<td>0.1935</td>
<td>0.3051</td>
<td>0.3063</td>
<td>0.8875</td>
<td>0.6357</td>
<td>0.3138</td>
</tr>
<tr>
<td>179</td>
<td>0.0302</td>
<td>0.0013</td>
<td>0.2207</td>
<td>0.2407</td>
<td>0.0712</td>
<td>0.0435</td>
<td>0.6008</td>
<td></td>
</tr>
<tr>
<td>D8S2613</td>
<td>0.0655</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>0.0014</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D8S2614</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D8S2615</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>0.009</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D8S2616</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D8S261</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Afm333th1</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D8S2617</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Each microsatellite marker is listed across the top and down the left side. Pair-wise linkage disequilibrium was calculated between each pair of marker loci using the EH program. The table shows the $p$-values that were determined by the $\chi^2$ statistic calculated with this program. LD is significant between two markers at $p<0.005$. 
Table 5.36 Linkage disequilibrium between paired microsatellite markers on chromosome 8p21.3-22 from the combined Scottish sample

<table>
<thead>
<tr>
<th>Intermarker distances (kb)</th>
<th>86</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>markers</td>
<td>D8S2615</td>
<td>D8S2616</td>
</tr>
<tr>
<td>D8S2615</td>
<td>0.0000</td>
<td>0.0002</td>
</tr>
<tr>
<td>D8S2616</td>
<td>86</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Each microsatellite marker is listed across the top and down the left side. Pair-wise linkage disequilibrium was calculated between each pair of marker loci using the EH program. The table shows the p-values that were determined by the $\chi^2$ statistic calculated with this program. LD is significant between two markers at $p<0.005$. 
Evidence of an allelic association in a region that has previously been implicated by linkage as harbouring a susceptibility locus for a disease provides further evidence for the presence of a disease susceptibility locus at the 8p21-22 region.

In this report, we attempted to narrow down the potential susceptibility region for schizophrenia on chromosome 8p21-22 by performing case-control allelic association studies. Markers, which had provided positive evidence for linkage with schizophrenia, were genotyped in cases and controls of Irish, Welsh, Scottish or English ancestry. A preliminary allelic association study was performed between markers D8S261 (37 cM), D8S136 (43 cM) and D8S1771 (50 cM) and schizophrenia. An allelic association was obtained with D8S261 and the disease.

According to the available mapping data at the time, 3 more markers (AFM333th1, D8S1145 and an MboII RFLP in the NAT1 gene) in close proximity (less than 1 cM) of D8S261 were examined for evidence of allelic association with schizophrenia. None of these markers was associated with the disease.

Finally, an allelic association study was performed with seven very closely placed microsatellite markers around D8S261, overall covering a region of ~700 kb. Two more markers, D8S2615 and D8S2616, 108 and 22 kb respectively centromeric to D8S261 demonstrated statistically significant allelic association with schizophrenia. The CLUMP program was used for estimating these association results. The identification of three markers in association with the disease as well as the stringent statistical analysis make it improbable that the result is a statistical error. The allelic association detected could to be caused from population stratification.
However, this seems unlikely as several markers at greater distances as well as at the vicinity of the markers that were in linkage disequilibrium with schizophrenia, had very similar allele frequencies between the cases and the controls (see Table 5.4).

Further support for this association finding comes from pairwise linkage disequilibrium analysis of this region, which showed that all the polymorphisms with the exception of D8S2618 and D8S2613 were in strong linkage disequilibrium with each other. Therefore, it is plausible that linkage disequilibrium exists between some of these marker loci and a disease gene in close proximity.

However, some of the microsatellite markers (D8S2614 and D8S2617) that demonstrated very strong linkage disequilibrium with the three primarily associated markers and especially AFM333th1, which is only 6 kb away from D8S261, failed to demonstrate linkage disequilibrium with the disease. A possible explanation can be the differences in distribution of allele frequencies between the marker locus and the disease locus. For the three primarily associated markers it seems that the alleles with low frequencies are associated with the disease. Therefore, if alleles with high frequencies at the non-associated marker loci are in linkage disequilibrium with the disease locus association will be less easy to detect. This seems to be particularly the case with AFM333th1. This observation demonstrates that for the fine localization of a susceptibility locus with a small contribution to disease, polymorphisms should be very closely spaced.

The failure to adequately replicate the association in an independent case-control sample is unlikely to be due to population stratification, since both groups of individuals were carefully matched for ethnicity and drawn from as ethnically a homogenous population as possible. However, it may be due to differences in the
patterns of disequilibrium of the two samples or to random variations in the inputs to genetic liability present in the two different samples.

After combining the results from the London (UK) and Scottish case-control samples, allelic association was still found between D8S2615, D8S2616, D8S261, and schizophrenia. This allelic association finding was obtained from the T3 statistic produced by CLUMP. The T3 statistic represents the largest of the $\chi^2$ statistics of 2x2 tables each of which compares one allele against the rest grouped together. However, the T1 statistic, which is the straightforward Pearson’s $\chi^2$ value of the ‘raw’ contingency table, produced a significant $p$-value only for D8S2616.

This result also indicates that the original allelic association finding in the London (UK) case-control sample was very strong and remained so even when the sample was increased in size by the addition of the Scottish samples.

The typing of more homogenous (genetic isolates) as well as heterogeneous populations and family samples is necessary in order to further validate this finding of allelic association.

Furthermore, the use of powerful methods to provide haplotype information will be essential using family data.

In addition, the genotyping of several single nucleotide polymorphisms in this area may help to provide further evidence of allelic association.

Overall, our data indicate that microsatellite markers have the power to detect linkage disequilibrium and can be used as tools for the fine localization of disease genes by disequilibrium mapping in regions previously implicated by linkage. Further validation of the data presented in this chapter will demonstrate the potential of such studies in accurately refining susceptibility regions in complex disorders even when a relatively heterogeneous (non-isolate) population sample is used.
CHAPTER 6

A CANDIDATE GENE STUDY OF HUMAN PERICENTRIOLAR MATERIAL 1 GENE (PCM1)
6.1 INTRODUCTION

As was described in the previous chapter, allelic association was observed between three neighbouring microsatellite markers, D8S2615, D8S2616, D8S261, and schizophrenia spanning an area of 108 kb on the 8p22 region. The region around the association was searched for candidate genes. The transcription map (GeneMap99) of human chromosome 8 at the National Centre for Biotechnology Information (http://www.ncbi.nlm.nih.gov/genemap99/map.cgi?CHR=8) and The Genome Channel browser (http://combio.orl.gov/tools/channel/index.html), which provides automated computational annotation of genomic sequence, were two excellent resources for this search. None of the genes already identified at the close proximity to the associated polymorphisms appeared to have an obvious involvement to the pathophysiology of schizophrenia.

