# MOLECULAR GENETICS OF THE 1q23.3 SCHIZOPHRENIA SUSCEPTIBILITY LOCUS

**Vinay Puri**

University College London

Molecular Psychiatry Laboratory
Department of Mental Health Sciences
Royal Free and University College London Medical School, The Windeyer Institute of Medical Sciences
46 Cleveland Street, London W1T 4JF

This thesis is submitted for the degree of PhD in Genetics at University College London
August 2007





ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

## *You do it for Love .*

You give of your heart, your love, and your life,
To a grandparent, parent, husband or wife.
You dress and you feed them, you clean up the mess.
For months, maybe years, you give of your best.
At times you can laugh, more often there's tears,
As you watch them decline year after year.
So often you feel like there's nobody there,
No one to talk to, there's no one who cares.
Then late at night, you'll sit and just cry,
"It all seems so hopeless, so why do I try?"
And then comes a voice, so soft and so clear,
You look all around you, but no one is near.
Again comes the voice, as soft as can be,
"You know why you try, just look and you'll see."
"You do it for love, you know that is true."
"This love that you have, will help see you through."
"You're not alone, there's someone who'll share."
"The burden you carry, I'll help you bear."
Then in the darkness, a warmth you can feel,
A soft gentle presence, you know it is real.
As you drop off to sleep, the angels above,
Echo the words...."You do it for love."



For all the carers.

# ACKNOWLEDGEMENTS

The Phd was supervised by Professor Hugh Gurling, Dr Andrew McQuillin provided supervision for laboratory methods. Professor David Curtis provided genetic analysis software and statistical genetic advice.

Patients and normal controls were recruited by a team of NHS doctors in London, Oxford and Reading organised by Professor Gurling.

I am grateful to my family, whose support was greatly appreciated.

# ABSTRACT

Family based linkage studies have confirmed that part of chromosome 1q23.3 contains a susceptibility gene for schizophrenia. This region was investigated by tests of allelic and haplotypic association in order to fine map a specific gene in the 1q23.3 region. Previously published studies claimed that the genes RGS4 and CAPON on 1q23.3 were associated with schizophrenia. For this research thesis multiple markers were genotyped at the RGS4 and CAPON loci in a London based case control sample, no evidence for association was found. Therefore further fine mapping was carried out in the region between the RGS4 and CAPON genes. Allelic and haplotypic associations with schizophrenia were found with three microsatellite and four SNP markers within the serine threonine kinase (UHMK1) gene. A replication study using an Aberdeen based case control sample also found statistically significant evidence of allelic and haplotypic association between UHMK1 and schizophrenia. Re-sequencing of the UHMK1 gene was carried out in those individuals who had inherited alleles and haplotypes associated with schizophrenia. Three genetic variants were found. Genotyping of the whole case control sample showed that these changes were not associated with schizophrenia. The previously reported associations between schizophrenia and RGS4 as well as CAPON could possibly be explained by linkage disequilibrium between UHMK1 and both CAPON and RGS4. Alternatively there could be two or even three susceptibility genes within the 700 Kb region. At present no potential aetiological base pair changes have been detected in any of the three genes. UHMK1 is known to be highly expressed in regions of the brain implicated in schizophrenia and was found to be significantly down regulated in mice treated with the antipsychotic drug Clozapine. Further confirmation of the involvement of this gene in schizophrenia is needed followed by further efforts to detect genetic variation in or next to the gene which has an effect on expression and function of UHMK1.

# TABLE OF CONTENTS

**ACKNOWLEDGEMENTS** ..... 3

**ABSTRACT** ..... 4

**TABLE OF CONTENTS** ..... 5

**TABLE OF TABLES** ..... 9

**TABLE OF FIGURES** ..... 10

**AIMS OF THESIS** ..... 12

**1.0 INTRODUCTION** ..... 13

*1.1 HISTORICAL BACKGROUND* ..... *14*
1.1.1 DIAGNOSIS OF SCHIZOPHRENIA ..... 16
1.1.1.1 SUBTYPES OF SCHIZOPHRENIA ..... 17
1.1.2 TREATMENT ..... 18
1.1.2.1 HOSPITALISATION ..... 19
1.1.2.2 PHARMACOTHERAPY ..... 19
1.1.2.2.1 DOPAMINE RECEPTOR ANTAGONISTS ..... 19
1.1.2.2.2 SEROTONIN-DOPAMINE ATYPICAL ANTIPSYCHOTICS ..... 20

*1.2 AETIOLOGY* ..... *22*
1.2.1 BRAIN MORPHOLOGY OF SCHIZOPHRENIA ..... 22

*1.3 GENERAL GENETICS* ..... *25*
1.3.1 FAMILY STUDIES ..... 26
1.3.2 TWIN STUDIES ..... 27
1.3.3 ADOPTION STUDIES ..... 27

*1.4 GENETIC TRANSMISSION MODELS* ..... *28*
1.4.1 SINGLE MAJOR LOCUS (SML) MODEL ..... 28
1.4.2 POLYGENIC OR OLIGOGENIC MODELS ..... 29
1.4.3 MULTIFACTORIAL-POLYGENIC (MFP) MODEL ..... 29
1.4.4 MIXED MODEL ..... 30
1.4.5 OTHER MODELS ..... 30

*1.5 METHODS OF MAPPING* ..... *32*
1.5.1 GENETIC MARKERS ..... 32
1.5.2 POSITIONAL CLONING ..... 34
1.5.3 LINKAGE ANALYSIS ..... 34
1.5.4 PARAMETRIC ..... 35
1.5.5 NON-PARAMETRIC ..... 36
1.5.6 ALLELIC AND HAPLOTYPIC ASSOCIATION STUDIES ..... 37

*1.6 THE PROBLEM OF HETEROGENEITY* ..... *40*

*1.7 LINKAGE RESULTS IN SCHIZOPHRENIA* ..... *41*

*1.8 ASSOCIATION RESULTS IN SCHIZOPHRENIA* ..... *42*
1.8.1 CHROMOSOME 5 ..... 42
1.8.2 CHROMOSOME 6 ..... 43
1.8.3 CHROMOSOME 8 ..... 43
1.8.4 CHROMOSOME 13 ..... 44

1.8.5 CHROMOSOME 22 ..... 44
1.8.6 CHROMOSOME 1 ..... 45

*1.9 CHROMOSOME 1Q23.3 AS A SCHIZOPHRENIA SUSCEPTIBILITY LOCUS ..... 47*

**2.0 MATERIALS AND METHODS ..... 48**

*2.1 SCREENING, ASSESSMENT AND ASCERTAINMENT OF SCHIZOPHRENIC CASES AND OF SUPER NORMAL CONTROLS ..... 48*

*2.2 CHEMICALS AND REAGENTS USED THROUGH OUT THE EXPERIMENTS ..... 49*
2.2.1 MILLI-Q SYSTEM ..... 49
2.2.2 REAGENTS FOR DNA EXTRACTION:- ..... 50
2.2.3 REAGENTS FOR DNA QUANTIFICATION ..... 52
2.2.4 REAGENTS FOR PCR AMPLIFICATION AND SEQUENCING REACTIONS ..... 52
2.2.5 REAGENTS FOR GEL ELECTROPHORESIS:- ..... 53

*2.3 DNA EXTRACTION FROM WHOLE BLOOD CELLS ..... 54*

*2.4 DNA QUANTIFICATION ..... 57*

*2.5 MICROSATELLITE AND SINGLE NUCLEOTIDE POLYMORPHISM (SNP) SELECTION FOR ASSOCIATION STUDIES ..... 58*
2.5.1 MICROSATELLITE SCREENING AND SELECTION ..... 58
2.5.2 SNP SELECTION FOR ASSOCIATION STUDIES ..... 60
2.5.3 CHOOSING SNPS ..... 61

*2.6 OPTIMISATION OF PRIMERS AND PCR ..... 62*
2.6.1 WHAT IS PCR? ..... 62
2.6.2 MASTERMIX FOR GENERAL OPTIMISATION AND AMPLIFICATION ..... 63
2.6.3 POLYMERASE CHAIN REACTION (PCR) CYCLING CONDITIONS ..... 66

*2.7 GEL ELECTROPHORESIS: AGAROSE AND THE USE OF LICOR DNA SEQUENCERS WITH POLYACRYLAMIDE ..... 68*
2.7.1 AGAROSE GEL ELECTROPHORESIS ..... 69
2.7.2 POLYACRYLAMIDE GELS ..... 69

*2.8 METHODOLOGY OF GENOTYPING AND SEQUENCING ..... 72*
2.8.1 METHOD OF GENOTYPING ..... 72
2.8.2 GENOTYPING WITH SAGA-GT ..... 73
2.8.3 GENOTYPING OF SNPs ..... 74
2.8.3.1 KBIOSCIENCE ..... 77
2.8.3.2 EPOCH BIOSCIENCES IN HOUSE SNP GENOTYPING ..... 80
2.8.4 SEQUENCING ..... 82
2.8.4.1 SANGER-COULSON (CHAIN TERMINATION SEQUENCING METHOD) 82

**3.0 STATISTICAL ANALYSIS ..... 87**

*3.1 CHECKHET ..... 87*

*3.2 HAPMAP ..... 88*
3.2.1 WHAT IS LINKAGE DISEQUILIBRIUM (LD)? ..... 88
3.2.1.1 LEWONTIN'S D' MEASUREMENT OF LD ..... 88
3.2.1.2 $r^2$ MEASUREMENT OF LD ..... 90
3.2.2 HARDY-WEINBERG EQUILIBRIUM. ..... 92
3.2.3 STATISTICAL TESTS FOR ASSOCIATION BETWEEN MARKER ALLELES AND DISEASE ..... 93
3.2.3.1 TESTING FOR ASSOCIATION WITH SNPs USING CHI SQUARE ..... 93
3.2.3.2 CLUMP TEST FOR ASSOCIATION OF MICROSATELLITES ..... 94

3.2.4 STATISTICAL ANALYSIS OF GENOTYPIC INFORMATION USING "GENECOUNTING" ..... 96
3.2.4.1 LDPAIRS ..... 97
3.2.4.2 SCANGROUP ..... 97
3.2.4.3 RUNGC ..... 97
3.2.5 POWER CALCULATIONS ..... 98
3.2.6 ODDS RATIO. ..... 99

**4.0 ATTEMPTED REPLICATION OF REGULATOR OF G-PROTEIN SIGNALLING 4 (RGS4) IN BRITISH CASE CONTROL SAMPLE ..... 101**

*4.1 GENERAL INTRODUCTION ..... 101*

*4.2 REGULATOR OF G-PROTEIN SIGNALLING 4 (RGS4) REPLICATION STUDY ..... 101*
4.2.1 INTRODUCTION ..... 101
4.2.2 METHOD FOR TESTS OF ALLELIC ASSOIATION BETWEEN RGS4 AND SCHIZOPHRENIA ..... 103
4.2.3 RESULTS FOR REPLICATION STUDY OF RGS4 ..... 104
4.2.4 RGS4 DISCUSSION ..... 105
4.2.5 RGS4 CONCLUSION ..... 107

**5.0 ATTEMPTED REPLICATION STUDY OF ASSOCIATION BETWEEN THE CAPON (NOS1AP) NITRIC OXIDE SYNTHASE 1 (NEURONAL) ADAPTOR PROTEIN AND SCHIZOPHRENIA ..... 113**

*5.1 GENERAL INTRODUCTION ..... 113*

*5.2 REPLICATION OF ORIGINAL CAPON STUDY ..... 114*

*5.3 INTRODUCTION ..... 114*

*5.4 GENERAL METHODS AND MATERIALS ..... 115*

*5.5 RESULTS AND DISCUSSION ..... 116*

*5.6 FURTHER PUBLICATIONS ON CAPON ..... 118*

**6.0 FINE MAPPING OF THE UHMK1 GENE ENCODING A SERINE/THREONINE PROTEIN KINASE, AS A NOVEL SCHIZOPHRENIA SUSCEPTIBILITY GENE ..... 124**

*6.1 INTRODUCTION ..... 125*

*6.2 METHOD ..... 125*

*6.3 RESULTS ..... 127*

*6.4 DISCUSSION ..... 130*

**7.0 CONFIRMATION OF THE GENETIC ASSOCIATION BETWEEN THE U2AF HOMOLOGY MOTIF (UHM) KINASE 1 (UHMK1) GENE AND SCHIZOPHRENIA ON CHROMOSOME 1Q23.3 ..... 138**

*7.1 INTRODUCTION ..... 138*

*7.2 BRIEF METHODS AND MATERIALS ..... 139*

*7.3 RESULTS ..... 140*

*7.4 DISCUSSION ..... 141*

*7.5 CONCLUSION ..... 143*

**8.0 FUNCTIONAL SCREENING FOR AETIOLOGICAL CHANGES INCREASING SUSCEPTIBILITY TO SCHIZOPHRENIA ..... 148**

*8.1 RE-SEQUENCING OF UHMK1 – IN SEARCH OF MUTATIONS ..... 148*
8.1.1 RE-SEQUENCING OF THE UHMK1 GENE ..... 149
8.1.1.1 MUTATION MECHANISMS ..... 149
8.1.2 METHOD OF SEQUENCING UHMK1 ..... 151
8.1.2.1 SAMPLE SELECTION FOR SEQUENCING ..... 151
8.1.3 RESULTS OF RE-SEQUENCING UHMK1 EXONS, 5'UTR AND 3'UTR ..... 152
8.1.4 TRANSCRIPTION ELEMENT SEARCH SYSTEM (TESS) ..... 158

**9.0 U2AF HOMOLOGY MOTIF (UHM) KINASE 1 (UHMK1), GENE, FUNCTION AND INTERACTIONS ..... 161**

*9.1 UHMK1 BASIC CHARACTERISTICS ..... 161*

*9.2 BRIEF GENE FUNCTION ..... 161*

*9.3 DETERMINING THE FUNCTIONAL ROLE OF UHMK1 ..... 162*
9.3.1 EXPRESSION OF UHMK1 ..... 162
9.3.2 UHMK1 WITH STATHMIN ..... 166
9.3.3 UHMK1 AND RNA METABOLISM ..... 167
9.3.3.1 UHMK1 INTERACTION WITH SPLICING FACTOR 1 ..... 168
9.3.4 UHMK1 and $p27^{Kip1}$ ..... 168
9.3.5 THE ROLE OF INTERACTION BETWEEN UHMK1, $P27^{KIP1}$ AND STATHMIN 169
9.3.6 UHMK1 INTERACTION WITH PAM ..... 170

*9.4 DESCRIPTION OF PATHWAYS OCCURRING IN Figure 9:5 ..... 173*

*9.5 CONCLUSION ..... 175*

**10.0 THESIS DISCUSSION ..... 176**

*10.1 FUTURE WORK ..... 177*
10.1.1 FURTHER REPLICATIONS ..... 177
10.1.2 RE-SEQUENCING ..... 178
10.1.3 GENOTYPING ..... 178
10.1.4 GENOME SCANS ..... 179
10.1.5 EXPRESSION VECTORS AND KNOCK-OUT/KNOCK-DOWN MOUSE MODELS 180
10.1.6 METHYLATION STATUS AT CpG ..... 180
10.1.7 MRI SCANS OF UHMK1 ASSOCIATED CASES ..... 181
10.1.8 INVESTIGATING EXPRESSION IN SCHIZOPHRENIC PATIENTS. ..... 181
10.1.9 BETTER UNDERSTANDING OF PROTEIN-PROTEIN INTERACTIONS ..... 182
10.1.10 GENE-GENE INTERACTIONS ..... 182
10.1.11 GENE SUSCEPTIBILITY AND SUB-TYPES OF SCHIZOPHRENIA ..... 182

*10.2 CONCLUSION ..... 183*

**11.0 ELECTRONIC REFERENCES: ..... 184**

**12.0 REFERENCES: ..... 185**

**13.0 APPENDIX ..... 198**

*13.1 RAW RESULTS OF LD BETWEEN ALL 55 MARKERS USED ..... 198*

*13.2 PRIMER SEQUENCES OF UHMK1 SEQUENCING ..... 200*

**Appendix 2 ..... 202**

*13.3 Published Papers ..... 202*

13.3.1 Failure to Confirm Genetic Association Between Schizophrenia and Markers on Chromosome 1q23.3 in the Region of the Gene Encoding the Regulator of G-Protein Signaling 4 Protein (*RGS4*) ..... 203
13.3.2 Failure to Confirm Allelic Association Between Markers at the CAPON Gene Locus and Schizophrenia in a British Sample ..... 208
13.3.3 Fine Mapping by Genetic Association Implicates the Chromosome 1q23.3 Gene UHMK1, Encoding a Serine/Threonine Protein Kinase, as a Novel Schizophrenia Susceptibility Gene. 211
13.3.4 Confirmation of Genetic association between the U2AF homology motif (UHM) Kinase 1 (UHMK1) gene and schizophrenia on chromosome 1q23.3 ..... 218

# TABLE OF TABLES

**Table 1:1 Examples of Positive and Negative Symptoms** ..... 16
**Table 1:2 Marker-disease association in a population.** ..... 38
**Table 1:3 The meta-analyses of linkage studies in schizophrenia Lewis et al. (2003) and Badner and Gershon (2002)** ..... 42
**Table 2:1 Primer tail sequences for genotyping and sequencing.** ..... 60
**Table 2:2 Showing the volumes (µl) of reagents in a single 12µl reaction and the common conditions used for primer optimisation and amplification.** ..... 65
**Table 2:3 PCR cycling conditions.** ..... 67
**Table 2:4 Sequencing cycling conditions for Sanger-Coulson chain termination method.** ..... 68
**Table 2:5 LICOR setup for genotyping gels and sequencing gels.** ..... 71
**Table 2:6 Table showing the master mix reagents and volumes required to sequence 16 individuals.** ..... 83
**Table 4:1 Tests of Allelic and Haplotypic Association with Schizophrenia at the RGS4 Gene Locus.** ..... 109
**Table 4:2 Allele frequencies of SNPs 4, 7 and 18 in the UCL case control sample in relation to the previous case control studies of RGS4 and schizophrenia.** ..... 110
**Table 4:3 Results of RGS4 previously reported genetic association studies.** ..... 111
**Table 5:1 Tests of Allelic and Haplotypic Association with Schizophrenia at the CAPON Gene Locus.** ..... 122
**Table 5:2 showing structure of Linkage disequilibrium between markers within CAPON with D', $r^2$ and p-value measurements** ..... 123
**Table 6:1 Microsatellite and SNP marker positions (UCSC March 2006 assembly) on 1q23.3 used for fine mapping of schizophrenia.** ..... 132
**Table 6:2 Tests of allelic association with schizophrenia at the 1q23.3 region** ..... 133
**Table 6:3 Haplotypic association with schizophrenia at the 1q23.3 region with markers within CAPON, SH2D1B, UHMK1 (KIST) and RGS4.** ..... 135
**Table 6:4 LD between all markers including RGS4 and CAPON (based on combined sample).** ..... 136
**Table 7:1 Tests of association between HAPMAP SNPs in HSD17B7 and schizophrenia in the UCL sample.** ..... 144
**Table 7:2 Tests for association of SNPs within UHMK1 in the UCL, Aberdeen and combined samples** ..... 145
**Table 7:3 Haplotypic association with schizophrenia with UHMK1 markers in UCL and Aberdeen samples.** ..... 146
**Table 8:1 the types and frequencies of different mutations leading to human disease (Botstein et al. 2003).** ..... 149
**Table 8:2 Table of DNA variants, position and location.** ..... 153
**Table 8:3 The results of typing detected DNA variants in the entire case control UCL sample.** ..... 157

Table 8:4 Wild type sequence and variant sequence that were analysed by TESS to determine if the binding efficiency of transcription factors were significantly altered..... 159
Table 13:1 Primer sequences used to sequence UHMK1 as described in section 8.1.2 ..... 200

# TABLE OF FIGURES

Figure 1:1 Pharmacological treatment of schizophrenia. ..... 21
Figure 1:2 A comparison a normal brain (left) and a brain and that from a schizophrenic patient (right). ..... 24
Figure 1:3 Lifetime risk of developing schizophrenia Gottesman (1991). ..... 26
Figure 1:4 Sib pair analysis ..... 36
Figure 1:5 Calculation of the 95% confidence interval of an odds ratio ..... 38
Figure 1:6 Calculation of the $\chi 2$ statistic using a 2x2 table ..... 38
Figure 2:2 Shows a tetranucleotide repeat marker that has been analysed using SAGA-GT software. Allele size calls for all the genotypes are shown using coloured "X" symbols. .. 74
Figure 2:3 microsatellite polymorphisms visulised with infra red signal detectin on a Licor polyacrilamide gel . two genetic loci are shown multiplexed on a single gel. ..... 74
Figure 2:4 Showing single base sequencing for 60 individuals on a single infra red wavelength, example of high throughput dectection on a rare SNP in this case a "C". ..... 76
Figure 2:5 KASPar method for genotyping a two allele SNP using allele-specific primers coupled with two Universal Amplifluor primers (Bengra et al. 2002) ..... 78
Figure 2:6 An alternative diagrammatic representation of two-allele SNP detection using allele-specific primers coupled with two Universal Amplifluor primers (Bengra et al. 2002)..... 79
Figure 2:8 An example of the result one would obtain by Sanger-Coulson chain termination sequencing method, with the image obtained from the LiCor DNA sequencer ..... 85
Figure 2:9 Example of a four base insertion-deletion (INDEL), a mutation which is easily seen by eye. ..... 86
Figure 3:1 Using a Punnett's square to derive the Hardy-Weinberg equation. ..... 92
Figure 4:1 Showing the location of SNPs and the GT microsatellite in the region of RGS4 (UCSC BLAT diagram) ..... 103
Figure 5:1 The figure shows the Eastwood et al (2005) interpretation on how expression levels of CAPON can lead to impaired NMDA receptor meditated glutamate neurotransmission, potentially causing schizohrenia. Nitric oxide synthase (NOS)-CAPON complexes result in reduced NMDA-NOS complexes leading to decreased NMDA receptor gated calcium influx and inactive NOS. ..... 119
Figure 6:1 Markers and gene positions in the 1q23.3 region between CAPON and RGS4. ..... 137
Figure 7:1 Linkage disequilibrium between all markers genotyped at CAPON, SH2D1B, UHMK1, HSD17B7 and RGS4 on chromosome 1q23.3, genotyped in the UCL dataset, produced by LocusView. ..... 147
Figure 8:2 Diagram from UCSC genome browser database showing the simple repeats covering the promoter region indicated by the thick black lines (March 2006 assembly). ..... 152
Figure 8:3 sequence of a four base insertion in schizophrenia in amplimer P6 in the promoter region of UHMK1. ..... 154
Figure 8:4 sequencing of an insertion/deletion in the promoter region near exon 1 of UHMK1 in controls. ..... 155
Figure 8:5 Sequence of a single nucleotide polymorphism SNP of A to C in schizophrenia caes. ..... 156
Figure 9:1 The location and features of UHMK1 ..... 161
Figure 9:2 (Figure taken from (Bieche et al. 2003)), showing the expression of UHMK1 across human tissue samples. ..... 163
Figure 9:3 Expression of UHMK1 mRNA in the rat embryo. ..... 164
Figure 9:4 In situ hybridisation with UHMK1 carried out in the mouse brain and highlighted by software, section extracted from the Allen Brain Atlas website. ..... 165

**Figure 13:1 LD plot of all 55 markers used in the UCL sample. Above is Cramer's V. Below are P-values.** ..... **198**
**Figure 13:2 Lod Score Values** ..... **199**
**Figure 13:3 LD Plot of all 55 markers. Above is D' values, below is r2. Distances between markers is on the left.** ..... **199**

# AIMS OF THESIS

- To screen a region of chromosome 1 that had previously been implicated by genetic linkage studies in families. The method employed would use the case control approach to detect evolutionarily determined linkage disequilibrium to fine map a novel schizophrenia susceptibility gene.
- To confirm or fail to replicate association with two genes previously implicated as involved in schizophrenia susceptibility. These genes were Regulator of G-protein Signalling 4 (RGS4) and Nitric Oxide (neuronal) Synthase 1 Adaptor protein (NOS1AP previously known as CAPON) present in the 1q23.3 susceptibility region.
- To re-sequence the exons, the splice site junctions, promoter as well as the 5' and 3' untranslated regions (UTRs) in cases selected for showing association with UHMK1 in order to find any potential aetiological base pair changes.
- To genotype any potential aetiological base pair changes in a large case-control sample in order to cofirm or reject involvement in genetic susceptibility to schizophrenia.
- To replicate the association in UHMK1 in an independent Aberdeen case-control sample.

# 1.0 INTRODUCTION

Schizophrenia is a common disorder with a lifetime prevalence of approximately 1% of the world population (SCZD [MIM 181500]), although it can be up to 5% if schizophrenia spectrum disorders are included (Robert Cancro 2005). Schizophrenia puts a huge burden on healthcare through out the world, for example in the United States it accounts for approximately 2.5% of health care cost (Bromet et al. 1999).

Schizophrenia is a collection of psychotic mental disorders consisting of variable clusters of symptoms. The variability of these symptoms is also associated with considerable prognostic variation. There is no characteristic neuropathology, such as neurofibrillary tangles that are found in Alzheimer disease, but abnormal brain morphology has been reported. Schizophrenia is mainly characterised by a disruption of processes of thought, perception, behaviour and reasoning as well as social and occupational deterioration. People affected with schizophrenia have difficulty differentiating what is generated by their illness and distinguishing it from reality.

In about a third of cases the disorder is chronic, perhaps with short periods with relative improvement but evidence of continuous deterioration. In a further one third there is moderate recovery between episodes with treatment, whilst in a third are treatment responsive and can also spontaneously remit without treatment. The onset can be sudden and insidious, and because of the relative early age of onset, schizophrenia is responsible for very high levels of morbidity. The mean age of onset of the disease is about 23 for males and a little later in females with a mean age of onset being 27.

A study ranking the severity of disabling health conditions in 14 countries, ranked Schizophrenia third overall, just behind quadriplegia and dementia (T Bedirhan Ustun et al. 1999). The detrimental effects of schizophrenia are experienced very

much by the families and carers of schizophrenics in terms of the emotional and financial support required and also the social stigma that it brings.

## 1.1 HISTORICAL BACKGROUND

The term "Schizophrenia" is less than 100 years old; however the disease itself has thought to have accompanied mankind through out its history. The disease has been known about since the ancient times, written documents have been found that identify schizophrenia as far back as ancient Egypt and India, in the second millennium before Christ (Okasha 1999). The evidence included many of the clinical symptoms commonly used to describe schizophrenia today, such as depression, hearing voices and thought disturbances, which were described in detail in the Book of Hearts of the Eber papyrus and also the Indian Ayurveda (Eve C. Johnstone et al. 1998), as well as Shakespeare and the Greeks.

In the past many cultures have attributed psychosis to possession by evil sprits. This idea is still common in parts of Africa and all over the world. Treatments varied from the basic exorcisms to the dangerous and lethal practice of drilling holes in the skull to release the spirits.

The first clinically characterised 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 1851, Falvet first described schizophrenia as a "Folie Circulaire" or cyclical madness. Twenty years later Hecker coined the term "Hebephrenia" or a silly undisciplined mind. In 1868 Kahlbaum documented both catatonic and paranoid disorders. Emil Kraepelin (1856-1926) combined these disorders into a single disease, which he called dementia praecox. There were four subtypes –

1. Simple, marked by slow social decline and withdrawal.
2. Paranoid, defined by persecutory delusions and fear.

3. Hebephrenic, marked by rambling and incoherent speech and incongruous affect.
4. Catatonic, characterized by a severely limited movement and expression.

The subtypes of schizophrenia in most diagnostic systems remain much the same today. Kraepelin believed that dementia praecox was primarily a disease of the brain, and particularly a form of dementia. Kraepelin named the disorder 'dementia praecox' (early dementia) to distinguish it from other forms of dementia (such as Alzheimer's disease) which typically occur relatively late in life. He used this term to emphasise the deterioration in mental abilities and the early age of onset of the disease. The Swiss psychiatrist, Eugen Bleuler, coined the term, "schizophrenia" (Split-mind) in 1911 to describe the fragmented thinking of people with the disorder. He was also the first to describe the symptoms as "positive" or "negative." Bleuler identified specific primary symptoms of schizophrenia to develop his theory about the internal mental schisms of patients. These symptoms included associational disturbances, especially looseness, affective disturbances, autism and ambivalence, summarised as the four A's: associations, affect, autism and ambivalence.

More recently attempts have been made to classify the disease by subdividing the symptoms into groups of "positive" and "negative" effects. Positive symptoms are described as an exaggeration of normal functions and the presents of something that should be absent, where negative symptoms were the loss of normal function possibly due to neuronal loss (Andreasen 1995). A list of positive and negative symptoms can be seen in Table 1:1. However this categorisation is partially unsuccessful because many cases of schizophrenia have an onset characterised by positive symptoms which later change to negative symptoms.

**Table 1:1 Examples of Positive and Negative Symptoms.**

| Positive symptoms | Negative symptoms |
|---|---|
| Hallucinations | Alogia (inability to speak) |
| Delusions | Affective blunting |
| Disorganised Speech/ formal thought disorder | Anhedonia (inability to experience pleasure) |
| Disorganised/bizarre/Catatontic behaviour | Avolition (lack of desire, motivation or persistence) |

Bleuler changed the name to schizophrenia as it was obvious that Kraepelin's name was misleading as the illness was not a dementia (it did not always lead to mental deterioration) and could sometimes occur late as well as early in life.

Since then, both Kraepelin's and Bleuler's categories of schizophrenia based on prominent symptoms have been further refined and continually updated by modern psychiatrists who have used operational criteria to improve the diagnosis of the schizophrenias in a reliable and valid diagnostic method. Nowdays the diagnosis of schizophrenia is as valid and reliable as most medical diagnoses and the operational criteria are both sensitive and specific.

### 1.1.1 DIAGNOSIS OF SCHIZOPHRENIA.

Modern diagnostic schemes were developed out of the earlier nosological theories of Kraepelin and Bleuler. The third edition of Diagnostic and Statistical Manual of mental disorders of the American Psychiatric Association's (DSM-III), diagnostic protocol is based on defining which category the patient fell into, disorganised, catatonic, paranoid, residual and undifferentiated. The first three categories were originally proposed by Kraepelin. These classifications are still employed in the DSM-IV (APA 1994) and also the International Classification of Disease, ICD-10 (WHO 1992). Both diagnostic schemes are comparative or operationally defined, taking a phenomenological and natural history approach to the symptoms and defining schizophrenia based on a clustering of clinical signs, symptoms and prognosis. Particular emphasis is placed on inappropriate affect and mood incongruent psychotic symptoms as a defining feature. 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. Below are (section 1.1.1.1) examples of the classifications of schizophrenia that the diagnostic protocols help to determine.

#### 1.1.1.1 SUBTYPES OF SCHIZOPHRENIA

Disorganized Type (formally known as hebephrenic).

A form of schizophrenia characterised by severe disintegration of personality including disorganised speech and childish mannerisms and bizarre behaviour; they tend to have an earlier age of onset, usually becomes evident during puberty; the most common diagnostic category in mental institutions, it has an unremittng course with poor prgnosis.

Catatonic Type

This subtype features gross psychomotor disturbances, i.e., stupor, negativism, mutism, rigidity, waxy flexibility, excitement, or posturing. They demostrate echolalia (repetition of words or phrases in a nonsensical manner) and echopraxia (mimicking the behaviours of others). Frequent fluctuation between these extreme physical states is common. During catatonic stupor or excitement, patients need careful supervision to prevent them from hurting themselves or others medical care is needed because of malnutrition, exhaustion, hyperpyrexia or self-inflected injury.

Paranoid Type

The paranoid type of schizophrenia is characterised by preoccupation with one or more delusions or more delusions of persecution or grandeur. Common are delusions that there is a conspiracy against the patient. Patients with paranoid schizophrenia are typically tense, suspicious, guarded, reserved and sometimes hostile or aggressive, but some can conduct themselves adequately in social situations, in addition they tend to have they symptoms at an older age than do patients with catatonic or disorganised

schizophrenia. One subtype of paranoid schizophrenia has a good prognosis enabling the patient to return to work with minamal drug treatment.

Undifferentiated Type

This type was originally designed as a catch-all category used when patients do not clearly fit into any one type, or fit into more than one type, this subtype is very common.

Residual Type

This type is commonly used when there is a history of an acute episode of schizophrenia, but at the time of presentation the patient does not manifest any of the associated psychotic or positive symptoms. However there is continued evidence for schizophrenia manifested in either negative symptoms or low grade symptoms. These may include odd behaviour, some abnormalities of thought processes, lack of volition and lack of self care.

Once the individual has been fully accessed and diagnosis has been made, a treatment program would need to be put into place, to try and resolve the symptoms. Once the symptoms are safely under control the subject may be reintegrated back into society with the correct support structure; this will be briefly discussed next.

### 1.1.2 TREATMENT

Although antipsychotic treatments are the mainstay of the treatment for schizophrenia, research has found that psychosocial interventions such as behavioural therapy and cognitive therapy and psychotherapy; can augment the clinical improvement. Most patients with schizophrenia benefit more from the combined use of antipsychotic drugs and psychosocial treatment than from either treatment alone. This is sanctioned in the UK national institute of clinical excellance guidelines which recomends ten cognitive therapy sessions for every in patient admission.

#### 1.1.2.1 HOSPITALISATION

Hospitalisation is induced during relapses. In-patient treatment can help in the stabilisation of the patient with medications, hospitalisation can ensure the patients' safety because of their potential suicidal or homicidal tendencies. Also provide basic needs such as food, clothing and shelter which some patients are unable to maintain for themselves. This should be carried out for their own protection. The role of the hospital and community team is to plan specific treatment strategies for the individual and to aid in helping the individual to rehabilitate and adjust, also to educate the patients and the family about schizophrenia.

The severity of the patients' illness should determine the length of admission rather then the availability of outpatient treatment or community care. After care for the patient should be organised whilst at hospital including day-care centres and home visits by nurses, occupational therapists, social workers and counsellors. Patients can often remain out of hospital and the quality of their daily lives can steadily improve (Benjamin J. Sadock et al. 2003).

#### 1.1.2.2 PHARMACOTHERAPY

Antipsychotic medications were introduced in the early 1950s and have revolutionised the treatment of schizophrenia. Some medications only ameliorate the symptoms and do not seem to alter the out come. Other antipsychotics such as clozapine and risperidone seem to hold back or reverse the disease better than others. The antipsychotic drugs include two major classes: dopamine receptor antagonists and serotonin-dopamine antagonists (SDAs).

##### 1.1.2.2.1 DOPAMINE RECEPTOR ANTAGONISTS

Dopamine receptor antagonists such as chlorpromazine and haloperidol, are effective forms of treatment of schizophrenia, particularly for the positive symptoms (e.g. delusions see Table 1:1) however, these drugs have two major short comings.

Firstly not all respond to treatment enough to return to normality and many still deteriorate. Second the dopamine antagonists were associated with a number of adverse side effects, such as akathisia and parkinsonian like symptoms of rigidity and tremor. The potential serious side effects included tardive dyskinesia and neuroleptic malignant syndrome.

#### 1.1.2.2.2 SEROTONIN-DOPAMINE ATYPICAL ANTIPSYCHOTICS

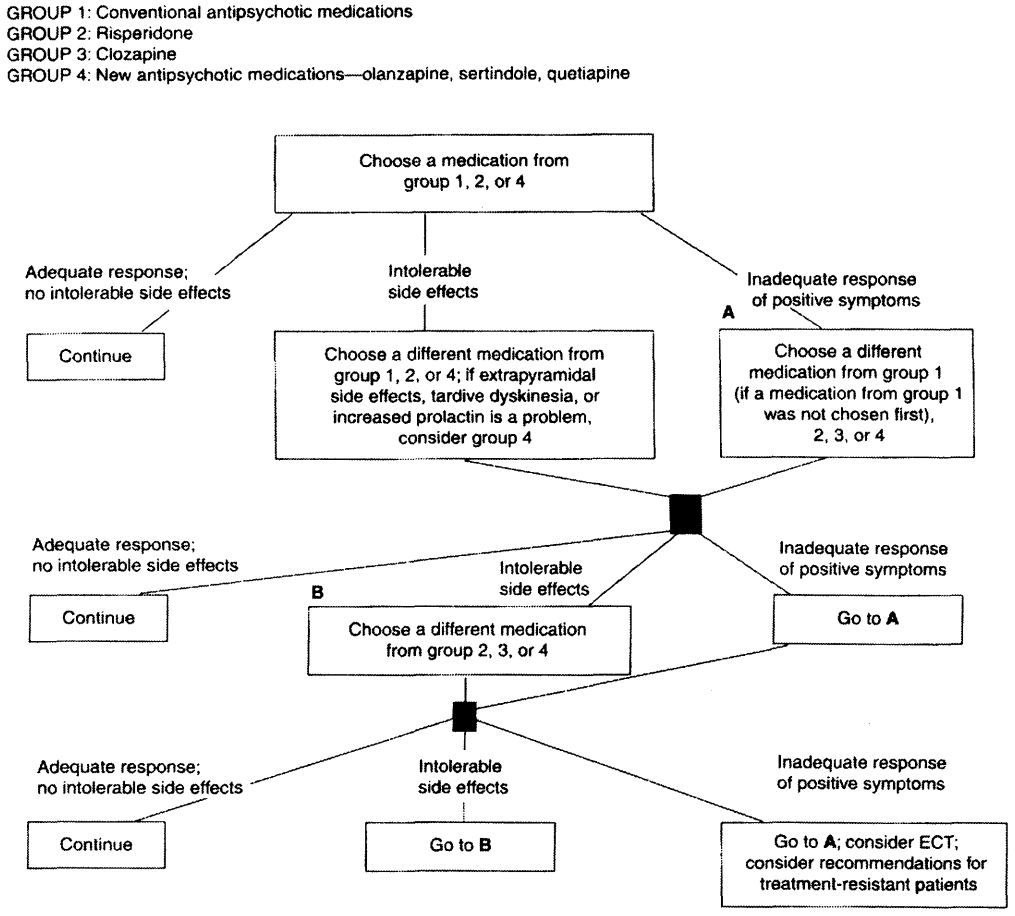
The atypical antipsychotics produce minimal or no extrapyramidal symptoms, interact with different subtypes of dopamine receptors than do the standard antipychotics and affect both the serotonin and glutamate receptors. They are more effective in treating negative symptoms of schizophrenia (e.g. withdrawal see Table 1:1) and also getting rid of delusion. Atypical antipsychotics, include risperidone, clozapine, olanzapine, sertindole, quetiapine and ziprasidone. These drugs have replaced the "typical" dopamine receptor antagonists as the drugs of first choice for treatment of schizophrenia.