The pericentriolar material 1 gene (PCM 1) has been mapped to chromosome 8p21.3-33 (Ohata et al., 1994). In addition, it is annotated in the Genome Channel and its sequence occupies part of clones AB020866 and AB020867 of the NT_000501 contig. Two of the associated polymorphisms, D8S2616 and D8S261 lie within the intronic sequence of this gene while the third polymorphism, D8S2615 resides approximately 75 kb upstream of the translation initiation codon of this gene.

Balczon et al, (Balczon & West, 1991) identified a ~220 kD centrosome autoantigen. A cDNA encoding the entire protein was isolated from a human fetal liver cDNA library. Analysis of the cloned sequence identified an open reading frame of 6,072 nucleotides encoding 2,024 amino acids. From the deduced amino acid sequence, the exact molecular mass of PCM1 was calculated to be 228,705 daltons (Balczon et al., 1994).
The centrosome functions as an organizing centre for cytoskeletal components, especially microtubules. This structure has a centriole pair at its centre, surrounded by fibrous material known as pericentriolar material. PCM1 is one of the molecules that constitute the pericentriolar material. In HeLa cells, PCM1 was reported to be associated with the centrosomes in the interphase but dissociated in metaphase. How the apparent dynamic nature of PCM1 contributes to the centrosome function is not clear at present but raises interesting questions regarding the regulation of the microtubule cytoskeleton. It has also been suggested that it may act as an inhibitor of microtubule growth by blocking microtubule-nucleating sites in the centrosome. Finally, it may be involved in other centrosomal phenomena such as maintenance of centrosomal integrity throughout the early phases of cell cycle and/or centriolar replication. It is widely accepted that centriolar replication begins near the G1/S boundary, continues through S phase and is completed during G2 phase. In good agreement, PCM1 at centrosomes is released into the cytoplasm on the entry to M phase and on the entry to interphase, this molecule is reconcentrated at centrosomes.

PCM1 was also identified as a component of centriolar satellites and fibrous granules, which are non-membranous organelles, which are accumulated around centrosomes and ciliary basal bodies, respectively, through their minus-end directed movement along microtubules (Kubo et al., 1999). The physiological function of centriolar satellites remains unclear. A possible role is in centriolar replication.

PCM1 resides in the region 8p21.3-22, which is commonly involved in loss of heterozygosity in hepatocellular carcinomas, colorectal cancers, and non-small cell lung cancers. Studies so far have failed to prove its involvement in cancer (Ohata et al., 1994).
Although, the biological function of PCM1 is not well elucidated its proximity to the associated polymorphisms make it a plausible candidate for susceptibility to schizophrenia. The study of the genetic variation in this gene will provide valuable information as to whether PCM1 or some other yet unidentified gene very proximal to PCM1 is involved in the pathophysiology of schizophrenia.

6.2 GENOMIC ORGANIZATION OF PCM1

The predicted coding sequence has a translation initiation codon (ATG) located at position 410 of PCM1 mRNA, which is preceded by a Kozak consensus (CCAXXATGG) initiation sequence (Kozak, 1984) and a stop codon (TGA) is located at position 6482 of the PCM1 mRNA (Figure 6.1).

In order to define the exon-intron genomic organization, “BLAST 2 sequences” program from the NCBI was used. This is a tool, which produces an alignment of two given sequences using the BLAST program. In this case, the two sequences were the mRNA sequence of PCM1 (Accession No. NM_006197) against the corresponding genomic sequence (clones AB020866 and AB020867 of the NT_000501 contig). The exon-intron junctions of the PCM1 transcript derived from these data is shown in Table 6.1. The predicted exon-intron boundaries are in excellent agreement with the “GT-AG rule” that is nucleotides at the exon-intron boundaries are not random but introns always seem to begin with GT and end with AG. In addition, the predicted exons are in excellent agreement with the exons generated by the GRAILEXP program used for gene prediction by the Genome Channel, which is a site that provides automated annotation of genomic sequence.
Figure 6.1 The longest open reading frame of Homo sapiens pericentriolar material 1 (PCM1) mRNA and the translated product. The coding sequence has a length of 6075 nt and the translated product is 2024 aa long. The translation initiation codon (atg) is at position 410 while the stop codon (tga) is at position 6482 of the PCM1 mRNA. Obtained from the National Institute of Biotechnology Information (NCBI) Open Reading Frame finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html)
[Text content extracted from the document]
Table 6.1 Intron-exon boundary sequences of PCM1*

<table>
<thead>
<tr>
<th>EXON</th>
<th>mRNA position</th>
<th>Splice acceptor</th>
<th>Splice donor</th>
<th>Intron size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5' UTR-505</td>
<td>5' UTR-ATGGCCACA</td>
<td>AACAATATGgtatgattc</td>
<td>1427</td>
</tr>
<tr>
<td>2</td>
<td>506-751</td>
<td>ttttaaagGATTGGGGT</td>
<td>TTAGATCGGgtttgaa</td>
<td>1359</td>
</tr>
<tr>
<td>3</td>
<td>752-1021</td>
<td>catatataagAGAAGCATT</td>
<td>AGCACGGGgtgataacg</td>
<td>694</td>
</tr>
<tr>
<td>4</td>
<td>1022-1192</td>
<td>cccctgcaagATTGTAAGC</td>
<td>AAAATCCTTgttaagtatt</td>
<td>7322</td>
</tr>
<tr>
<td>5</td>
<td>1193-1373</td>
<td>attttccagGCGAGAGAT</td>
<td>ATGATTCTGgtatgac</td>
<td>3251</td>
</tr>
<tr>
<td>6</td>
<td>1374-1484</td>
<td>tctttccagTTGTTGCA</td>
<td>GATTTCAGgtaacctag</td>
<td>2250</td>
</tr>
<tr>
<td>7</td>
<td>1485-1700</td>
<td>tatttaaagCCTCCAGCT</td>
<td>ATTCATCATgtgatgaa</td>
<td>2283</td>
</tr>
<tr>
<td>8</td>
<td>1701-1884</td>
<td>cttcttcagCCTCTCCAC</td>
<td>AAAACTCCAgttaaacat</td>
<td>950</td>
</tr>
<tr>
<td>9</td>
<td>1885-2058</td>
<td>gtttatcagGAAGTTAAA</td>
<td>TAACATTTCGgttaagaact</td>
<td>487</td>
</tr>
</tbody>
</table>