Finding the correct drug to use to treat an individual with schizophrenia, is not strictly straightforward and depends on a number of factors, such as the method of medication (tablet or injection), unwanted side effects, and the effectiveness of the drug (see Figure 1:1).

At each point drug treatments are chosen on the basis of:

- Past response
- Side effects
- Patient preference
- Planned route of administration

**Figure 1:1 Pharmacological treatment of schizophrenia.**




practice guideline for the treatment of patients with schizophrenia (Lehman et al. 1997; Benjamin J. Sadock et al. 2003)

## 1.2 AETIOLOGY

The causes of the schizophrenias are not fully understood but some factors including biological, psychological and sociological are known to be relevant in the aetiology. Genes have been consistently shown to be an important biological risk factor for the development of schizophrenia. Further, schizophrenia-like psychoses may develop as a result of demonstrable organic diseases (Johnstone et al. 1987). Head injuries may also result in a schizophrenia-like psychosis and some types of epilepsy are associated with schizophrenia.

Numerous attempts have been made to implicate different environmental risk factors as possible causes for the disease including winter-spring season of birth, perinatal or obstetrical complications, urban birth and rearing, and viral infection. However, none of these factors have showed consistent or large effects on disease's risk (Done et al. 1991). The most consistent risk factor for schizophrenia is a genetic susceptibility.

### 1.2.1 BRAIN MORPHOLOGY OF SCHIZOPHRENIA

Schizophrenia is not only characterised by psychopathological features but also by physiological and anatomical changes. Brain changes, first convincingly demonstrated following the introduction of non-invasive imaging by Johnstone et al (1976) have been consistently found to include the finding of enlarged ventricular spaces and reduced cortical volumes particularly involving temporal lob structures (Shenton et al. 1992). A meta-analysis of regional brain volumes found differences between normal controls and schizophrenic patients. Patients had a smaller cerebral volume, a bilaterally reduced volume of medial temporal lobe structures and a greater ventricular volume (Wright et al. 2000).

More recently, positron emission tomography scans suggested an activation hypofrontality (Liddle et al. 1992). Thompson et al. (2001) have shown accelerated gray matter loss in very early-onset schizophrenia with MRI. Repeated MRI scans over 5 years found that the deficits progressed anteriorly into temporal lobes,

engulfing sensorimotor and dorsolateral pre-frontal cortices and frontal eye fields in some schizophrenics. Kovelman and Scheihel (1984) also found abnormanlly arranged pyramidal cells in the hippocampus (Kovelman et al. 1984).

The histological findings in schizophrenia are less convincing but include aberrantly located or clustered neurons particularly in lamina II of the entorhinal cortex (Jakob et al. 1986) and in the neocortical white matter (Akbarian et al. 1996). These abnormalities perhaps are indicative of an early neurodevelopmental anomaly affecting neuronal migration, survival and connectivity. Although there have been several positive reports of the above findings they should not be accepted without question (Harrison et al. 2005). Other findings include the cell bodies of pyramidal neurons in the hippocampus and in the neocortex are smaller (Zaidel et al. 1997; Pierri et al. 2001).

In summary, the neuropathology of schizophrenia has been suggested to consists of alterations in various neural microcircuitry ranging from the dendritic tree to the cell body and axon to the synaptic terminal (Harrison et al. 2005). Some investigators have tried to bring together all the neuropathological findings by explaining schizophrenia as a disorder of the synapse (McGlashan et al. 2000; Frankle et al. 2003).

The development of the MRI led to more definitive findings of brain abnormalities in schizophrenics. A review was carried out by Shenton et al (2001) on over 193 MRI studies between 1988 and august 2000. This showed multiple brain regions displayed morphological abnormalities in schizophrenics. These include ventricular enlargement in 73% of studies, also a preferential involvement of the medial temporal lobe structures 74%, which include the amygdala, hippocampus, and the parahippocampal gyrus, and neocortical temporal lobe regions (superior temporal gyrus) which forms the limbic system, this is important because it controls emotions. There was additional evidence that the hippocampus contained disorganised neurons (Benjamin J. Sadock et al. 2003). When the gray and white matter of the superior

temporal gyrus was combined 67% of studies reported abnormalities. There was also evidence for frontal lobe abnormalities in 59% of cases and parietal lobe abnormalities in 60%. In addition there was strong evidence for subcortical abnormalities but more equivocal evidence for cerebellar abnormalities in 31% of the studies reviewed (Shenton et al. 2001).

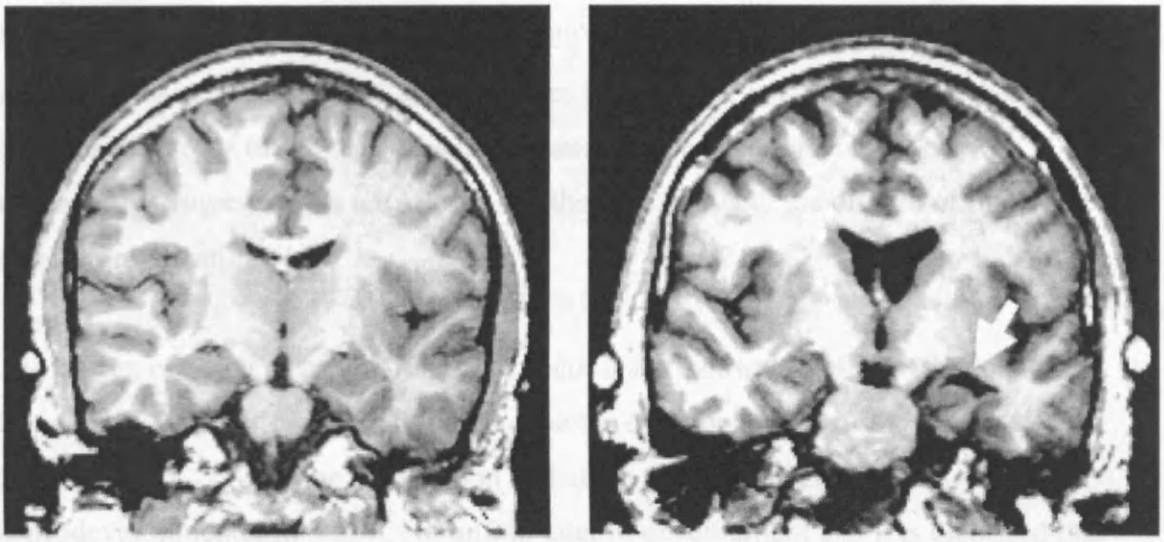


**Figure 1:2 A comparison a normal brain (left) and a brain and that from a schizophrenic patient (right).**

Figure 1:2 shows a "coronal 1.5mm slice of a normal control (left panel) and a schizophrenic patient (right panel). Note the increased Cerebrospinal fluid (CSF) (black) in the left Sylvian fissure in the patient image (right panel, viewers right), as well as the increased CSF in the left temporal horn which surrounds the amygdala (see white arrow) and tissue reduction in the left superior temporal gyrus. The lateral ventricles are also enlarged in the patient image as can be seen by the black CSF regions in the centre of the image. Contrast this with the slice at approximately the same neuroanatomical level for the normal control (left panel)" (Shenton et al. 2001).

The basal ganglia and cerebellum have also been of great theoretical interest for at least two reasons. First, many patients with schizophrenia display odd movements, even in the absence of medication induced movements such as tardive dyskinesia. The odd movements include an awkward gait, facial grimacing, and stereotypes.

Since the basal ganglia and cerebellum are involved in the control of movement, dysfunction in these areas are implicated in the pathophysiology of schizophrenia.

Second, many neurological disorders can have psychosis as an associated symptom. Movement disorders involving the basal ganglia (like Huntington's disease) are the ones most commonly associated with psychosis in affected patients. Furthermore the basal ganglia and cerebellum are reciprocally connected to the frontal lobes. There are inconclusive reports about abnormalities in the basal ganglia; many studies have shown an increase of D2 receptors in the caudate, the putamen and the nucleus accumbens. However, it is not known whether this is due to the disease or due to the effect of medication.

The timing of when these abnormalities occur is not known, but there is evidence that a subset of brain abnormalities may change over the course of the illness (Shenton et al. 2001). One explanation is that the abnormalities are neurodevelopmental in origin, but unfold later in development and it is theorised that excessive pruning of synapses occurs during development and is abnormal in schizophenia. Thus setting the stage for the development of the symptoms contributing to schizophrenia.

The most plausible synthesis of the genetic, neuropathological and epidemiological evidence outlined briefly above is that schizophrenia is a chronogeneic neurodevelopmental disorder (Weinberger 1995) in which genetically timed abnormalities are expressed during development.

## 1.3 GENERAL GENETICS

The evidence for genetic transmission of schizophrenia is compelling. The proportion of variance due to genetic effects of the "heritability" of the disease has been calculated to be about 81% (Sullivan et al. 2003).

### 1.3.1 FAMILY STUDIES

The seminal work of Gottesman (1991) calculated the lifetime morbid risk of schizophrenia corrected for age. Which show that the closer the relation is to the person with schizophrenia the more likely the individual is to develop schizophrenia.

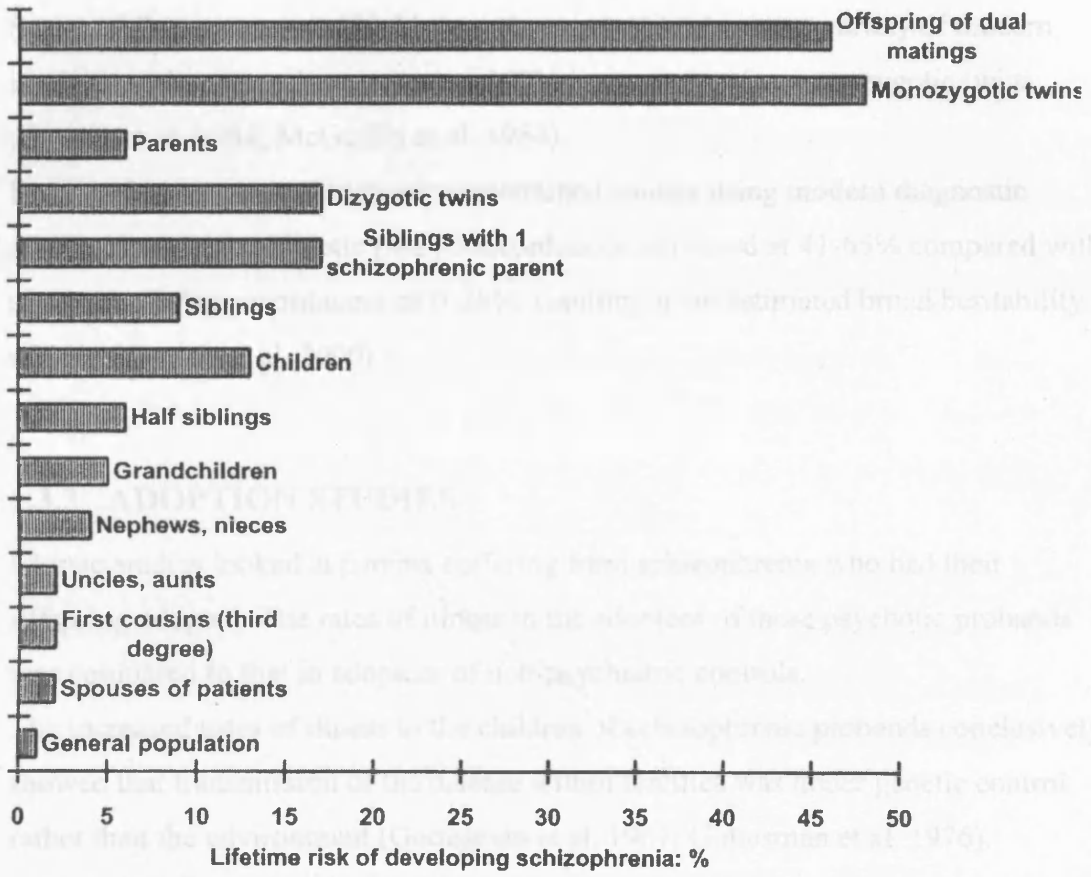



**Figure 1:3 Lifetime risk of developing schizophrenia Gottesman (1991).**

Family studies have provided an important validation of Kraepelin's original formulation of dementia praecox and manic depressive psychosis as separate disorders. These studies have repeatedly shown that biological relatives of parents with schizophrenia have increased risk of schizophrenia and schizophrenia spectrum disorders, whereas biological relatives with parents with major affective disorders have an increased risk for the affective disorder. The separation is not complete but it is supportive of the two disorders as independent disease entities.

### 1.3.2 TWIN STUDIES

Early studies (conducted before world war II in germany) found concordance in monozygotic twins to be in excess of 65%; these investigations were criticised for having potential sources of bias. The figures for dizygotic twins range from about 6% to 20%. Further studies performed more recently have reproduced the finding of greater concordance in monozygotic versus dizygotic twins. The Maudsley Twin Series of Gottesman and Shields have been reassessed using a variety of modern methods and produced an estimate of 67% concordance for monozygotic twins (Farmer et al. 1984; McGuffin et al. 1984).

Finally, five recent systematically ascertained studies using modern diagnostic criteria report monozygotic (MZ) concordances estimated at 41-65% compared with dizygotic (DZ) concordances of 0-28%, resulting in an estimated broad heritability of 85% (Cardno et al. 2000).

### 1.3.3 ADOPTION STUDIES

Classic studies looked at parents suffering from schizophrenia who had their offspring adopted. The rates of illness in the adoptees of these psychotic probands was compared to that in adoptees of non-psychiatric controls.

The increased rates of illness in the children of schizophrenic probands conclusively showed that transmission of the disease within families was under genetic control rather than the environment (Gottesman et al. 1967; Gottesman et al. 1976).

A study looking at the effect of being raised by a psychotic adoptive parent was undertaken by (Wender 1974). In this cohort study, adoptees with schizophrenic biological parents raised by normal adoptive parents were compared with a group of adoptees born to normal parents and adopted by parents who subsequently became schizophrenic. The adoptees who had schizophrenic biological parents had high rates of spectrum disorders, whereas those who had normal biological parents had low rates.

## 1.4 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 (Winokur et al. 1982; McGue et al. 1989a; McGue et al. 1989b) while the purely environmental models cannot adequately explain the results of adoption studies (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 (1986a) and Baron (1986b).

### 1.4.1 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 (Elston et al. 1970). 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 1971; Kidd 1973), while other analyses found it insufficient, (Baron et al. 1982; Risch et al. 1984; McGue et al. 1985).

### 1.4.2 POLYGENIC OR OLIGOGENIC MODELS

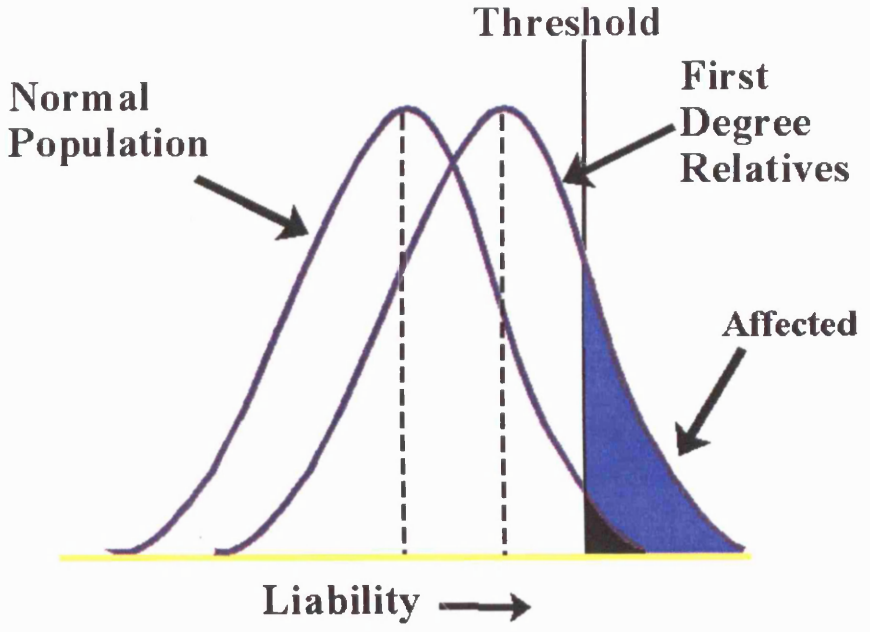
A polygenic (or oligogenic) model is one where two or more disease alleles at two or more distinct loci are needed before a disease is expressed. Such models can be used to explain the recurrence of many common diseases. However, methodology in the area is undeveloped. Further understanding of how two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

Risch (1990a) 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 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.

### 1.4.3 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. The closer the degree of relationship to the affected individual the greater the pre-existing liability the individual has. Since first degree relatives have 1/2 their genes in common with their affected relative, first degree relatives of affected will be 1/2 of the way between the mean for affected and the population mean. A new normal curve with the same variance is then plotted using the mean for first degree

relatives. The threshold does not move, so the overlap of the threshold will give the probability of recurrence to first degree relatives of affected. As can be seen a greater proportion of the first degree relatives are affected compared with the general population, due to prior liability factors they have inherited.




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 et al. 1967; Kidd 1973; Rao et al. 1981; McGue et al. 1985) but it has also been rejected in several other studies (Matthysse et al. 1976; Baron et al. 1982; Tsuang et al. 1983).

### 1.4.4 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).

### 1.4.5 OTHER MODELS

Other models include the two-locus theory 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.
In order to overcome the limitations of segregation analysis, geneticists turned to linkage and/or association (linkage disequilibrium) studies using several different types of genetic markers (see below for full description). Investigators believe these forms of study do have enough power to detect the susceptibility genes underlying a common disorder (Risch et al. 1996).
Finally, it may be that a disease – particularly complex ones – results from interactions between different genes. These interactions can be direct and additive or multiplicative or other interactions could be epistatic in which a major gene effect is modified by another locus, this may be one of the reasons for the difficulties in finding the genes involved in the disease. If the effect of one locus is altered or masked by effects at another locus, power to detect the first locus is likely to be reduced and elucidation of the joint effects at the two loci will be hindered by their interaction. If more than two loci are involved, the situation is likely to be further

complicated by the possibility of complex multi-way interactions among some or all of the contributing loci (Cordell 2002).

## 1.5 METHODS OF MAPPING

### 1.5.1 GENETIC MARKERS

RFLPs (Restriction Fragment Length Polymorphisms) were first used in 1978 (Kan 1978) and became the first genetic markers to be used in a successful study to find the cause of a disease (Huntington's disease) (Gusella et al. 1983). They are based on single base pair change that create or obliterate a cleavage sites for specific restriction enzymes. 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) in linkage studies.
The heterozygosity of the genotyping markers was greatly increased by the identification of VNTRs (Variable Number of Tandem Repeats) in 1987 (Nakamura et al. 1987) and also the minisatellite markers. This class of marker is made of a specific set of consensus sequences that vary between 14 and 100 base pairs in length for a VNTR and even more coplex repeates for minisatellites. They are remarkably polymorphic, with a high heterozygosity rates (usually > 0.6) in the population. However, there are only a small number of VNTRs available and their distribution is rather limited in the genome, often tending to cluster toward human telomeres. Minisatellites tened to be too polymorphic with producing the same fragment sizes but having different sequences.
STRs (Short Tandem Repeats) or microsatellites were initially described by Weber and May (1989) and Litt and 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. Following the success of the dinucleotide repeat markers, other polymorphic markers namely microsatellites were isolated and characterized. They are usually tri- and tetra-nucleotide repeat markers.

Finally, in recent years, attention has been re-focused back 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). Previously SNPs were analysed with restriction enzymes to produce RFLPs. On average, they occur 1 per 300-1000 base pairs (Collins et al. 1997), of which restriction enzymes can not detect all of the SNPs. However now technology has advanced, it enables us to analyse and characterise the SNPs in a variety of different ways. 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 have been developed to provide a powerful tool for human genetic studies. The International HapMap project is one of several such ventures (2003).

### 1.5.2 POSITIONAL CLONING

Positional cloning – involves identifying disease genes through their position in the genome rather than through their function (Collins 1992). The power of positional cloning lies in the fact that it is conducted in the absence of a priori knowledge as to the disease pathogenesis. It is seen by some as the "pure" geneticists approach. The methods adopted for positional cloning are linkage followed by association analyses.

### 1.5.3 LINKAGE ANALYSIS

Linkage analysis relies upon the ability to detect the co-segregation of marker alleles with those of the disease gene. For a marker to be "informative" for a particular meiosis, the individual concerned must be heterozygous at both the marker and disease locus. Therefore, the usefulness of a given genetic marker in linkage analysis or "informativeness" depends on the frequency with which it is heterozygous within a population. Microsatellites gave investigators polymorphic, regularly spaced markers across the whole genome. Linkage studies investigate the departure from independent assortment of a marker and the disorder by estimating the amount of recombination between the marker and the disease. This recombination results from cross-over between the homologous pair of chromosomes during meiosis. Offspring where cross-over has occurred are known as recombinants: offspring where non crossing-over has occurred are known as non-recombinants.

Linkage involves calculating how much the recombination fraction (the number of recombinants divided by the total number of offspring) is significantly different from 0.5, the value expected on the null hypothesis of no linkage.

If two loci are on different chromosomes or are very widely separated on the same chromosome, then independent assortment takes place and the recombination fraction would be 0.5 or ½. That is there would be a 50:50 chance of offspring being either recombinants or non-recombinants.

When two loci are close together (or linked), the assortment is no longer independent and recombination fractions of less than ½ would be seen.

Morton (1955) showed that calculating the lod score represented the most efficient statistic for evaluating pedigrees for linkage.
A lod score (or log of the odds score) is the common log of the likelihood that the recombination fraction has a certain value, θ' divided by the likelihood that θ is ½:

$$\text{Lod}(\theta') = \log_{10}\{\text{likelihood}(\theta = \theta') / \text{likelihood}(\theta = 0.5)\}$$

Traditionally a lod of 3 or more is taken as "significant" evidence for linkage (although this is does equate with the meaning of statistical significance in other contexts).
Linkage analysis conventionally was undertaken using a two-point analysis, where the co-segregation of the disease and a single marker are studied. This method was extended to multi-point linkage analysis (Ott 1991) where multiple markers lying reasonably close together on the same chromosome are studied. This allows for greater precision in the positioning of the disease locus. It also allows the investigator to rule out a particular locus as being responsible for causing the disease.

### 1.5.4 PARAMETRIC

Standard lod score analysis is called parametric because it requires a precise genetic model, detailing the mode of inheritance, gene frequencies and penetrance of each genotype. For non-Mendelian conditions this method has be accused of causing problems but simulation of two locus transmission and uncertain penetrance have shown it to be a robust method (Vieland et al. 2003). The parametric method has a lot more power to detect linkage than the sib-pair non-parametric approach when all the parameters of the model are known and specified correctly. Power is decreased, often severely when the wrong assumptions about the genetic model are made. In most complex traits the true degree of each parameter is not known, this would make the results of the model unreliable.

### 1.5.5 NON-PARAMETRIC

Model free or non-parametric methods of linkage analysis look for alleles or chromosomal segments that are shared by affected individuals (Risch 1990b; Risch 1990c; Risch 1990d).

A common alternative approach is to examine allele-sharing between pairs of affected relatives; an example of this is the sib-pair method. Taking pairs of affected siblings, we would expect that by chance they would share two alleles of a DNA marker 25% of the time, one allele 50% and no alleles 25%. However if the marker is linked to the disease gene then alleles will be shared between affected sib pairs more often than expected. If parents are also genotyped then the inheritance of the marker alleles can be studied directly (identity-by-descent, IBD analysis), but even if the parents are unavailable one can use population allele frequencies to estimate whether increased allele-sharing is occurring (identity-by-state, IBS analysis). The strength of evidence in favour of linkage can be given by a chi-squared statistic or by a maximum likelihood score (MLS), the latter being similar to a lod score.

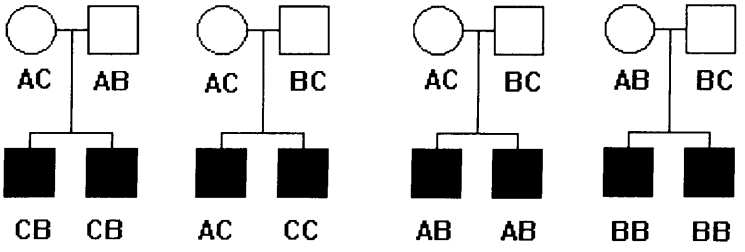



**Figure 1:4 Sib pair analysis**

(courtesy Dave Curtis's website: http://www.smd.qmul.ac.uk/statgen/dcurtis.html). The second sib pair shares one allele IBD, while all the others share both alleles, suggesting a recessive gene may be quite closely linked to the marker.

The lod score method, which requires specification of values for transmission model parameters, is termed parametric, whereas tests which do not involve model specification are termed non-parametric. The latter consist mainly of tests for increased allele-sharing between affected relatives. They are sometimes regarded as

being more appropriate for studying complex diseases (although there is controversy about this), but they are poor at providing a precise location of the disease gene compared to the lod score method.

Linkage studies rely on studying sets of related affected subjects and are capable of detecting a disease gene over a relatively large range. Approximately 300-400 markers are sufficient to carry out a screen of the whole genome. Thus linkage studies provide initial localisations for disease genes without any prior knowledge of the function or chromosomal location.

### 1.5.6 ALLELIC AND HAPLOTYPIC ASSOCIATION STUDIES

Case control and family based genetic association studies represent the final common pathway for all genetic studies regardless of the initial design and are necessary for the ultimate identification of the alleles that confer vulnerability to specific phenotypes.

Association studies play a critical role in the analysis of genetically complex traits in fine mapping a region already implicated by linkage studies.

Association studies look for an allelic association between a marker and a disease allele. Linkage disequilibrium is the reason for observing allelic association between two adjacent markers or between a marker allele and disease allele (Jorde 1995). Linkage disequilibrium between alleles mainly reflects the recombination history in the evolution of that haplotype. Therefore, recently acquired mutations, or those in founder or isolated populations with limited chromosome diversity, are likely to show linkage disequilibrium that might extend over long distances. However, as a result of a number of different factors including regional variability in recombination patterns, recent population admixture and local chromosomal composition linkage disequilibrium can vary significantly within and between populations.

Comparison is made of the frequency of marker phenotypes in a sample of patients and a sample of healthy controls. The statistical analysis is a 2x2 contingency table as shown below. See Table 1:2.

**Table 1:2 Marker-disease association in a population.**

| Marker | n affected | n unaffected |
|---|---|---|
| Present | a | b |
| Absent | c | d |

In case-control studies, the odds ratio can be calculated. The odds of an event happening are the ratio of the probability that it happens to the probability that it does not. The odds ratio is the ratio of the two odds and can be calculated from *ad/bc*. The 95% confidence interval for the log odds ratio is obtained as 1.96 standard errors either side of the estimate. The standard error of the log odds ratio is estimated by the square root of the sum of the reciprocals of the four frequencies as shown in the formula in Figure 1:5.

**Figure 1:5 Calculation of the 95% confidence interval of an odds ratio**

$$\text{SE}(\log \text{OR}) = \sqrt{\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}\right)}$$

With a 2x2 table it is also possible to calculate the $\chi^2$ statistic using the formula shown in Figure 1:6 below where there is one degree of freedom.

**Figure 1:6 Calculation of the χ2 statistic using a 2x2 table**

$$\frac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(b+d)(a+c)}$$

Association may be found because of:-

1) Direct causation - such that having an allele results in susceptibility to the disease. Possession alone may not necessarily lead to the development of the disease, but it increases the likelihood.

2) Population stratification – the population contains several genetically distinct subsets, and both the disease and a particular allele happen to be more frequent in one subset. Lander & Schork (1994) gave an example of the association in the San Francisco Bay area between HLA-A1 and ability to eat with chopsticks. HLA-A1 is more frequent among Chinese than among Caucasians.

3) Type I error – association studies test a large number of markers, even without true effects 5% will be significant at the p=0.05 level.

4) Linkage disequilibrium truly exists between the marker and a disease locus.

5) Natural selection – possession of the marker allele of the disease may lead to increased survival against adverse conditions. Therefore their children are more likely to inherit the disease and the allele associated with it will be more likely to be seen in a population of cases.

## 1.6 THE PROBLEM OF HETEROGENEITY

In the search for genes involved in the causes of complex diseases, investigators have suffered from the problem of heterogeneity at all stages.

In linkage studies, the problem of locus heterogeneity has been the greatest cause of difficulties in identifying regions of importance within families. There may be two or more loci at which mutations result in similar phenotypes. The mutations at the different loci may have different modes of transmission. Charcot-Marie-Tooth, tuberous sclerosis and retinitis pigmentosa all result from a number of distinct mutations of different chromosomes. Charcot-Marie-Tooth exists in an X-linked form and two autosomal dominant forms the loci of which are on different chromosomes. If homogeneity had been assumed and linkage analysis carried out on a collection of families containing a mixture of the two autosomal forms the results could have been misleading and linkage might have been overlooked.

It may be that in a minority of cases schizophrenia are due to a major locus but the majority are polygenic. Indeed Murray (1985) divided the disease into genetic and non-genetic forms; the non-genetic forms being those with no family history, however this does not equate to lack of genetic susceptibility as low penetrance alleles may be present in unaffected family members. Bleuler (1978) found that 60% of patients have no history of schizophrenia in first or second degree relatives but these stuies were not age corrected. Those without a family history maybe due to environmental factors and would therefore be classified as phenocopies. However twin and adoptin studies show that most schizophrenics have a genetic effect.

The adoption of non-parametric linkage analysis was devised to overcome difficulties of being able to define:-1) the model of inheritance or 2) the degree of penetrance. The results of parametric and non-parametric linkage analysis have resulted in several regions being identified where association studies have been used in an effort to narrow the area even further. The linkage studies strongly supported heterogeneity for schizophrenic genes and chromosomes were duplicated.

Locus heterogeneity is now the problem facing SNP association studies. For the power of association studies decreases dramatically as soon as there is more than one susceptibility locus influencing schizophrenia.

The question is, are most susceptibility-alleles ancient common polymorphisms (the so-called Common-Disease/Common-Variant model) or are they a heterogeneous collection of rare recent mutations, like most Mendelian diseases?
Wright et al. (2003) explain that in a review of the genetics of breast cancer, it can be seen that the two familial breast cancer genes, BRCA1 and BRCA2, there are at least 1200 and 1400 different mutations of large effect, each of which are rare – except in the founder populations. By contrast, out of the five or six common SNPs within BRCA1 and BRCA2 coding regions only one has been shown to exert a marginal (1.3-fold) increase in breast cancer risk. Wright et al. (2003) conclude that the common SNPs in the coding region are less interesting "precisely because they have little or no functional effect either on disease or on reproductive fitness."
Reich (2001) looks at population genetics and various examples of common diseases including Alzheimer's disease (APOEε4 ) to conclude that the common disease/common variant model is a good predictor of disease alleles in complex disorders. However studies of hypercholesterolemia favoured multiple rare variants as being aetiological in three genes This view is also supported by the findings of Lohmueller et al. (2003).

## 1.7 LINKAGE RESULTS IN SCHIZOPHRENIA

Thus far there have been two meta-analyses of linkage studies undertaken in schizophrenia (Badner et al. 2002) and (Lewis et al. 2003). The methods used by Lewis et al. are set out by Levinson et al., (2003).
There are two areas where both studies agree, chromosome 8p and 22q.

**Table 1:3 The meta-analyses of linkage studies in schizophrenia Lewis et al. (2003) and Badner and Gershon (2002).**

| Lewis et al. | Badner & Gershon |
|---|---|
| 1q | |
| 2q | |
| 3p | |
| 5q | |
| 6q | |
| 6p | |
| 8p | 8p |
| 10p | |
| 11q | |
| | 13q |
| 14p | |
| 15q | |
| 16q | |
| 18q | |
| 20q | |
| 22q | 22q |

Lewis et al. (2003) most strongly favoured 2q, but also found that the number of loci meeting the aggregate criteria for significance was much greater than the number of loci expected by chance ($p<0.001$). Support was also obtained for regions on chromosomes 1q, 3p, 11q, 6p, 5q, 22q, 8p, 20q and 14p. It is worth mentioning that the region most strongly supported by the evidence of Lewis et al. (2003) on chromosome 2q, is not one that had received strong support previously.

## 1.8 ASSOCIATION RESULTS IN SCHIZOPHRENIA

### 1.8.1 CHROMOSOME 5

There is evidence of a schizophrenia susceptibility gene on chromosome 5 (Schwab et al. 1997; Straub et al. 1997; Devlin et al. 2002). Pimm et al. (2005) had found significant evidence of linkage disequlibrium of the 5' end of the gene Epsin 4 (5q23.1-33.3) and also two single nucleotide polymorphisms (SNPs) within Epsin 4 (a clathrin-associated protein enthoprotin). This association was further supported in a replication by Tang et al. (2006) in a sample of 308 Han Chinese family trios,

however Liou et al. (2006) failed to find independent evidence of association to Epsin 4.

Petryshen et al (2005) and Ikeda et al (2005) have also found some evidence of association between GABAA receptor gene cluster on 5q31-35 being involved in schizophrenia although the results are weak and are suggestive. Epsin 4 is so far the most promising candidate gene on chromosome 5.

### 1.8.2 CHROMOSOME 6

Dysbindin (DTNBP1)

The protein Dysbindin is encoded by a gene on chromosome 6p linkage between 6p and schizophrenia was originally found by Straub et al. (2002) and the region was fine mapped to implicate DTNBP.

The finding has been replicated by the several groups including Williams et al. (2004); Kirov et al. (2004); and Numakawa et al. (2004). Raybould et al. (2005) found no evidence for Dysbindin being involved in bipolar disorder.

Finally, Breen et al. (2006) found a positive association with dysbindin and bipolar affective disorder.

### 1.8.3 CHROMOSOME 8

NRG1 - More than 11 individual studies have reported a positive association of Neuregulin (NRG1) with schizophrenia, the original one being Stefansson et al. (2002); two have reported negative findings. Law et al. (2006) have recently reported an interesting variation in mRNA in post-mortem hippocampi associated with a single SNP within a risk haplotype.

Green et al. (2005), showed association ($p=0.003$) with a core haplotype in the NRG1 locus with bipolar disorder.

PCM1 - Other genes on chromosome 8p have been implicated. One of these (PCM1) has been fine mapped on chromosome 8p. This was found in both UK and USA samples (Gurling et al. 2006) but further confirmations are needed.

Frizzled, a transmembrane receptor essential in neurodevelopment was shown to be associated by Yang et al. (2003); the finding has not been replicated by Wei et al. (2004) and Hashimoto et al. (2005).

PPP3CC (protein phosphatase 3 (formerly 2B), catalytic subunit, gamma isoform (calcineurin A gamma)) also on 8p22, was found to show allelic and haplotypic association with schizophrenia by Gerber et al (2003); a finding not replicated by Kinoshita et al. (2005).

### 1.8.4 CHROMOSOME 13

Chumakov et al. (2002) obtained a strong LD signal from haplotypes within a 5Mb candidate gene on 13q. The signal localised around two overlapping genes, G30 and G72. Analysis of G72 revealed an interaction with D-amino acid oxidase (DAO, DAAO also known as DAOA). Hattori et al. (2003) found both individual SNP markers and haplotypes to be associated with bipolar disorder. Chen et al. (2004) also found an association with bipolar disorder. A large case control study found evidence of association for select SNPs and a haplotype with both schizophrenia and bipolar disorder Schumacher et al. (2004).

### 1.8.5 CHROMOSOME 22

Velocardiofacial and DiGeorge syndromes are associated with learning disability and usually arise from small, relatively frequent deletions on the long arm of chromosome 22 (22q11 deletion syndromes).

The associated phenotype is highly variable with congenital heart defects occurring in approximately three-quarters of patients. Nearly 90% have a 3Mbp deletion encompassing 30 genes. The relative risk of schizophrenia in people with 22q11 deletion syndromes is around 25-30.

Two genes in the region stand out through linkage and association studies as possible candidates for psychiatric outcomes. These are catechol-O-methyltransferase (COMT) and proline dehydrogenase (PRODH). Shifman et al (2002) found a highly significant association between schizophrenia and a COMT

haplotype in a large case-control sample in Ashkenazi Jews ($P=9.5x10^{-8}$), and also a resulting lower expresson of the gene in the human brain . shifman now report errors in their genotyping and meta-analysis of COMT in schizophrenia (Munafo et al. 2005). Li et al (2004) analyzed the PRODH gene in patients with schizophrenia and their families from Sichuan Province in China, comprising 528 family trios and sib pairs. They found association of schizophrenia with 2 haplotypes consisting of the 1945T-C and 1852G-A variants (global p = 0.006) and the 1852G-A and 1766A-G variants (global p = 0.01).

### 1.8.6 CHROMOSOME 1

DISC1 and DISC2.

The presence of a schizophrenia susceptibility locus on chromosome 1q42.1 was first suggested by the observation of linkage between schizophrenia and other psychiatric disorders with a balanced translocation involving chromosomes 1 and 11 [t(1;11)(q42.1;q14.3)] (St Clair et al. 1990; Blackwood et al. 1998; Millar et al. 2000). Investigators in Edinburgh identified an individual with a reciprocal translocation between chromosomes 1 and 11 (+(1;11)). The segregation of this re-arrangement was followed through an extended pedigree. Clinical studies found schizophrenia, major depression and bipolar disorder in translocation carriers but not in members of the family that were non-carriers (Blackwood et al. 2001).
Further analysis showed the chromosome 1 breakpoint directly disrupted two over-lapping genes, termed disrupted in schizophrenia 1 and 2 (DISC1 and DISC2). DISC2 coded on the opposite DNA strand to DISC1, DISC2 is transcribed but not translated and is possibly an RNA gene with some regulatory function (Millar et al. 2001).

Subsequently two linkage studies found significant evidence for linkage at 1q23.3, which is centromeric to the locus implicated by the cytogenetic abnormality (Brzustowicz et al. 2000; Gurling et al. 2001). Another study combined allelic association, cytogenetic and family linkage data on chromosome 1 using heterochromatic C-band variants in the 1q22.1-23 region and found cosegregation of

a 1qH (C-band) variant and Duffy blood group alleles with schizophrenia in a single family (Kosower et al. 1995).

RGS4 maps to one of the two regions showing linkage of chromosome 1q23.3, it was targeted for genetic analysis following a microarray based gene expression study in which decreased RGS4 expression was found in schizophrenic post-mortem brain (Chowdari et al. 2002). Several replications have been reported (Morris et al. 2004) and (Williams et al. 2004). But a meta-analysis was inconclusive (Talkowski et al. 2006).

CAPON (also known NOS1AP) Brzustowicz et al. (2004) first reported an association of novel candidate gene CAPON (now known as NOS1AP) on 1q23.3. This gene was identified in 24 Canadian familial-schizophrenia pedigrees, and encodes a protein nitric oxide synthase 1 (neuronal) adaptor protein which is thought to be involved in neuronal signalling.

Other linkage studies from Finland found significant evidence for linkage to schizophrenia in a position that was telomeric to 1q23.3 but proximal to 1q42.1 (Ekelund et al. 1997; Hovatta et al. 1999; Ekelund et al. 2000). However, the marker that was found to have the highest parametric lod in the earlier Finnish linkage study (Hovatta et al. 1999) was also the marker closest to the more centromeric linkage region at 1q23.3. Two further analyses from Finland in independent family samples found a lod of 3.21 (Ekelund et al. 2000) and then 2.7 at the more distal position of 1q42 near the DISC1 locus (Ekelund et al. 2004). Three further linkage studies found less conclusive but supportive evidence for linkage to schizophrenia at 1q23.3 with a lod of 1.7 at a position 20 cM proximal to DISC1 (Cai et al. 2002). A Taiwanese linkage study reported weakly positive lods at both the RGS4 region at 1q23.3 and at 1q42.1 (Hwu et al. 2003). A linkage study based on a combined UK/USA sample supported linkage at 1q23.3 with a lod of 1.80 with the marker D1S196, which is the same marker that showed a lod of 3.20 in the UK/Icelandic genome scan (Shaw et al. 1998; Gurling et al. 2001). At present the

linkage studies lead one to suspect that there are two schizophrenia susceptibility loci in the 1q region, sixty four megabases apart at 1q23.3 and 1q42.

Recently, Mah et al. (2006) have undertaken a genome-wide scan using over 25,000 SNPs located within approximately 14,000 genes. They have found a marker on 1q32 within a novel candidate gene (PLXNA2).

## 1.9 CHROMOSOME 1Q23.3 AS A SCHIZOPHRENIA SUSCEPTIBILITY LOCUS

The 1q23.3 locus was chosen to be further investigated as it had been implicated in a meta-analysis linkage study with a rank of 7$^{th}$ through weighted analysis (Lewis et al. 2003). The region was also implicated in our own sample (Gurling et al. 2001). It was therefore clearly an interesting site upon which to focus further attention. Two promising genes RGS4 (Chowdari et al. 2002) and CAPON (Brzustowicz et al. 2004) have been implicated by association in the 1q23.3 region in a small sample. To confirm or reject these genes as candidates for schizophrenia susceptibility, a replication study in a British case control sample would be carried out alongside fine mapping a broader area to see if the surrounding regions are associated with schizophrenia or not.

# 2.0 MATERIALS AND METHODS

## 2.1 SCREENING, ASSESSMENT AND ASCERTAINMENT OF SCHIZOPHRENIC CASES AND OF SUPER NORMAL CONTROLS

DNA samples from 450 volunteers with schizophrenia (70% male, 30% female, mean age 44.8) and 450 controls (46% male, 54% female, mean age 36.5) had been collected and were used in our case control sample (figure correct from 2004 to 2006). All subjects were included only if both parents were of English, Irish, Welsh or Scottish descent and if three out of four grandparents were of the same descent. One grandparent was permitted to be of Caucasian European origin but not of Jewish or non-EU ancestry, based on the EU countries before the recent enlargement (2004). The ancestry information was recorded in an Ancestry Questionnaire and it was confirmed from the family histories recorded in medical records. The ancestry selection criteria is not meant to be discriminatory, but was carried out to ensure that the observed genetic differences are disease related and not ancestry related. UK National Health Service (NHS) multicentre and local research ethics committee approval was obtained and all subjects signed an approved consent form after reading an information sheet. All 450 schizophrenic research subjects had previously been diagnosed and assessed by NHS psychiatrists as part of routine clinical diagnosis and treatment. Subjects with short-term drug-induced psychoses, learning disabilities, head injuries schizoaffective bipolar disorder or schizo-mania and other symptomatic psychoses were excluded. Schizophrenic subjects were recruited for research interviewing on the basis of having an International Classification of Diseases version 10 (ICD10) diagnosis of schizophrenia recorded in the medical case notes. The Schizophrenia and Affective Disorders Schedule- Lifetime Schedule (SADS-L) interview (Spitzer et al. 1977) was completed by a research psychiatrist who interviewed all cases and controls. Cases were defined as having schizophrenia with the SADS-L interview according to the "probable" level of the Research

Diagnostic Criteria (RDC). The "supernormal" control subjects were selected based for not having a personal history of any RDC defined mental disorder derived from the SADS-L interview and for not having a family history of schizophrenia, alcoholism or bipolar disorder at interview or recorded in case notes.

Volunteers' blood was collected in 8ml EDTA or citrate plastic tubes and was stored at -80$^0$C until DNA extraction. For patient confidentiality, all samples were anonymomised.

## 2.2 CHEMICALS AND REAGENTS USED THROUGH OUT THE EXPERIMENTS

The following sections contain information on all of the regents used in the experiments for each particular technique.

### 2.2.1 MILLI-Q SYSTEM

The water used to make up all reagents (except PCR master mixes which used ultra pure water (Sigma, W4502) which is known to be DNase and RNase free) was pure water produced by the Milli-Q system, referred from now on as Milli-Q water.

The Milli-Q water system works by taking in mains tap water and filtering it. The first step of the filtration is through the reverse osmosis filter which removes the ions and main contaminants present in the water. This distilled (deionised) water is then passed to a main holding tank which houses an Ultra Violet (UV) light source. The UV light is there to destroy and prevent any micro-organisms from contaminating the water and also break down any organic compounds present by photo-oxidation. The water then moves to another resin filter to remove any remaining water contaminants present, the water finally moves to a small final filter to polish the water before the water is finally dispensed for use.

The Main reagents used through out the experiment are listed below.

### 2.2.2 REAGENTS FOR DNA EXTRACTION:-

- 0.5M ethylene diamine tetraacetate acid (EDTA) (BDH, product number 100935v)

186.1g was added to 800mL of Milli-Q water, and dissolved by adjusting to pH 8.0 using sodium hydroxide pellets (Sigma S5881) the solution was then sterilised by autoclaving at 15 pounds per square inch (15 PSI) at $121^0$C for 15 minutes, using autoclave tape as an indicator that the process has taken place.

- 2M Tris-HCL

121.1 grams of Tris base (BDH, 103156X) was dissolved in 400ml of Milli-Q water, the pH was adjusted to pH 8.0 with concentrated hydrochloric acid (BDH 10125), and the final solution was brought up to 500ml before being sterilised by autoclaving.

- TE Buffer.

This solution is made up by combining 10mM of Tris-HCL with 1mM of EDTA. Hence 5ml of 2M Tris-HCL and 2ml of 0.5M EDTA were mixed and the solution was brought up to 1000ml with Milli-Q water. The solution was then sterilised by autoclaving.

- Proteinase K Buffer

50mM of Tris-HCL, 50mM EDTA and 100mM of NaCl (Sigma S7653) is required. Hence 25ml of 2M Tris-HCL, 100ml of 0.5M EDTA, and 25ml of 4M NaCl were mixed and the solution was brought up to 1000ml with Milli-Q water before sterilising by autoclaving.

- 10X Lysis buffer

5.84g of NaCl was mixed with 37.22g of EDTA in 1000ml of Milli-Q water. Solution was then autoclaved. (EDTA is an important component as it removes magnesium ions that are essential for preserving the structure of cell envelope, and inhibits cellular enzymes that could degrade DNA.)

- 3M Sodium Acetate (pH 5.2)

  246.09g of anhydrous sodium acetate (Sigma S2889) was dissolved in 800ml of Milli-Q water. The pH was adjusted to 5.2 with Glacial acetic acid (BDH, 10001) the volume was brought up to 1000ml and then autoclaved.

- 10% Sodium dodecyl sulfate SDS

  This was made by dissolving 10g of SDS (Sigma L4509) in 100ml of Milli-Q water.

- 20mg/ml Proteinase K enzyme solution (BDH 39509)

- Buffered Phenol (Sigma P4457) mixed with a few flakes of indicator dye 8-hydroxyquinoline (Sigma H6878) this dyes the phenol yellow and becomes brown once oxidised indicating that the phenol has been oxidised and therefore no longer useable.

- Chloroform (BDH, 100776B)

- Isoamyl alcohol (IAA) (Sigma, S5881) also known as 3-methyl-1-butanol

- PVPP (Polyvinylpolypyrrolidine [Sigma P-6755])

- Ethanol (BDH, 10107)

- Virkon (VWR 148-0202), used to deactivate blood and cleanse work surfaces, (10g dissolved in 1 litre of water). Once mixed with waste blood and left to stand for 5 minutes it was then safe to dispose of down the sink.

### 2.2.3 REAGENTS FOR DNA QUANTIFICATION

- Picogreen (molecular probes, Invitrogen P7581)

- Calf thymus DNA 50µg of dehydrated DNA (Sigma D3664)

### 2.2.4 REAGENTS FOR PCR AMPLIFICATION AND SEQUENCING REACTIONS

**PCR amplification:** reagents supplied by bioline in a BioTaq Red polymerase kit (BIO-21041)

- 10X solution $NH_4$ reaction Buffer
- 50mM magnesium chloride
- 25mM solution dNTPs
- Taq polymerase red 1U/µl.

- 5M Betaine (Sigma B2629), 29.29g dissolved in 50ml of ultra pure water (for stock)

- Ultra Pure Water (Sigma, W4502)

- For Geneotyping (M13 Forward IRD 800/700) also added (MWG, Ebersberg Germany).

**Template Clean-up**

microCLEAN (Microzone 2MCL-50), removes mastermix reagents Enzymes and primer dimer to leave clean DNA template (for full method of use see section 2.8.4).

**Sequencing reaction**

Reagents used, SequiTherm EXCEL™ II (Microzone, SE9101LC)

Clean DNA Template.
3.5X SequiTherm Buffer.
SequiTherm Taq Polymerase.
Termination mix di-deoxyNTP.
M13F 700 and M13R 800 10pmol/µl (MWG).
Ultra pure water.

## 2.2.5 REAGENTS FOR GEL ELECTROPHORESIS:-

- Low resolution electrophoresis: 1% agarose (Sigma, A9539) with 1x TBE (National diagnostics EC-860)

- SequaGel XR (National Diagnostics EC-842) polyacrylamide gels for high resolution genotyping and sequencing were created as follows.

10% Ammonium Persulfate (APS) (Sigma, A9164) mixed with a set volume of the buffered acrylamide solution to allow polymerisation (800µL of 10% APS to every 100mL of buffered Sequagel solution). The gel was then caste between the 2mm gap of two glass plates of 25cm length for genotyping or a 44cm length for sequencing. The gel was allowed to set for two hours

- TBE (Tris-Borate-EDTA) (National diagnostics EC-860) electrolyte buffer.

- Fuchsin: (Sigma, S5881)

This loading buffer was used for sequencing and genotyping and was made with 0.05g of Fuchsin and 50ml of Formamide (Sigma, F9037).

- Ethidium bromide (Sigma, E215) used in agarose gels to intercalate with double stranded DNA in order to visualise DNA bands under Ultraviolet light

- Hyperladder IV (Bioline, London, UK, BIO-33029) molecular weight for agarose gels sizes from 100bp to 1000bp increasing in 100bp increments

- MicroSTEP20a (Microzone, 1N700/80020a) molecular weight for polyacrylamide gels, both IRD 700 and 800 molecular weights are used molecular weight markers were as follows:- 70, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 310, 330, 340, 350, 360, 380 and 400 base pairs.

## 2.3 DNA EXTRACTION FROM WHOLE BLOOD CELLS

Genomic DNA was extracted from whole blood samples using a standard cell lysis, proteinase K digestion, phenol/chloroform ethanol precipitation method. It is a two day method where in the first day the leukocytes are isolated and then lysed to release the genomic material, followed by overnight digestion of proteins with proteinase K. The second day involves the isolation and extraction of genomic DNA.

DNA was extracted from fresh blood if it was practical to do so. Otherwise the blood was stored at -80$^0$C.

**Day 1.**

If the blood was frozen, the required blood was removed from cold storage at -80°C, and was allowed to gently thaw out in a water bath at 30-35°C. This was done to prevent the cells from lysing from shock, and preventing the release of damaging enzymes such as DNAase from damaging the genomic DNA.

The fully thawed blood was then transferred to 50ml centrifuge tubes, which were then toped to 50ml with lysis buffer. The tube was inverted several times to ensure mixing. This allowed the lysis buffer to break up the whole red blood cells (erythrocytes) whilst leaving the DNA containing white blood cells (leukocytes) intact.

The lysate was then spun in a balanced centrifuge at 3000rpm (2000g) at 4°C for 15 minutes to pellet and isolate the white blood cells. The supernatant was disposed of into Virkon disinfectant leaving the white blood cell pellet behind.

The pellet was re-suspended in 50ml of lysis buffer to remove any remaining red blood cells and again centrifuged at 3000rpm for 15 minutes. The supernatant was again disposed of in Virkon.

The cleaned white blood cell pellet was re-suspended in 500µl of 10% sodium dodecyl sulfate (SDS) which would break open the cells by disrupting the lipid membrane, hence releasing its contents including the genomic DNA into the solution. In addition 10ml proteinase K buffer was added to provide stable optimum conditions for the proteinase enzyme. To the lysate, 50µl of 20mg/ml proteinase K was mixed in. The presence of the enzyme would break down complex protein allowing for their efficient removal and also to deactivate any damaging enzymes such as DNAase which would harm the targeted DNA. This lysate was incubated in a water bath at 55°C (just below the optimum temperature of proteinase K at 65 °C) on a shaking platform overnight, to allow for the reaction to occur as far as possible.

**Day 2.**

5ml of buffered phenol along with 5ml of chloroform (1:24 Isoamyl alcohol : chloroform) was mixed with the cell lysate. The phenol-chloroform mixture is used to dissolve the lipids and precipitate the proteins this would leave the nucleic acids in an aqueous layer, white coagulated protein would be left at the interphase after the mixture was centrifuged.

The mixture is centrifuged at room temperature for 15 min at 3000rpm which would allow the mixture to separate out into the organic solvent (phenol-chloroform) layer at the bottom of the centrifuge tube (containing lipids), the protein interphase, and on top the aqueous layer containing the nucleic acid. This aqueous layer is transferred into a fresh labelled 50ml centrifuge tube for DNA precipitation, and the organic solvents are disposed of safely according to UCLs' health and safety disposal of hazardous solvents protocol.

As a note, it is notoriously difficult to extract the aqueous layer up to the interphase without disturbing the layer itself. So we used in addition to the phenol-chloroform mix, 1 gram of PVPP (Polyvinylpolypyrrolidine) mixed with 5ml of TE. Once centrifuged with the organic mix, it will hold down the protein interphase along with the organic solvents allowing one to remove the maximum amount of the aqueous layer without contamination.

From this aqueous layer DNA was precipitated by adding 1500µl of 3M Sodium Acetate then adding 30ml of absolute ethanol. The tube was then inverted gently until the DNA precipitates out of solution into a condensed white clump.

The DNA clump was removed with a sterile glass or plastic rod and then washed in 70% ethanol to remove as much sodium acetate as possible (in order not to interfere with any Polmerase Chain Reactions (PCR)). The DNA clump is then transferred into a labelled 1.5ml micro screw tube in 500µl of TE. The DNA samples were

stored away from light at room temperature for about a week. This was to allow the DNA to dissolve into solution before quantification.

## 2.4 DNA QUANTIFICATION

All DNA samples were quantified with Picogreen (Molecular Probes), by fluorimetry. (FLA-3000 scanner (Fuji))

All the extracted DNA samples were quantified to allow for standardisation. This standardisation was done in order to have known amounts of DNA for use in genotyping and sequencing.

The DNA samples were quantified by transferring 2µl of each sample into a fresh labelled 2ml eppendorf tube containing 78µl of TE. This was thoroughly mixed by votexing, form which 10µl of each sample was transferred to a (known recorded location) flat bottomed well plate (96 well plate) containing 90µl of TE. Along side the samples to be quantified was a series of wells containing known concentrations of DNA derived from calf thymus. This was diluted to produce samples at 0, 1, 10, 20, 50, 75 and 100ng/µl. This was done in order to produce standard curve from which the concentration of the samples could be derived from.

The DNA samples were mixed with 100µl of Picogreen (150µl Picogreen dissolved in 30ml of TE) which is a fluorescent dye which specifically binds to double stranded DNA. Once the picogreen is mixed, the plate containing the sample is scanned by fluorimetry. The fluorescence is directly proportional to the quantity of DNA present. From the results, the DNA was then diluted to form two 1.5ml micro-screw tubes with a concentration of 25ng/µl for direct use, the rest is labelled up as stock with its known concentration. All DNA is stored below $5^0$C in a dark place to maintain its quality.

## 2.5 MICROSATELLITE AND SINGLE NUCLEOTIDE POLYMORPHISM (SNP) SELECTION FOR ASSOCIATION STUDIES

Genetic markers such as polymorphic Microsatellites and Single nucleotide polymorphisms (SNPs) are used to detect association between alleles and the disease. Microsatellites are useful for detecting distant linkage disequilibrium; however they are not very common in the genome. Further fine mapping of a region was then carried out by SNPs which tend to be more common (approximately 1 SNP in 300 to 1000 bases). Selection of markers, genotyping and analysis of the data are described below.

### 2.5.1 MICROSATELLITE SCREENING AND SELECTION

Microsatellites are short tandem repeat polymorphisms (STRPs) which consist of a short repetitive sequence typically of one to four bases. The most common type is the $(CA)_n$ dinucteotide repeat, tetra nucleotide repeats are more rare. What is useful about microsatellites is that they are common and dispersed throughout the genome and found even within genes. The size (number of repeats) of these microsatellites has altered though evolutionary history increasing or decreasing their repeat size. These polymorphic microsatellites contain information which enables us to map and screen the genome, to carry out linkage and association studies which, enables one to fine map and identify a candidate gene. The theory is that during the same evolutionary time scale, the mutation event which produced a schizophrenia susceptibility/causative gene, also was the same time point that a nearby, linked microsatellite mutated. This would produce an associated allele linking the gene with that particular allele size, which would co-segregate with the susceptibility/causative gene. In addition other events could have occurred where the mutation occurred on the evolutionary background of a pre-existing polymorphism.

Finding a polymorphic microsatellite marker is fairly hit and miss. The best way to find them is to use bioinformatic databases to find established microsatellites such as The Human Genome Database GDB, and University of California Santa Cruz (UCSC) Genome Browser. In order to find microsatellites in regions where there are not many established microsatellites, bioinformatic databases can be used in order to search for simple repeats to find non established microsatellites. Dinucleotide and tetranucleotide repeats are sort, and a minimum repeat copy size of approximately 10 is used to select the putative microsatellite to increase ones chances for selecting a polymorphic microsatellite.

Once the desired sequences are selected, primers are designed to amplify the sequence. This was carried out using an internet program called Primer3 (developed by MIT Whitehead institute (Rozen et al. 2000)) this program uses algorithms to pick primer sequences with sequence specificity (avoiding repetitive sequence) forward and reverse primers with similar melting temperatures (Tm), and with a low probability of forming hairpin loops, which would inhibit amplification.

With the desired primers picked, M13 tails were added to the 5 prime end of a single primer for genotyping, or a forward and reverse M13 sequence to each primer for sequencing. M13 tails are used to detect the PCR product through fluorescence, by also adding a second M13 complementary primer with a particular wavelength dye to incorporate itself into the product.

The M13 wave length used were M13 forward 800 and M13 forward 700 for genotyping to allow one to genotype more than one marker of similar size at the same time. For sequencing, M13 forward 700 and M13 Reverse 800 were used to capture the sequence of both strands.

**Table 2:1 Primer tail sequences for genotyping and sequencing.**

| | Primer Sequence of tail | |
|---|---|---|
| **Genotyping M13F 700/800** | M13F CACGACGTTGTAAAACGAC | |
| **Sequencing M13F 700, M13R 800.** | M13F CACGACGTTGTAAAACGAC | M13R GGATAACAATTTCACACAGG |

With the M13 tail sequences added to the 5 prime end of the primer. The entire primer sequence was checked to see if a secondary structure such as a hairpin loop would impede amplification. This was checked by using the NetPrimer (Premier Biosoft international) program. A score under 70 is undesirable, and hence the primer with the best score for the M13 tail was chosen. Microsatellites where chosen to target specific loci as described below.

The primers were ordered from MWG biotech (Ebersberg, Germany) at a concentration of 100 picomoles/micolitre.

### 2.5.2 SNP SELECTION FOR ASSOCIATION STUDIES

The problem with microsatellites is that they are not very densely interspersed within the human genome. Therefore fine mapping requires other types of markers in the implicated region. In these instances one can turn to single nucleotide polymorphisms (SNPs) to further fine map the area. These are bi-allelic systems which are far more common than microsatellites and are approximately present in a ratio of one SNP per 300 bases to 1000 bases of sequence. This enable one to fine map a region with greater density and also enable one to select SNPs within a gene of interest. As a result this greatly enhances the resolution of the implicated region and hence helps to implicate a susceptibility gene.

The ability to genotype densely populated SNP regions, has led to a new era of whole genome association studies. Where the whole genome can be mapped using densely populated SNPs enabling one to refine and enhance the resolution of the

association study. When this project was started whole genome association did not exist and typing multiple SNPs was very costly.

### 2.5.3 CHOOSING SNPS

For fine mapping a particular area one can chose SNPs form an online database such as HapMap (HapMap 2003). HapMap is an online consortium consisting of typed SNPs in subjects from different ethnic groups. The aim of this project was to identify and genotype as many common SNPs in the genome as possible. This would enable researchers involved in association studies to pick out the most appropriate SNPs in their specific chromosomal region and specific ethnic population. A number of programs enable the selection of "Tag SNPs". These are SNPs that are likely to carry identical genetic information to that of nearby SNPs in complete linkage disequilibrium. This enables one to choose SNPs which represent other SNPs in the region without duplicating information thereby extracting as much information about the region as possible. The data generated by HapMap can be interpreted by using a piece of software known as HAPLOVIEW (Developed at the Whitehead Instiute) (Barrett et al. 2005). With this software, one is able to visualise the Linkage Disequilibrium (LD) makeup of a chosen region graphically, and judge which SNPs are useful to type, based on their haplotype structure, also by using the tagger function (set to its default setting) which helps to pick up the best tagged SNPs with different information. One is also able to use their D' and $r^2$ values as seen graphically on HAPLOVIEW (D' and $r^2$ in section 3.2.1) in brief D' and $r^2$ values are both different measures of LD and are both measured in a scale between zero (for no LD) to one (for complete LD) if both the values are "one" then the SNP is said to be genetically identical and share the same information.

An alternative method of selecting SNPs is through sequencing of selected individuals in the sample which one thinks are significantly associated with a putative region or susceptibility gene (through statistical haplotypic analysis of previous implicated markers (section 3.2.4.3). One may come across undefined (non-database) polymorphisms within a targeted region, which in turn can then be

genotyped in the entire case-control sample to see if this marker is also associated with the disease. This would help to add further evidence for the involvement of a putative susceptibility gene, help to identify the most likely individuals for targeted sequencing for causative mutations, or even may be implicated in directly influencing the expression of the gene, which may cause a non-synonymous mutation of an exon or disrupt intron splicing or binding of transcription factors.

Genotyping the SNPs in the selected sample can be carried out a number of ways and are further described in section 2.8.3. The methods used to type SNPs in brief are, Sending the higher frequency SNPs to Kbioscience Ltd (Hoddesdon, UK) to genotype on in our sample. This company have developed a cost effective optimised KASPar assay system to type SNPs. Another method used in this thesis is single base sequencing (mini sequencing), where the nucleotide which carries the SNP is solely used in chain termination sequencing (see Figure 2:4), this high throughput method was used to identify and genotype rare SNPs which Kbiosciences Ltd are less effective in formatting because they screen for SNPs and optermise initially in a small number of samples. Often we are aware of SNPs that do not observe because of their low frequency.

## 2.6 OPTIMISATION OF PRIMERS AND PCR

### 2.6.1 WHAT IS PCR?

PCR is a method to amplify a specific target DNA sequences present within a source of complex DNA. To permit specific amplification of the targeted sequence, information about the sequence is required to produce two specific primers that enclose the targeted sequence these primers are generally 15-25 nucleotides long for specific targeting and amplification, the design of primers is mentioned in section 2.5.1.

Lyophilised primers are first rehydrated in TE to produce a concentration of 100 pmoles/µl. Before each use they are votexed and briefly spun down. Newly designed primers need to be optimised to find the optimal master mix concentrations and PCR

conditions to produce the best results and a successful amplification of the targeted region. Optimisation is carried out on a few DNA samples that have been excluded from the study and PCR products are visualised on agarose and polyacrylimide gels. The components of the mastermix and the PCR conditions used (both of which can be vaired) will be discussed next.

### 2.6.2 MASTERMIX FOR GENERAL OPTIMISATION AND AMPLIFICATION

For optimisation of primers, four common conditions were used to amplify the target region (Table 2:2). These conditions can be altered and manipulated to produce the optimal conditions to amplify the target region specifically in conjunction with choosing PCR conditions.

For amplifying and genotyping a microsatellite, the relevant M13 Forward 700 or 800 is added to the master mix, to produce a product at the relevant frequency. Once amplified Fuchsin loading buffer was added (equal volume). The amplified samples were then denatured at 95°C for 3 minuets and then loaded on the polyacrylamide gel and visualised by the LICOR DNA sequencers.

However for sequencing, M13F/R is not added at this stage. The first stage is to amplify the target for sequencing without integrating any of the IRD 700/800 dye. The template is then cleaned up and set up for the next PCR phase, the Sanger-Coulson chain termination sequencing method using the M13F/R primers for PCR amplification (Section 2.8.4.1).

**Figure 2:1 Intergation of the M13 sequence into the amplimer and subsequent amplification of the target sequence with the M13 primers containing the infra red dyes for visualisation by LICOR DNA**

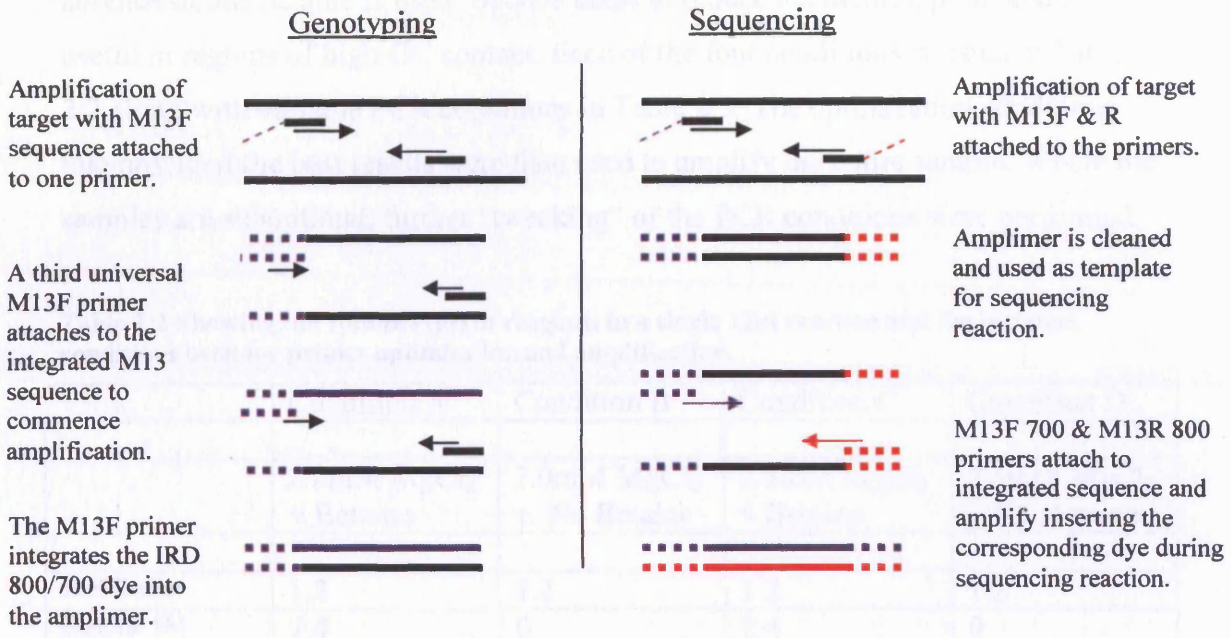



Figure 2.1 shows the difference between the genotyping step where M13F primer is solely used in a single step PCR reaction. The Sequencing first step is to integrate the M13 forward and reverse sequence and amplify the target, before it is cleaned. The template is then amplified with the M13F/R primers in the sequencing reaction integrating the IRD dye and the dNTPs, the reactions are terminated upon integration of ddNTPs.

For PCR optimisation four Master Mix conditions were commonly used, the four Master Mix optimisation conditions used are shown in Table 2:2. The magnesium ion concentration is varied which influences the specificity of the PCR reaction and also acts as an co-enzyme to the Taq polymerase. In addition the presence or absence of 5M betaine is used. Betaine helps to reduce the melting point and is useful in regions of high GC content. Each of the four conditions is listed in Table 2:2 along with variable PCR conditions in Table 2:3. The optimisation conditions that provided the best results were then used to amplify the entire sample. Where the samples are suboptimal, further "tweaking" of the PCR conditions were performed.

**Table 2:2 Showing the volumes (µl) of reagents in a single 12µl reaction and the common conditions used for primer optimisation and amplification.**

| | Condition A | Condition B | Condition C | Condition D |
|---|---|---|---|---|
| | | | | |
| | 2.0mM $MgCl_2$ + Betaine | 2.0mM $MgCl_2$ + No Betaine | 2.5mM $MgCl_2$ + Betaine | 2.5mM $MgCl_2$ + No Betaine |
| | | | | |
| 10X Buffer | 1.2 | 1.2 | 1.2 | 1.2 |
| Betaine 5M | 2.4 | 0 | 2.4 | 0 |
| 50mM $MgCl_2$ | 0.48 | 0.48 | 0.6 | 0.6 |
| 25mM dNTP | 0.1 | 0.1 | 0.1 | 0.1 |
| F primer 100pmol/µl | 0.036 | 0.036 | 0.036 | 0.036 |
| R primer 100pmol/µl | 0.036 | 0.036 | 0.036 | 0.036 |
| M13 800 100pmol/µl | | | | |
| M13 700 100pmol/µl | 0.036 | 0.036 | 0.036 | 0.036 |
| Taq 1U/µl | 0.19 | 0.19 | 0.19 | 0.19 |
| DNA 25ng/µl | 2.0 | 2.0 | 2.0 | 2.0 |
| Water | 5.52 | 7.92 | 5.39 | 7.79 |
| | | | | |
| Total Volume (µl) | 12.0 | 12.0 | 12.0 | 12.0 |

### 2.6.3 POLYMERASE CHAIN REACTION (PCR) CYCLING CONDITIONS

Once the primers are added to the master mix containing all the required reagents for DNA amplification (such as dNTPs, Taq, magnesium and buffer and also the genomic template) the PCR can begin.

First the DNA template is denatured for an extended period of time to separate the double stranded helix and allow for the template to become linear.

The PCR runs in a three step cycle of:-

1. Denaturation, which typically occurs at 93-95$^0$C
2. Reannealing where the primers bind specifically to their complementary sequence, the annealing temperature usually depends on the melting temperature ($T_m$) of the expected duplex this tends to be approximately 5$^0$C below the expected $T_m$.
3. DNA synthesis, extension of the complementary strand initiated by the annealed primer which occurs at 70-75$^0$C.

This runs for approximately 25-35 cycles and is finished by a final extended hold at 72$^0$C to complete all synthesis and extensions of the targeted region. After about 25 cycles approximately $2^{26}$ copies of the targeted region would be made.

The PCR thermal cycler was an MWG-HT Primus 96.
Usually only three standard PCR programs are required to amplify a specific region successfully. They are Standard 55$^0$C, Standard 60$^0$C and Touch Down. These programs can be tweaked to adjust the annealing temperature and by adding or removing one or two cycles to obtain a clean specific amplification of the target region. The cycles are set out in Table 2:3. Also set out in Table 2:4 are the sequencing PCR conditions (sequitherm) with consist of a longer extension time to allow for full extension length variations before termination.

**Table 2:3 PCR cycling conditions.**

| **Standard 55$^0$C** | **Standard 60$^0$C** | **Touch Down** |
|---|---|---|
| Lid Heated to 105$^0$C | Lid Heated to 105$^0$C | Lid Heated to 105$^0$C |
| Products are denatured at | Products are denatured at | Products are denatured at |
| 94$^0$C for 5 minutes | 94$^0$C for 5 minutes | 94$^0$C for 5 minutes |
| | | |
| 35 cycles of:- | 35 cycles of:- | 3 cycles of:- |
| 94$^0$C – 30 seconds | 94$^0$C – 30 seconds | 94$^0$C – 30 seconds |
| 55$^0$C – 30 seconds | 60$^0$C – 30 seconds | 63$^0$C – 30 seconds |
| 72$^0$C – 30 seconds | 72$^0$C – 30 seconds | 72$^0$C – 30 seconds |
| | | |
| Hold at 72$^0$C – 10 minutes | Hold at 72$^0$C – 10 minutes | 3 cycles of |
| Store at 4$^0$C | Store at 4$^0$C | 94$^0$C – 30 seconds |
| | | 60$^0$C – 30 seconds |
| | | 72$^0$C – 30 seconds |
| | | |
| | | 3 cycles of |
| | | 94$^0$C – 30 seconds |
| | | 57$^0$C – 30 seconds |
| | | 72$^0$C – 30 seconds |
| | | |
| | | 3 cycles of |
| | | 94$^0$C – 30 seconds |
| | | 54$^0$C – 30 seconds |
| | | 72$^0$C – 30 seconds |
| | | |
| | | 3 cycles of |
| | | 94$^0$C – 30 seconds |
| | | 51$^0$C – 30 seconds |
| | | 72$^0$C – 30 seconds |
| | | |
| | | 20 cycles of |
| | | 94$^0$C – 30 seconds |
| | | 48$^0$C – 30 seconds |
| | | 72$^0$C – 30 seconds |
| | | |
| | | Hold at 72$^0$C – 10 minutes |
| | | Store at 4$^0$C |

The Sequencing PCR Protocol which consists of a longer extension time than standard PCR conditions to enable random chain termination is shown below in Table 2:4.

**Table 2:4 Sequencing cycling conditions for Sanger-Coulson chain termination method.**

**Sequatherm PCR cycle:**

Lid Heated to $105^0$C
Products are denatured at $94^0$C for 5 minutes

30 cycles of:-
$92^0$C – 30 seconds
$60^0$C – 30 seconds
$70^0$C – 1 minute
End

Store at $4^0$C.

## 2.7 GEL ELECTROPHORESIS: AGAROSE AND THE USE OF LICOR DNA SEQUENCERS WITH POLYACRYLAMIDE

Once amplification has occurred through PCR, the size of the amplimer can be seen by gel electrophoresis. The size of the amplimer can be deduced by running a molecular weight alongside it.

Gel electrophoresis is used to separate DNA fragments according to size. The gel always runs in a negative to positive direction with the DNA loaded near the negative cathode. This is because DNA carries a negative charge due to the phosphodiester backbone and will hence migrate to the positive anode. As the DNA migrates through the gel matrix the fragments separate out according to size with the small fragments migrating fast and hence further due to less impediment of its migration through the gels matrix. The more concentrated the gel matrix, and the further the product has to migrate, the finer the resolution of fragment separation becomes.

### 2.7.1 AGAROSE GEL ELECTROPHORESIS

For low resolution of the amplimer, to judge the size of the overall band(s), one would use agarose gel electrophoresis. A 1% agarose gel is created by dissolving agarose powder in 1x TBE (the electrolyte used). A molten solution is created by heating the solution in a microwave until the agarose powder is fully dissolved. Ethidium bromide is then added and mixed with the cooling molten agarose to make a final concentration of 0.5µl/ml. The molten agarose is poured into a sealed casting plate and a comb is inserted to cast wells within the agarose. The gel is then allowed to cool to set.

Once set, the comb is removed to form the wells and the gel is immersed in an electrolyte (1x TBE). The samples are loaded into the wells with a molecular ladder running along side (such as hyperladder IV, Bioline). The molecular ladder will enable one to judge the size of the amplimer along with acting as a positive control. The voltage is held at a set value (the higher the voltage the faster the migration) for a length of time where one deems the product has migrated far enough through the gel to be properly resolved by eye. As the product migrates through the gel the ethidium bromide intercalates with the DNA, will allows the bands to fluoresce when the gel is viewed (safely) under Ultra Violet light.

### 2.7.2 POLYACRYLAMIDE GELS

Polyacrylamide gels are used to resolve DNA fragments to less than one base difference. The gels are prepared as in section 2.2.5, using amonium persulfate (APS) to polymerise the acrylamide (800µl of 10% APS to every 100ml of buffered Sequa gel solution). The gel apparatus consists of two glass plates separated by using two 2mm thick spacers. The gel is then poured into the gap. A single well is prepared at the top of the gel. Once the gel has set (in approximately 2 hours), the surplus polyacrylamide is cleaned from the plates, a sharktooth comb is inserted to form the wells that allow sample loading.

The gel is loaded vertically into the LI-COR DNA sequencer, that contains a trough at the top of the gel that holds 500ml of electrolyte (1x TBE) and stands in a second trough containing another 500ml of the electrolyte. This allows a current to be applied through the gel for electrophoresis to occur. A scanning laser at the bottom of the gel continually scans for both 700nm and 800nm wavelengths. The signal is then converted into a graphical image to view. The electrophoresis condition varies depending on the size of the gel and whether it is for genotyping or sequencing as seen in Table 2:5.

The LICOR set up is usually as follows:-

**Table 2:5 LICOR setup for genotyping gels and sequencing gels.**

| Settings | Genotyping | Sequencing |
|---|---|---|
| Plate size (cm) | 25 | 40 |
| Volts (V) | 1200 | 2000 |
| Current (Ma) | 25 | 35 |
| Power (W) | 30 | 45 |
| Temperature ($^0$C) | 45 | 45 |
| Signal filter | 3 | 3 |
| Signal speed | 4 | 4 |

The LICOR used through out this project was the "LICOR DNA Sequencer Long Reader 4200".

The samples are first mixed with an equal volume of Fuchsin loading buffer and denatured at 95$^0$C for approximately three minutes to allow the DNA strands to become single stranded. The gel is preheated to 45$^0$C (to help prevent the formation of secondary structure and hence allow the linear DNA run in proportion to their length) before loading the gel.

## 2.8 METHODOLOGY OF GENOTYPING AND SEQUENCING

Once the polymorphic markers are amplified for genotyping or the implicated region is amplified for sequencing, the next stage is to analyse the gel. The methodology for the process of genotyping, sequencing and the analysis of the data will be discussed in this section.