*intronic sequence is indicated in lower-case letters, exonic sequence in upper-case
Table 6.1 Intron-exon boundary sequences of PCM1* (continues)

<table>
<thead>
<tr>
<th>EXON</th>
<th>mRNA position</th>
<th>Splice acceptor</th>
<th>Splice donor</th>
<th>Intron size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2059-2216</td>
<td>ttcttctagAAATCCACA</td>
<td>CTTCTTTAGgtatgactg</td>
<td>119</td>
</tr>
<tr>
<td>11</td>
<td>2217-2393</td>
<td>ttccagcagATTTGTCGAT</td>
<td>ATGGTACAGGtaaatatt</td>
<td>2257</td>
</tr>
<tr>
<td>12</td>
<td>2394-2604</td>
<td>gcttttcagGATGATGAT</td>
<td>AAAGGCAAGGtatgtaaa</td>
<td>148</td>
</tr>
<tr>
<td>13</td>
<td>2605-2734</td>
<td>ttattgtagAGAGAAATT</td>
<td>GACCTACAGGtaattatg</td>
<td>1605</td>
</tr>
<tr>
<td>14</td>
<td>2735-2875</td>
<td>ccttttttagCTGTCAGCT</td>
<td>GATAATGAGGtatgtaaa</td>
<td>927</td>
</tr>
<tr>
<td>15</td>
<td>2876-3066</td>
<td>aaaaaatagTTGTGGTCA</td>
<td>AACAGAAAAAgtaaagag</td>
<td>1277</td>
</tr>
<tr>
<td>16</td>
<td>3067-3267</td>
<td>aatcactagAAGCATGAC</td>
<td>TAAAAATAGgtagtttc</td>
<td>1231</td>
</tr>
<tr>
<td>17</td>
<td>3268-3484</td>
<td>cttggttagGTGGAAGAA</td>
<td>GACCAGCAGGtaaatattt</td>
<td>110</td>
</tr>
<tr>
<td>18</td>
<td>3485-3630</td>
<td>tttcaatagACTCTATCT</td>
<td>TGTTACAGAGGtaatgt</td>
<td>507</td>
</tr>
</tbody>
</table>

*intronic sequence is indicated in lower-case letters, exonic sequence in upper-case
Table 6.1 Intron-exon boundary sequences of PCM1* (continues)

<table>
<thead>
<tr>
<th>EXON</th>
<th>mRNA position</th>
<th>Splice acceptor</th>
<th>Splice donor</th>
<th>Intron size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>3631-3821</td>
<td>tctttgtagGTTGAAACA</td>
<td>TTGCACCAGgtagtgac</td>
<td>2421</td>
</tr>
<tr>
<td>20</td>
<td>3822-3993</td>
<td>actgattagGTATGAATT</td>
<td>GAAACCAAGGttacctgt</td>
<td>2584</td>
</tr>
<tr>
<td>21</td>
<td>3994-4352</td>
<td>attggtaagGAATAAAAA</td>
<td>AAAACATCAAgtaagtt</td>
<td>7906</td>
</tr>
<tr>
<td>22</td>
<td>4353-4517</td>
<td>ttactcaagGATATGAAA</td>
<td>TATCTTCAGGtttattct</td>
<td>4691</td>
</tr>
<tr>
<td>23</td>
<td>4518-4690</td>
<td>gtattccagAAACTGGGA</td>
<td>GCATTGCAGGtatctgg</td>
<td>340</td>
</tr>
<tr>
<td>24</td>
<td>4691-4819</td>
<td>cttattnagGACATAGTA</td>
<td>ACTGATGATGtaagctg</td>
<td>3734</td>
</tr>
<tr>
<td>25</td>
<td>4820-4940</td>
<td>gtaacttagGAAACTTTT</td>
<td>ATGATCTAGGtaagcaga</td>
<td>1554</td>
</tr>
<tr>
<td>26</td>
<td>4941-5039</td>
<td>caaagtagGTAACACCG</td>
<td>ATTTAGAAGGtttatatt</td>
<td>1782</td>
</tr>
<tr>
<td>27</td>
<td>5040-5236</td>
<td>aaaaatttagAAACTCCCG</td>
<td>TTTTTGAAGGttaagcaat</td>
<td>12628</td>
</tr>
</tbody>
</table>

*intronic sequence is indicated in lower-case letters, exonic sequence in upper-case
Table 6.1 Intron-exon boundary sequences of PCM1* (continues)

<table>
<thead>
<tr>
<th>EXON</th>
<th>mRNA position</th>
<th>Splice acceptor</th>
<th>Splice donor</th>
<th>Intron size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>5237-5371</td>
<td>tttttcctcagGAGCACATG</td>
<td>ATATTACAGGtaagagtt</td>
<td>3148</td>
</tr>
<tr>
<td>29</td>
<td>5372-5569</td>
<td>tttaatagGATTCACCTG</td>
<td>GCACAGAGGtaaataac</td>
<td>863</td>
</tr>
<tr>
<td>30</td>
<td>5570-5629</td>
<td>gtttttaagGCTAAAAGG</td>
<td>GAAGACAAGGatgtgtgtct</td>
<td>566</td>
</tr>
<tr>
<td>31</td>
<td>5630-5759</td>
<td>tggtctcagGACAAGGAT</td>
<td>TGTCTATTAgtaagttta</td>
<td>301</td>
</tr>
<tr>
<td>32</td>
<td>5760-5920</td>
<td>tcaacttagATTTGTCTA</td>
<td>GATGAACAGGatattcccg</td>
<td>2132</td>
</tr>
<tr>
<td>33</td>
<td>5921-5992</td>
<td>cccctctcagGTCTACAA</td>
<td>CTAGTAAAGtaagaat</td>
<td>562</td>
</tr>
<tr>
<td>34</td>
<td>5993-6250</td>
<td>tgaatacagATGACAAAGA</td>
<td>AATGAATGATgtatgtac</td>
<td>10528</td>
</tr>
<tr>
<td>35</td>
<td>6251-6343</td>
<td>acattacagGAAGCAGAA</td>
<td>TATTCAGAGgtatggagcc</td>
<td>89</td>
</tr>
<tr>
<td>36</td>
<td>6344-6458</td>
<td>attccaaagGCAGATCTA</td>
<td>AAGAACCTGGtaagagtt</td>
<td>976</td>
</tr>
<tr>
<td>37</td>
<td>6459-3'UTR</td>
<td>gttttttcagAAACGTTGG</td>
<td>AGTATAAGGAGAAGAGAGAAGG</td>
<td>3' UTR</td>
</tr>
</tbody>
</table>

*intronic sequence is indicated in lower-case letters, exonic sequence in upper-case
The data indicate that the PCM1 gene is composed of 37 exons spread over an area of approximately 92 kb of genomic sequence and cover perfectly the whole 6075 nt of the PCM1 open reading frame.

Furthermore, Nucleotide Identify X (NIX, http://menu hgmp mrc ac uk/menu-bin/Nix/Nix.pl) analysis of 20 kb of genomic sequence upstream of the translation initiation codon identified a 868 nt putative CpG island (67%GC) encompassing part of the 5' end of the PCM1 transcript. Blast 2.0 homology search revealed perfect match of part of this CpG island with a CpG island isolated by the Sanger Centre (Cross et al., 1994). Potential promoter sites were also predicted around this CpG island.

A potential consensus polyadenylation signal is present 63 nt downstream of the stop codon.

Exon 37

gttttcagAAACGGTGGGAGCCAGAGTATA
TGAGATGTCTTCTCAGAGGCTCAT
CTAACTCTGTCTTACATACTCAATGCATATATGAAAACAAATACTAAATA
ACATCTGATCTGTATAAAAATGTGAA

PCM1 is represented by several ESTs in the Unigene database (http://www.ncbi.nlm.nih.gov/UniGene) and has been assigned the Unigene identification number Hs.75737. Expression data available in Unigene show that PCM1 is expressed in a wide variety of tissues (adrenal gland, aorta, bone, brain, breast, CNS, colon, eye, germ cell, head and neck, heart, kidney, lung, lymph, muscle, ovary, pancreas, parathyroid, placenta, prostate, smooth muscle, spleen, stomach, testis, thyroid, tonsil, uterus, whole embryo), including brain.
The open reading frame of PCM1 encodes a putative protein of 2024 amino acids. A consensus nucleotide-binding site extending from amino acids 1167-1174 was found. This amino acid stretch of ARILSGKT corresponds to the consensus ATP/GTP binding motif that has been identified in several ATPases, kinases and GTP-binding proteins. In addition, it has the potential of forming coiled-coils.

6.3 ANALYSIS OF PCM1 DNA VARIATION BY AUTOMATED BI-DIRECTIONAL DNA SEQUENCING OF AMPLIFIED EXONS

For this analysis, 19 schizophrenia cases from the sample used in the association study together with 2 cases from each British multiply affected schizophrenia family that showed positive lods on chromosome 8p21-22 (26 affected individuals) and 10 unrelated healthy controls were used. In order to increase the chances of detecting a mutation in linkage disequilibrium with schizophrenia, affected individuals that carried the alleles with higher frequencies in the cases than in the controls in the allelic association study that was described in the previous chapter for the microsatellite markers D8S261, D8S2616 and D8S2615 were chosen.

PCM1 consists of 37 exons spanning about 92 kb of genomic sequence. Each exon is however relatively small ranging in size from about 60 to 360 bp. Marker D8S2616 lies within the intronic sequence of PCM1 between exons 4 and 5, while D8S261 is between exons 19 and 20.

The strategy chosen for mutation and DNA variation screening was automated bi-directional sequencing of PCR amplified exons. The whole exonic sequence as well as more than 100 bp of intronic sequence on either side of each exon were PCR
amplified using M13 tailed primers as described in Materials and Methods (see section 2.3.5.6.2). The PCR product was then sequenced using the Thermo Sequenase kit. Each sequencing reaction contains two different M13 primers labelled with different IR dyes (IRD700 and IRD800) in order to obtain sequence from both DNA strands of the PCR product (bi-directional sequencing). In this way, both strands of the PCR products can be completely sequenced. This approach is particularly useful for mutation detection as both forward and reverse strands are assessed simultaneously for the presence of a variation. In addition, detection of heterozygotes is highly accurate. Furthermore, this sequence-based method can detect the specific location of each mutation. Finally, the use of the same sequencing primers for all PCR products makes bi-directional tailed primer sequencing economically feasible. Therefore, direct sequencing on the LI-COR was preferred to a combination of SSCP and sequencing as a more reliable and faster approach.

LI-COR's dual-dye automated DNA sequencer (Model 4200 IR² System) was used to electrophorese and analyse the forward and reverse sequences of the bi-directional sequencing reactions in parallel. Sequence data were interpreted using the automated base-calling algorithms of LI-COR's Base ImagIR™ software.

Automatic alignment was performed by LI-COR's AlignIR™ software. The software aligns the sequences for the forward and reverse strand from all the samples and produces a consensus. The consensus was generated by a minimum method that considers all base letters and generates an IUPAC ambiguity code for combination of bases at that position. The consensus sequence is shown in a row below the last sample sequence while an annotation row below makes it easy to see ambiguities and mismatches in the consensus sequence. When an ambiguity or a mismatch was encountered at a certain position, windows containing the chromatogram files (.scf
files) for both forward and reverse strands for all the samples could be opened in order to assess the validity of the base calls. Wrong base calls were edited in the AlignIR window by just highlighting the sequence and typing the appropriate letter or the IUPAC ambiguity code.

Heterozygous sites were identified by scanning the assembled sequence traces for: (i) the presence of a drop in fluorescence peak height at a position when compared to the respective peak height for all individuals that are homozygous at the position and (ii) presence of another base (a second peak) that accompanies the drop in fluorescence peak height.