### 2.8.1 METHOD OF GENOTYPING

As discussed in section 2.5 and 2.6 polymorphic markers were chosen and amplified in our schizophrenic sample and matched control sample (each marker in a 96 well plate format), along with an M13 forward primer (700 or 800 wavelengths) to allow more than one marker to be loaded on a gel simultaneously. The products were then pooled and mixed with an equal volume of loading buffer (Fuchsin). The sample was loaded (on to a vertical polyacryamide gel) using a 64 well loading comb, in addition eight molecular weight (MW) markers were run along the sample. The MW were evenly spaced to enable differentiation between the polymorphic marker alleles and also to show what the gel is doing to enable one to call the alleles correctly and give the correct size. The gel was analysed by the LICOR by using a dual laser scanning laser.

In addition to analysing the entire case control sample, a minimum of 17% of all the samples are repeated in separate microtitre plates as a cross check, to ensure the alleles were called correctly and the common allele size was kept constant between gels to prevent any gel shifts. This helped to improve genotyping accuracy and reliability.

### 2.8.2 GENOTYPING WITH SAGA-GT

SAGA-GT is a software program (developed by LICOR) which is used to allow one to efficiently genotype gel images produced by the LICOR. It has the ability to load up the sample identifications, detect the lanes which they are in, and the location of each marker with the aid of the molecular weights. The program detects the uniformity of the gel with aid of the spaced molecular weights and produces a "desmile line" to allow correct calling of allele sizes. To some extent SAGA-GT is able to automatically call the genotypes, however each individual genotype was checked by eye and corrected if deemed necessary. As a result the software helps to reduce human error.

Both the entire sample of cases and controls were genotyped with the aid of SAGA-GT. The genotypes were then read by a second individual blind to their diagnosis. Any discrepancies between callers resulted in the individual in being re-amplified for the marker and re-genotyped, until the discrepancy was resolved.

The genotyping data was then extracted for each marker from SAGA-GT and entered into a database in Microsoft Access. This database contains a list of all individual identification codes and has a program to enable one to check for conflicts between repeat genotyped individuals. Conflicts were rechecked on SAGA-GT to determine the reason for the conflict and amended.

The genotype data was then ready for statistical analysis to check for allelic association to schizophrenia using an optimised $Chi^2$ method "CLUMP". CLUMP employs an empirical Monte Carlo test for significance and which does not require further correction for multiple alleles (Sham et al. 1995). Further statistical analyses are described in section 3.0.

**Figure 2:2 Shows a tetranucleotide repeat marker that has been analysed using SAGA-GT software. Allele size calls for all the genotypes are shown using coloured "X" symbols.**

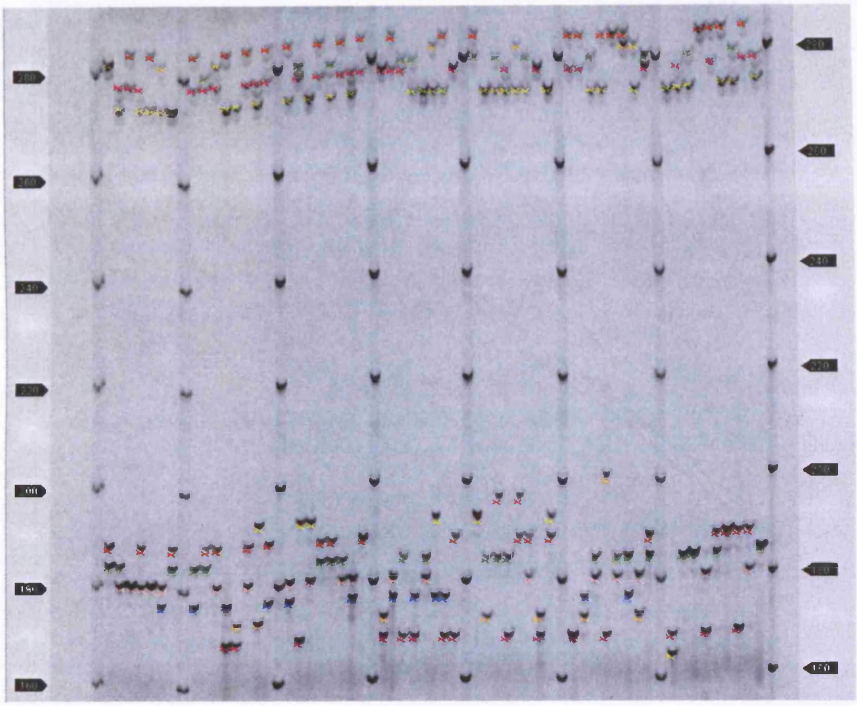


**Figure 2:3 microsatellite polymorphisms visulised with infra red signal detectin on a Licor polyacrilamide gel . two genetic loci are shown multiplexed on a single gel.**

### 2.8.3 GENOTYPING OF SNPs

SNPs were genotyped in several ways as shown before. Common established SNPs were sent off to Kbiosciences Ltd (Hoddesdon, UK) to be genotyped. They have developed and optimised a cost effective way of genotyping SNP using an adapted Fluoresence resonance energy transfer (FRET) reporter system, which they named

KASPar. DNA had been sent to the company with 17% of the samples duplicated to test for reproducibility and accuracy of the genotypes, the results were sent back for analysis.

Occasionally, especially for markers that amplified poorly, we had to use a high throughput sequencing method. It is a method devised to type rare low frequency SNPs found through sequencing in our entire case control population that Kbiosciences were unable to produce an assay for. It was carried out by sequencing a single base in which the SNP would occur, alongside a known full sequence positive control. This allowed one to sequence 120 individuals on a single polyacryamide gel using both IRD 700 & 800 wavelengths.

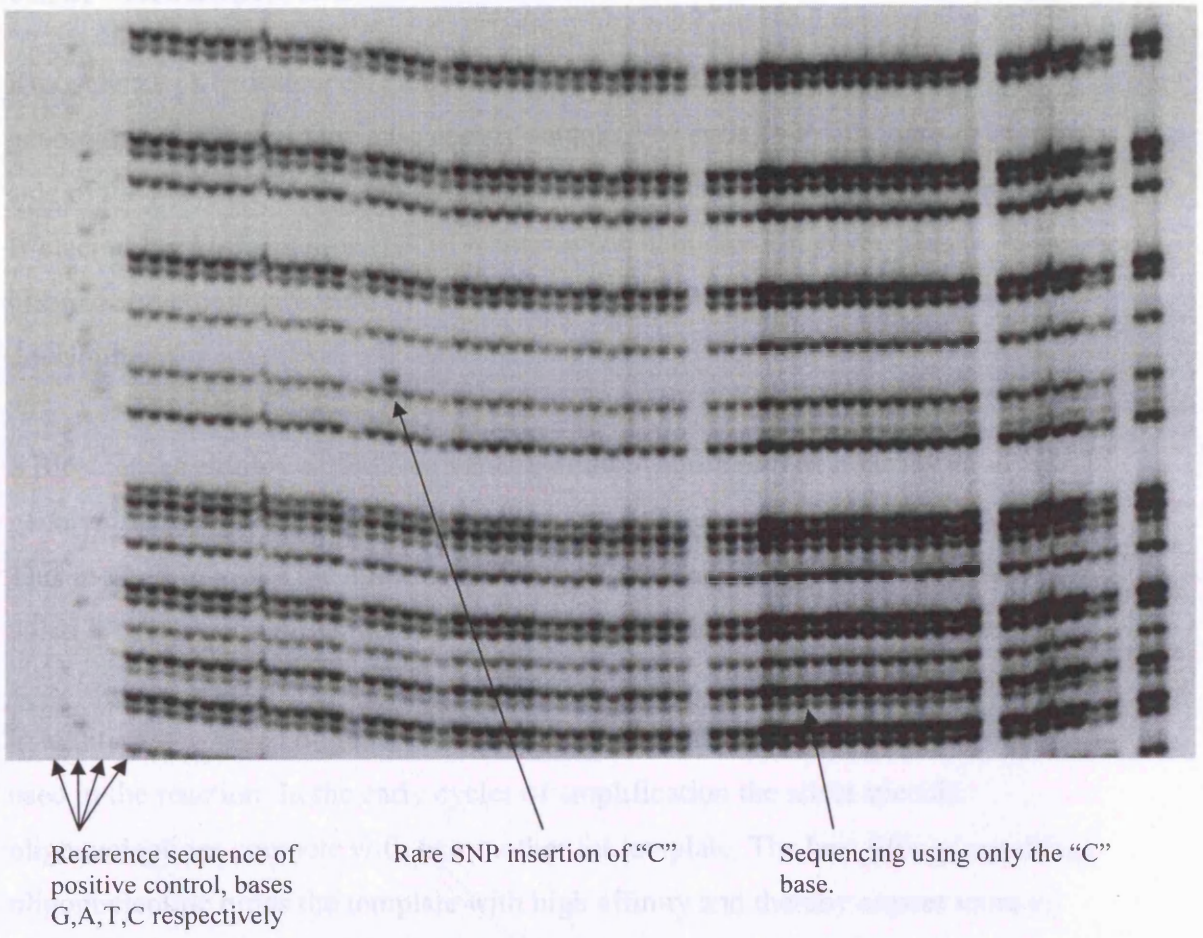



**Figure 2:4 Showing single base sequencing for 60 individuals on a single infra red wavelength, example of high throughput dectection on a rare SNP in this case a "C".**

In all methods, 17% of the sample was duplicated to check for reproducibility and accuracy of the genotypes. The genotype data was collated in the required format and checked for Hardy-Weinberg equilibrium, and the chi square between cases and controls was assessed.

Further statistical analysis was carried out on the data for linkage disequilibrium between markers and and for haplotypic association with schizophrenia as described in section 3.2.4.

### 2.8.3.1 KBIOSCIENCE

Kbioscience (KBiosciences, Hoddesdon, UK) were employed to undertake genotyping of SNPs in the case control sample. For each SNP fifty base pairs either side of the polymorphism were selected from dbSNP in The National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/). This information along with the SNP name was sent to KBiosciences for SNP assay development.

KBiosciences employ a modified version of the Amplifluor (Myakishev et al. 2001) genotyping method (Millipore, MA, US).
This method involves the allele specific amplification of SNP alleles using two tailed locus specific oligonucleotides and a standard locus specific reverse primer.

In addition two tailed oligonucleotides labelled with different fluorescent dyes are used in the reaction. In the early cycles of amplification the allele specific oligonucleotides compete with one another for template. The best fitting (matching) oligonucleotide binds the template with high affinity and thereby creates more of this allele's template for subsequent amplification.
In the later cycles the appropriate tailed fluorescent oligonucleotide is incorporated. The end point fluorescence is then detected using a fluorescent plate reader. There are four different possible outcomes each with different fluorescent intensities: low intensity for both fluorescent dyes indicating that no amplification has occurred; high intensity for one fluorescent dye and low intensity for the other indicating a homozygote template for the appropriate SNP allele; the opposite scenario; and finally moderate intensities for both fluorescent dyes indicating a heterozygous template for the SNP. See Figure 2:5 and Figure 2:6.

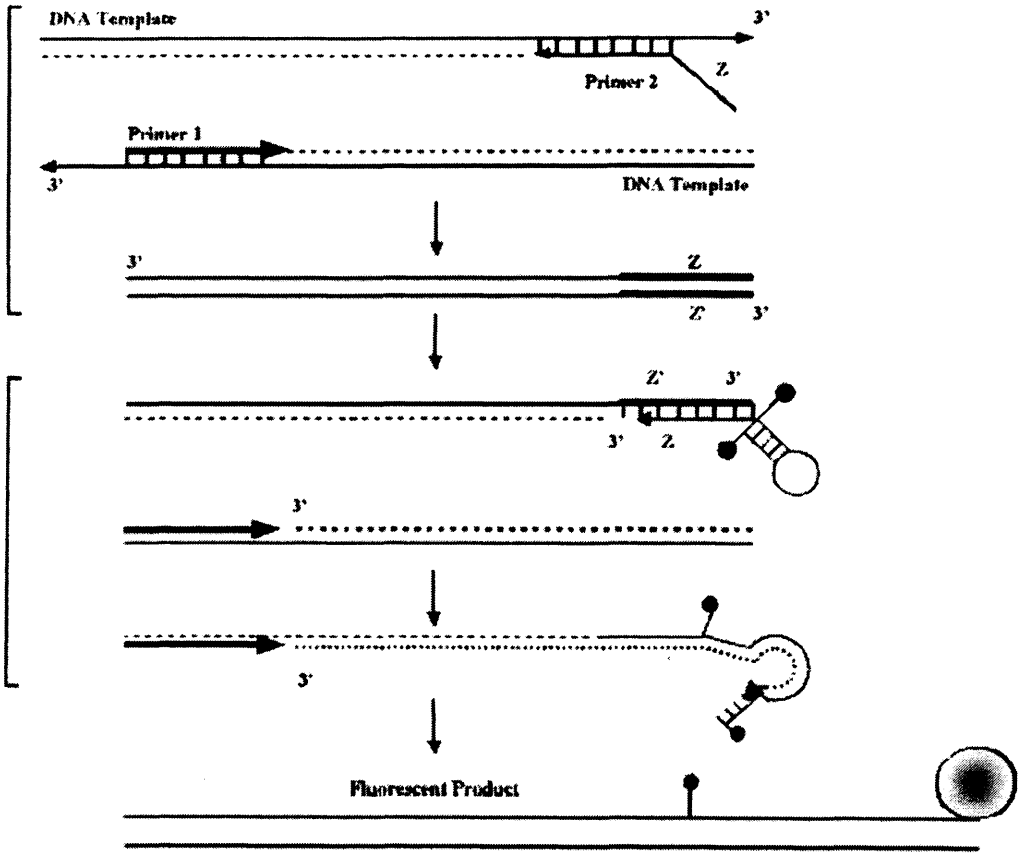



**Figure 2:5 KASPar method for genotyping a two allele SNP using allele-specific primers coupled with two Universal Amplifluor primers (Bengra et al. 2002).**

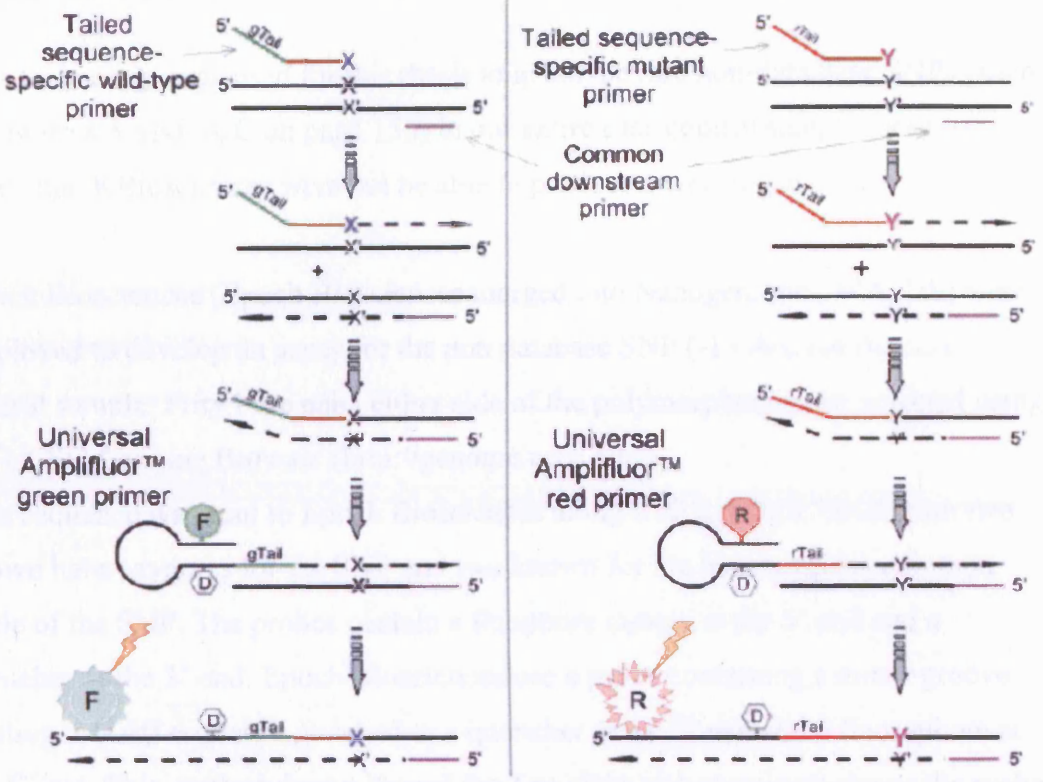



**Figure 2:6 An alternative diagrammatic representation of two-allele SNP detection using allele-specific primers coupled with two Universal Amplifluor primers (Bengra et al. 2002).**

### 2.8.3.2 EPOCH BIOSCIENCES IN HOUSE SNP GENOTYPING

This technology was used for this thesis to genotype rare non-data base SNPs (such as Figure 8:5 SNP A/C on page 153) in our entire case control sample these were SNPs that KBiosciences were not be able to produce a working assay for.

Epoch Biosciences (Epoch Biosciences merged into Nanogen, Inc., WA, US) were employed to develop an assay for the non database SNP (-15 A/C) in the case control sample. Fifty base pairs either side of the polymorphism were selected using the UCSC Genome Browser (http://genome.ucsc.edu/).
This sequence was sent to Epoch Biosciences along with genomic DNA from two known heterozygotes for the SNP and two known for the homozygotes common allele of the SNP. The probes contain a fluophore moiety at the 5' end and a quencher at the 3' end. Epoch Biosciences use a probe containing a minor groove binding (MGB) sequence attached to a quencher at the 5' end and a fluorophore at the 3' end. This method does not need the Taq DNA polymerase to cleave the probe. It simply emits fluorescence once hybridized to the sequence. Further, it is protected from 5' digestion by Taq DNA polymerase during the the amplification stages.
Each MGB probe has a different fluorescent reporter dye. The DNA duplexes are denatured over a time course and the decrease in the fluorescent signal of each probe is measured. Comparison of the two-colour melting curves allows for differentiation of sequence variants of the SNP.

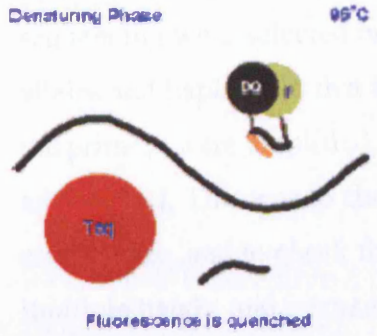



*Figure 1 – denaturing phase – the probe is unattached and no fluorescence is emitted.*

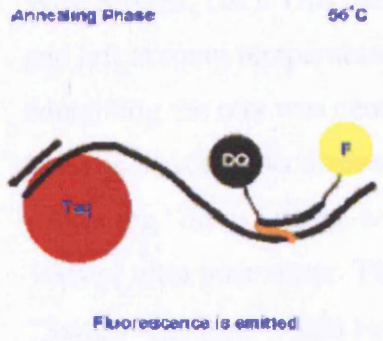



*Figure 2 – the annealing phase – the probe is attached in the minor groove and fluorescence is emitted*

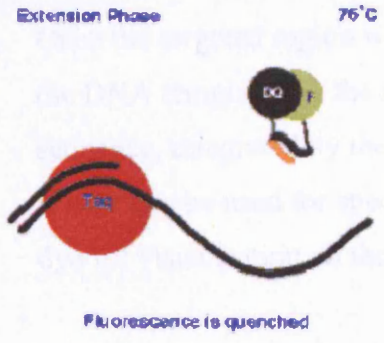



*Figure 3 – the extension phase – the probe is unattached and no fluorescence is emitted.*

## 2.8.4 SEQUENCING

Primers for sequencing are optimised as described in section 2.6. DNA samples for sequencing were selected on the basis that cases of schizophrenia had inherited alleles and haplotypes that had previously shown association to schizophrenia. Once the primers were amplified, a proportion of the PCR products were analysed on an agarose gel. This was to check that the individuals which had been amplified, are the correct size, and to check for non specific amplification as shown by the presence of multiple bands, and intensity of primer dimer.

The amplimer for each individual was then cleaned from the master mix reagents and purified from any primer dimer by the use of microCLEAN (Microzone Ltd, West Sussex, UK). This reagent was mixed in equal quantities to the PCR reaction and left at room temperature for approximately 5 minutes. Then the eppendorf containing the mix was centrifuged at 14000g for 7 minutes to pellet the template. The supernatant was disposed of and the tubes briefly spun to pull down any remaining "dregs" which were then removed. The pellet was then re-suspended in 3-10µl of ultra pure water. This targeted region was then used as a template for the "Sanger-Coulson" chain termination sequencing method.

### 2.8.4.1 SANGER-COULSON (CHAIN TERMINATION SEQUENCING METHOD)

Once the targeted region was amplified and cleaned, sequencing could take place on the DNA template. As the amplimer contains an integrated M13Forward/Reverse sequence, integrated by the primers used in the initial amplification. An M13F/R primer can be used for specific sequencing and simultaneous incorporation of the dye for visualisation on the LICOR.

The M13F/R primer during the PCR reaction binds to the specific template and allows extension and synthesis of the complementary strand by the sequatherm DNA

polymerase. The advantage of using M13F IRD700 and M13R IRD800 is that both strands can be sequenced at the same time and used as a crosscheck for one another to assess whether novel bands (DNA variants) were artefacts or real polymorphisms.

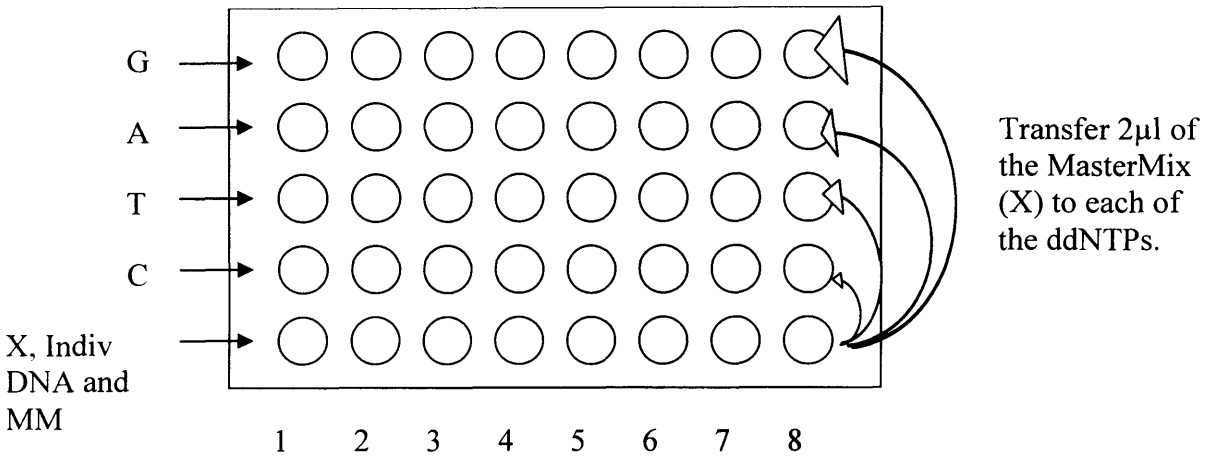
In sequencing the usual dNTPs are used along with a small amount of dideoxynucleotide (like ddATP) which is incorporated just as efficiently into a growing polynucleotide strand but prevents further synthesis. This is because the dideoxynucleotide lacks a hydroxyl group at the 3 prime position of the sugar component and this prevents further incorporation into the polynucleotide. Therefore chain termination occurs wherever a dideoxynucleotide is incorporated. As a result the reaction can run with each individual ddNTP and will result with terminations at every base of the sequence showing exactly where each base belongs in a sequence.

**Table 2:6 Table showing the master mix reagents and volumes required to sequence 16 individuals.**

| Sequencing reagent for MasterMix | Volume |
|---|---|
| | |
| 3.5x Buffer | 75µl |
| M13F 700 (10pmol/µl) | 1.33µl |
| M13R 800 (10pmol/µl) | 1.33µl |
| Sequitherm Polymerase | 11µl |
| Water | 27µl |
| | |

3µl of the cleaned DNA template is mixed with 6.8µl of the sequencing master mix. 2µl of this is then mixed with 2µl of each ddNTP as shown in the figure below. The sequencing reaction is then PCR amplified by the Sequatherm program as stated in Section 2.6.3. Table 2:4. This is carried out on a 96 well plate and enables one to sequence 16 individuals at a time on a single polyacrylamide gel.

**Figure 2:7 Layout of the sequencing reaction, showing how the Master Mix and DNA are mixed with each ddNTP separately.**




After PCR an equal volume of Fuchsin loading buffer is added to the product. The amplified product is aliquotted to a relevant plate format to enable loading with a multichannel pipette contiguously which will keep all the same nucleotides aligned.

The sequencing product is then loaded on a 44cm agarose gel, and analysed by the LICOR scanner on the preset sequencing parameters (Section 2.7.2 Table 2:5). The resultant sequencing image is read and checked against the database sequence to make sure the correct region is sequenced and also to check for any abnormalities. The sequence is examined for any polymorphisms and mutations both present in the database and not. These abnormalities are cross-checked with the reverse strand sequence to make sure they were not an artefact of the gel image. Any abnormalities were checked in random controls to see if they appear in the same frequency as the cases or more importantly not. In total 32 of the most significantly associated individuals were sequenced (these were chosen from haplotypes with the most significant markers section 3.2.4.3). The presence of these potential aetiological base pair changes were then investigated by sequencing with an equal number of random controls.

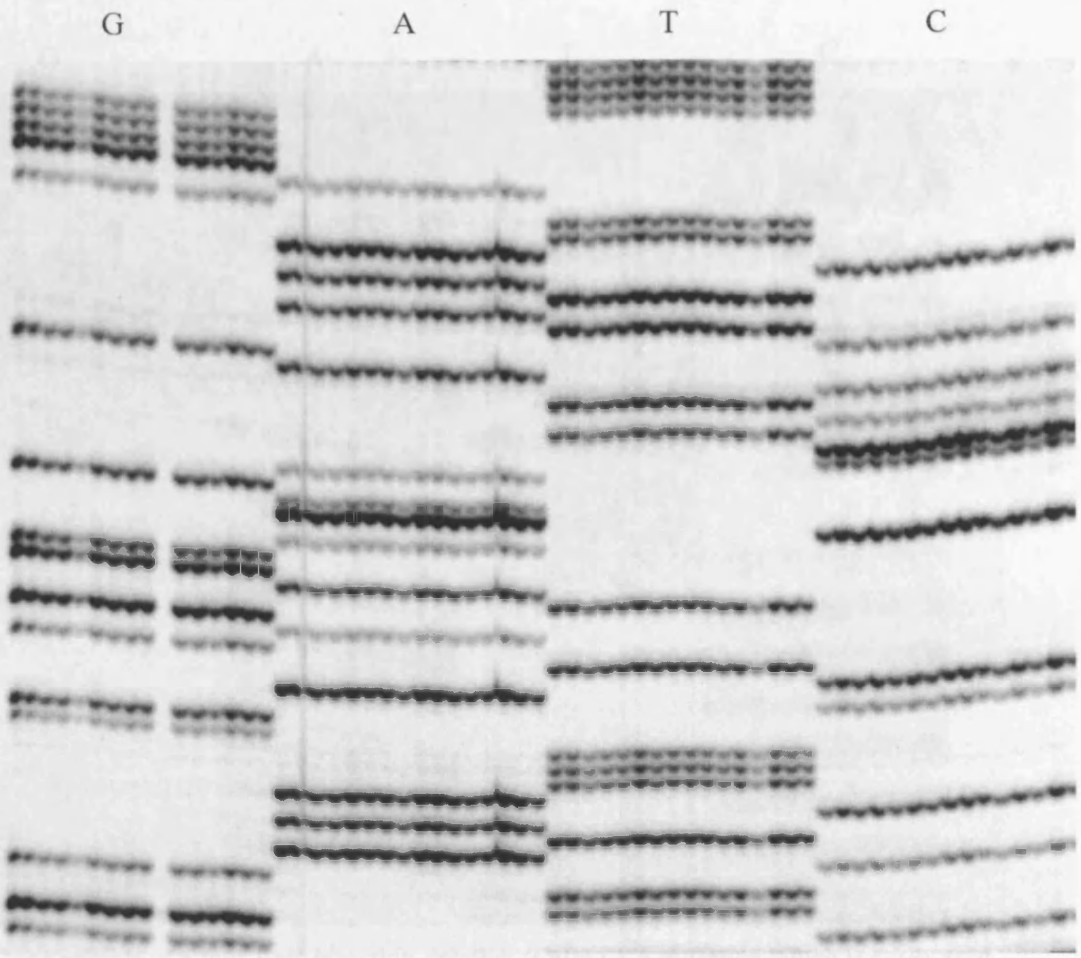



**Figure 2:8 An example of the result one would obtain by Sanger-Coulson chain termination sequencing method, with the image obtained from the LiCor DNA sequencer**

SNPs or any other abnormalities discovered in sequencing of interest were further typed as described in section 2.8.3.

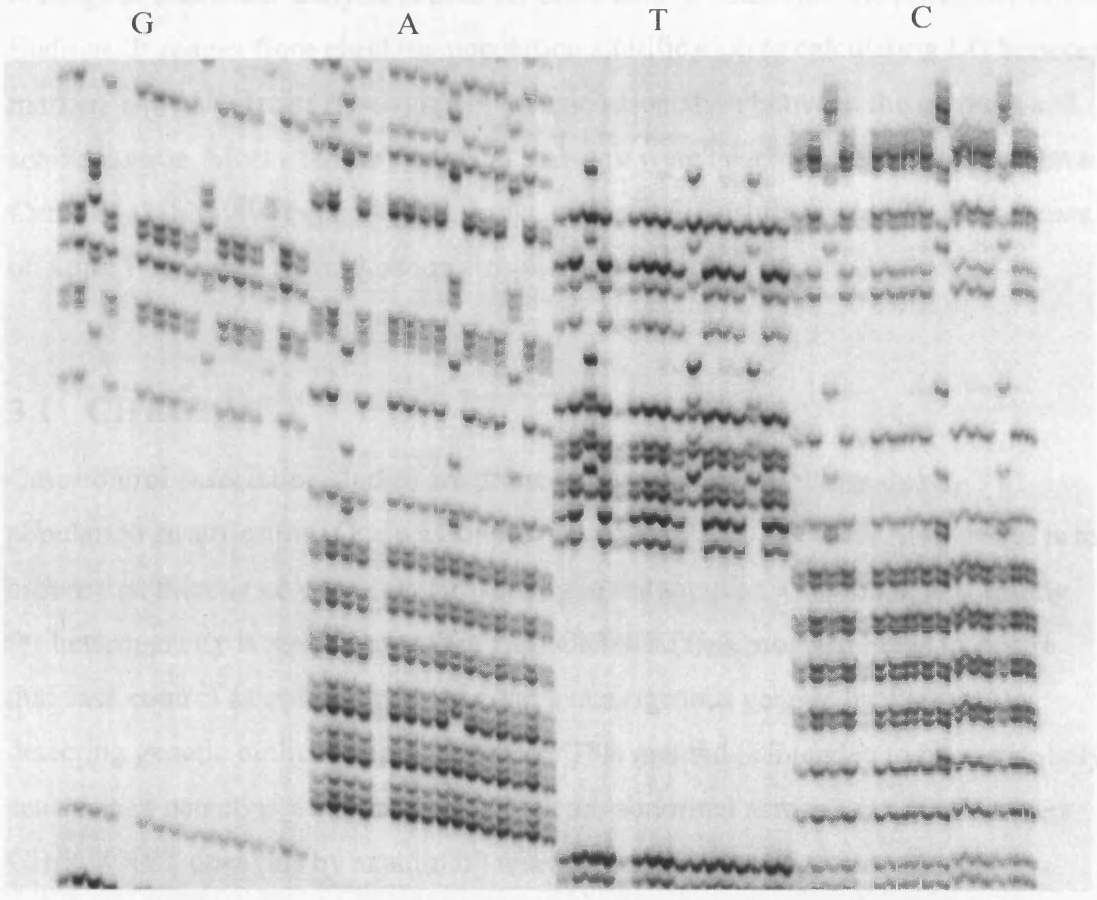



**Figure 2:9 Example of a four base insertion-deletion (INDEL), a mutation which is easily seen by eye.**

# 3.0 STATISTICAL ANALYSIS

A range of statistical analysis is used for these data to determine the reliability of our findings. It ranges from checking population stratification to calculating LD between markers and calculating the degree of genetic association between the markers and schizophrenia. Most of these analytical software were developed by Professor Dave Curtis et al (http://www.smd.qmul.ac.uk/statgen/dcurtis/software.html). Department of Adult Psychiatry Royal London Hospital.

## 3.1 CHECKHET

Case control association studies are prone to false positive findings due to population stratification. One way of preventing these false positive discoveries is to ensure that there is no overall heterogeneity in the samples. One method of testing for heterogeneity is a program called CHECKHET. This program helps to ensure that case control samples are drawn from a homogenous genetic background by detecting genetic outliers from the sample. This method is intended to be particularly sensitive in detecting a few subjects which are abnormal across a range of markers. CHECKHET does this by examining whether a subject tends to have normal genotypes across a range of markers using a log likelihood ratio test of heterogeneity as a measure of abnormality. The tests are summed over all markers and permutation tests are then used to identify subjects that have exceptionally large scores. These samples are then excluded from the analysis. The program depends on genotypes rather than allelic frequencies it does not make the assumption that all markers are in Hardy-Weinberg equilibrium (Curtis et al. 2002).

Data from fifteen genetic markers at chromosomal loci not thought to be involved in schizophrenia were genotyped in a subset of the sample (200 cases and 300 controls) and analyzed using CHECKHET. This test detected two schizophrenia cases with abnormal genotypes and these were excluded from further study before any of the study markers were genotyped (Curtis et al. 2002).

## 3.2 HAPMAP

As described in section 2.5.3, HAPMAP was a project that aimed to identify all common SNPs in the human genome across four different populations. This project was designed to aid association studies by enabling one to fine map a region of interest by selecting tgging SNPs. The program Tagger was used to identify "tagging" SNPs. Tagger selects SNPs with varying genetic information, indicated by the different $r^2$ LD vaules. This allows for the best statistical selection of single SNPs form ones in high LD. This method proves the most economical method to fully genotype a region with known SNPs.

### 3.2.1 WHAT IS LINKAGE DISEQUILIBRIUM (LD)?

Linkage disequilibrium is often termed "allelic association." When alleles at two distinctive loci occur in gametes more frequently than expected (rather from other causes such as population stratification or epistasis) given the known allele frequencies and recombination fraction between the two loci, the alleles are said to be in linkage disequilibrium. Evidence for linkage disequilibrium can be helpful in mapping disease genes since it suggests that the disease causing change and genetic marker may be very close to one another. LD is also commonly used to describe the relationship between genetic markers. LD reflects the genetic distance between markers, it also can be disrupted by the recombination rate and the presence of new mutations.

There are many measures of LD. The most common used through out this project are D' and $r^2$ values which will be discussed next.

#### 3.2.1.1 LEWONTIN'S D' MEASUREMENT OF LD

Lewontin's D' (Lewontin 1964a; Lewontin 1964b) is derived from D, which is a LD coefficient derived from haplotypic frequencies as displayed below. Using a biallelic

system such as a SNP the haplotypic frequencies can be simply estimated between two markers.

D measures the deviation of haplotype frequencies from the equilibrium state. LD occurs when D is significantly greater than zero. Considering two linked SNPs with alleles (A,a) and (B,b) resulting in four possible haplotypes: AB, Ab, aB, ab. "D" can be calculated as in the equation below

| | | SNP1 | | |
|---|---|---|---|---|
| SNP2 | | A | A | |
| | B | fAB | faB | fB |
| | b | fAb | Fab | fb |
| | | fA | Fa | |

f = frequency

As a result the LD coefficient D could be worked out as follows:-

$$D = f(AB) - f(A) \times f(B)$$

The problem with the D coefficient is that is was dependent of the marginal allele frequencies in the contingency table. This disqualifies D as a measure of LD because the data is dependent and cannot be compared for different SNPs and different populations (Devlin et al. 1995). However D can be normaised to D' allowing the measurement to be comparable across SNPs and populations.

D' is defined as the absolute ratio of D compared with its maximum value, $D_{max}$, when $D \geq 0$, or compared with its minimal value, $D_{min}$, when $D < 0$ (modal result taken), D' = 1 denotes complete LD, and historical recombination results in the decay of D' towards zero.
This means that D' is equal to 1 if alleles at adjacent marker loci are in LD with each as much as is possible given that they have different population allele frequencies. D' values less than 1 imply that the maximum extent of LD between two marker alleles has not been observed. For example, D' = 0.87 represents a strong LD between SNP alleles; similarly D' = 0.12 represents weak LD between SNP alleles.

D′ has the same range of values regardless of the frequencies of the SNPs compared (Lewontin 1988). Its sign (positive or negative) depends on the arbitrary choice of the alleles paired at the two loci.

The normalised measure of D (D') is worked out as follows:

$$D' = \frac{D}{\sqrt{p(1-p)q(1-q)}}$$

Where $p$ and $q$ the major and minor allele frequencies as described in the Hardy-Weinberg equation (section 3.2.2)

Where $|D'| = 1$ denotes complete LD,

Where D' = 0 it shows no LD (linkage equilibrium).

And where $|D'| < 1$ it shows various degree of LD between the markers (higher the number the stronger the LD).

The limitations of using D' include its insensitivity to distinguish between different degrees of LD, a $|D'| = 1$ can reflect 2 to 3 haplotypes present. There is also upward bias for small to moderate sample sizes, and it can take extreme values when at least one allele frequency is small (Devlin et al. 1995).

Another measure of LD is $r^2$ this will be discussed in the next section.

### 3.2.1.2 $r^2$ MEASUREMENT OF LD

$r^2$ is another measurement of LD (Pritchard et al. 2001). It is thought to over come the problems of D', as $r^2$ is able to detect LD in small sample sizes and also is far more reliable for detecting LD for low allele frequencies compared to D'.

The equation for $r^2$ is as follows:

$$r^2 = \frac{(f(AB) - f(A)f(B))^2}{fA.fa.fB.fb} \text{ or } \frac{D^2}{fA.fa.fB.fb}$$

Where f means frequency of the alleles.

When $r^2 = 1$ the markers are said to be in perfect LD. Observations of one marker provides complete information about the other marker, making them both redundant. Therefore they are genetically identical and one can be used as a "Tag SNP" for the other in association studies.

$r^2 = 1$ only when no recombination has occurred between the markers and the allele frequencies are identical. It also corresponds to a situation where two haplotypes are present out of a possible four.

In general $r^2$ is more useful for dividing closely located SNPs into blocks, if the purpose is to identify "tagging SNPs". However $r^2$ represents statistical association at the population level as well as incorporating LD and has no direct relationship with recombination like D'. $r^2$ depends on marker allele frequencies and can be difficult to interpret when comparing multiple markers in a region (Hedrick 1987). D' is directly related to recombination fraction and its generalization to more than two loci is the only measure of LD not sensitive to allele frequencies.

Hence when judging LD both D' values and $r^2$ values are used to describe the LD relationship between markers and are used in selecting the most appropriate SNPs with the use of HAPMAP and HaploView.

### 3.2.2 HARDY-WEINBERG EQUILIBRIUM.

With the data from genotyping of microsatalites and SNPs, the information is checked to make sure that they follow Hardy-Weinberg equilibrium before any further statistical analysis is considered. This is to check if the alleles have been genotyped correctly.

The equation for Hardy-Weinberg equilibrium is as follows:

$$p^2 + 2pq + q^2 = 1$$

Where *p* is the major allele and *q* is the minor allele.

The concept of Hardy-Weinberg equilibrium is used to understand the genetic characteristics of populations (Hardy 1908; Weinberg 1908). They independently stated that if p is the frequency of one allele (A) for a biallelic locus, then the HWE-expected frequency will be $p^2$ for the AA genotype, 2pq for the Aa genotype and $q^2$ for the aa genotype. The three genotypic proportions should sum to 1, as should the allele frequencies.
The derivation of the Hardy-Weinberg equation is appreciated by use of a Punnett's square shown in Figure 3:1 below.

**Figure 3:1 Using a Punnett's square to derive the Hardy-Weinberg equation.**

| | | Females | |
|---|---|---|---|
| | | A (p) | a (q) |
| Males | A (p) | AA ($p^2$) | Aa (pq) |
| | a (q) | Aa (pq) | aa ($q^2$) |

The final three possible genotypic frequencies in the offspring become:

$$f(AA) = p^2,\ f(Aa) = 2pq \text{ and } f(aa) = q^2.$$

Testing for HWE is used for quality control of large-scale genotyping and is an important method to identify systematic genotyping errors in unrelated individuals

(Gomes et al. 1999). Genotyping errors are known specifically to affect certain genetic measurements such as LD, upon which association studies depend.

Tests for deviation of Hardy-Weinberg equilibrium is carried out by a simple chi squared test using the observed genotyped frequencies for the data and the expected genotype frequencies from the equation. Once the data is found to be in Hardy-Weinberg equilibrium it can be used to test for association using the statistical analysis as will be described in the following sections. However if the data is not in Hardy-Weinberg equilibrium, the data will have to be re-checked to verify if the calls are correct and the data is in the correct format.

### 3.2.3 STATISTICAL TESTS FOR ASSOCIATION BETWEEN MARKER ALLELES AND DISEASE

To test for association between the polymorphic marker and disease, a simple Chi square test is employed to test if alleles are more frequent in cases compared to controls.

#### 3.2.3.1 TESTING FOR ASSOCIATION WITH SNPs USING CHI SQUARE

The data from SNP genotyping is first checked for accuracy by comparing the data with duplicates (17% of the sample is duplicated). The data is then checked to ensure that it follows Hardy-Weinberg equilibrium. Once the data has satisfied these criteria, the genotypic information is checked for association. Allelic association is tested for by forming a table with the allele frequencies for both cases and controls then performing a chi square test with one degree of freedom. Genotypic association is also tested for with a similar table representing each genotype frequency and performing a chi square test with two degrees of freedom.

In association studies allelic association is the most reliable test of association and genotypic association is not generally trusted in the absence of allelic association, because in the absence of allelic association there is a high probability that the result

is due to discrepant genotyping results rather than a genuine result. The reason behind this, is if one genotype predominates then you would expect the allele frequency to also show allelic association

### 3.2.3.2 CLUMP TEST FOR ASSOCIATION OF MICROSATELLITES.

Allelic tests of association for microsatellites were tested using a program called "CLUMP" (Sham et al. 1995). The data is first also tested for accuracy by cross checking the genotypes with duplicates which make up a minimum of 17% of the sample. The data is then tabulated for allele frequencies for the microsatellite in both cases and controls. The data is then fed in to CLUMP to perform a number chi square tests for association.

As stated by the creators, CLUMP is a program developed to assess the significance of the departure of the observed values from the expected values conditional on the marginal totals within a 2 x N contingency table. The significance is assessed using a Monte Carlo approach, by performing repeated simulations on randomly generated tables with the same marginal values using simulated data, and counting the number of times the chi square value is associated with the table under consideration by chance. This means that the significance levels assigned is unbiased (accuracy dependent on the number of simulations) and also means that no special considerations need to be made for small expected values (Sham et al. 1995).

CLUMP is used to calculate maximal chi square values by clumping together columns into a new two by two table in a way which is designed to maximise the chi square value. The method produces an "inflated" chi square value but does not cause any problems in interpretation because its significance is assessed using a Monte Carlo method which takes into account the increased chance of a false positive result.

The output of CLUMP generates four chi square results and their significance is evaluated by testing how many times the value produces is exceeded by chance with the Monte Carlo simulation method. The four results produced are as follows.

1. Pearson's $\chi^2$ statistic of the 'raw' 2-by-N contingency table originally supplied. The statistic is referred to as "**T1**" the Chi Square value is worked out for the original supplied table. However if there are small expected values (less than 5) in some cells then these values might not follow the expected distribution of a chi-squared statistic with N-1 degrees of freedom. However the significance can be reliably accessed using the Monte Carlo simulations.

2. The $\chi^2$ statistic of a table with rare alleles grouped together to prevent small expected cell counts. If any cell has a value less than 5, then its column is clumped together with the column of the next smallest value. This process is repeated until all columns have values of 5 or more. The resulting statistic is referred to as "**T2**" the significance is assessed by Monte Carlo simulations and the degrees of freedom is N-1 (where N is the number of resulting columns after they were clumped).

3. A 2-by-2 table obtained by comparing one column of the original table against the total of all other columns. (The largest of the $\chi^2$ statistics of 2x2 tables each of which compares one allele against the rest grouped together). This tests the hypothesis that there is one particular column with a number of cells deviating form the expected values. For each column in turn, all the other columns are clumped together to assess the chi square value. The columns, which produces the maximal chi square value is used and the results are outputed. This statistic is "**T3**". Columns containing values less than 5 are not considered and the columns are clumped together with the next

column with the lowest value. In this case the T2 table would be used. The significance of the result is assessed with Monte Carlo simulations.

4. A 2-by2 table obtained by clumping the columns of the original table to maximise the chi squared value. This method compares any combination of columns (alleles) against the rest to produce the maximum chi square value possible and produces the largest of all 2-by-2 contingency tables. Its significance is assessed using Monte Carlo simulations. This produces the T4 statistic.

From the output the most significant statistic is chosen to show the degree of allelic association of the marker.

In order to run the CLUMP program, one needs to give the number of columns (alleles) the table has and the number of permutations to perform, and also provide a random number to start calculations in order to produce the pseudo-random tables for Monte Carlo simulations. The number of permutation should be significantly large to produce a satisfactorily accurate estimate of the true significance that is achieved, about 10 000 (9999 in the programme as it starts at 0) or more permutations are generally used.

### 3.2.4 STATISTICAL ANALYSIS OF GENOTYPIC INFORMATION USING "GENECOUNTING"

Ascertaining allelic association between a single marker and the disease is one method of extracting information from genomic data. However one can analyse a combination of neighbouring markers to extract more information to find out the relationship between markers (to judge the strength of LD between them) and to try increase the power of information generated for association to the disease by analysing different haplotypes between markers which are in LD with one another. Having a large number of markers which form a positive haplotype is useful, as the

information can be used to help select the most significantly associated individuals, who contain those haplotypes. These individuals can be chosen for sequencing candidate regions to increase ones chance of finding an abnormality in the genomic sequence.

The program which was used to ascertain the LD relationship between markers and haplotype significance was as sub-program of "GENECOUNTING" (Zhao et al. 2000; Zhao et al. 2002; Curtis et al. 2006) and the functions will be discussed next.

#### 3.2.4.1 LDPAIRS

LDPAIRS reports pair-wise linkage disequilibrium statistics between all pairs of markers. This program allows one to calculate LD between any set of selected markers, it will calculate a p value for the statistical significance of the test for linkage disequilibrium between the markers and from the estimated haplotype frequencies it will calculate a mesure of LD called Cramer's v, which is equivalent to the square root of the $r^2$ measure of LD (Bishop et al. 1975) (section 3.2.1.2). The program in addition will also work out the absolute value of the D' statistic (section 3.2.1.1) between the commonest alleles at each locus.

#### 3.2.4.2 SCANGROUP

SCANGROUP compares haplotype frequencies in cases and controls of haplotypes built from subsets of markers, consisting either of groups of contiguous markers or of all possible combinations of a specified number of markers. This program is useful to detect if any of genotyped plates have been swapped over which would produced aberrant results.

#### 3.2.4.3 RUNGC

RUNGC estimates haplotype frequencies between unrelated subjects and has the ability to deal with multiallelic markers and markers with missing genotypes.

RUNGC compares haplotype frequencies in cases and controls, tests for significant differences using a likelihood ratio test and permutation test.

The program out-puts the frequencies of haplotypes and groups them side by side for comparison. This allows one to pick out hapotypes with the largest discrepancies (i.e over representation in cases) in order to pick out the most implicated haplotype or a range of implicated haplotypes. Also by adding a second output parameter (.hap) the RUNGC program is able to output grouped individuals containing their haplotypes enabling one to select individuals with the most significant hapotypes for specific sequencing of those "most at risk individuals".

### 3.2.5 POWER CALCULATIONS.

Power calculations are carried out to find out the power of a sample and its ability to detect low frequency alleles in the population. It is known that the larger the sample the greater the power it has to detect rare allelic variants and a better chance the sample has to perform a replication association study of an implicated region in a smaller sample. Power calculations can be calculated by using online resources such as the "Genetic Power Calculator" (http://pngu.mgh.harvard.edu/~purcell/gpc/), (Purcell et al. 2003).

We have calculated that the sample used in this thesis (about 450 cases and 450 controls) is large enough to detect a marker allele frequency difference of 5% in controls and 10% in cases with a power of 0.99 at $p<0.05$ and of 0.91 at $p<0.001$ assuming the minor marker allele frequency is less than 10%. With a high frequency biallelic marker present in 45% of controls and 55% of cases the power of the sample is 0.90. If a high frequency marker is employed with only a five percent allele frequency difference between cases and controls, for example 45% versus 50%, our sample had a power of only 0.41

### 3.2.6 ODDS RATIO.

The odds ratio is a way of comparing whether the probability of a certain event is the same for two groups. An odds ratio of 1 implies that the event is equally likely in both groups. An odds ratio greater than one implies that the event is more likely in the first group. An odds ratio less than one implies that the event is less likely in the first group.

Shown below is the typical 2 by 2 table.

| | Allele | |
|---|---|---|
| | T | G |
| Control | a | b |
| Case | c | d |

Where a & c = the number of T alleles observed in the controls and cases, and where b & d = the number of G alleles observed.

By definition the odds ratio (OR) is;

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

This can be reduced to:

$$\text{OR} = \frac{ad}{bc}$$

Confidence Limits can be worked out for the odd ratio by first working out the standard error (SE)

$$\text{SE}(\ln \text{OR}) = \sqrt{\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}\right)}$$

Confidence intervals for odd ratio is tested at the for a two sided 95% confidence interval where $Z_L$ is hence 1.96

The confidence limits for odds ratio is calculated as follows

$$\exp(\ln(OR) \pm Z_L \times SE).$$

# 4.0 ATTEMPTED REPLICATION OF REGULATOR OF G-PROTEIN SIGNALLING 4 (RGS4) IN BRITISH CASE CONTROL SAMPLE

## 4.1 GENERAL INTRODUCTION

Regulator of G-protein Signalling 4 (RGS4) has been implicated as a schizophrenia susceptibility gene in several studies. Mirnics et al (2001) first implicated the gene, by finding RGS4 mRNA expression had significantly decreased in schizophrenic brain samples using microarray technology. This was then followed by a significant association finding by Chowdari et al (2002). The finding was interesting as the gene is situated in genetic locus 1q23.3, a linkage hotspot for schizophrenia susceptibility. The association has been replicated in part by a number of independent groups; however their results were not completely conclusive. We attempted to replicate the original finding by Chowdari et al (2002) in our own case control sample.

## 4.2 REGULATOR OF G-PROTEIN SIGNALLING 4 (RGS4) REPLICATION STUDY.

### 4.2.1 INTRODUCTION

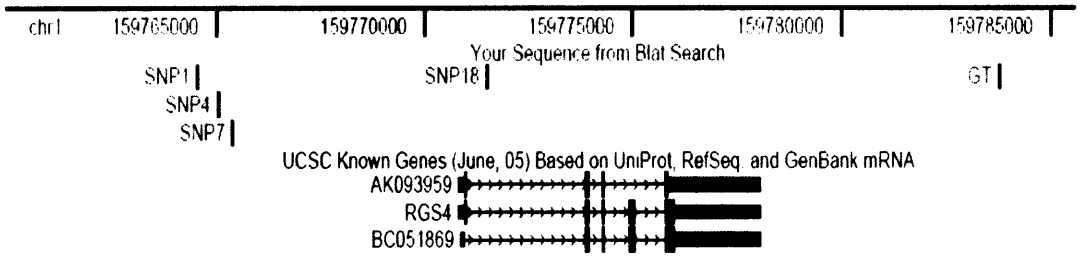
Regulator of G-protein signalling (RGS) belongs to a diverse family of proteins modulating the G-protein signalling pathways. The RGS domain accelerates GTP hydrolysis, deactivates the molecule to the inactive GDP-bound form, and as a result quenches signal transduction through the involved G-protein coupled receptors (Sobell et al. 2005). The function of RGS4 is to regulate the activity of GTPase, which is coupled with several receptors involved in the signal transduction of dopamine, glutamate and several other neurotransmitters (Saugstad et al. 1998;

Taymans et al. 2003; Taymans et al. 2004). These receptors are thought to play some sort of role in schizophrenia.

Regulator of the G-protein signalling 4 (RGS4) is an attractive candidate gene for schizophrenia and is close to the 1q23.3 region implicated by linkage studies, ((Brzustowicz et al. 2000; Blackwood et al. 2001; Gurling et al. 2001; Ekelund et al. 2004)). RGS4 had initially been implicated through decreased gene expression in post mortem schizophrenia brain samples using microarrays (Mirnics et al. 2001). Tests of allelic and haplotypic association with schizophrenia at the RGS4 locus were first carried out using three methods. These were case control, family and trios sample designs with samples from Pittsburgh, the NIMH (USA), and New Delhi. (Chowdari et al. 2002). Thirteen RGS4 SNPs were genotyped. Family-based association analyses using the transmission disequilibrium test (TDT) found transmission disequilibrium for individual SNPs as well as for haplotypes composed of four SNPs in the 5' flanking sequence and first intron of RGS4 (SNPs 1 (rs10917670), 4 (rs951436), 7 (rs951439) and 18 (rs2661319)). The alleles and haplotypes which were associated with schizophrenia differed between the Pittsburgh and NIMH samples. Comparison of the Pittsburgh cases with two independent groups of unrelated controls did not reveal significant associations, raising the possibility that the TDT results might reflect methodological problems or producing false positive results. Two subsequent studies involving cases and unrelated controls of caucasian ethnicity did however reveal associations with these four SNPs. Williams et al (2004) studied the association between the four SNPs and schizophrenia in a large case (n = 709) control (n = 710) sample from Wales, UK. Modest but significant associations were found between alleles of SNPs 4 (rs951436) and 18 (rs2661319) and schizophrenia. An Irish study (Morris et al. 2004) genotyped the same four SNPs in a sample of 249 cases and 231 controls from the Republic of Ireland and found significant but not very strong allelic association with two different SNP markers (SNP 1 (rs10917670) and SNP 7 (rs951439)). More recently a family-based study (Chen et al. 2004) also found evidence for association of RGS4 with schizophrenia in the Irish Study of High Density Schizophrenia

Families (ISHDSF). Single marker TDT for the four SNPs showed modest association for SNP 18 (rs2661319) and a trend for SNP 4 (rs951436). The goal of the present study for this thesis was to attempt to replicate the original reported associations between RGS4 markers and schizophrenia.

**Figure 4:1 Showing the location of SNPs and the GT microsatellite in the region of RGS4 (UCSC BLAT diagram).**




## 4.2.2 METHOD FOR TESTS OF ALLELIC ASSOIATION BETWEEN RGS4 AND SCHIZOPHRENIA

Markers chosen for genotyping were the SNP markers 4 (rs951436), 7 (rs941439) and 18 (rs2661319) (Chowdari et al. 2002). SNP 1 is in strong LD with SNP 7. $D'$ and $r^2$ values close to 1 between both SNPs have been reported in the earlier studies of RGS4 and schizophrenia from Caucasian origin. According to Morris et al (2004) $D'$ was more than 0.98 and $r^2$ was more than 0.96 in both cases and controls. This was confirmed in Wales ($D'$ was 0.987 in Cases and 0.981 in controls, and $r^2$ was 0.96 in cases and 0.963 in controls) (Williams et al. 2004). Therefore SNP 7 should carry the same information as SNP 1. In addition we genotyped a polymorphic microsatellite $(GT)_{14}$ which is approximately 7Kb distal from RGS4 for an additional test of association. The SNPs were genotyped using sequence data available from the University of Pittsburgh website (http://www.pitt.edu/~nimga/research/RGS4). The microsatellite $(GT)_{14}$ marker was obtained from the University of California at Santa Cruz (http://www.genome.ucsc.edu/) and was at database position 159783591-159783618 base pairs on chromosome 1q23.3. The marker was genotyped using the following primer sequences:

Forward: CTGTAATCCCAACAACACGG.

Reverse: CACTCAGCATAGGCAGGACA.

PCR amplification of the microsatellite marker was carried out using an M13 tailed primer and a second non-tailed primer. A third universal M13 sequence primer labelled with IRD 700 or IRD 800 was used to hybridise against the M13 tailed locus specific primer. The microsatellite marker was genotyped by Licor and SAGA-GT as previously described in section 2.8. Single nucleotide polymorphisms were determined by the Amplifluor SNP genotyping method as modified by KBiosciences (KBiosciences: http://www.Kbioscience.co.uk/). 17% of samples on each microtitre plate were reduplicated in order to detect error and confirm the reproducibility of genotypes section 2.8.3.1. The data were then analysed to confirm Hardy Weinberg equilibrium (HWE). Next the data were analysed for allelic association with schizophrenia using CLUMP and $Chi^2$ tests section 3.2.3.2.

### 4.2.3 RESULTS FOR REPLICATION STUDY OF RGS4

None of the markers (SNPs 4 (rs951436), 7 (rs951439) and 18 (rs2661319)) which were previously found to be associated with schizophrenia (Chowdari et al. 2002) could be implicated in the present sample. The $(GT)_{14}$ microsatellite approximately 7Kb distal from RGS4 also failed to show allelic association with schizophrenia (Table 4:1). Tests of haplotypic association using GENECOUNTING and permutation testing also did not produce any evidence for association at this locus with any of the SNPs and the microsatellite marker as shown in Table 4:1.
Taking the example of SNP 4, the most consistently associated SNP, our sample only had the power of 0.53 to detect the effect size observed by (Williams et al. 2004).

A comparison between the alleles frequencies of the SNPs markers in our sample and the previous case control studies was shown in Table 4:2.

After these results were analyzed and published (Rizig et al. 2006) a number of new papers on the results of association of RGS4 with schizophrenia had been published. The results of known published papers are shown in Table 4:3 including the more recent meta-analysis (Talkowski et al. 2006) of RGS4. The meta-analysis consisted of results from all studies using trios, family samples and case controls samples.

From the overview of all results (Table 4:3), it is seen that the association results for RGS4 over all is very weak. Single SNP association from the majority of reports are weak or non-significant and haplotypic associations where reported positive are in different haplotypic combinations and marker numbers, or were negative. Even the published meta-analysis of 2160 families and 3486 cases with 3755 controls (Talkowski et al. 2006) shows that no individual markers were significantly associated in the family samples and haplotypic association were found not to be significant. In the case control samples, all SNPs were found to be non-significant apart from SNP 4 which had shown modest association, however there was no significantly associated haplotype.

### 4.2.4 RGS4 DISCUSSION

The sample employed here had greater power than the original sample of Chowdari et al (2002) and also greater power than the case control sample employed in the Irish study (Morris et al. 2004) as noted in Section 3.2.5. The Irish study only found associations with p values of about 0.04 with each of the two SNPs labelled 1 and 7 (rs941439). In the case control association study from Wales (Williams et al. 2004) two different SNP markers (4 (rs951436) and 18 (rs2661319)) showed association with p values of 0.017 and 0.038.

It is possible that the original study (Chowdari et al. 2002) may have exaggerated the effect of association by confounding it with linkage which is known to be present in the region where association is claimed. The methods implemented in GENEHUNTER and TRANSMIT for detecting allelic association do not always

eliminate the possibility of linkage being confounded with association especially when the commonest allele is found to be associated with the disease as in the case of the Pittsburgh and NIMH samples and where multiple cases within a sibship rather than just one case in a family are used to test for association. The authors do not specify whether they have used the option of using the most stringent test of association in family data where only one case per family is used in the TRANSMIT analyses. Such a procedure would have the effect of possibly removing some of the confounding of linkage with association. In any case the single marker evidence for transmission disequilibrium in the original Pittsburgh sample (p=0.05) was modest with SNP 4. In the analysis of the NIMH sample the authors also do not state what options for the global test of association were used for the TRANSMIT analyses. Different SNP markers were found to be associated, but the p values were stronger at 0.01, 0.003 and 0.005 for markers 1(rs10917670), 4 (rs951436) and 18 (rs2661319) respectively.

However the Irish and Welsh studies do not suffer from the problems that tests of association using family data can suffer from, but the significance of the associations between RGS4 SNPs and schizophrenia were not found to be strong. The discrepancy between our results and those of the Pittsburgh, Welsh and Irish studies could be due to locus heterogeneity with different proportions of disease alleles being present in different geographical regions.

However replication of the results through a more powerful meta-analysis has also failed to find clear positive independent marker or haplotypic association with family samples and no haplotypic association with case control samples (Talkowski et al. 2006). The meta-analysis collectively found modest association of over transmission of both common haplotypes consisting of SNPs 1-4-7-18 (haplotypes A-T-A-A & G-G-G-G). Talkowski et al (2006) had attributed the phenomenon to biological, statistical, molecular and population factors. This could happen as a result of recurrent mutations or transfer of liability alleles between haplotypes by recombination which accounts for the common haplotypes, also by allelic

heterogeneity or contribution of multiple loci to susceptibility or by a single rare mutation on a background of two common haplotypes.

It was also claimed by Talkowski et al (2006) that lack of consistent results could be attributed to lack of quality control over samples and genotyping errors could mimic biased transmission of common alleles/haplotypes. Over all the weight of the data which is inconsistent, the lack of good positive association with a large meta-analysis data and the failure to replicate the positive results in our own case/control sample greatly weakens RGS4 as a putative candidate susceptibility gene for schizophrenia.

A more recent study by Ishiguro et al (2006), in a large Japanese population (1918 cases, 1909 controls) also failed to find association with the four SNPs previously implicated by Chowdari at al 2002 (Table 4:3) Guo et al (2006) had also found no association with the four SNPs and also in two additional SNPs (rs2842030 and rs2344671) in a Han Chinese population (Table 4:3). In addition to the replication study, Guo et al had also carried out their own meta-analysis study which consisted of 3062 cases and 3564 controls, in which they had found no evidence of association with RGS4 to the disease.

Both meta-analysis and new data, shows that RGS4 is still not a strong or likely susceitibility gene for schizophrenia. Its expression may be altered due to drug effects or interaction with genes directly involved with the disease, but it does not seem to be a likely susceptibility gene for schizophrenia.

### 4.2.5 RGS4 CONCLUSION

Consideration should be given to the possibility that the observed allelic association could implicate other genes which are close enough to show linkage disequilibrium with the RGS4 marker alleles. Our findings, in addition to the two meta-analyses carried out (Guo et al. 2006; Talkowski et al. 2006) diminish support for the notion

that RGS4 gene polymorphisms influence susceptibility to schizophrenia. They also suggest that more research is needed to prove that the RGS4 gene is implicated in the genetic susceptibility to schizophrenia.

**Table 4:1 Tests of Allelic and Haplotypic Association with Schizophrenia at the RGS4 Gene Locus.**

| Marker | Marker location in base pairs on C1 | bases from previous marker | Allelic bases or fragment sizes and observed allele frequencies below | | | | | | | Chi$^2$ | P$^{ab}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| SNP 4 (rs951436) | 159765000 | | G | T | | | | | | 0.06 | 0.80[a] |
| Con | | | 445 | 413 | | | | | | | |
| Scz | | | 426 | 386 | | | | | | | |
| SNP 7 (rs951439) | 159765349 | 349 | G | A | | | | | | 0.021 | 0.88[a] |
| Con | | | 512 | 346 | | | | | | | |
| Scz | | | 491 | 327 | | | | | | | |
| SNP 18 (rs2661319) | 159771435 | 6086 | G | A | | | | | | 0.077 | 0.78[a] |
| Con | | | 470 | 394 | | | | | | | |
| Scz | | | 447 | 385 | | | | | | | |
| $(GT)_{14}$ | 159783591 | 12156 | 274 | 276 | 278 | 280 | 282 | 284 | 286 | 2.126 | 0.60[T3] |
| Con | | | 262 | 1 | 215 | 107 | 84 | 175 | 12 | | |
| Scz | | | 217 | 1 | 177 | 94 | 71 | 131 | 17 | | |

| Breakdown of the estimated three SNPs markers-haplotypes with frequency > 0.01 | T-A-A | T-A-G | T-G-A | G-G-G |
|---|---|---|---|---|
| Con | 0.377 | 0.024 | 0.079 | 0.519 |
| Scz | 0.382 | 0.015 | 0.077 | 0.52 |

| | | |
|---|---|---|
| Global haplotypic association combining all SNPs | Empirical p value | 0.43[b] |
| Global haplotypic association combining all SNPs with microsatellite markers $GT_{14}$ | Empirical p value | 0.79[b] |

[a] p value for 2x2 chi squared (1 df)
[T3] Most significant p value from CLUMP
[b] global haplotypic association permutation test p value based on 9,999 permutations

Table 4:2 Allele frequencies of SNPs 4, 7 and 18 in the UCL case control sample in relation to the previous case control studies of RGS4 and schizophrenia.

| | | Rizig et al (Caucasian) | | Williams et al (Caucasian) | | Morris et al (Caucasian) | | Corderio et al (Brazilian) | | Guo et al (Han Chinese) | | Ishiguro et al (Japanese) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Marker | Polymorphism | Con (n=450) | Scz (n=450) | Con (n=710) | Scz (n=709) | Con (n=231) | Scz (n=196) | Con (n=576) | Scz (n=271) | Con (n=288) | Scz (n=288) | Con (n=1909) | Scz (n=1918) |
| SNP 4 (rs951436) | T/G | 0.481 | 0.475 | 0.468 | 0.514 | 0.524 | 0.464 | 0.592 | 0.586 | 0.45 | 0.51 | 0.486 | 0.51 |
| SNP 7 (rs951439) | G/A | 0.597 | 0.6 | 0.606 | 0.587 | 0.539 | 0.609 | 0.567 | 0.574 | 0.51 | 0.51 | 0.508 | 0.51 |
| SNP 18 (rs2661319) | A/G | 0.456 | 0.463 | 0.395 | 0.407 | 0.496 | 0.446 | 0.576 | 0.556 | 0.47 | 0.48 | 0.453 | 0.485 |

**Table 4:3 Results of RGS4 previously reported genetic association studies.**

| | Author | Year | Sample type | Sample size | Statistics | P-value |
|---|---|---|---|---|---|---|
| | **Chowdari, K.V *et al*** | **2002** | **Pittsburgh Case- control** | **Cases = 59, controls = 89** | **TDT (transmit)** | |
| SNP 1 | | | | | | 0.055 |
| SNP 4 | | | | | | 0.05 |
| SNP 7 | | | | | | >0.05 |
| SNP 18 | | | | | | >0.05 |
| Haplotype[A] | | | | | **TDT (global)** | 0.035 |
| | | | **NIMH family sample** | **25 families** | **TDT (transmit)** | |
| SNP 1 | | | | | | 0.01 |
| SNP 4 | | | | | | 0.003 |
| SNP 7 | | | | | | >0.05 |
| SNP 18 | | | | | | 0.005 |
| Haplotype[A] | | | | | **TDT (global)** | 0.016 |
| | | | **Indian trios** | **269 case parent trio + 72 sib pair families** | **TDT (transmit)** | |
| SNP 1 | | | | | | >0.05 |
| SNP 4 | | | | | | >0.05 |
| SNP 7 | | | | | | >0.05 |
| SNP 18 | | | | | | >0.05 |
| Haplotype[A] | | | | | **TDT (global)** | 0.078 |
| | **Morris, D.W *et al*** | **2004** | **Irish Case-Control** | **249 cases 231 controls** | **Chi squared** | **Values are excluding cases with schizoaffective disorders** |
| SNP 1 | | | | | | 0.042 |
| SNP 4 | | | | | | 0.082 |
| SNP 7 | | | | | | 0.041 |
| SNP 18 | | | | | | 0.149 |
| Haplotype[A] | G-G-G-G | | | | **FASTEHPLUS** | 0.044 |
| | **Williams, N.M *et al*** | **2004** | **UK born Caucasins** | **709 cases 710 controls** | **Chi squared** | |
| SNP 1 | | | | | | 0.51 |
| SNP 4 | | | | | | 0.017 |
| SNP 7 | | | | | | 0.32 |
| SNP 18 | | | | | | 0.038 |
| Haplotype[A] | 1-4 | | | | **EHPLUS** | 0.016 |
| | 1-4-7-18 | | | | | 0.266 |
| | **Chen X *et al*** | **2004** | **Irish families (ISHDF)** | **247 pedigrees** | **FBAT** | |
| SNP 1 | | | | | | 0.35 |
| SNP 4 | | | | | | 0.11 |
| SNP 7 | | | | | | 0.43 |
| SNP 18 | | | | | | 0.04 |
| Haplotype[A] | 1-4-18 | | | | **SIMWALK2 (global)** | 0.04 |
| | 1-4-7-18 | | | | | 0.20 |
| | **Zhang *et al*** | **2005** | **Han Chinese** | **322 families** | **PDTPHASE** | |
| SNP 1 | | | | | | 0.88 |
| SNP 4 | | | | | | 0.72 |
| SNP 7 | | | | | | 0.24 |
| SNP 18 | | | | | | 0.27 |
| Haplotype[A] | 1-4-7-18 | | | | | 0.72 |
| | | | **Scottish Case -control** | **580 cases 620 Controls** | **COCAPHASE** | |
| SNP 1 | | | | | | 0.04 |
| SNP 4 | | | | | | 0.082 |
| SNP 7 | | | | | | 0.011 |
| SNP 18 | | | | | | 0.033 |
| Haplotype[A] | 4-7 | | | | | 0.01 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | 1-4-7-18 | | | | | 0.10 |
| | **Sobell *et al*** | **2005** | **Western/Northern European Caucasians** | **568 cases 689 controls** | **Chi square** | |
| SNP 1 | | | | | | 0.11 |
| SNP 4 | | | | | | 0.39 |
| SNP 7 | | | | | | 0.11 |
| SNP 18 | | | | | | 0.52 |
| Haplotype[A] | | | | | **Modified EM** | >0.05 |
| | **Rizig *et al*** | **2006** | **British Caucasians** | **450 Cases 450 Controls** | **Chi Square** | |
| SNP 4 | | | | | | 0.80 |
| SNP 7 | | | | | | 0.88 |
| SNP 18 | | | | | | 0.078 |
| Haplotype[A] | | | | | **GENECOUNTING** | 0.43 |
| | **Liu *et al*** | **2006** | **Taiwaneese** | **218 nuclear Families** | **TRANSMIT** | |
| SNP 4 | | | | | | 0.237 |
| SNP 7 | | | | | | 0.356 |
| SNP 18 | | | | | | 0.397 |
| Haplotype[A] | | | | | | 0.277 |
| | **Ishiguro *et al*** | **2006** | **Japanese** | **1918 Cases 1909 Controls** | | |
| SNP 1 | | | | | | 0.92 |
| SNP 4 | | | | | | 0.91 |
| SNP 7 | | | | | | 0.27 |
| SNP 18 | | | | | | 0.43 |
| Haplotype[A] | | | | | | n/s |
| | **Guo *et al*** | **2006** | **Han Chinese** | **288 Cases 288 Controls** | **SHEsis** | |
| SNP 1 | | | | | | 0.64 |
| SNP 4 | | | | | | 0.11 |
| SNP 7 | | | | | | 0.35 |
| SNP 18 | | | | | | 0.80 |
| rs2842030 | | | | | | 0.35 |
| Rs2344671 | | | | | | 0.35 |
| Haplotype[A] | | | | | **SHEsis Global haplotype** | 0.97 |
| **Meta-analysis** | | | | **3062 Cases 3564 Controls** | | n/s |
| | **Talkowski *et al*** | **2006** | | **2160 Families** | | **Combined results global results** |
| | **Meta-analysis** | | | | | |
| SNP 1 | | | | | | n/s |
| SNP 4 | | | | | | n/s |
| SNP 7 | | | | | | n/s |
| SNP 18 | | | | | | n/s |
| Haplotype[A] | | | | | | n/s |
| | | | | **3486 Cases 3755 Controls** | | |
| SNP 1 | | | | | | n/s |
| SNP 4 | | | | | | 0.01 |
| SNP 7 | | | | | | n/s |
| SNP 18 | | | | | | n/s |
| Haplotype[A] | | | | | | n/s |

[A] the most significant haplotype.
n/s not significant

# 5.0 ATTEMPTED REPLICATION STUDY OF ASSOCIATION BETWEEN THE CAPON (NOS1AP) NITRIC OXIDE SYNTHASE 1 (NEURONAL) ADAPTOR PROTEIN AND SCHIZOPHRENIA

## 5.1 GENERAL INTRODUCTION

Another gene in the 1q23.3 region had been associated with schizophrenia by fine mapping. The gene originally known as CAPON was implicated by Brzustowicz *et al* (2004) who had also implicated the region by linkage analysis using one of the original markers that was used for the association study D1S1679 (Brzustowicz et al. 2000). The same region was implicated in two linkage studies that found significant evidence for linkage with lod scores of 6.35 and 3.20 (Brzustowicz et al. 2000; Gurling et al. 2001). A third study also found evidence supportive of linkage to schizophrenia at 1q23.3 (Shaw et al. 1998).

The region of strongest linkage was further examined in 24 Canadian familial schizophrenia pedigrees. 14 microsatellites and 15 SNPs were genotyped between D1S1653 and D1S1677 of which 2 microsatellites and 6 SNPs had shown significant evidence of association. All of the markers exhibiting significant association with schizophrenia fell within the genomic extent of the gene for carboxyl-terminal PDZ ligand of neuronal nitric oxide synthase (*CAPON*) (Brzustowicz et al. 2004) now known as Nitric Oxide Synthase 1 (neuronal) Adaptor Protein (*NOS1AP*).

CAPON's location and identification through association analysis first, rather than through expression analysis and its neuronal expression and function made the gene an attractive candidate for schizophrenia susceptibility.

CAPON (NOS1AP) is involved in signal transduction in the N-methyl-D-aspartate (NMDA) receptor mediated glutamate neurotansmision system, abnormalities of which have long been proposed to be involved in schizophrenia. It functions by competitively binding with neuronal nitric oxide synthase (nNOS) which prevents nNOS from forming a complex with the NMDA receptor hence preventing calcium reflux and signal transduction. Over expression of CAPON is thought to lead to impaired NMDA receptor mediated neurotransmission, this highlights the potential importance of this pathway in the aetiology of schizophrenia.

## 5.2 REPLICATION OF ORIGINAL CAPON STUDY

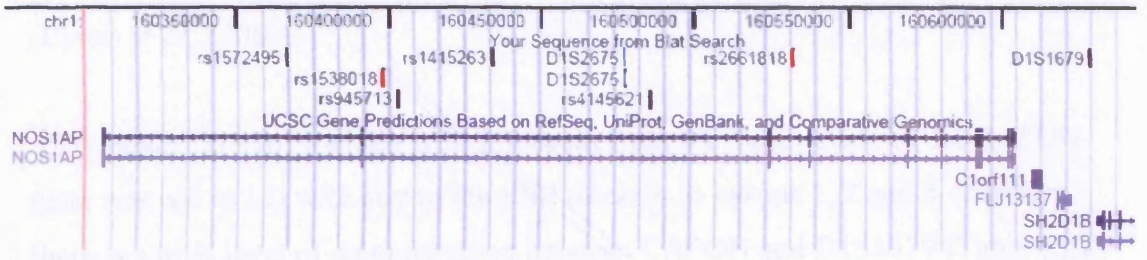
We attempted to replicate the original association implicated by Brzustowicz et al, (Brzustowicz et al. 2004) in our UK case control sample using their associated markers. The experimental set up and results are shown in the following sections.

## 5.3 INTRODUCTION

The Canadian family sample which showed linkage to schizophrenia at 1q23.3 (Brzustowicz et al. 2004) was used to test for allelic association with schizophrenia using the methods implemented in PSEUDOMARKER (Goring et al. 2000; Brzustowicz et al. 2004). This approach claims to be able to detect allelic association in a family sample by correcting for the confounding effects of linkage (Goring et al. 2000). The authors stated that their sample size was equivalent to a collection of forty five independent trios. Evidence for significant allelic association with schizophrenia ($p=0.015$) was found using two microsatellite and six SNP markers at the CAPON locus. A Chinese case control study showed significant association between a single SNP at CAPON and schizophrenia but failed to find association with any of the six SNP markers which had been implicated in the Canadian sample (Zheng et al. 2005). A Spanish study found evidence for positive extended transmission disequilibrium (ETDT) with schizophrenia using the marker D1S1679 ($p=0.004$) at 1q23.3 which is 23 kilobases distal to the 3 prime end of the CAPON

gene. ETDT analysis is not able to separate the effects of linkage from allelic association (Sham et al. 1995; Rosa et al. 2002). The marker D1S1679 was not associated with schizophrenia (p=0.9) in the Canadian study. Positive allelic association between markers at the RGS4 gene on 1q23 and schizophrenia has been reported. The RGS4 gene is 700 kilobases distal to CAPON (Chowdari et al. 2002; Chen et al. 2004; Morris et al. 2004; Williams et al. 2004). Although, both CAPON, D1S1679 and RGS4 are in the same region it is unlikely that the positive results reported at the CAPON locus are due to linkage disequilibrium with RGS4 which is 700 Kb away and vice versa, but linkage disequilibrium between alleles at D1S1679 and CAPON is possible as is LD between RGS4 and a susceptibility locus in the middle of the 700kb region between CAPON and RGS4.