6.3.1 Analysis of exon 4 and its exon/intron boundaries

Exon 4 was screened first for mutations and DNA variation because marker D8S2616 that showed evidence of allelic association with schizophrenia lies within the intronic sequence of PCM1 between exons 4 and 5.

A sequence of 507 bp was PCR amplified using M13 tailed primers. The sequence of the primers has as follows:

5'-GGATAACAATTTCACACAGG-CCAAGTGTCTTTGGTTATCTTCG-3'
M13 reverse primer (-21) forward PCR primer

5'-CACGACGTGTAAAACGACAGTCCGAACATCCTCCTCCT-3'
M13 forward primer (-29) reverse PCR primer

This product included the whole of exon 4 (170 bp) while the rest of the amplified sequence was intronic sequence from both sides of this exon. PCR products were subsequently sequenced bi-directionally. Sequence was obtained for all 10 controls and 45 cases.

Sequence analysis by LI-COR’s AlignIR™ software followed. A single nucleotide polymorphism was identified in the intronic sequence that was 3’ to the
exon. This is an A to G substitution that occurred in both patients and controls. Overall, 6 GG homozygotes, 1 AA homozygote and 3 heterozygotes were identified in the samples from the control individuals while 23 GG homozygotes, 3 AA, homozygotes and 19 heterozygotes were identified in the individuals with schizophrenia. The DNA variation occurs at nucleotide position 80254 in the AB020866 genomic clone. The AlignIR report that contains the novel DNA variant identified in PCM1 using sequence analysis is presented in figure 6.2. The chromatogram traces for the homozygous nucleotide A, homozygous nucleotide G and heterozygous R is presented in figure 6.3.

A candidate single nucleotide polymorphism was identified in the coding sequence at position 1100 of the PCM1 mRNA (NM_006197) and 80123 of the AB020866 clone. It is a G to A substitution with relatively low frequency as only GG homozygotes and a few heterozygotes were observed. More specifically, 7 heterozygotes were observed among the cases while there were no heterozygotes in the controls. The rest of the individuals were GG homozygotes. This is a very interesting polymorphism or mutation as it is likely to produce an amino acid change in the protein sequence from alanine (gct) to threonine (act). More individuals will have to be examined in order to validate this polymorphism. The AlignIR report that contains this DNA variant is presented on figure 6.4. The chromatogram traces for the homozygous nucleotide G and heterozygous R is presented in figure 6.5.
Figure 6.2 LI-COR AlignIR alignment report for a novel DNA variant identified in intronic sequence 3' to exon 4 of PCM1 (position 80254 of the AB020866 clone). The sequence of clone AB020866 was the reference genomic sequence obtained from the Entrez nucleotide database and can be seen at the first line (ex4gnm). The variation occurs at nucleotide position 80254 of the AB020866 clone. The consensus sequence obtained from AlignIR can be seen at the last line (Consensus). Heterozygotes, R, are highlighted in yellow while homozygotes for C are in green and for A in pink.
Consensus ACATTATT TTTGTRGTAT CTCAGTGTTG ATTACTGAT GACATTCTGA TGCCACCCCA
Figure 6.3 Chromatograms of the sequence containing a novel DNA variant identified in intronic sequence 3' to exon 4 of PCM1 (position 80254 of the AB020866 clone).

(a) Trace view of the homozygous nucleotide that was the most common among the individuals sequenced (G peak).

(b) Trace view of the homozygote for the alternative allele (A peak).

(c) Trace view of the heterozygous position containing both G and A peaks.
Figure 6.4 LI-COR AlignIR alignment report for a novel DNA variant identified in exon 4 of PCM1 (position 80123 of the AB020866 clone). The sequence of the cDNA of PCM1 (NM_006197) can be seen at the first line (ex4cDNA) while the clone AB020866 was the reference genomic sequence and can be seen at the second line (ex4gnm). Both sequences were obtained from the Entrez nucleotide database. The variation occurs at nucleotide position 1100 of the cDNA and 80123 of the AB020866 clone. The consensus sequence obtained from AlignIR can be seen at the last line (Consensus). Heterozygotes, $R$, are highlighted in yellow while homozygotes for $G$ are in green.
Figure 6.5 Chromatograms of the sequence containing a novel DNA variant identified in exon 4 of PCM1 (position 80123 of the AB020866 clone).

(a) Trace view of the homozygous nucleotide that was the most common among the individuals sequenced (G peak).

(b) Trace view of the heterozygous position containing both G and A peaks.
6.3.2 Analysis of exon 5 and its exon/intron boundaries

Exon 5 was screened for mutations and DNA variation at this stage because marker D8S2616 that showed evidence of allelic association with schizophrenia lies within the intronic sequence of PCM1 between exons 4 and 5.

A sequence of 595 bp was PCR amplified using M13 tailed primers. The sequence of the primers has as follows:

\[
\begin{align*}
5'\text{-} & \text{GGATAACAATTTCACACAGG-TGAGCCATTGATTATG-3'} \\
& \text{M13 reverse primer (-21) forward PCR primer} \\
5'\text{-} & \text{CACGACGTTGTAAAACGAC-AGTTGTCCCTGCAACCT-3'} \\
& \text{M13 forward primer (-29) reverse PCR primer}
\end{align*}
\]

This product included the whole of exon 5 (180 bp) while the rest of the amplified sequence was intronic sequence from both sides of this exon. PCR products were subsequently sequenced bi-directionally. Sequence was obtained for 5 controls and 42 cases.

Sequence analysis by LI-COR’s AlignIR™ software followed. Two single nucleotide polymorphisms were identified in the intronic sequence that was 5’ to exon 5. One is a C to T substitution occurring 171 bp from the start of exon 5 at position 87366 of the AB020866 genomic clone. In the individuals examined, 1 TT homozygote and 5 heterozygotes were identified among the cases while the rest were CC homozygotes including all 5 controls. The AlignIR report that contains this DNA variant is presented on figure 6.6. The chromatogram traces for the homozygous nucleotide G and heterozygous R is presented in figure 6.7.

The second single nucleotide polymorphism is a G to A substitution occurring 30 bp 5’ to the start of the exon at nucleotide position 87507 of the AB020866 genomic clone. 2 GG homozygotes and 3 heterozygotes were identified in the controls examined while 11 GG homozygotes, 10 AA homozygotes and 21 heterozygotes were
identified in the cases. The AlignIR report that contains this DNA variant is presented on figure 6.8. The chromatogram traces for the homozygous nucleotide G and heterozygous R is presented in figure 6.9.