**Figure 5:0 Showing the location of SNPs and the microsatellites in the region of CAPON/NOSA1P (UCSC May 2004 assembly, BLAT diagram).**




## 5.4 GENERAL METHODS AND MATERIALS

The positive Canadian study markers were genotyped in a British ancestrally matched sample of DNA from 450 cases of schizophrenia 450 controls. SNPs and microsatellite markers were genotyped using primer sequences available from the NCBI databases (http://www.ncbi.nlm.nih.gov/). Genotyping was checked by a second independent person blind to diagnosis. Any discrepant genotypes were repeated. SNPs were determined by a modified Amplifluor method (KBiosciences, Hoddesdon, UK). 17% of genotypings were reduplicated in a separate microtitre plate to detect error (section 2.8.3.1). Markers with lack of Hardy Weinberg Equilibrium were rejected and genotyping repeated. The data were analysed for allelic association with schizophrenia using CLUMP (Sham et al. 1995). Haplotypic

association with schizophrenia were tested using GENECOUNTING and an empirical permutation test of significance (Zhao et al. 2000; Zhao et al. 2002). LD relationship between markers were also analysed using the same software. The full method is described in the method section 2.0.

## 5.5 RESULTS AND DISCUSSION

Reduplication of SNPs showed an accuracy rate of 99.4%. None of the markers implicated in the Canadian study were found to show allelic, genotypic or haplotypic association with schizophrenia (Table 5:1). Six SNPs in introns 1, 2 and 3 of CAPON were found to be in strong LD with D1S2675 (all at $p<0.00007$) (LDPAIRS p value from Cramer's V) (Table 5:2). It should be noted here that the more familiar D prime and $r^2$ statistics cannot be used for measure of LD with multiallelic microsatellite markers. Cramer's V (Bishop *et al.* 1975) is calculated and is output along with the associated p-value in order to show LD with multiallelic markers (Curtis *et al.* 2006).

The marker D1S1679 which is 23 kilobases from the 3 prime end of the CAPON gene was not in LD with any of the SNP markers in introns 1, 2 and 3, therefore there is a high level of recombination between CAPON and D1S1679 (Table 5:2). Introns 4, 5, 6 and 7 and exons 4 to 8 of CAPON are contained within a distance of only 67 kilobases and allelic association over this part of the gene should be detectable with D1S1679. Both the Canadian and the UK samples were negative for this marker.

The Chinese case control study which also claims to have implicated CAPON in schizophrenia (Zheng et al. 2005) only found association with a single SNP marker (rs348624). Other markers in linkage disequilibrium with rs348624 such as rs905721 should also have shown some evidence for association but did not. In addition rs348624 was not significant in the original study conducted by Brzustowicz et al (2004). The Chinese study did not genotype the marker D1S2675

that showed association (p=0.036) in the Canadian sample but which was negative in the UK study. One explanation for the different outcomes may be that different proportions of CAPON susceptibility alleles were present each sample. Alternatively it may be that CAPON does not influence susceptibility to schizophrenia and it is a false positive result. The study by Zheng et al (2005) might also be false positive because of the lack of support for association from adjacent markers. Finally, our own results may represent a false negative finding.

We used two microsatellite markers flanking the five prime end of CAPON gene and in the second intron and this suggests that additional markers would also fail to show association. There were equal proportions of successful schizophrenia and control genotypes obtained for the SNP and microsatellite markers except for marker D1S2675 for which there was a 19% excess of controls over cases. Further genotypes were not obtained because the sample already had more power to exclude association between schizophrenia and CAPON than the original study. The lack of evidence of association between schizophrenia and CAPON is further supported by our finding that D1S2675 shows strong linkage disequilibrium with all the SNPs found to be associated with schizophrenia within CAPON and that the microsatellite marker D1S1679 also shows strong linkage disequilibrium with the SNP markers that are in the 3 prime region of CAPON up to 210.35 kilobases distal from D1S1679. Brzustowicz et al (2004) had sequenced all the exons of the CAPON gene, but had found no aetiological mutations, however an aetiological mutation may still be present elsewhere. There could be a mutation affecting gene-splicing, the 5' UTR and the 3' UTR affecting the promoter region, transcription factor binding, RNA stability and enhancer elements. There is a possibility that association with CAPON is more revelent in the Chinese population than the European population therefore association with the SNP marker, rs348624, should be replicated in a Chinese rather than a European sample. Our findings however, diminish support for the notion that the CAPON gene influences susceptibility to schizophrenia in European samples.

## 5.6 FURTHER PUBLICATIONS ON CAPON

Recent research by Xu et al (2005) from Brzustowicz's group, examined the expression of CAPON, to test if gene expression is altered in schizophrenic patients and bipolar patients compared to controls (Xu et al. 2005). Two alternative transcripts were found in a fetal brain cDNA library by using exon 10 as a probe. Exon 10 was used as a probe on the brains as it contains an nNOS-binding PDZ domain which should therefore hybridise to neuronal transcripts containing the domain. One transcript was the original full length transcript, the second was a shortened transcript containing just exons 9 and 10.

Expression of the transcripts were assessed in the dorsolateral prefrontal cortex (which is thought to be a pathological hotspot) that has been identified in studies of schizophrenia (Xu et al. 2005) in 35 schizophrenic, 35 bipolar and 35 "control" brains. They have found that there was increased expression of the short form of the transcript in both the schizophrenics and bipolar brains when compared to the controls. In addition they had also found the individuals associated with the 3 most associated SNPs (rs1415263, rs4145621 and rs2661818) implicated to the disease also led to increased expression of the short form of the transcript. As the short form lacked the phosphotyrosine binding domain which is needed to bind to the amino terminal targets Dextras1 and Synapsin, the role of the protein would be limited to the competitive inhibition of the binding of other ligands to the PDZ domain of nNOS and PSD93 or PSD95.

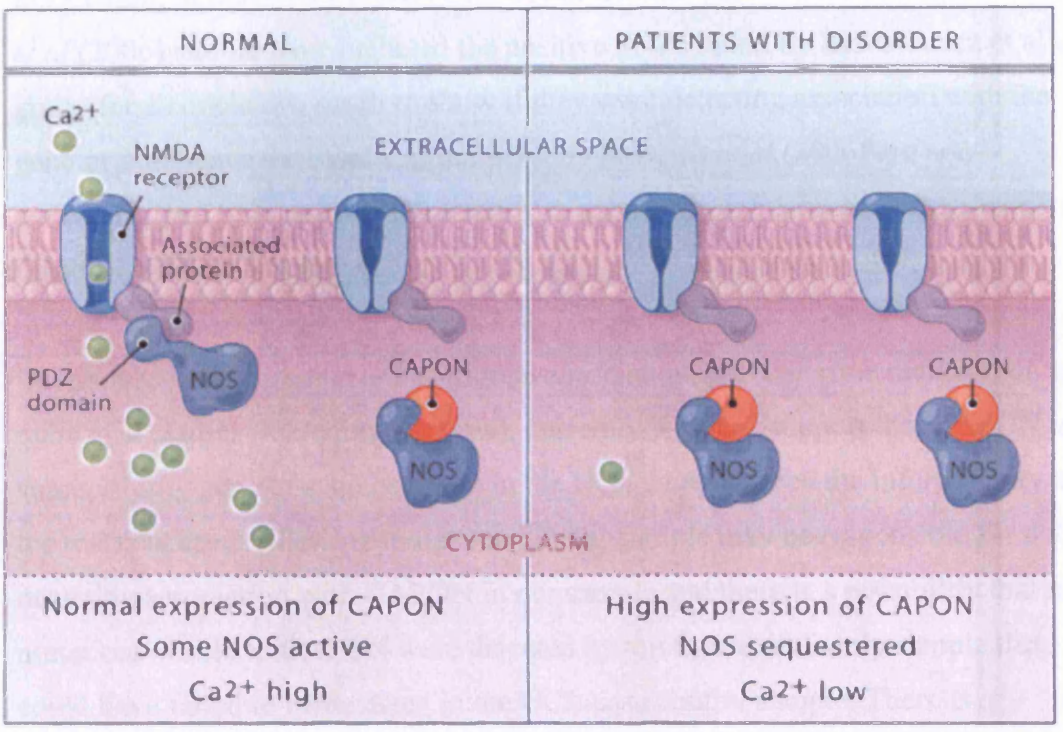



**Figure 5:1 The figure shows the Eastwood et al (2005) interpretation on how expression levels of CAPON can lead to impaired NMDA receptor meditated glutamate neurotransmission, potentially causing schizohrenia. Nitric oxide synthase (NOS)-CAPON complexes result in reduced NMDA-NOS complexes leading to decreased NMDA receptor gated calcium influx and inactive NOS.**

However, this study had a small sample size, lacked statistical significance and was not replicated of the results in another brain series. In addition with lack of evidence of other brain regions and histo-chemical studies such as in-situ hybridisation to show altered expression as confirmatory evidence, the evidence seems weak. Also the resulting protein from the truncated transcript would also need to be assessed if it was functional, as this could cause structural alterations to the protein, leaving the protein inert.

In addition, Miranda et al (2006) had reported a putative association with CAPON. The association was not a replication using the same positive SNPs as implicated by Bruzustowicz et al (2004), but only with the use of a single microsatellite (D1S1679, p=0.019) in 110 Columbian family trios. The original study including our own (Brzustowicz et al. 2004; Puri et al. 2006) did not find association with D1S1679. It

is possible that Miranda et al's results may have confounded linkage with association, or detected LD with CAPON or another gene within the region. Miranda *et al* (2006) should have included the positive SNPs found by Brzustowicz et al's group for a conclusive result to show if they were detecting association with the gene or not, hence the results of the work by Miranda *et al* (2006) are non-conclusive and premature in implicating or supporting evidence in the association of CAPON with schizophrenia.

The failure to fully replicate the original association with the same markers, of the published studies (including our own), currently weakens support that CAPON is a susceptibility gene of schizophrenia in the UCL sample given the informativity of the markers used. Allelic heterogeneity in our sample may be responsible for not detecting association with CAPON in our sample and there is a possibility that rare mutations within in CAPON were detected by the Canadian family sample that could have failed to be detected in the UCL case control sample. There is a speculative possibility by fully fine mapping CAPON with Tagged SNPs in a larger case control sample; one may be able to detect association with CAPON. However we have failed to detect association with the markers in which the original study had.

Taking the example of rs1415263, the most associated SNP in the Brzustowicz et al (2004) study, our sample had the power of approximately 0.8 to detect association.

Our failure to replicate the CAPON and RGS4 associations could be due to genetic heterogeneity or might arise from the previous positive studies actually picking up evidence for allelic association due to a susceptibility locus placed in between CAPON and RGS4. In our data there was no detectable LD between any of the markers within CAPON and any of the markers within RGS4 (see Table 6:4). It should be noted that the study by Brzustowicz et al which found association between schizophrenia and CAPON gene marker alleles (Brzustowicz et al. 2004) had a sample size equivalent to only forty five cases and forty five controls and had a

power of 11% to detect a 10% frequency difference, at p=0.05, for a marker allele frequency of 0.43 in controls and 0.53 in cases. By contrast our sample had a power of 83% to detect such a frequency difference. Thus there is a risk that the original CAPON result has arisen by chance or is due to the confounding of linkage with association because of the strong evidence of linkage to the CAPON region in this family sample. The study by Chowdari et al (2002) implicating RGS4 has been subjected to a meta-analysis in a very large sample (Talkowski et al. 2006). Despite the size of the sample, significant associations with individual SNPs and haplotypes were not observed in the combined meta-analysis. The global analysis revealed significant transmission distortion (p = 0.0009) for two very common haplotypes. Separate analyses of 3486 cases and 3755 control samples (eight samples) detected only a significant association with RGS4 SNP 4 (rs951436) (p = 0.01).
In the original study when the US patients were contrasted to two population-based control samples no significant differences were observed. Thus the evidence for involvement of RGS4 in schizophrenia derives more from transmission distortion in families than it does from allele frequency differences in case control samples.

With the lack of evidence for association with Regulator of G-Protein Signalling 4 (RGS4), (section 4.0) and also Nitric Oxide Synthase 1 (neuronal) Adaptor Protein (NOS1AP/CAPON) with the British case control sample, this leaves the possibility that the linkage signal in the 1q23-22 region may be due to another gene in the region. Hence the region should be further fine mapped to detect this signal with a better resolution using an even larger case control sample.

Table 5:1 Tests of Allelic and Haplotypic Association with Schizophrenia at the CAPON Gene Locus.

| Marker | Marker location in base pairs on C1 | Bases from previous marker | % genotyped | Allelic bases or fragment sizes and observed allele frequencies below | | | | | | | | | | | Chi² | P[a,b] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rs1572495 | 158830959 | | | C | T | | | | | | | | | | 0.409 | 0.5223[a] |
| Con | | | 97.60% | 797 | 81 | | | | | | | | | | | |
| Scz | | | 86.70% | 715 | 65 | | | | | | | | | | | |
| rs1538018 | 158862139 | 31180 | | G | C | | | | | | | | | | 0.414 | 0.5199[a] |
| Con | | | 93.10% | 631 | 207 | | | | | | | | | | | |
| Scz | | | 86.70% | 598 | 182 | | | | | | | | | | | |
| rs945713 | 158867328 | 5189 | | C | T | | | | | | | | | | 0.023 | 0.8782[a] |
| Con | | | 95.60% | 321 | 539 | | | | | | | | | | | |
| Scz | | | 86.70% | 294 | 486 | | | | | | | | | | | |
| rs1415263 | 158897701 | 30373 | | C | T | | | | | | | | | | 1.238 | 0.2659[a] |
| Con | | | 97.80% | 549 | 331 | | | | | | | | | | | |
| Scz | | | 85.80% | 502 | 270 | | | | | | | | | | | |
| D1S2675 | 158941619 | 43918 | | 168 | 174 | 176 | 182 | 184 | 186 | 188 | 190 | 192 | 194 | 196 | 4.533 | 0.1703[T3] |
| Con | | | 75.60% | 16 | 2 | 1 | 6 | 94 | 343 | 88 | 107 | 18 | 4 | 1 | | |
| Scz | | | 56.40% | 17 | 3 | 1 | 3 | 70 | 278 | 61 | 58 | 10 | 4 | 3 | | |
| rs4145621 | 158950346 | 8727 | | C | T | | | | | | | | | | 0.762 | 0.3828[a] |
| Con | | | 93.80% | 360 | 484 | | | | | | | | | | | |
| Scz | | | 78.00% | 284 | 418 | | | | | | | | | | | |
| rs2661818 | 158996472 | 46126 | | G | C | | | | | | | | | | 0.245 | 0.6208[a] |
| Con | | | 94.40% | 394 | 456 | | | | | | | | | | | |
| Scz | | | 87.10% | 346 | 438 | | | | | | | | | | | |
| D1S1679 | 159093422 | 96950 | | 164 | 168 | 172 | 176 | 180 | 184 | 188 | 192 | 196 | 200 | | 8.908 | 0.115[T4] |
| Con | | | 97.80% | 7 | 159 | 102 | 113 | 144 | 170 | 132 | 41 | 11 | 1 | | | |
| Scz | | | 96.00% | 12 | 119 | 92 | 139 | 133 | 165 | 154 | 41 | 9 | 0 | | | |
| 8 Marker Haplotypes | | | | | | | | | | | | | | Empirical p value | | 0.441[b] |

[a] significance p from 2x2 chi squared 1 df or most significant p value from CLUMP Monte Carlo T1, T2, T3 or T4 as shown, [b] haplotype permutation test empirical p, based on 99,999 permutations

Table 5:2 showing structure of Linkage disequilibrium between markers within CAPON with D', $r^2$ and p-value measurements.

| Marker | Position | Distance between markers (bp) | Absolute value of D' | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Rs1572495 | rs1538018 | rs945713 | rs1415263 | D1S2675 | rs4145621 | rs2661818 | D1S1679 |
| rs1572495 | 160365925 | | | 0.156 | 0.662 | 0.209 | 0.016 | 0.028 | 0.411 | 0.081 |
| rs1538018 | 160397105 | 31180 | 0.00757 | | 0.134 | 0.007 | 0.225 | 0.017 | 0.101 | 0.054 |
| rs945713 | 160402294 | 5189 | 0.07129 | 0.00941 | | 0.667 | 0.392 | 0.608 | 0.538 | 0.09 |
| rs1415263 | 160432667 | 30373 | 0.00740 | 0.00004 | 0.42250 | | 0.784 | 0.869 | 0.677 | 0.003 |
| D1S2675 | 160476585 | 43918 | 0.05570 | 0.09734 | 0.22753 | 0.44756 | | 0.679 | 0.392 | 0.228 |
| rs4145621 | 160485312 | 8727 | 0.00010 | 0.00012 | 0.31923 | 0.62568 | 0.42903 | | 0.769 | 0.019 |
| rs2661818 | 160531438 | 46126 | 0.02045 | 0.00260 | 0.21437 | 0.32490 | 0.19184 | 0.50980 | | 0.024 |
| D1S1679 | 160628388 | 96950 | 0.02103 | 0.01538 | 0.01904 | 0.01613 | 0.01638 | 0.02103 | 0.01850 | |

$r^2$ value

p values

| | rs1572495 | rs1538018 | rs945713 | rs1415263 | D1S2675 | rs4145621 | rs2661818 | D1S1679 |
|---|---|---|---|---|---|---|---|---|
| rs1572495 | | 0.01194 | <0.00001 | 0.01235 | 0.00007 | 0.75171 | 0.00002 | 0.0926 |
| rs1538018 | | | 0.00552 | 0.80679 | <0.00001 | 0.70815 | 0.12939 | 0.18255 |
| rs945713 | | | | <0.00001 | <0.00001 | <0.00001 | <0.00001 | 0.07619 |
| rs1415263 | | | | | <0.00001 | <0.00001 | <0.00001 | 0.11205 |
| D1S2675 | | | | | | <0.00001 | <0.00001 | 0.57474 |
| rs4145621 | | | | | | | <0.00001 | 0.08822 |
| rs2661818 | | | | | | | | 0.1334 |
| D1S1679 | | | | | | | | |

# 6.0 FINE MAPPING OF THE UHMK1 GENE ENCODING A SERINE/THREONINE PROTEIN KINASE, AS A NOVEL SCHIZOPHRENIA SUSCEPTIBILITY GENE

Linkage studies by us and others have confirmed that chromosome 1q23.3 is a susceptibility locus for schizophrenia (Brzustowicz et al. 2000; Gurling et al. 2001). Based on this information several research groups published evidence that markers within both the RGS4 and CAPON genes, which are 700Kb apart, independently showed allelic association with schizophrenia (Chowdari et al. 2002; Brzustowicz et al. 2004). Tests of allelic association with both of these genes in our case control sample were negative (Puri et al. 2006; Rizig et al. 2006) section 4.0-5.0. Therefore we carried out further fine mapping in the 1q23.3 locus in the 700Kb region between the RGS4 and CAPON genes.

Microsatellites were first designed across the region and genotyped in the case control sample, three of which detected positive allelic association, implicating three genes SH2D1B, UHMK1 and HSD17B7. The region between SH2D1B and UHMK1 was more positively associated and hence further fine mapped with SNPs to ascertain which was the most likely gene to be associated with schizophrenia. In the following sections a description of the fine mapping of the region along with the results will be presented, leading to an initial implication of a susceptibility gene. This is followed by sequencing results, further evidence of association using tagging SNPs and in addition, an attempt to replicate our findings in an independent case control sample from the university of Aberdeen donated by Professor St Clair.

## 6.1 INTRODUCTION

Association studies of schizophrenia at 1q23.3 have implicated both CAPON (Brzustowicz et al. 2004; Zheng et al. 2005)and RGS4 (Chowdari et al. 2002; Chen et al. 2004; Morris et al. 2004; Williams et al. 2004). In the British case-control sample employed in this study we failed to replicate either of these associations (Puri et al. 2006; Rizig et al. 2006). The distance between CAPON and RGS4 is approximately 700Kb. In order to investigate whether the positive results at CAPON and RGS4 might actually be due to linkage disequilibrium (LD) with a closely linked gene, we carried out fine mapping in the region with a further set of markers as reported below

## 6.2 METHOD

SNPs rs164126, rs164123, rs3121196, rs351453, rs164128, rs10494370, rs7513662, rs423227, rs6427680 were chosen from the International HapMap Project and previously unformatted microsatellite repeats were identified from the UCSC Genome Browser (May 2004 assembly). These new repeats were formatted by us for genotyping. The primers were designed with Primer3 (Rozen et al. 2000). The microsatellite markers are now listed in the GDB database with the D segment numbers D1Z12, D1Z13, D1Z14, D1Z15, D1Z16, D1Z17, D1Z18 and D1Z19. The positions and primer sequences of these markers are shown in Table 6:1. PCR amplification of the microsatellite markers was carried out using an M13 tailed primer and a second non-tailed primer. A third universal M13 sequence primer labelled with IRD 700 or IRD 800 was used to hybridise against the M13 tailed locus specific primer. Microsatellite marker fragment sizes were separated and visualised with either of two infrared dyes IRD700 and IRD800 on dual argon laser LiCor 4200L sequencers. Genotyping was carried out with the SAGA-GT genotyping software (LiCor) and checked by eye. Allele calling by SAGA-GT was checked by a second independent rater blind to diagnosis. Any genotypes discrepant between the two raters were PCR-amplified and run again. SNPs were determined by the Amplifluor SNP genotyping method as modified by KBiosciences. 17% of

samples on each microtitre plate were reduplicated in order to detect error and confirm the reproducibility of genotypes. The data were then analysed to confirm Hardy Weinberg equilibrium (HWE). Markers with lack of HWE in the control group were rejected and genotyping was repeated. Before association analysis the genotype data for each 96 well microtitre plate were analysed for linkage disequilibrium with closely linked markers to ensure that LD relationships were similar for each plate.

Next this data were analysed for allelic association with schizophrenia using CLUMP (section 3.2.3.2), which employs an empirical Monte Carlo test of significance and which does not require correction for multiple alleles (Sham et al. 1995).

The genotypes were then analysed for haplotypic association with schizophrenia using GENECOUNTING, which computes maximum likelihood estimates of haplotype frequencies from phase unknown case control data, and the significance of any overall haplotype association with schizophrenia was computed with a permutation test (Zhao et al. 2000; Zhao et al. 2002; Curtis et al. 2006). GENECOUNTING was also used to calculate pair-wise linkage disequilibrium between all markers.

## 6.3 RESULTS

Twenty nine markers in the 1q23.3 region were genotyped as shown in Figure 6:1. These included markers at the CAPON and RGS4 loci which had previously failed to show any evidence for association with schizophrenia (Puri et al. 2006; Rizig et al. 2006) in section 4.0 and 5.0. The first markers to be genotyped were microsatellites within the 700 Kb region between CAPON and RGS4. As shown in Table 6:2, three of these produced evidence for association with schizophrenia: D1Z12, D1Z13 and D1Z18. D1Z12 (significant at p=0.011) is a $(CA)_{20}$ dinucleotide repeat 35.3Kb distal of CAPON and situated directly within the genomic extent of the gene SH2D1B [MIM608510]. SH2D1B, also known as EAT2, regulates signal transduction through receptors expressed on the surface of antigen-presenting cells (Morra et al. 2001). D1Z13 (significant at p=0.014) is situated between two genes, 47.6Kb distal of SH2D1B and 64.3Kb proximal to UHMK1 [MIM608849]. UHMK1 (U2AF homology motif (UHM) kinase 1) is a serine/threonine-protein kinase also known as KIS or Kist (Kinase interacting with stathmin). This gene is highly expressed in the brain and other parts of the central nervous system and less so in other tissues (Maucuer et al. 1995; Maucuer et al. 1997; Maucuer et al. 2000; Boehm et al. 2002; Bieche et al. 2003). A third microsatellite marker D1Z18 was also found to be associated with schizophrenia (p=0.049). D1Z18 is located 90.2 Kb proximal to the five prime end of RGS4 and is 166.3 Kb distal to the gene encoding the enzyme 17-beta-hydroxysteroid dehydrogenase (HSD17B7 [MIM606756]). D1Z18 is also proximal to two overlapping predicted genes MGC48998 and BC040018, both with unknown function. A BLAST search for conserved domains in the NCBI Conserved Domain Search database did not reveal known functional domains. We found no evidence for association between RGS4 and schizophrenia in our sample (Rizig et al. 2006) (section 4.0).

SNP markers were genotyped in the SH2D1B and UHMK1 genes because these genes are in close proximity to the microsatellite markers D1Z12 and D1Z13. The results from these SNPs are shown in Table 6:2 (single marker table) and Table 6:3

(haplotypic association table). No evidence for association with schizophrenia was found from SNPs within SH2D1B (rs164126, rs164123, rs3121196, rs351453, rs164128). Furthermore, there was no evidence for haplotypic association with schizophrenia when these SNPs were combined with the marker D1Z12 (empirical permutation p = 0.123) or when these SNPs were combined with both D1Z12 and D1Z13 (empirical permutation p = 0.233). Tests of haplotypic association by combining D1Z12 and D1Z13 into a two locus haplotype were not formally significant (empirical permutation p=0.065).
By contrast, SNPs genotyped within UHMK1 demonstrated significant association (rs10494370, p = 0.004, odd ratio 1.64 (95% CI 1.16 - 2.32) and rs7513662, p = 0.043, odds ratio 0.81 (95% CI 0.66 - 0.99)) and two additional SNPs showed trends towards association (rs423227 p = 0.079 and rs6427680 p = 0.069). Evidence for haplotypic association at UHMK1 was found when data from these four SNPs were analysed as a single haplotype using GENECOUNTING (empirical permutation p = 0.0086), the most significant individual haplotype (consisting of rs10494370-rs7513662-rs423227-rs6427680) over expressed in cases was haplotype 1 (GATT) present in 6.5% of controls and 10.2% of cases and haplotype 2 (AACT) present in 10.8% controls and 14.1% in cases.

We found positive allelic association after the eighth marker was genotyped. Tests of experiment wide significance for marker rs10494370 remained significant following bonferroni correction (alpha=0.006) for multiple tests. Tests of haplotypic association were also significant for UHMK1 (p=0.009) using empirical permutation tests, which make it unnecessary to further correct for both multiple alleles and multiple markers

There was no evidence for haplotypic association when D1Z13 was included with the SNP data (empirical p = 0.337). D1Z13 did not show evidence for LD with SNP markers at UHMK1. However, D1Z13 showed good evidence for LD with markers within SH2D1B (see Table 6:4). Tests of haplotypic association in all neighbouring regions including at the CAPON and RGS4 loci were negative as shown in Table

6:3. Thus the evidence from these results points to UHMK1 as the most likely susceptibility gene for schizophrenia.

## 6.4 DISCUSSION

The positive results at CAPON and RGS4 could be that both are due to allelic association with aetiological base pair changes in or near the UHMK1 locus that has been implicated by us. This idea gains support from the fact that we detected significant linkage disequilibrium between markers at UHMK1 with markers at both RGS4 and CAPON with significant P values but unfortunately not with D' or $r^2$ (p= 0.02 and 0.00084 respectively Appendix Section 13.0 ). There is thus at least a possibility that the family-based methods used previously were unable to distinguish the effects of these three loci, implicating too broad a region.

The UHMK1 gene is a novel kinase with an RNA recognition site which has been shown to be highly expressed in the brain (Bieche et al. 2003). Five SNPs within the neighbouring gene SH2D1B (also known as EAT2, which regulates signal transduction through receptors expressed on the surface of antigen-presenting cells (Morra et al. 2001)) showed no allelic association with schizophrenia. The gene HSD17B7, which is proximal to the schizophrenia-associated marker D1Z18, is near to RGS4 on the distal side of UHMK1. HSD17B7 encodes the enzyme 17-beta-hydroxysteroid dehydrogenase (oxidizes or reduces estrogens and androgens in mammals and regulates the biologic potency of these steroids) (Krazeisen et al. 1999). SH2D1B, UHMK1 and HSD17B7 are all plausible candidates for influencing susceptibility to schizophrenia. However, our haplotype analyses for SNPs in the SH2D1B do not implicate this gene (Table 6:3). The allelic and haplotypic analyses of SNPs within UHMK1 suggest that this gene itself is likely to be increasing susceptibility to schizophrenia (Table 6:3).

The schizophrenic subjects we studied were routine clinical cases from United Kingdom NHS services in London and Southern England. They were not selected for having a positive family history of schizophrenia, although many of them did. The finding of association between schizophrenia and markers on chromosome 1q23.3 suggests that genetic effects detected by linkage studies of large multiply affected pedigrees may also be operative in a significant proportion of patients not necessarily having a family history of schizophrenia. The sample in which we have found positive association with UHMK1 has also shown allelic association with markers at the EPSIN 4 gene on chromosome 5q (Pimm et al. 2005). This locus had also been implicated by prior linkage studies (Schwab et al. 1997; Straub et al. 1997; Gurling et al. 2001; Devlin et al. 2002) and further supports the conclusion that genetic effects found in families by linkage analysis may also operate in unrelated cases of schizophrenia drawn from general hospital patients.

**Table 6:1 Microsatellite and SNP marker positions (UCSC March 2006 assembly) on 1q23.3 used for fine mapping of schizophrenia.**

| Marker name | Polymorphism | Position of polymorphism (bp) | Left primer | Right primer |
|---|---|---|---|---|
| rs164126 | C/T | 159097317 | | |
| rs164123 | C/T | 159099312 | | |
| rs3121196 | A/G | 159104376 | | |
| D1Z12 | $(CA)_{20}$ | 159105192 | GCCAAGACATGGAATCAACC | ATGCCACGTATGAGTGAGACC |
| rs351453 | A/G | 159107574 | | |
| rs164128 | G/T | 159117169 | | |
| D1Z13 | $(CA)_{21}$ | 159161203 | TCCAAGAAGCCCAATGAAAC | ACTCCTGGGCTCAAACAATC |
| D1Z14 | $(TTTA)_{10}$ | 159164066 | TCCATTCCTTCTCACCCATC | GCTGAGGCAGGAGAATCACT |
| rs10494370 | A/G | 159200571 | | |
| rs7513662 | A/G | 159211803 | | |
| rs423227 | C/T | 159217015 | | |
| rs6427680 | C/T | 159221876 | | |
| D1Z15 | $(TG)_{22}$ | 159332631 | CCTCTGAACCATTCCCATGT | GCAAACAGGAAAGGGAGGAT |
| D1Z16 | $(GT)_{13}$ | 159469256 | GCTTCTGATGCACATTCGAG | TGAAGAGGGACTCAGAGGGATA |
| D1Z17 | $(TC)_{12}$ | 159525259 | TGAGCGATAGAGTCAGGATTCA | AGCACAGAGCCAAACACCTT |
| D1Z18 | $(CA)_{23}$ | 159680585 | CCCAGCCTTCCCTATTGTCT | GTTTTCCAACCCCTGTCAGA |
| D1Z19 | $(CA)_{19}$ | 159705995 | TTGCTGTCTTGCCCTACAGA | CCTCATCATGGGCAGAGAAT |

**Table 6:2 Tests of allelic association with schizophrenia at the 1q23.3 region.**

| Marker | Marker location (bp on C1) | bp from prior marker | Allelic bases or fragment sizes with observed allele frequencies below | | | | | | | | | | | | | | | | Chi$^2$ | p[a] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rs164126 | 159097317 | | C | T | | | | | | | | | | | | | | | 0.30 | 0.58 |
| Con | | | 561 | 321 | | | | | | | | | | | | | | | | |
| Scz | | | 516 | 312 | | | | | | | | | | | | | | | | |
| rs164123 | 159099312 | 1995 | C | T | | | | | | | | | | | | | | | 0.00 | 0.98 |
| Con | | | 180 | 698 | | | | | | | | | | | | | | | | |
| Scz | | | 175 | 677 | | | | | | | | | | | | | | | | |
| rs3121196 | 159104376 | 5064 | A | G | | | | | | | | | | | | | | | 0.67 | 0.41 |
| Con | | | 631 | 217 | | | | | | | | | | | | | | | | |
| Scz | | | 619 | 233 | | | | | | | | | | | | | | | | |
| D1Z12 | 159105192 | 816 | 230 | 232 | 234 | 236 | 238 | 240 | 242 | 244 | 246 | 248 | 250 | 252 | | | | | 9.82 | 0.01[T3] |
| Con | | | 288 | 0 | 1 | 1 | 2 | 106 | 237 | 21 | 37 | 141 | 39 | 5 | | | | | | |
| Scz | | | 217 | 1 | 1 | 1 | 1 | 95 | 211 | 16 | 61 | 128 | 43 | 1 | | | | | | |
| rs351453 | 159107574 | 2382 | A | G | | | | | | | | | | | | | | | 0.67 | 0.41 |
| Con | | | 652 | 222 | | | | | | | | | | | | | | | | |
| Scz | | | 612 | 228 | | | | | | | | | | | | | | | | |
| rs164128 | 159117169 | 9595 | G | T | | | | | | | | | | | | | | | 0.27 | 0.60 |
| Con | | | 394 | 462 | | | | | | | | | | | | | | | | |
| Scz | | | 348 | 430 | | | | | | | | | | | | | | | | |
| D1Z13 | 159161203 | 44034 | 212 | 214 | 216 | 218 | 220 | 222 | 224 | 226 | 228 | 230 | 232 | 234 | 236 | 238 | 240 | 242 | 26.2 | 0.01[T2] |
| Con | | | 35 | 3 | 1 | 39 | 186 | 2 | 21 | 20 | 84 | 118 | 67 | 138 | 32 | 21 | 11 | 0 | | |
| Scz | | | 37 | 5 | 2 | 19 | 172 | 10 | 27 | 24 | 59 | 117 | 48 | 127 | 40 | 19 | 18 | 4 | | |
| D1Z14 | 159164066 | 2863 | 200 | 204 | 208 | 212 | 216 | 220 | 221 | 224 | | | | | | | | | 1.15 | 0.89[T3] |
| Con | | | 18 | 164 | 9 | 123 | 170 | 319 | 1 | 32 | | | | | | | | | | |
| Scz | | | 15 | 136 | 7 | 115 | 145 | 245 | 1 | 24 | | | | | | | | | | |
| rs10494370 | 159200571 | 36505 | A | G | | | | | | | | | | | | | | | 8.04 | 0.004[b] |
| Con | | | 810 | 58 | | | | | | | | | | | | | | | | |
| Scz | | | 748 | 88 | | | | | | | | | | | | | | | | |

Table 6:2 Continued.