All 4 single nucleotide polymorphisms identified in and around exons 4 and 5 of the PCM1 were examined for possible alteration in a restriction enzyme cutting site but none seemed to produce or obliterate one.
Figure 6.6 LI-COR AlignIR alignment report for a novel DNA variant identified in intronic sequence 5' to exon 5 of PCM1 (position 87366 of the AB020866 clone). The sequence of clone AB020866 was the reference genomic sequence obtained from the Entrez nucleotide database and can be seen at the first line (ex5gnm). The variation occurs at nucleotide position 87366 of the AB020866 clone. The consensus sequence obtained from AlignIR can be seen at the last line (Consensus). Heterozygotes, Y, are highlighted in yellow while homozygotes for C are in blue and for T in red.
> x14r  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x15f  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x15r  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x16f  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x16r  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x17f  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x17r  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x18f  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x18r  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x19f  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x19r  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x20f  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x20r  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x21f  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x21r  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x22f  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x22r  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x23f  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x23r  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x24f  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x24r  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x25f  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x25r  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x26f  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x26r  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x27f  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
> x27r  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA

Consensus  CAGGCTATAT ATATAATTTCT GTGTCATCAT TCTGTTCTTA CTCCTTTTTTC TTGAAGGTCTA
Figure 6.7 Chromatograms of the sequence containing a novel DNA variant identified in intronic sequence 5' to exon 5 of PCM1 (position 87366 of the AB020866 clone).

(a) Trace view of the homozygous nucleotide that was the most common among the individuals sequenced (C peak).

(b) Trace view of the homozygote for the alternative allele (T peak).

(c) Trace view of the heterozygous position containing both C and T peaks.
Figure 6.8 LI-COR AlignIR alignment report for a novel DNA variant identified in intronic sequence 5' to exon 5 of PCM1 (position 87507 of the AB020866 clone). The sequence of clone AB020866 was the reference genomic sequence obtained from the Entrez nucleotide database and can be seen at the first line (ex5gnm). The variation occurs at nucleotide position 87507 of the AB020866 clone. The consensus sequence obtained from AlignIR can be seen at the last line (Consensus). Heterozygotes, R, are highlighted in yellow while homozygotes for C are in green and for A in pink.
Consensus TTGATAATTG AAACCGCAAC ATGGATTTTA ATAGTCARAT ACAAGTATG TTGGTCTTAT
Figure 6.9 Chromatograms of the sequence containing a novel DNA variant identified in intronic sequence 5' to exon 5 of PCM1 (position 87507 of the AB020866 clone).

(a) Trace view of the homozygous nucleotide that was the most common among the individuals sequenced (G peak).

(b) Trace view of the homozygote for the alternative allele (A peak).

(d) Trace view of the heterozygous position containing both G and A peaks.
6.4 DISCUSSION

The PCM1 gene was chosen as a candidate gene for schizophrenia due to its proximity to the markers showing evidence of allelic association with this disease in our case-control association study on chromosome 8p21-22.

PCM1 is part of the centrosome assembly. Although, possible roles in the maintenance of centrosome integrity and the regulation of the microtubule cytoskeleton have been mentioned its function has not been elucidated. Its protein structure bears similarities to myosins, structural proteins and proteins involved in motility and/or transport (microtubule binding proteins). It is particularly interesting that the gene DISC1, which is disrupted by the translocation breakpoint on chromosome 1 that cosegregates with schizophrenia and other types of psychiatric disorder in a single large Scottish pedigree, has also similarities to structural proteins such as myosins (Millar et al., 2000). As is described by Millar et al. many such structural proteins are implicated in processes such as axon guidance, synaptogenesis, functioning of the synapse and intracellular transport along axons and dendrites. It is possible that PCM1 has a similar role implicating this protein in the development of the nervous system and/or neuronal activity and therefore being involved in the aetiology of schizophrenia.

Furthermore, PCM1 interacts with a brain-specific protein, huntingtin-associated protein 1 (HAP1). HAP1 binds to huntingtin in a glutamine repeat length-dependent manner as well as interacts with cytoskeletal, vesicular and motor proteins. In this way, it acts as an adaptor protein in order to mediate interactions among all these different molecules (Engelender et al., 1997).
The fact that PC M1 interacts with proteins that are directly involved with neuronal function indicates that it may be involved with a wide network of proteins responsible for the normal functioning of neuronal cells.

In order to identify mutations as well as DNA sequence variation the method of fluorescence bi-directional DNA sequencing was employed. DNA sequencing is the most sensitive method for finding DNA polymorphisms and mutations. In addition to its sensitivity, full DNA sequence analysis provides complete knowledge of the type, position and context of every variation, regardless of whether it is a single nucleotide substitution or an insertion/deletion variation.

So far, two of the exons, exon 4 and exon 5 together with at least 100 bp of intronic sequence on either side of each exon have been screened for single nucleotide polymorphisms. These exons were screened first because marker D8S2616 that showed evidence for allelic association with schizophrenia lies in the intronic sequence between these two exons.

In exon 4, a G to A transition was identified at the 3' intronic sequence of this exon in both patients and controls. It seems to be a frequent variant as homozygotes for each of the alternative alleles are present in the individuals screened as well as a number of heterozygotes containing these alternative alleles. In addition, a single nucleotide polymorphism was identified in exon 4. It is again a G to A transition but the relative frequency of this variant seems to be low as it was identified by infrequent heterozygotes among homozygotes. This polymorphism is particularly interesting as all heterozygote individuals identified were among the cases. Furthermore, it appears to produce an amino acid change, from an alanine to a threonine, in the predicted protein sequence of pericentriolar material 1. More individuals will have to be
screened in order to validate this polymorphism and check its potential as a functional variant.

Two single nucleotide polymorphisms were identified in the intronic sequence 5' to exon 5. Both substitutions are transitions, one from a C to a T and the other from a G to an A and are 140 bp apart.

Although only a small part of the genomic sequence of PCM1 has been screened so far, the results are consistent with other studies of human DNA sequence variation, which have identified on average 1 variable site every 217 bp (Halushka et al., 1999). Furthermore, all the nucleotide substitutions identified in this study were transitions, which is in agreement with other studies (Halushka et al., 1999), where A/G substitutions have been shown to have the highest prevalence among functional genes (Gojobori et al., 1982).