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rs7513662 | 159211803 | 11232 | A | G | | | | | | | | | | | | | 4.12 | 0.043[c] |
| Con | | | 537 | 317 | | | | | | | | | | | | | | |
| Scz | | | 561 | 269 | | | | | | | | | | | | | | |
| rs423227 | 159217015 | 5212 | C | T | | | | | | | | | | | | | 3.08 | 0.08 |
| Con | | | 95 | 757 | | | | | | | | | | | | | | |
| Scz | | | 119 | 733 | | | | | | | | | | | | | | |
| rs6427680 | 159221876 | 4861 | C | T | | | | | | | | | | | | | 3.30 | 0.07 |
| Con | | | 436 | 436 | | | | | | | | | | | | | | |
| Scz | | | 385 | 459 | | | | | | | | | | | | | | |
| D1Z15 | 159332631 | 110755 | 312 | 314 | 316 | 318 | 320 | 322 | 324 | 326 | 328 | | | | | | 2.74 | 0.49[T3] |
| Con | | | 2 | 44 | 36 | 444 | 45 | 9 | 207 | 22 | 1 | | | | | | | |
| Scz | | | 0 | 30 | 20 | 367 | 32 | 3 | 150 | 16 | 2 | | | | | | | |
| D1Z16 | 159469256 | 136625 | 276 | 278 | 280 | 282 | 284 | 286 | | | | | | | | | | |
| Con | | | 2 | 1 | 42 | 612 | 93 | 2 | | | | | | | | | 1.824 | 0.38[T3] |
| Scz | | | 1 | 1 | 53 | 618 | 110 | 3 | | | | | | | | | | |
| D1Z17 | 159525259 | 56003 | 212 | 214 | 216 | 218 | | | | | | | | | | | 1.23 | 0.69[T4] |
| Con | | | 0 | 418 | 415 | 5 | | | | | | | | | | | | |
| Scz | | | 1 | 336 | 342 | 3 | | | | | | | | | | | | |
| D1Z18 | 159680585 | 155326 | 235 | 241 | 249 | 251 | 253 | 255 | 257 | 259 | 261 | 263 | 265 | 267 | 269 | 273 | 21.7 | 0.049[T1] |
| Con | | | 35 | 0 | 22 | 31 | 61 | 93 | 148 | 104 | 181 | 66 | 41 | 14 | 0 | 0 | | |
| Scz | | | 27 | 2 | 17 | 28 | 41 | 120 | 136 | 131 | 187 | 50 | 30 | 9 | 1 | 1 | | |
| D1Z19 | 159705995 | 25410 | 372 | 374 | 376 | 378 | 380 | 382 | 384 | 386 | 388 | | | | | | 3.29 | 0.67[T4] |
| Con | | | 0 | 26 | 138 | 42 | 370 | 150 | 86 | 35 | 15 | | | | | | | |
| Scz | | | 1 | 23 | 106 | 40 | 276 | 108 | 81 | 30 | 11 | | | | | | | |

[a] significance p from 2x2 chi squared 1 df or most significant p value from CLUMP Monte Carlo T1, T2, T3 or T4 as shown. [b]Odds ratio1.64 upper limit 2.32, lower limit 1.16 [c]Odds ratio 0.81 upper limit 0.99. lower limit 0.66

Positions from UCSC March 2006 assembly

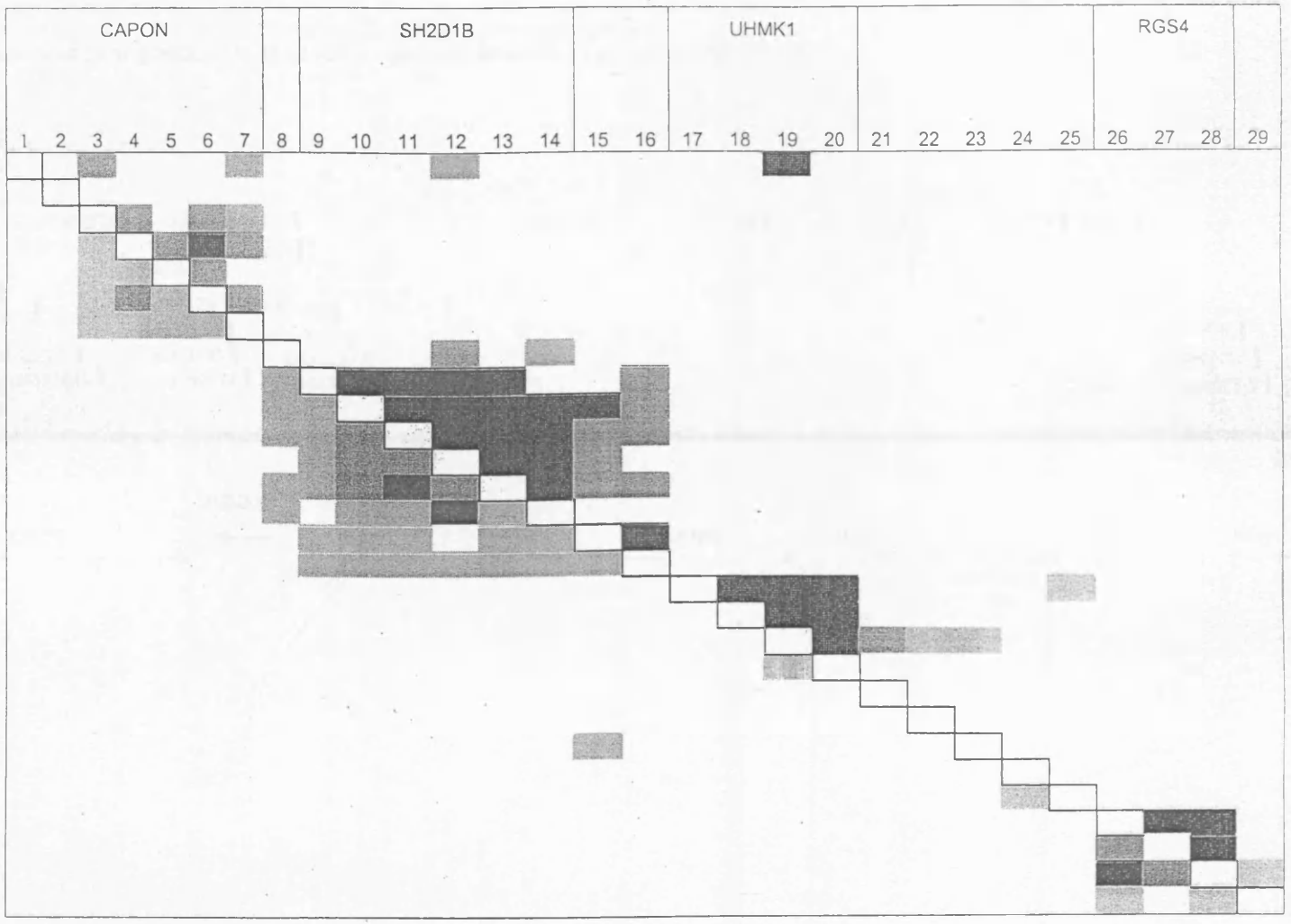
**Table 6:3 Haplotypic association with schizophrenia at the 1q23.3 region with markers within CAPON, SH2D1B, UHMK1 (KIST) and RGS4.**

| Location | Haplotype | Empirical p Value[a] |
|---|---|---|
| Markers within CAPON[b] | rs1572495- rs1538018- rs945713- rs1415263- D1S2675- rs4145621- rs2661818- D1S1679 | p = 0.441 |
| Markers within SH2D1B | rs164126- rs164123- rs3121196- D1Z12- rs351453- rs164128 | p = 0.123 |
| Markers within SH2D1B with D1Z13 | rs164126- rs164123- rs3121196- D1Z12- rs351453- rs164128- D1Z13 | p = 0.233 |
| Marker D1Z12 with D1Z13 | D1Z12- D1Z13 | P = 0.065 |
| Markers within UHMK1 | rs10494370- rs7513662- rs423227- rs6427680 | p = 0.0086 |
| Markers within UHMK1 with D1Z13 | D1Z13-rs10494370- rs7513662- rs423227- rs6427680 | p = 0.337 |
| Markers within UHMK1 | rs423227- rs6427680 | p = 0.081 |
| Markers within RGS4[b] | rs951436- rs941439-rs2661319 | p = 0.431 |

[a] haplotype permutation test empirical p, based on 99,999 permutations
[b] (Puri et al. 2006; Rizig et al. 2006)

Table 6:4 LD between all markers including RGS4 and CAPON (based on combined sample).

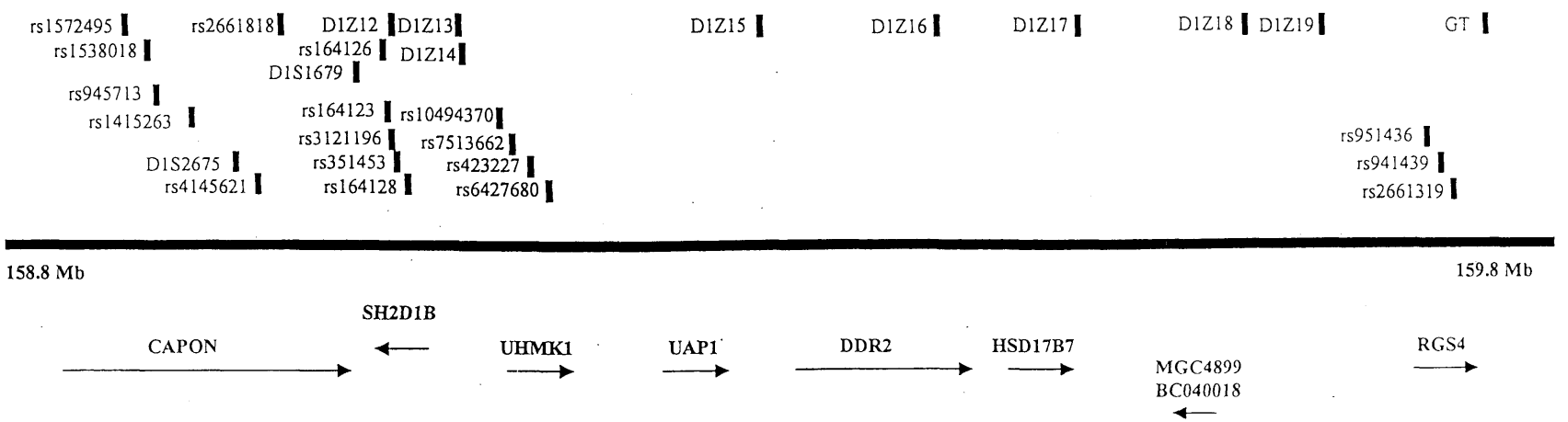
| Marker | Distance from previous marker (bp) | |
|---|---|---|
| rs1572495 | | 1 |
| rs1538018 | 31180 | 2 |
| rs945713 | 5189 | 3 |
| rs1415263 | 30373 | 4 |
| D1S2675 | 43918 | 5 |
| rs4145621 | 8727 | 6 |
| rs2661818 | 46126 | 7 |
| D1S1679 | 96950 | 8 |
| rs164126 | 3895 | 9 |
| rs164123 | 1995 | 10 |
| rs3121196 | 5064 | 11 |
| D1Z12 | 816 | 12 |
| rs351453 | 2382 | 13 |
| rs164128 | 9595 | 14 |
| D1Z13 | 44034 | 15 |
| D1Z14 | 2863 | 16 |
| rs10494370 | 36505 | 17 |
| rs7513662 | 11232 | 18 |
| rs423227 | 5212 | 19 |
| rs6427680 | 4861 | 20 |
| D1Z15 | 110755 | 21 |
| D1Z16 | 136625 | 22 |
| D1Z17 | 56003 | 23 |
| D1Z18 | 155326 | 24 |
| D1Z19 | 25410 | 25 |
| rs951436 | 59005 | 26 |
| rs941439 | 349 | 27 |
| rs2661319 | 12156 | 28 |
| GT | 6086 | 29 |




D' above the diagonal and $r^2$ below the diagonal

| | |
|---|---|
| D' ≤ 1 | r ≤ 1 |
| D' < 0.8 | r< 0.8 |
| D' < 0.6 | r < 0.6 |

Figure 6:1 Markers and gene positions in the 1q23.3 region between CAPON and RGS4.



# 7.0 CONFIRMATION OF THE GENETIC ASSOCIATION BETWEEN THE U2AF HOMOLOGY MOTIF (UHM) KINASE 1 (UHMK1) GENE AND SCHIZOPHRENIA ON CHROMOSOME 1Q23.3

## 7.1 INTRODUCTION

Our original genetic association studies did not find evidence for both the CAPON and RGS4 loci in the UCL case control sample (Puri et al. 2006; Rizig et al. 2006) section 4.0-5.0. Our original study of UHMK1 (Puri et al. 2007) section 6.0, found positive allelic association with two neighbouring microsatellites (D1Z12 and D1Z13) and two SNPs within UHMK1 (rs10494370 and rs7513662), in addition two SNPs showed a trend towards association (rs423227 and rs6427680). A third microsatellite (D1Z18) located between the steroid dehydrogenase gene HSD17B7 and RGS4 was also associated.

The weakly associated microsatellite marker D1Z18 is close to the hydroxysteroid (17-beta) dehydrogenase 7 gene (HSD17B7). We have now genotyped tagging SNPs derived from HapMap to test whether HSD17B7 may be involved in the susceptibility to schizophrenia and have also genotyped HapMap SNPs in the UHMK1 gene selected to cover the 5' and 3' untranslated regions (UTRs). We have also carried out linkage disequilibrium (LD) analyses of the whole 700Kb region including markers in CAPON, UHMK1, RGS4 and HSD17B7.

Markers localised within UHMK1 and known to show association in the UCL sample have also been genotyped in a replication case control sample from Aberdeen.

## 7.2 BRIEF METHODS AND MATERIALS

The methods and DNA samples employed were described previously in sections 2.0-3.0. The UCL case control sample consisted of DNA from 450 volunteers with schizophrenia and 450 controls. Selection and screening were carried out as stated in section 2.1. For the replication study using the Aberdeen sample, DNA from 858 cases of schizophrenia and 591 controls was available. The cases were diagnosed using a combination of case note inspection and in the majority of cases (n=717) SCID interview. All met the DSM 111R diagnosis of schizophrenia by OPCRIT. The controls were drawn from the same Scottish population. They were recruited as volunteers via general practices and screened for absence of psychiatric illness. They were sex matched.

Tagging SNPs where ascertained from the international HapMap Project (in the CEU population) with the use of the Haploview (3.32) Tagger function (Barrett et al. 2005; de Bakker et al. 2005), on its default setting (aggressive tagging use of 2 and 3 marker haplotypes, with $r^2$ threshold of 0.8 and a LOD threshold for multi marker tests of 3.0). All SNP markers were genotyped by the KASPar method which is a modification of the Amplifluor method (KBiosciences, Hoddesdon, UK). 17% of samples on each microtitre plate were reduplicated in order to detect error and confirm the reproducibility of genotypes. The data was then analysed to confirm Hardy Weinberg equilibrium (HWE). Markers with lack of HWE in the control group were rejected and repeated. The genotype data was analysed separately by 96-well DNA plate using SCANGROUP to test for plate by plate differences in haplotype frequencies (Curtis et al. 2006). Some genotyping errors can show up as rare haplotypes occurring on just a single plate whereas true haplotypic associations with will be manifest as certain haplotypes being preponderant among cases or controls spread across a number of plates.

Genotypic and allelic associations of individual markers were tested for using chi square tests. Tests for haplotypic association were carried out using GENECOUNTING, which computes maximum likelihood estimates of haplotype frequencies from phase-unknown data, and the empirical significance was obtained using permutation tests (Zhao et al. 2000; Zhao et al. 2002; Curtis et al. 2006). GENECOUNTING was also used to calculate pair-wise linkage disequilibrium between all markers, and their relationship was visualised on LocusView 2.0 (Broad Institute) (Petryshen et al. 2003).

## 7.3 RESULTS

The SNPs at HSD17B7 (Figure 7:1) did not demonstrate allelic or haplotypic association with schizophrenia (Table 7:1). Results from typing additional markers around UHMK1, covering the 5' and 3' UTRs along with results from the five SNPs previously found to be associated with schizophrenia (section 6.0) (Puri et al. 2007) are shown in Table 7:2. Two of the new SNPs, rs6604863 and rs10753578, were found to be associated with schizophrenia (p=0.02 and p=0.017 respectively). A third marker, rs6704428, showed a trend towards allelic association (p=0.089). A seven marker haplotype including the original SNPs described in Puri et al (2007) and the new SNPs described here produced an empirical significance of p=0.0033 (Table 7:3). These findings from the UCL sample provide additional evidence for UHMK1 as a schizophrenia susceptibility gene.
The following markers were genotyped in the Aberdeen sample: rs10494370, rs7513662, rs423227, rs6427680, rs6604863, rs10753578 and rs6704428. Two of these SNPs yielded evidence for association, rs7513662 (p=0.0087) and rs10753578 (p=0.022) (Table 7:2). Allele A of rs7513662 was associated with schizophrenia in both the UCL and Aberdeen samples. However the distribution of allele frequencies for rs10753578 is in the opposite direction to that found in the UCL sample. This may indicate that the aetiological nucleotide change is present on different haplotype background in the two samples or that the disease haplotype frequencies are different in the north and south of the UK. The four SNPs originally implicated in the UCL sample produced evidence for haplotypic association in the Aberdeen sample (empirical permutation p=0.0135) and the presence of two associated SNPs in the Aberdeen sample confirms UHMK1 as a gene increasing susceptibility to schizophrenia. Of relevance is also the fact that haplotypic association with the original four marker haplotype (rs10494370, rs7513662, rs423227 and rs6427680) (Puri et al. 2007) has been confirmed and replicated (permutation p=0.0135), in addition a haplotype analysis of all seven markers in the Aberdeen sample was replicated producing an empirical p value of 0.0004. Amongst the other haplotypes tested with less than seven markers, the most significant in the Aberdeen sample comprised markers rs7513662 and rs10753578 (permutation p=0.0002; see Table 3).

Data from the UCL and Aberdeen samples were combined and analysed together. Evidence for association was stronger in the combined sample for three SNPs, rs7513662, rs6427680 and rs6694863 (p=0.0007, 0.0252 and 0.015 respectively). Haplotype association for the

original four markers was also significant (empirical p=0.011). However the strongest evidence for association came from haplotype analysis of all seven markers (empirical p=0.00005) (Table 7:3).

Global haplotypic analyses of the UCL and Aberdeen samples data showed that the same marker loci were associated in both samples. However inspection of the specific subgroups of alleles at these loci showed some differences between the sample. In the UCL sample, the four locus SNP haplotypes consisting of rs10494370-rs7513662-rs423227-rs6427680 (haplotypes UHMK1F and UHMK1G) were elevated in cases relative to controls (UHMK1F 16.2% cases vs. 13% controls; UHMK1G 10.2% cases vs. 6.5% controls). In the Aberdeen sample, haplotypes UHMK1D and UHMK1E were elevated in cases relative to controls (32.7% cases and 29.1% controls; 13.8% cases and 11.4% controls respectively). However, in the analysis of the combined samples all four haplotypes were elevated in cases relative to controls (see table 3). For the 7 SNP haplotype analysis (rs10494370-rs7513662-rs423227-rs6427680-rs6694863-rs10753578-rs6704428) the elevated haplotypes in the UCL sample were UHMK1B (12.2% cases 11.7% controls,) and UHMK1C (12.8% cases and 10.9% controls). In the Aberdeen sample elevated haplotypes were UHMK1A (30.2% cases and 24.8% controls) and UHMK1B (13.8% cases and 11.4% controls). Haplotype UHMK1B is elevated in both individual samples. However, in the combined analysis all three haplotypes (UHMK1A, UHMK1B and UHMK1C) are more common in cases (Puri et al. Submitted).

## 7.4 DISCUSSION

The failure to detect association between markers in HSD17B7 with schizophrenia suggests that this gene is unlikely to be involved in schizophrenia susceptibility in the British population. The weakly positive result from D1Z18 could be attributed to detecting long distance LD with aetiological base pair changes in UHMK1 or another gene in the region. Superficially we have managed to replicate association between UHMK1 markers and schizophrenia in second case control sample. However careful attention to the detail of which marker alleles and haplotypes were associated with schizophrenia in the original UCL and then in the Aberdeen sample show some differences which need explaining. The differences could be explained by the fact that the disease haplotype frequencies are

different in the north and south of the UK. Indeed when allele frequencies of the seven SNPs typed at the UHMK1 locus in the UCL and Aberdeen samples were compared, the differences were greater between the two control samples as opposed to the schizophrenia samples.

As noted above two previous studies have implicated the microsatellite D1S1679 by showing association or linkage with schizophrenia. D1S1679 is 105.9Kb from the 5' end of UHMK1 and displays significant LD with markers rs7532188, D1Z14, rs164160, rs164167, rs164168 and rs164171 with p values $\leq 1x10^{-5}$ (13.0 Appendix Figure 13:1) in the likely promoter region of UHMK1. The evidence for association between schizophrenia and D1S1679 might therefore be due to LD with aetiological base pair changes in UHMK1.

UHMK1 (U2AF homology motif kinase 1) is a serine/threonine-protein kinase with an RNA recognition site and it has been shown to be highly expressed in the brain (Bieche et al. 2003). Both the gene and its protein are highly conserved between mouse, rat and human. UHMK1 is also known to be highly expressed in most parts of the mouse brain, particularly in the amygdala and hippocampus, according to the Allen Brain Atlas. The protein has the ability to phosphorylate classical in vitro substrates such as myelin basic protein and synapsin 1 (Maucuer et al. 1997), as well as stathmin, a complex signal relay protein. Stathmin is phosphorylated in response to many signals such as hormone growth and differentiation factors, neurotransmitters or upon activation of T lymphocytes. It has also been proposed as a key regulator of microtubule dynamics (Maucuer et al. 1997). UHMK1 was formerly named as KIS (kinase interacting with stathmin) and had originally been discovered by a yeast two hybrid screen using stathmin as bait (Maucuer et al. 1995). The stathmin gene locus has been knocked out in the mouse, and it was found that mice homozygous for stathmin gene deletion lacked instinctive fear and had other behavioural abnormalities. The stathmin knockout also had weak memories for past aversive experiences such as those in fear conditioning tests (Shumyatsky et al. 2005). The gene is highly enriched in the lateral nucleus of the amygdala. Because of the close association of UHMK1 protein with stathmin protein a knockout of UHMK1 may also produce similar behavioural abnormalities. A study comparing anterior cingulate cortex gray matter proteomes between schizophrenia and controls found that stathmin was increased by 1.8 fold change in a schizophrenia cohort (Clark et al. 2006). It has also been found that the

expression of UHMK1 is significantly down-regulated in mice treated with the antipsychotic drugs clozapine and haloperidol Rizig et al (2009)

## 7.5 CONCLUSION

Association between UHMK1 and schizophrenia has been successfully confirmed in a case control sample from Aberdeen. There was increased statistical significance when the sample was combined with the UCL sample. Further independent replications of association are now needed in as many populations as possible. Resequencing, expression analysis and several types of conditional and knockout transgenic mice are now needed to further explore the normal and abnormal biological functions of the UHMK1 gene.

**Table 7:1 Tests of association between HAPMAP SNPs in HSD17B7 and schizophrenia in the UCL sample.**

| Marker | Marker location (UCSC March 2006 assembly) | bp from prior marker | Allelic bases or fragment sizes with observed allele frequencies below | | $Chi^2$ | P[a] |
|---|---|---|---|---|---|---|
| rs1892125 | 161031373 | | A | C | | |
| UCL Controls | | | 142 | 764 | 2.11 | 0.14 |
| UCL Schizophrenia | | | 118 | 772 | | |
| rs1780019 | 161033100 | 1727 | G | A | | |
| UCL Controls | | | 193 | 679 | 1.22 | 0.27 |
| UCL Schizophrenia | | | 173 | 693 | | |
| rs2684881 | 161033792 | 692 | A | T | | |
| UCL Controls | | | 22 | 912 | 0.66 | 0.42 |
| UCL Schizophrenia | | | 16 | 868 | | |
| rs11589262 | 161037271 | 3479 | G | A | 0.41 | 0.52 |
| UCL Controls | | | 336 | 594 | | |
| UCL Schizophrenia | | | 333 | 553 | | |
| rs10917598 | 161040138 | 2867 | C | G | 0.81 | 0.37 |
| UCL Controls | | | 304 | 610 | | |
| UCL Schizophrenia | | | 309 | 567 | | |
| rs1039874 | 161043150 | 3012 | C | T | 0.51 | 0.47 |
| UCL Controls | | | 47 | 893 | | |
| UCL Schizophrenia | | | 40 | 890 | | |
| rs12402864 | 161048148 | 4998 | A | G | | |
| UCL Controls | | | 277 | 593 | 1.29 | 0.26 |
| UCL Schizophrenia | | | 302 | 576 | | |
| rs2805053 | 161048411 | 263 | G | A | | |
| UCL Controls | | | 458 | 452 | 0.06 | 0.81 |
| UCL Schizophrenia | | | 450 | 434 | | |
| 8 SNP haplotype | | | | | Global empirical p value | 0.87 |

[a] two tailed significance p from 2x2 chi squared 1 df

**Table 7:2 Tests for association of SNPs within UHMK1 in the UCL, Aberdeen and combined samples .**

| Marker | Marker location (UCSC March 2006 assembly) | bp from prior marker | Bases with observed allele frequencies | | | | $Chi^2$ | P[a] |
|---|---|---|---|---|---|---|---|---|
| rs10494370* | 160735537 | | G | | A | | | |
| UCL Controls | | | 58 | (0.07) | 810 | (0.93) | 8.04 | **0.004** |
| UCL Schizophrenia | | | 88 | (0.11) | 748 | (0.89) | | |
| Aberdeen Controls | | | 110 | (0.10) | 1030 | (0.90) | 0.85 | 0.36 |
| Aberdeen Schizophrenia | | | 141 | (0.09) | 1493 | (0.91) | | |
| Combined Controls | | | 168 | (0.08) | 1840 | (0.92) | 1.12 | 0.29 |
| Combined Schizophrenia | | | 229 | (0.09) | 2241 | (0.91) | | |
| Mantel Haenszel meta analysis calculated odds ratio 1.11 (CI 0.87-1.43) | | | | | | | | |
| rs7513662* | 160746769 | 11232 | G | | A | | | |
| UCL Controls | | | 317 | (0.37) | 537 | (0.63) | 4.12 | **0.043** |
| UCL Schizophrenia | | | 269 | (0.31) | 561 | (0.69) | | |
| Aberdeen Controls | | | 419 | (0.37) | 727 | (0.63) | 6.88 | **0.0087** |
| Aberdeen Schizophrenia | | | 518 | (0.32) | 1112 | (0.68) | | |
| Combined Controls | | | 736 | (0.37) | 1264 | (0.63) | 11.34 | **0.00075** |
| Combined Schizophrenia | | | 787 | (0.32) | 1673 | (0.68) | | |
| Mantel Haenszel meta analysis calculated odds ratio 1.23 (CI 1.06-1.44) | | | | | | | | |
| rs423227* | 160751981 | 5212 | C | | T | | | |
| UCL Controls | | | 95 | (0.11) | 757 | (0.89) | 3.08 | 0.08 |
| UCL Schizophrenia | | | 119 | (0.14) | 733 | (0.86) | | |
| Aberdeen Controls | | | 151 | (0.13) | 969 | (0.87) | 0.031 | 0.86 |
| Aberdeen Schizophrenia | | | 216 | (0.13) | 1414 | (0.87) | | |
| Combined Controls | | | 246 | (0.12) | 1726 | (0.88) | 1.01 | 0.314 |
| Combined Schizophrenia | | | 335 | (0.13) | 2147 | (0.87) | | |
| Mantel Haenszel meta analysis calculated odds ratio 1.09 (CI 0.88-1.35) | | | | | | | | |
| rs6427680* | 160756842 | 4861 | C | | T | | | |
| UCL Controls | | | 436 | (0.5) | 436 | (0.5) | 3.30 | 0.07 |
| UCL Schizophrenia | | | 385 | (0.46) | 459 | (0.54) | | |
| Aberdeen Controls | | | 554 | (0.48) | 600 | (0.52) | 1.76 | 0.18 |
| Aberdeen Schizophrenia | | | 730 | (0.45) | 876 | (0.55) | | |
| Combined Controls | | | 990 | (0.49) | 1036 | (0.51) | 5.01 | **0.025** |
| Combined Schizophrenia | | | 1115 | (0.46) | 1335 | (0.54) | | |
| Mantel Haenszel meta analysis calculated odds ratio 1.14 (CI 0.99-1.32) | | | | | | | | |
| rs6694863 | 160771556 | 14714 | A | | C | | | |
| UCL Controls | | | 72 | (0.08) | 868 | (0.92) | 5.37 | **0.02** |
| UCL Schizophrenia | | | 46 | (0.05) | 868 | (0.95) | | |
| Aberdeen Controls | | | 74 | (0.06) | 1074 | (0.94) | 1.26 | 0.26 |
| Aberdeen Schizophrenia | | | 91 | (0.05) | 1583 | (0.95) | | |
| Combined Controls | | | 146 | (0.07) | 1942 | (0.93) | 5.86 | **0.015** |
| Combined Schizophrenia | | | 137 | (0.05) | 2451 | (0.95) | | |
| Mantel Haenszel meta analysis calculated odds ratio 1.34 (CI 1.0-1.8) | | | | | | | | |
| rs10753578 | 160788440 | 16884 | A | | G | | | |
| UCL Controls | | | 157 | (0.18) | 729 | (0.82) | 5.7 | **0.017** |
| UCL Schizophrenia | | | 201 | (0.22) | 703 | (0.78) | | |
| Aberdeen Controls | | | 272 | (0.23) | 886 | (0.77) | 5.23 | **0.022** |
| Aberdeen Schizophrenia | | | 332 | (0.20) | 1336 | (0.80) | | |
| Combined Controls | | | 429 | (0.21) | 1615 | (0.79) | 0.049 | 0.83 |
| Combined Schizophrenia | | | 533 | (0.21) | 2039 | (0.79) | | |
| Mantel Haenszel meta analysis calculated odds ratio 0.98 (CI 0.82-1.16) | | | | | | | | |
| rs6704428 | 160795129 | 6689 | G | | A | | | |
| UCL Controls | | | 51 | (0.06) | 853 | (0.94) | 2.89 | 0.089 |
| UCL Schizophrenia | | | 69 | (0.08) | 835 | (0.92) | | |
| Aberdeen Controls | | | 93 | (0.08) | 1051 | (0.92) | 2.56 | 0.11 |
| Aberdeen Schizophrenia | | | 109 | (0.07) | 1557 | (0.93) | | |
| Combined Controls | | | 144 | (0.07) | 1904 | (0.93) | 0.019 | 0.89 |
| Combined Schizophrenia | | | 178 | (0.07) | 2392 | (0.93) | | |
| Mantel Haenszel meta analysis calculated odds ratio 0.98 (CI 0.74-1.28) | | | | | | | | |

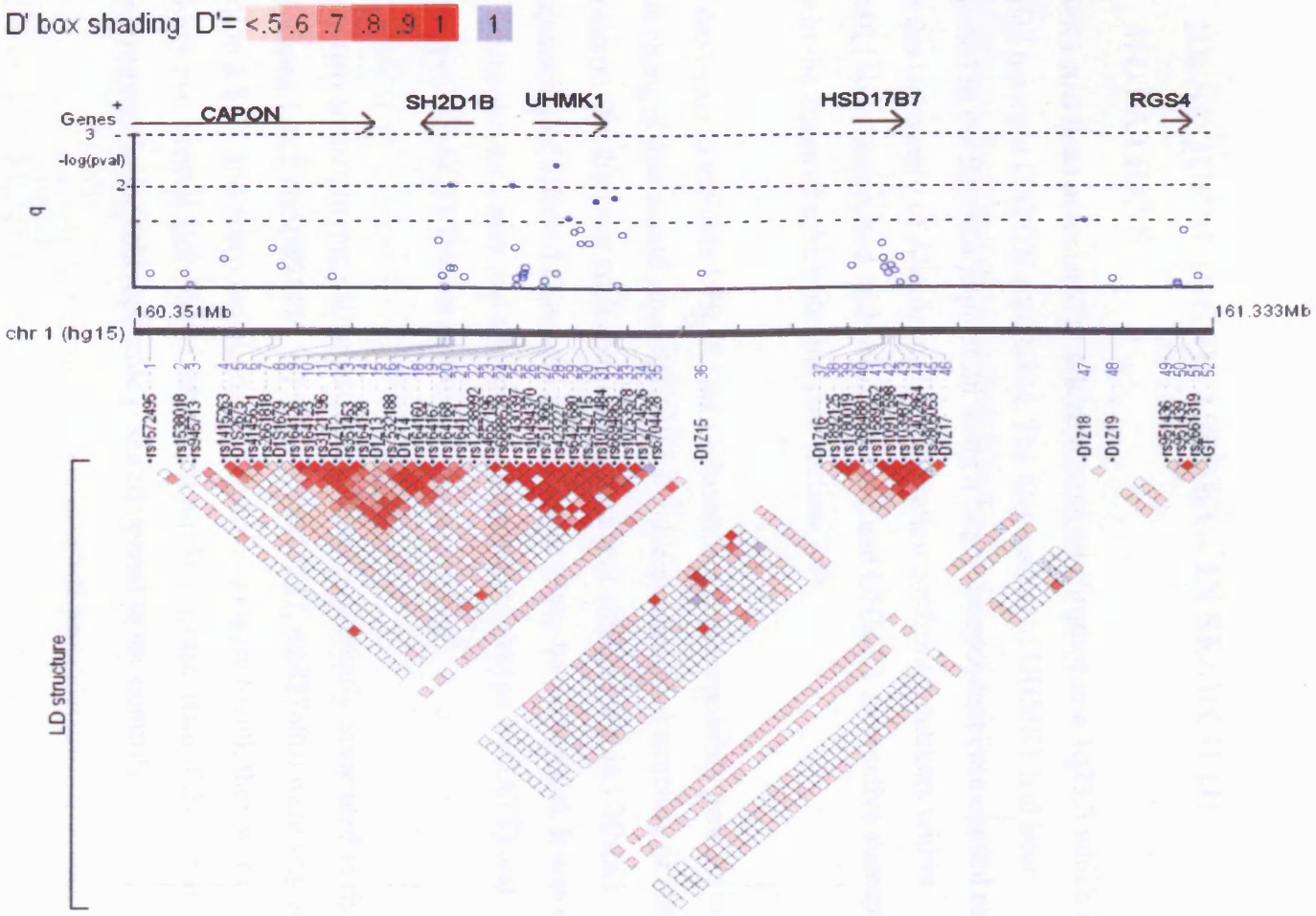
[a] two tailed significance p from 2x2 chi squared 1 df

*Original results in section 6.0 Table 6:2 (Puri et al. 2007).

Table 7:3 Haplotypic association with schizophrenia with UHMK1 markers in UCL and Aberdeen samples.

| Haplotype Composition | Alleles | Haplotype name | UCL | | | Aberdeen | | | Combined | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Cases | Controls | Global empirical significance | Cases | Controls | Global empirical significance | Cases | Controls | Global empirical significance |
| rs10494370-rs7513662-rs423227-rs6427680-rs6694863-rs10753578-rs6704428 | A-A-T-T-C-G-A | UHMK1A | 26.9% | 30.2% | | 30.2% | 24.8% | | 29.1% | 27.1% | |
| | A-A-T-C-C-G-A | UHMK1B | 12.2% | 11.7% | 0.0033 | 13.8% | 11.4% | 0.0004 | 13.2% | 11.7% | 0.00005 |
| | A-A-C-T-C-A-A | UHMK1C | 12.8% | 10.9% | | 12.4% | 12.8% | | 12.7% | 11.9% | |
| rs10494370-rs7513662-rs423227-rs6427680 | A-A-T-T | UHMK1D | 30.1% | 33.0% | | 32.7% | 29.1% | | 31.9% | 30.8% | |
| | A-A-T-C | UHMK1E | 12.3% | 12.3% | 0.0086 | 13.7% | 11.6% | 0.0135 | 13.2% | 12.0% | 0.011 |
| | A-A-C-T | UHMK1F | 16.2% | 13.0% | | 13.1% | 13.2% | | 13.4% | 12.2% | |
| | G-A-T-T | UHMK1G | 10.2% | 6.5% | | 8.3% | 9.6% | | 8.9% | 8.2% | |
| rs7513662-rs10753578 | A-G | UHMK1H | 45.1% | 45.4% | 0.024 | 48.3% | 40.9% | 0.0002 | 47.1% | 42.6% | 0.0004 |
| | A-A | UHMK1I | 22.3% | 17.5% | | 19.9% | 22.6% | | 20.9% | 20.6% | |
| rs423227-rs6427680-rs6704428 | C-T-A | UHMK1J | 14.1% | 10.6% | 0.067 | | | | | | |
| rs10494370-rs7513662-rs6694863-rs10753578 | A-A-C-A | UHMK1K | 16.2% | 13.0% | 0.004 | | | | | | |
| rs7513662-rs6427680 | A-T | UHMK1M | | | | | | | 54.3% | 51.3% | 0.0006 |
| | A-C | UHMK1N | | | | | | | 13.6% | 12.0% | |
| rs7513662-rs6427680-rs6694863 | A-T-C | UHMK1O | | | | | | | 54.3% | 51.1% | 0.00005 |
| | A-C-C | UHMK1P | | | | | | | 13.6% | 11.8% | |

**Figure 7:1 Linkage disequilibrium between all markers genotyped at CAPON, SH2D1B, UHMK1, HSD17B7 and RGS4 on chromosome 1q23.3, genotyped in the UCL dataset, produced by LocusView.**



# 8.0 FUNCTIONAL SCREENING FOR AETIOLOGICAL CHANGES INCREASING SUSCEPTIBILITY TO SCHIZOPHRENIA

## 8.1 RE-SEQUENCING OF UHMK1 – IN SEARCH OF MUTATIONS

UHMK1 had been associated to schizophrenia on chromosome 1q23.3 which is situated between CAPON and RGS4. The association of UHMK1 had been replicated in the Scottish population using a larger independent case control sample from the University of Aberdeen. The replication confirmed markers within UHMK1 were associated with schizophrenia, and UHMK1 as putative susceptibility gene in the cause of schizophrenia (see section 7.0).

The next step to validate UHMK1 as a schizophrenia susceptibility gene, is to find DNA changes that would cause UHMK1 to have an abnormal structure or abnormal expression. To discover such aetiological base pair changes within UHMK1 resequencing of affected cases and in controls needs to be carried out. It was noticed that schizophrenic cases had a higher frequency of haplotype 1 (GATT) and haplotype 2 (AACT), than in controls.

Thirty two schizophrenic patients with the most significantly associated at risk haplotypes 1 & 2 (rs10494370, rs7513662, rs423227, rs6427680) were sequenced (section 2.8.4), and when mutations (DNA variants) were found, they were screened in thirty two control individuals selected at random, to ascertain if the variant was more frequent in the schizophrenic cases compared to the controls.

## 8.1.1 RE-SEQUENCING OF THE UHMK1 GENE

### 8.1.1.1 MUTATION MECHANISMS

Before re-sequencing the UHMK1 gene was commenced, it was necessary to consider where any aetiological base-pair change was most likely to be located. Botstein & Risch (2003), have reviewed the past successes of linkage studies to delineate the different types of mutation which result in human disease and their relative frequencies:-

**Table 8:1 the types and frequencies of different mutations leading to human disease (Botstein et al. 2003).**

| Change | Number | % of total |
|---|---|---|
| | | |
| deletion | 6,085 | 21.8 |
| Insertion/duplication | 1,911 | 6.8 |
| Complex rearrangement | 512 | 1.8 |
| repeat variations | 38 | 0.1 |
| Missense/nonsense | 16,441 | 58.9 |
| Splicing | 2,727 | 9.8 |
| Regulatory | 213 | 0.8 |
| Total | 27,027 | 100 |

The data indicate that most Mendelian clinical phenotypes are associated with alterations in normal coding sequence proteins: so far, very few (0.8%) are associated with regulatory changes. Linkage studies have proved very successful in the identification of both complex disease genes, as well as rare disease-associated mutations. They have clear inheritance patterns.

Clearly, it is important to consider where the mutations causing schizophrenia or the other complex disorders are likely to be found?

Botstein & Risch (2003) favour the coding regions on the basis that the moderate to low risk polymorphisms found so far mainly appear in the coding regions of genes. Outside of the coding regions the mechanisms by which diseases arise are complex and varied. Most genes encoding for proteins are organised into multiple exons, which must be spliced to produce the mRNA that is translated into protein. A 5' promoter element, contiguous with the transcription start site, is required to

assemble the protein complex necessary for RNA synthesis (Levine et al. 2003). For many genes, the region immediately upstream of the minimal promoter contains sufficient transcription factor binding sites to direct correct expression of the gene – called regulatory promoters. Many genes also require multiple *cis*-acting distant genomic elements for correct expression to occur. These "enhancers" can be located upstream, within introns, or downstream of the transcription unit (which is made up of the transcribed exons and introns from the promoter to the polyadenylation site (Kleinjan et al. 2001)). The genomic regions harbouring regulatory elements can stretch as much as 1Mb in either direction from the transcription unit (Pfeifer et al. 1999) and (Kimura-Yoshida et al. 2004). Some or all of these elements may reside within the introns of neighbouring genes, often with function unrelated to the regulated gene ((Kleinjan et al. 2001; Lettice et al. 2002)). From a regulatory viewpoint, genes can be grouped into three classes:-

Housekeeping genes required for the functioning of most or all cells; these are generally ubiquitously expressed and have promoters that are active in all cells without needing enhancers.

Tissue-specific genes, which play a specific role in the particular function of the differentiated cell-type; these genes are regulated through one or a few specific enhancers.

Developmental regulator genes, which function in specific tissues at defined time-points in development – sometimes at critically defined levels – and have to be strictly inactive in all other tissues and time points. These genes require multiple enhancer elements that all need to be fitted into the *cis* region surrounding the gene. Thus far most genes, in which disturbance of long-range control have been observed, are key developmental regulators (Kleinjan et al. 2005).