The aim of the work described in this chapter was to optimise the conditions for high throughput sequencing and SNP identification for PCM1 in order for it to be studied as a candidate gene for schizophrenia. In the future, our aim is to identify the pattern of sequence variation of PCM1 in the entire coding and a portion of non-coding DNA of this gene. The collection of cSNPs will be used to test the hypothesis that such changes are involved in the aetiology of schizophrenia. In addition, cSNPs will be used together with SNPs in the non-coding regions of PCM1 as markers for linkage disequilibrium studies, serving as surrogates for unrecognised neighbouring functional SNPs.
CHAPTER 7

GENERAL DISCUSSION
Schizophrenia is not inherited as a single-gene disorder and does not show a simple pattern of inheritance. Environmental and genetic factors interact in a complex fashion to produce disease susceptibility and expression but very little is known about any environmental factors. However, many observations indicate the presence of multiple genetic determinants of schizophrenia.

The available evidence from previous linkage studies can be interpreted as follows: A simple homogeneous major gene effect has been disproven. The presence of multiple susceptibility loci has been proven because of the obvious presence of heterogeneity of linkage. However, evidence of heterogeneity fails to confirm or disconfirm major or minor genes or polygenic transmission.

It seems likely that there are indeed genes of major effect operating in a subset of schizophrenics particularly those with high familial loading (Kendler & Diehl, 1993). In others polygenic transmission may be responsible but must also display heterogeneity of linkage.

In addition, there are more than one studies providing evidence for linkage in certain chromosomal regions. One such region is chromosome 8p21-22 (Blouin et al., 1998, Brzustowicz et al., 1999, Pulver et al., 1995). Results from a genome scan as well as a linkage analysis study of 23 multiply affected families from UK and Iceland performed by our group provided some evidence for linkage of this chromosomal location and schizophrenia (Gurling et al., 1999, Gurling et al., 2000, submitted, Kalsi, 1996). In an attempt to provide further support for these findings, a sample of 16 families was collected and analysed. However, this sample did not show any linkage between chromosome 8p21-22 and schizophrenia. The simplest interpretation of these differences is random variation. As other linkage studies have demonstrated (eg. the collaborative replication study of chromosomes 3, 6 and 8),
results are often of lesser magnitude when new data are added to the sample (Levinson et al., 1996).

Another issue in the linkage analysis studies of schizophrenia, and generally on disorders with unknown mode of transmission, is the method of analysis chosen. In this study as well as in our previous studies on chromosome 8p21-22, parametric linkage analysis was used because it is generally a more powerful and robust method than non-parametric methods for detecting linkage despite errors or simplifications in the analysing model, as long as both a dominant and a recessive model are used (Abreu et al., 1999, Dumer et al., 1999, Greenberg et al., 1998).

Another method that can be used as a tool to identify disease susceptibility genes in complex disorders and does not depend on the mode of transmission is the method of allelic association. Many investigators claim that this method has the power to detect genes of moderate or small effect such as those that could be increasing liability to a proportion of cases of schizophrenia (Risch & Merikangas, 1996). Such an approach can be used for candidate gene studies or for fine mapping in order to narrow down susceptibility regions indicated by previous linkage findings. Both of these studies were performed in this project.

In schizophrenia, there are few compelling candidate genes due to ignorance of the pathophysiology of this disorder (Crowe, 1993). A way to increase the prior probability of detecting an association is by studying positional candidates (genes that reside on chromosomal regions where positive linkage has been found). In this study, the two candidate genes studied, PNOC and CHRNA2, map to chromosome 8p21, a region implicated by linkage to schizophrenia (Blouin et al., 1998, Brzustowicz et al., 1999, Levinson et al., 1996, Pulver et al., 1995). The DNA polymorphisms used were anonymous polymorphisms (microsatellite markers) that were close to the
candidate genes in order to detect linkage disequilibrium with the actual disease causing variant(s). However, no significant differences in the allele frequencies of any of the markers studied near PNOC and CHRNA2 were observed in our population of schizophrenic patients in comparison with ethnically matched control individuals.

Examining the candidacy of these two genes was not pursued any further as fine mapping of the 8p21-22 region undertaken simultaneously with the candidate gene studies indicated a very interesting allelic association between schizophrenia and marker D8S261 in a case-control sample of UK ancestry. A region of ~700 kb around D8S261 was further investigated with closely spaced microsatellite markers. Two more markers, D8S2615 and D8S2616, 108 and 22 kb respectively centromeric to D8S261 demonstrated statistically significant allelic association with schizophrenia. A replication of these findings in two case-control samples of Scottish ancestry was attempted. The first case-control Scottish sample did not show any evidence of allelic association with any of the previous associated markers and schizophrenia. The second case-control Scottish sample showed evidence of a significant allelic association only between D8D2616 and schizophrenia. Combination of both the Scottish samples failed to demonstrate significant allelic association between any of the marker loci and the disease. Finally, when the UK and Scottish case-control samples were combined significant allelic association was maintained only between D8S2616 and schizophrenia.

Association studies are prone to false positives, sometimes due to population stratification, but probably more often due to a combination of multiple testing and the low prior odds of finding allelic association (Owen et al., 1997). However, in this study several of the markers tested at greater distances as well as at close proximity to
the markers that were in linkage disequilibrium with schizophrenia were found to have very similar allele frequencies in the cases and the controls. This suggests that population stratification is unlikely to operate in our case-control sample. Further support of these positive association findings comes from pairwise linkage disequilibrium for each of the markers used in this study that provides significant evidence of linkage disequilibrium between the markers that showed positive association with the disease as well as with their neighbouring markers.

The failure to detect significant allelic association with the case-control sample of Scottish origin may be a false negative since larger samples than those used in this study may be required to have adequate power to achieve replication of an association.

The marginal significance and the failure to replicate may also result from the fact that the markers studied are unlikely to be the causative polymorphisms but are rather serving as surrogates for unrecognised neighbouring functional polymorphisms.

In addition, it is possible that conflicting studies contain different balances of subtypes, known or unknown, of the disease, which may make replication difficult.

Furthermore, if the locus in the region studied acts epistatically with another locus elsewhere in the genome, then allelic associations looking at just one of these loci may lack power and reproducibility.

Finally, if there are multiple mutations in a single gene with a similar effect on the susceptibility, then the complex pattern of association may cancel out each other when examined alone, but may be revealed if haplotypes are used. The apparent lack of consistent replication with any single polymorphism might eventually be resolved by finding an association with a different haplotype by combining adjacent markers.
In order to validate these association findings, family based association studies will need to be performed. Although such studies have lower power than case-control allelic association studies due to both intrinsic and smaller sample size problems (usually samples are difficult to collect), if a true association exists then such problems will be offset by higher prior odds. In addition, it is easier to construct haplotypes with these types of studies, which make the use of parental information, than with case-control studies.

The microsatellite markers that were found to be in linkage disequilibrium with schizophrenia, D8S2615, D8S2616 and D8S261 cover an area of 108 kb on the 8p22 region. Two of them, D8S2616 and D8S261 lie within the intronic sequence of the human pericentriolar material 1 gene (PCM1) while the third polymorphism, D8S2615 resides approximately 75 kb upstream of the translation initiation codon. Therefore, this gene is considered a positional candidate for susceptibility to schizophrenia. The study of the DNA sequence variation of this gene is underway. The single nucleotide polymorphisms identified will be used in further association studies. In this way, more information will be obtained from many markers within and around this gene, which will be used to obtain haplotypes. Powerful methods of analysing haplotype data will need to be developed, as handling all this information will be difficult since many possible combinations of haplotypes will be possible. Such new methods will confirm chromosome 8p association and may also to help to identify which of the many gene variants are responsible for disease susceptibility.

Genome-wide association studies have been proposed as a method of identifying genes in complex disorders. These can be direct or indirect (Collins et al., 1997). The direct method searches for allelic association between all the functional variants in every single gene and the disorder while the indirect method tests for
linkage disequilibrium between markers and susceptibility variants. At this point in time both such studies are not feasible due to mainly practical considerations such as not all the genes in the genome or all the functional polymorphisms in these genes have been characterized for the direct approach or the density of the SNPs required to be able to detect linkage disequilibrium has not been elucidated for the indirect approach. Moreover, both these methods require the genotyping of a vast amount of markers in a large number of individuals and hence fast and reliable methods of genotyping are needed as well as statistical methods that can cope with large amounts of data.

Furthermore, chip and array technology already allows the expression of thousands of genes to be measured simultaneously (Ramsay, 1998). Experiments that will assess the differential expression of genes between schizophrenics and controls are likely to help identify susceptibility genes.

Large-scale expression studies and SNP mapping together with the availability of the complete human sequence may form the bridge between positional localization to large regions and the systematic identification of disease genes and their function.

Although gene expression and SNP mapping pose multiple and serious problems if used in genome-wide strategies, the problems become much more manageable when applied to limited chromosomal regions, such as those already defined by genome-wide screens for genetic linkage. Thus, these new technologies will readily form a bridge between genetic linkage and gene identification.

Another approach to candidate gene identification and dissection of the pathophysiological mechanisms in complex disorders is the development of suitable animal models. Although it is difficult to develop meaningful animal models for
psychiatric disorders such as schizophrenia that involve higher cognitive functions; some efforts are being made.

The velo-cardiofacial syndrome (VCFS) locus on chromosome 22q11, known to underlie a specific defect in craniofacial development is a region implicated to harbour a susceptibility gene for schizophrenia by several linkage studies (Blouin et al., 1998, Shaw et al., 1998). Psychotic symptoms are common in VCFS patients (Murphy et al., 1999) and facial dysmorphology is common in schizophrenics (Bassett & Chow, 1999). Transgenic mice deleted for the syntenic region of human chromosome 22, which is the mouse chromosome 16 are being produced (Lindsay et al., 1999). These animals need to be examined closely for neuroanatomical and behavioural phenotypes of possible relevance to schizophrenia. In addition, behavioural studies are being conducted in mice with mutations in genes within this region (Gogos et al., 1998, Gogos et al., 1999).

Genetically altered mice expressing only 5% of the normal levels of the N-methyl-D-aspartate (NMDA) receptor central subunit (NR1) are said to develop 'negative symptoms' commonly associated with schizophrenia, such as withdrawal, poor social functioning and lack of volition and these behaviours are ameliorated by the antipsychotic drugs haloperidol and clozapine (Mohn et al., 1999). The human chromosomal region containing this gene, 9q34.3, showed some evidence for linkage to schizophrenia in a genome scan of African-Americans (Kaufmann et al., 1998).

In the long term, genome studies combined with functional studies in developmental biology and neurobiology will lead to a better understanding of the aetiology and will allow the validation and characterization of existing and novel biochemical pathways pertinent to the inherited susceptibility to schizophrenia and psychiatric disorders in general. A first beginning in our understanding of the
pathogenesis of schizophrenia will require that we initially identify at least one predisposing gene, which will then allow us to begin to identify other interacting proteins and their genes.

In the future, the identification of relevant polymorphisms in individual patients, perhaps using DNA microarrays, might allow the prediction of the risk of developing schizophrenia and hence might allow disease-prevention strategies to be directed more accurately. The presence of polymorphisms might also prove useful in predicting the clinical course of the disease, its severity and response to therapy. Ultimately, the end product of genetic research will result in the development of novel pharmacological treatment strategies for schizophrenia.
REFERENCES


Gurling, H. M. D., Kalsi, G., Blaveri, E., McQuillin, A., Read, T., Murphy, P., Butler, R., Brynjolfsson, J., Sigmundsson, T., Petursson, H. & Curtis, D. (1999). Initial genome wide parametric genetic linkage analysis of schizophrenia and schizophrenia spectrum disorders finds lod scores above 3.00 on four chromosomes at 1Q22-23, 5Q22-35, 8P21-23 and 11Q14-24. a further lod above 3.00 at 4Q21-31 was found within a single family. Molecular Psychiatry 4, S4.


Williams, J., Spurlock, G., Holmans, P., Mant, R., Murphy, K., Jones, L., Cardno, A., Asherson, P., Blackwood, D., Muir, W., Meszaros, K., Aschauer,


APPENDIX 1

Pedigree structures and diagnoses of the 16 English families

Key to pedigree diagrams
- Core definition of schizophrenia
- Spectrum definition of schizophrenia DOSS
- Bulimia Nervosa
F621