The effects of regulatory elements reaching a long distance are clearly illustrated by callipyge mutation in sheep. The callipyge (CLPG) phenotype is characterised by hindquarter muscle overgrowth that only affects heterozygotes with paternal inheritancy of the CLPG mutation. By use of DNA from the mutants and from the mosaic founder individual, the CLPG mutation was identified as a single-base

substitution in a region of strong sequence conservation 33 Kb upstream of one of the genes involved (Freking et al. 2002).
Clearly, finding such mutations and the effects of such mutations is difficult. A striking example of a small deletion or mutation in a *cis* element is provided by the case of preaxial polydactyl (PPD) and sonic hedgehog in which a single nucleotide substitution located 1Mb from the causative gene produces a severe genetic defect (Lettice et al. 2003). With regard to *UHMK1* the decision was made to focus attention upon the exons and their surrounding splice sites and the 5′ region.

## 8.1.2 METHOD OF SEQUENCING UHMK1

### 8.1.2.1 SAMPLE SELECTION FOR SEQUENCING

Samples were selected for sequencing following analysis of the results of the haplotypic association (see section 6.3). It can be seen that certain haplotypes generated from this analysis had a marked increase in frequency in cases compared to controls, and these were selected for use in identifying samples likely to contain aetiological base pair changes, in this case haplotype 1 and haplotype 2 were most significant. Samples from the schizophrenic cases were chosen if they contained one or both of the haplotypes in either the heterozygous or homozygous state. This resulted in the selection of 44 samples. Forty of these were selected for PCR amplification and 32 of the amplified PCR products were bi-directionally sequenced to screen for aetiological base pair changes.

Sequencing of the exons included approximately 100 bases of the intronic region to take into account any abnormal intron/exon splice site variations. Primers were designed (Section 2.8.4) to amplify between 100 to 600 bases, and large regions were covered by sequencing overlapping regions. Primers which failed to amplify were redesigned with different annealing positions (list of primer sequences used is seen in appendix 1 section 13.2). It was found that the CpG island was very difficult to amplify, this was eventually overcome by reducing the size of the amplimer.

Sequencing commenced using the Sanger-Coulson Chain termination method as described in the Methods Section 2.8.4.1.

**Figure 8:1 Schematic diagram of UHMK1 and the regions sequenced.**

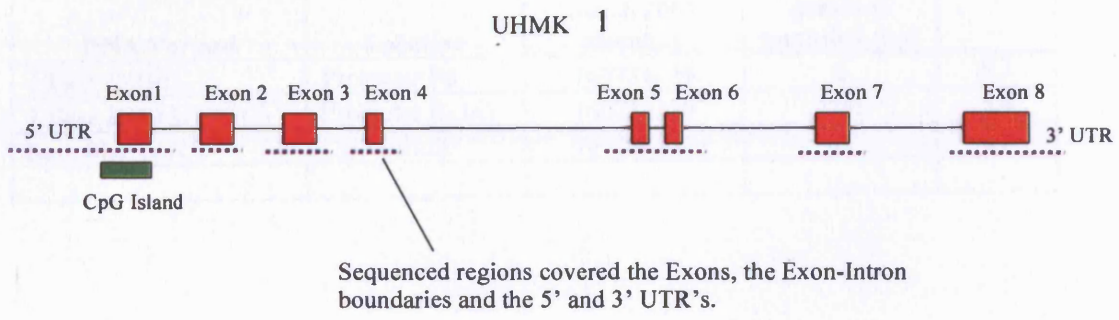
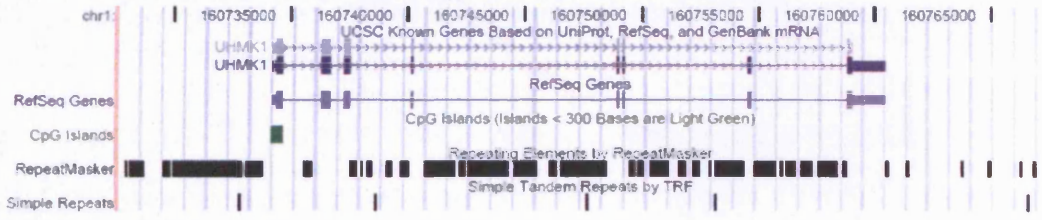



**Figure 8:2 Diagram from UCSC genome browser database showing the simple repeats covering the promoter region indicated by the thick black lines (March 2006 assembly).**




## 8.1.3 RESULTS OF RE-SEQUENCING UHMK1 EXONS, 5'UTR AND 3'UTR

Certain promoter regions could not be amplified (even with alterations to PCR cycling conditions, maser mix alterations and primer re-design) due to the region consisting of repeats such as LTRs (long terminal repeats), LINES (long interspersed nuclear elements) and SINES (short interspersed nuclear elements). This resulted in no amplification or amplification of many different regions due to the lack of binding specificity (see Figure 8:2). However, over 1Kb of the promoter region and

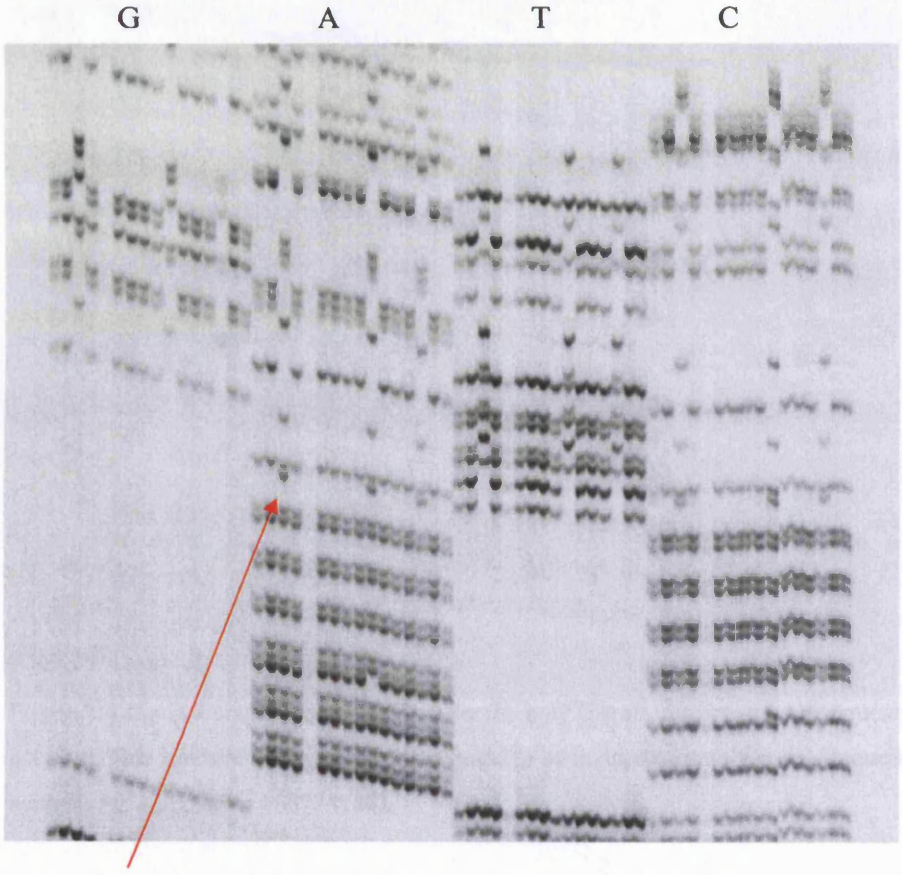
all of the exons were successfully sequenced and screened for mutations (see Figure 8:1).

None of the exons were found to contain DNA variants in the 32 schizophrenics that were sequenced. However, a number of DNA Variants were found in the promoter region which will be discussed next.

**Table 8:2 Table of DNA variants, position and location.**

| DNA Variant | Location | Position UCSC march 2004 assembly | Distance from previous mutation (bp) | Figure |
|---|---|---|---|---|
| 4 Base INDEL | Promoter P6 | 160731824 | 0 | 8.3 |
| 1 Base INDEL ( ± /G) | Promoter Ex1a3 | 160734180 | 2356 | 8.4 |
| Rare SNP (A/C) | CpG Ex1a3 | 160734272 | 92 | 8.5 |
| | | | | |

**Figure 8:3 sequence of a four base insertion in schizophrenia in amplimer P6 in the promoter region of UHMK1.**




As seen in Figure 8:3 a four base insertion has shifted the bases up by four positions after the AACC repeat. This insertion deletion was later found to be approximately equal in cases and controls (position 160731824).

**Figure 8:4 sequencing of an insertion/deletion in the promoter region near exon 1 of UHMK1 in controls.**

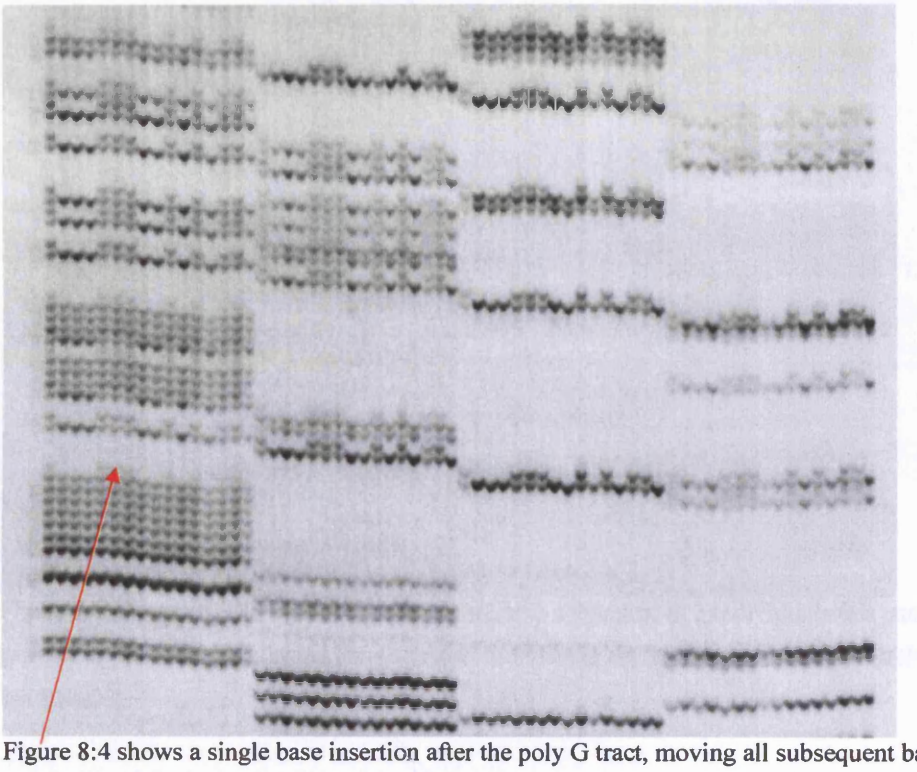


Figure 8:4 shows a single base insertion after the poly G tract, moving all subsequent bases up one position. This insertion deletion was later found to be in approximately equal frequencies in cases and controls (±/G position 160734180).

**Figure 8:5 Sequence of a single nucleotide polymorphism SNP of A to C in schizophrenia caes.**

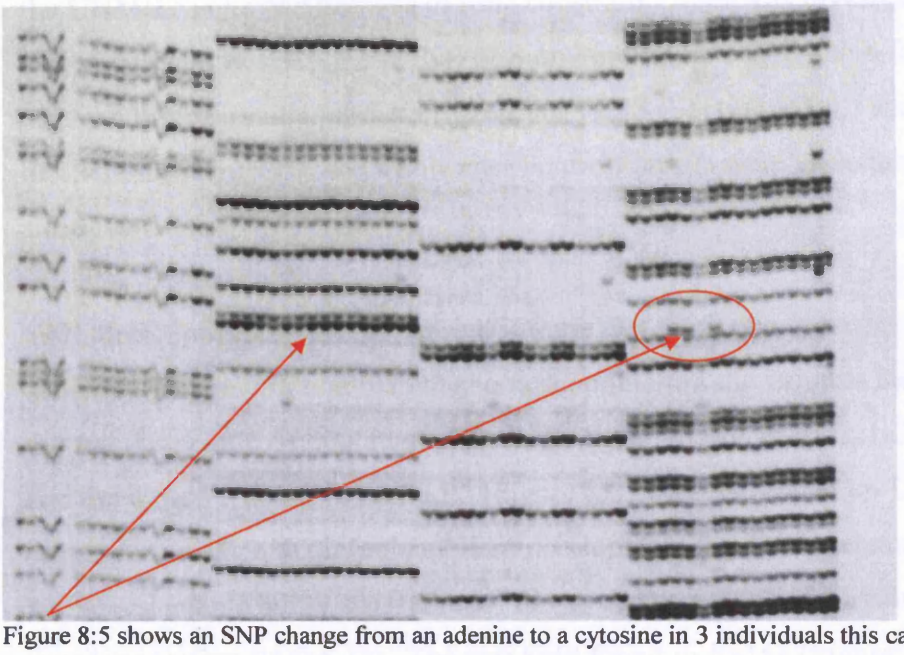


Figure 8:5 shows an SNP change from an adenine to a cytosine in 3 individuals this can be seen parallel to the adenine calls and shows that all individuals are heterozygous for the SNP (A/C position 160734272).

As seen in Table 8:2 DNA variants were found in the promoter and CpG island located close to exon 1. The variants consisted of a 4 base pair insertion deletion in the promoter region (Figure 8:3), and 2.4Kb down stream a single base pair insertion deletion (±/G IN/DEL) Figure 8:4. In addition there was a low frequency adenine to cytosine change in the CpG island only 92 bases away from the single base IN/DEL (Figure 8:5).

Of interest was the fact that the rare SNP (A/C position: 160734272, Figure 8:5) was found in only 3 of the chosen 32 schizophrenic patients and in none of the 32 randomly chosen controls. However when the controls were screened a high frequency single base IN/DEL (±G 160734180, Figure 8:4) was seen only 92 bases from the SNP location that was not present in the Schizophrenic samples. The 4 base IN/DEL further up the promoter (location 160731824, Figure 8:3) was observed in only half of the schizophrenics and in most of the controls as well.

Re-sequencing of the exons and the promoter region of the UHMK1 gene, led to the discovery of three non-database variants. Because of its location in the 5′ region of the UHMK1 gene and because of time constraints upon this investigator, both the effects of the non-dbSNP and INDELS on binding of transcription factors (carried out by a bioinformatic prediction program TESS (Section 8.1.4)) and the genotyping of allele frequencies in the whole case control sample were undertaken simultaneously.

With these intriguing results, to find if these DNA variants were more prevalent in schizophrenic population than the control population the variants had to be typed in the full case control population. The single base INDEL from Ex1a3 (Figure 8:4) and the 4 base INDEL found in P6 (Figure 8:3) was genotyped on LiCOR DNA sequencers using a 40cm polyacrylamide sequencing gels to increase the resolution and hence enable better separation of the product sizes. The individuals were genotyped using SAGA-GT. The rare SNP found in Ex1a3 (Figure 8:5) in the CpG island was genotyped using EPOCH probes in a qPCR machine (method described in section 2.8.3.2). The results are shown below.

**Table 8:3 The results of typing detected DNA variants in the entire case control UCL sample.**

| **Variants** | | | | **chi$^2$** | **P-Value** |
|---|---|---|---|---|---|
| **4 Base INDEL P6** (Figure 8.3) | | Deletion | Insertion | 0.008 | 0.930 |
| | Controls | 346 | 428 | | |
| | Cases | 363 | 453 | | |
| | | | | | |
| **1 Base INDEL Ex1a3** (Figure 8.4) ±/G | | Deletion | Insertion | 1.920 | 0.166 |
| | Controls | 577 | 233 | | |
| | Cases | 618 | 214 | | |
| | | | | | |
| **SNP Ex1a3** (Figure 8.5) A/C | | A | C | 1.193 | 0.275 |
| | Controls | 928 | 8 | | |
| | Cases | 925 | 13 | | |

Table 8:3 shows that none of the observed DNA variants were significantly associated with schizophrenia in the entire sample. Although the results look disappointing, there is a possibility that the SNP in the CpG island (Ex1a3 A/C

Figure 8:5) could be present in a particular sub-type of schizophrenia, or become positive if the sample size is increased. This could be possible due to the complexity of the disease, which is why it was easily picked up in the most significantly associated individuals chosen with the significant haplotypes. There is a possibility that the sample does not have the power to detect a low frequency SNP.

Once the DNA variants were found, the variants in the promoter region were screened bioinformatically with a program called transcription element search system (TESS) before the variants were fully genotyped in the whole sample. TESS compares variant with the wild type sequence, to see in any transcription factor binding sites were significantly disrupted.

### 8.1.4 TRANSCRIPTION ELEMENT SEARCH SYSTEM (TESS)

The novel DNA variants identified by re-sequencing the promoter region of the UHMK1 gene were examined by TESS in an attempt to identify altered transcription factor binding.

Twenty five base pairs either side of the non database SNPs were selected and analysed in the transcription factor binding site search engine, TESS (transcription element search system). (http://www.cbil.upenn.edu/cgi-bin/tess/tess?RQ=WELCOME).
Both the normal and variant sequences were submitted to TESS and the analysis recorded. The numerical values shown in the output from TESS predict in a logarithmic scale the strength of the binding for various transcription binding factors.

**Table 8:4 Wild type sequence and variant sequence that were analysed by TESS to determine if the binding efficiency of transcription factors were significantly altered.**

| Mutation type | Sequence | |
|---|---|---|
| 4 Base INDEL P6 (Figure 8:3) | GAGTGAGACTTTGTCTCAAACAAACAAACCAACCAACCAACCTATCTTGT | WT |
| | GAGTGAGACTTTGTCTCAAACAAACcaaaAAACCAACCAACCAACCTATCTTGT | INS |
| 1 Base INDEL Ex1a3 (Figure 8:4) | TCTCTTCTAGCCCCGCCCCTTCTGAGCCCCCCCTCTTCGGCCTGTATGATA | WT |
| | TCTCTTCTAGCCCCGCCCCTTCTGAGcCCCCCCCTCTTCGGCCTGTATGATA | INS |
| SNP Ex1a3 (Figure 8:5) | CACGGCTTCCGGTGTCATGGCTGCTtGAAGTCCCGGGAGTCGGTGAGGCGG | WT |
| | CACGGCTTCCGGTGTCATGGCTGCTgGAAGTCCCGGGAGTCGGTGAGGCGG | |

WT = Wild type sequence

INS = insertion

The results showed that neither the rare SNP in Ex1a3 (Figure 8:5) nor the 1 base IN/DEL in Ex1a3 (Figure 8:4), significantly altered the predicted binding efficiencies of any transcription factors. However, the wild type sequence which contains the 4 base IN/DEL P6 (Figure 8:3), demonstrates significant binding capabilities to transcription factors Freac-6 and Ste11 both with a LOD value of 17.00. When the 4-base insertion is added to the sequence as seen in Table 8:4, the two transcription factors disappear. Freac-6 (Forkhead-related transcription factor 6) is involved in the regulation of embryonic development in humans, and Ste11 is a Serine/threonine protein kinase required for cell-type-specific transcription and signal transduction involved in pheromone response and pseudohyphal/invasive growth pathways in yeast, it has no known function in humans.

After genotyping in the entire UCL case control sample, none of the variants found by sequencing were associated with schizophrenia. However because some of the variants may have low penetrance it is still just possible that they have aetiological significance. Only a much larger sample could show this. It might be worth while to test the effects of the mutated sequences on transcription factor binding using an Electrophoretic Mobility Shift Assay (EMSA), which is a technique described by Fried et al (1981). This is a method where a sequence of interest is compared to its

wild type, if a sequence binds to a protein from a nuclear extract such as a transcription factor it would migrate more slowly through a gel compared to a DNA fragment where no linking has occurred. This would be the first sign in finding out if a mutated nucleotide sequence prevents protein-nucleotide interaction compared to the wild type. However based on the low significance of these results and time constraints, this was not carried out.

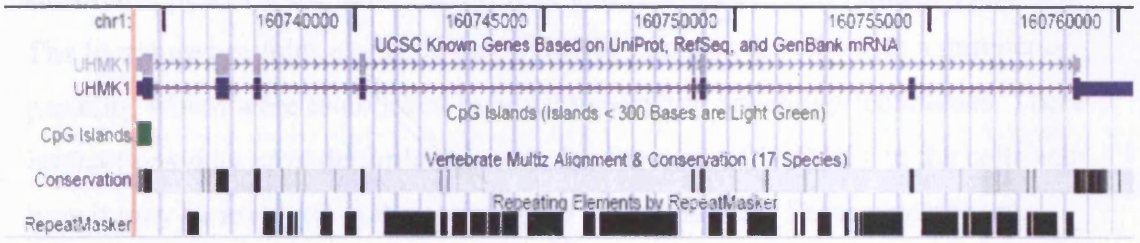
There are many other complexities that affect gene expression that could to be explored. For example the methylation state of the promoter region (particularly the CpG island) would be interesting to look at because all of the sequence variants were found in the promoter region. A larger region of the 5' and 3' UTR should be sequenced because distant mutations can effect gene expression as previously discussed. The introns should also be fully sequenced because they are known to be important in gene regulation, for example mutations in introns can cause incorrect splicing resulting in "exon skipping" or premature termination of the protein.

# 9.0 U2AF HOMOLOGY MOTIF (UHM) KINASE 1 (UHMK1), GENE, FUNCTION AND INTERACTIONS

## 9.1 UHMK1 BASIC CHARACTERISTICS

The U2AF Homology Motif (UHM) Kinase 1 gene is located on chromosome 1q23.3, genomic position 160734279-160760468 and is transcribed on the positive strand. The gene consists of 8 exons and has a CpG island in the 5' end that overlaps exon 1.

**Figure 9:1 The location and features of UHMK1**




The transcript contains 2901 bases (UCSC March 2006 assembly) and has an ATG start codon within the sequence ccaacaccgATGg that has weak homology to the Kozak sequence (Maucuer et al. 1997). The transcript is translated into the UHMK1 protein which contains 419 amino acids with a molecular weight of 46.4 KiloDaltons.

## 9.2 BRIEF GENE FUNCTION

UHMK1 is a serine/threonine-protein kinase, which phosphorylates CDKN1B/$p27^{Kip1}$ when serum is present, thus controlling CDKN1B subcellular location and cell cycle progression in G1 phase. The gene may be involved in trafficking and/or processing of RNA as it contains an RNA recognition motif (by similarity) (provided by UCSC March 2006 assembly). UHMK1 does not fit into

any known kinase subfamily because it has a catalytic core that does not share homology with any known kinase. It is also the only know kinase that contains an RNA recognition motif. Therefore UHMK1 belongs to a new subfamily of kinases.

UHMK1 is ubiquitously expressed, but expression is much more abundant in the nervous system during development and in the brain during adulthood. The protein has the ability to phosphorylate in vitro classical substrates such as myelin basic protein and synapsin 1 (Maucuer et al. 1997). UHMK1 is expressed in neurons during development and also in mature neurons (Maucuer et al. 1995; Maucuer et al. 1997; Maucuer et al. 2000; Boehm et al. 2002; Bieche et al. 2003).

## 9.3 DETERMINING THE FUNCTIONAL ROLE OF UHMK1

UHMK1 is a complex multifunctional protein that interacts with a wide rage of substrates, however the action and true role of this protein still remains unknown. The literature provides evidence that UHMK1 interacts closely with a number of proteins, which were established by yeast two hybrid screening techniques. These interactions may provide further clues to the function of UHMK1 in the cells and how it may be involved in the aetiology of schizophrenia. These proteins are:-

- Stathmin
- Splicing Factor 1 (SF1)
- $P27^{Kip1}$
- Peptidyl-glycine α-amidating monooxygenase (PAM).

Each protein will be discussed below along side an interpretation of the complex function of the UHMK1 gene product.

### 9.3.1 EXPRESSION OF UHMK1

In a study which quantified normalised UHMK1 mRNA levels by real time RT-PCR in a panel of tissues from adult and developing rat, and a collection of human tissues. It was found that UHMK1 expression is greater in nervous tissues in both rat

and human, and expression outside the nervous system seem to be evenly distributed with a ratio of 10 between the kidney and pancreas (Bieche et al. 2003) Figure 9:2. This shows that UHMK1 may have a particular function in the nervous system during development of the human nervous system and in the maintenance of the adult brain.

**Figure 9:2 (Figure taken from (Bieche et al. 2003)), showing the expression of UHMK1 across human tissue samples.**

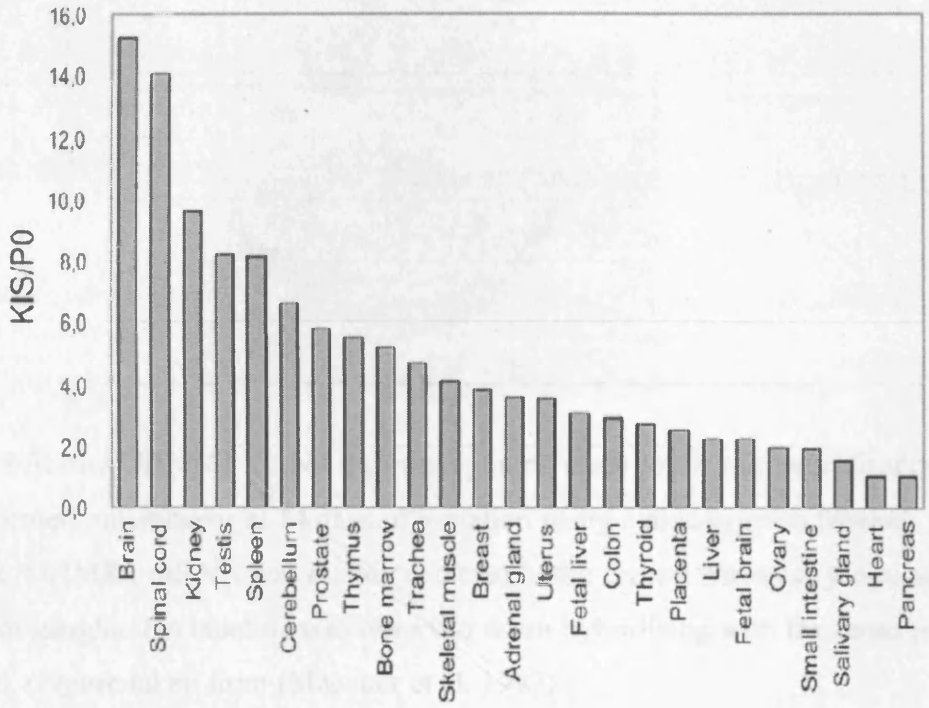



Figure 9:2 shows UHMK1 mRNA is ubiquitously expressed in human tissues. The higher levels are detected in the brain and spinal cord. UHMK1 mRNA was quantified using real-time RT-PCR in a series of RNA samples from human tissues. Results were normalised to RPLP0 signal and then to the lowest signal measured (pancreas).

In situ hybridisation experiments in developing rat embryos showed that UHMK1 mRNA was strongly expressed in the developing nervous system, Figure 9:3. The role for UHMK1 expression in the developing nervous system was further supported

by the finding that UHMK1 expression in human also increases during nervous tissue ontogenesis (Bieche et al. 2003).

**Figure 9:3 Expression of UHMK1 mRNA in the rat embryo.**

Figure 9:3 Shows UHMK1 mRNA expression in rat embryo. In situ hybridisation was preformed on embryos at 14 days of gestation using a digoxygenin labeled riboprobe. UHMK1 mRNA was mainly detected in the central nervous system and dorsal root ganglia. No labeling was observed when hybridising with the sense probe as control. (Figure taken from (Maucuer et al. 1997)

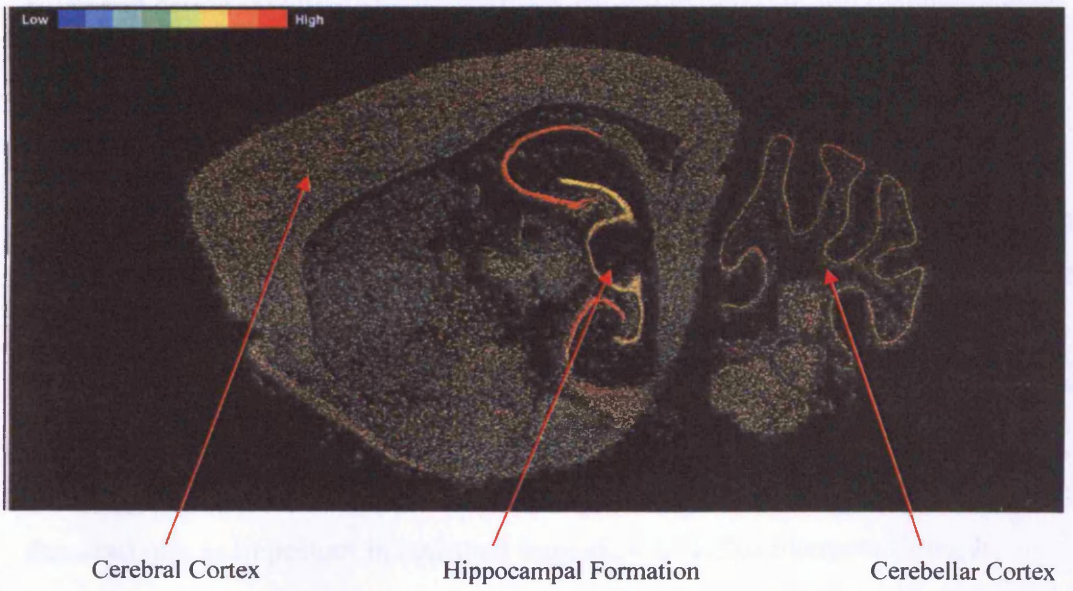
Research by Maucuer et al (1997) has also shown that during nervous system maturation, the levels of UHMK1 mRNA decreased, and then increased again in the adult brain (Maucuer et al. 1997).

In-situ hybridisation of rat brain tissue were carried out to observe the distribution of UHMK1 expression by Bieche et al (2003). The results showed differential expression in the cortical layers, as well as variable expression in regions of the hippocampus. Particularly strong expression was observed in the substantia nigra compacta, and nuclei of the brain stem, MA3 (medullar accessory occulomotor

nucleus) and red nucleus, motor trigeminal nucleus (Mo5), mesencephalic trigeminal nucleus (Me5), pontine reticular nucleus caudal (PnC), trapezoid body (Tz), Superior olive (SO), Vestibular nucleus, gigantocellular reticular nucleus (Gi) and facial nucleus (Bieche et al. 2003)(results not shown).

In addition, by observation of in situ hybridisation (ISH) of the sagittal section of the mouse brain (Paul Allen Brain Atlas) shown in Figure 9:4. High expression of UHMK1 in the hippocampal formation was shown. Also ISH showed that UHMK1 is expressed at lower levels through out the brain, and confirms expression patterns described by Maucuer et al (1997) and Bieche et al (2003).

**Figure 9:4 In situ hybridisation with UHMK1 carried out in the mouse brain and highlighted by software, section extracted from the Allen Brain Atlas website.**




Extensive analysis firmly establishes a ubiquitous but preferentially neural expression of UHMK1 gene in the rat and human. It has been shown to be highly expressed in the mature brain. Both the gene and its protein are highly conserved in the mouse, rat and human (Bieche et al. 2003).

Although UHMK1 is highly expressed in the brain its precise role is still unknown. The functional domains of the protein and the recently discovered interacting

proteins, provide only hints to its function, as described next, previous experiments with Stathmin which interacts with UHMK1 have been carried out.

### 9.3.2 UHMK1 WITH STATHMIN

UHMK1 has also been designated with the gene names KIS and KIST. KIST stands for "Kinase interacting with Stathmin." This interaction had originally been found by yeast two hybrid screening system using stathmin as bait (Maucuer et al. 1995). Stathmin is a small ubiquitously expressed cytoplasmic phosphoprotein that is enriched in neurons. It is hypothesised to play a role in the relay and integration of diverse intracellular signalling pathways and networks. Stathmin is phosphorylated in response to many signals such as hormone growth and neuronal differentiation factors, neurotransmitters or upon activation of T lymphocytes. It has also been proposed as a key regulator of microtubule dynamics in particular during the cell cycle and hence involved in cell proliferation and differentiation (Maucuer et al. 1997; Maucuer et al. 2000; Boehm et al. 2002).

UHMK1 phosphorylates stathmin on serine residues in different positions from those sites already known to be phosphorylated by other proteins in vivo (Maucuer et al. 1997). This may indicate another regulatory role of stathmin in a cell type or a situation that has not currently been explored. It is now known that UHMK1 has preferential expression in the developing nervous system and in the mature brain, and this expression is correlated with stathmin expression and related neuronal proteins which interact with UHMK1 (Maucuer et al. 1997). Studies have shown that stathmin is important in neuronal migration and also interacts through phosphorylation with brain derived neurotrophic factor (BDNF), with the possibility that it may be important in the development of cortical neurons (Cardinaux et al. 1997; Giampietro et al. 2005).

The stathmin gene locus has been subjected to "knock out" transgenic experiments in the mouse. It was found that mice homozygous for the stathmin gene deletion

lacked instinctive fear and had other behavioural abnormalities. The stathmin knockout also had poor memory for past aversive experiences such as those in fear conditioning tests (Shumyatsky et al. 2005). The gene is highly enriched in the lateral nucleus of the amygdala. Because of the interaction of UHMK1 protein with stathmin protein, UHMK1 might also be expected to produce similar behavioural abnormalities. Some of these abnormalities might be symptoms of schizophrenia.

A separate investigation by Liedtke et al (2002) showed that aging stathmin deficient mice developed an axonopathy of the central and peripheral nervous systems. The pathological hallmark of the early axonal lesions was a highly irregular axoplasm predominantly affecting large, heavily myelinated axons in the motor tracts. As the lesions progressed, degeneration of axons, dysmyelination, and an unusual glial reaction were observed. These findings further support the essential role that stathmin plays in the maintenance of axonal integrity (Liedtke et al. 2002).

A study comparing anterior cingulate cortex gray matter proteomes between patients with schizophrenia and controls found that stathmin was increased significantly by 1.8 fold in the schizophrenia cohort (Clark et al. 2006). This is possibly relevant because UHMK1 is directly associated with stathmin and we know that the stathmin knock out mouse exhibits behavioural abnormalities that are compatible with schizophrenia.

### 9.3.3 UHMK1 AND RNA METABOLISM

As stated earlier UHMK1 (U2AF homology motif kinase 1) is a serine/threonine-protein kinase with an RNP-type RNA recognition motif with an intriguing homology to the C-terminal motif of the splicing factor U2AF. U2AF is known to specifically bind to polypyrimidine (T,C) tracts associated with 3' splice sites (Maucuer et al. 1997).

UHMK1 is thought to be involved in RNA processing. Over-expression of UHMK1 in HEK293 fibroblastic cells demonstrated that the protein is present in the cytoplasm and enriched in the nucleus. This may reflect the shuttling of the kinase in relation to its function in regulating RNA associated factors. Thus UHMK1 may be implicated in the trafficking and/or splicing of RNAs and phosphorylating RNA associated proteins, and may ultimately have a role in the control of gene expression, given its RNA recognition motif and phosphorylation potential (Maucuer et al. 1997; Maucuer et al. 2000). UHMK1 also been shown to have a narrow substrate specificity because it preferentially phosphorylates serine or threonine residues flanked by a carboxy-terminal proline.

#### 9.3.3.1 UHMK1 INTERACTION WITH SPLICING FACTOR 1

The yeast two hybridisation system has shown that splicing factor 1 (SF1), interacts with UHMK1. UHMK1 phosphorylates the two serine residues adjacent to proline residues at positions 80 and 82. The phosphorylated SF1 protein increases its affinity to the $U2AF^{65}$ motif of UHMK1 and forms a complex to the pre-mRNA (Manceau et al. 2006). SF1 binds to the branch point pre-mRNA consensus sequence (BPS) near the 3' splice site and facilitates binding of the $U2AF^{65}$ motif to the adjacent polypyrimidine tract forming the initial spliceosome. SF1 is displaced from the spliceosome by the ATP-dependent entry of the U2 small nuclear ribonucleoprotein particle (snRNP), whose SF3b155/SAP155 protein subunit interacts with $U2AF^{65}$ and whose RNA component (U2 snRNA) anneals with the BPS. This first ATP-dependent step of 3' splice site recognition represents a critical step in the regulation of pre-mRNA splicing (Manceau et al. 2006). Protein kinase (PKG) is a potential regulator of this step, by inhibiting the initial SF1-$U2AF^{65}$ interaction by phosphorylation of SF1 serine(20) in the $U2AF^{65}$ interacting domain.

### 9.3.4 UHMK1 and $p27^{Kip1}$

UHMK1 was found to regulate the cell cycle by responding to mitogens and negatively regulating cdk inhibitor $p27^{kip1}$, phosphorylating it on serine 10 and promoting its nuclear export into the cytoplasm. This resulted in the promotion of

cell cycle progression from growth arrest (quiescence $G_0$ through to $G_1$) (Boehm et al. 2002).

Relatively low levels of UHMK1 mRNA were detected in the rat embryo, but an increase was observed in the brain around birth and postnatal development. Similarly Bieche et al (2003) has found a higher mRNA level in adult human brain as compared to foetus. Thus its been proposed that UHMK1 may inhibit $p27^{Kip1}$ activity in most tissues and during development but it plays additional functions and phosphorylates different substrates in the mature adult nervous system (Bieche et al. 2003). This potentially links UHMK1 in neural differentiation and function on one side and the control of cell cycle on the other.

### 9.3.5 THE ROLE OF INTERACTION BETWEEN UHMK1, $P27^{KIP1}$ AND STATHMIN

New research has identified the conection between the two proteins that closely interact with UHMK1, stathmin and cytoplasmic $p27^{Kip1}$. Iancu-Rubin et al (2005) has implicated them in relation to cell migration and proliferation.

Cytoplasmic $p27^{Kip1}$ was shown to play a role in the regulation of cell migration (Denicourt et al. 2004). It has been shown in two studies that stathmin is necessary for the migration of neurons in vivo and in vitro. Jin et al (2004) demonstrated that inhibition of stathmin mRNA caused inhibited migration of newly formed neurons of the olfactory system in adult rat brain. In addition Giampietro et al (2005) showed that migration of immortalised neurons is decreased when stathmin expression is down regulated and is increased when stathmin is over expressed.

Baldassarre et al (2005) showed that $p27^{Kip1}$ inhibits cell migration and $p27^{Kip1}$ interferes with the ability of stathmin to sequester tubulin, leading to increased microtubule polymerisation. Over expression of stathmin has the effect of increasing cell migration, and hence when $p27^{Kip1}$ has phosphorylated stathmin it leads to a

reduction of cell migration. Although $p27^{Kip1}$ and stathmin were originally discovered as important regulators of the eukaryotic cell cycle, both are now thought to be involved in the process of cell migration. A dual role for these proteins in cell migration and proliferation should not be surprising as microtubules are known to be important in both processes (Iancu-Rubin et al. 2005).

### 9.3.6 UHMK1 INTERACTION WITH PAM

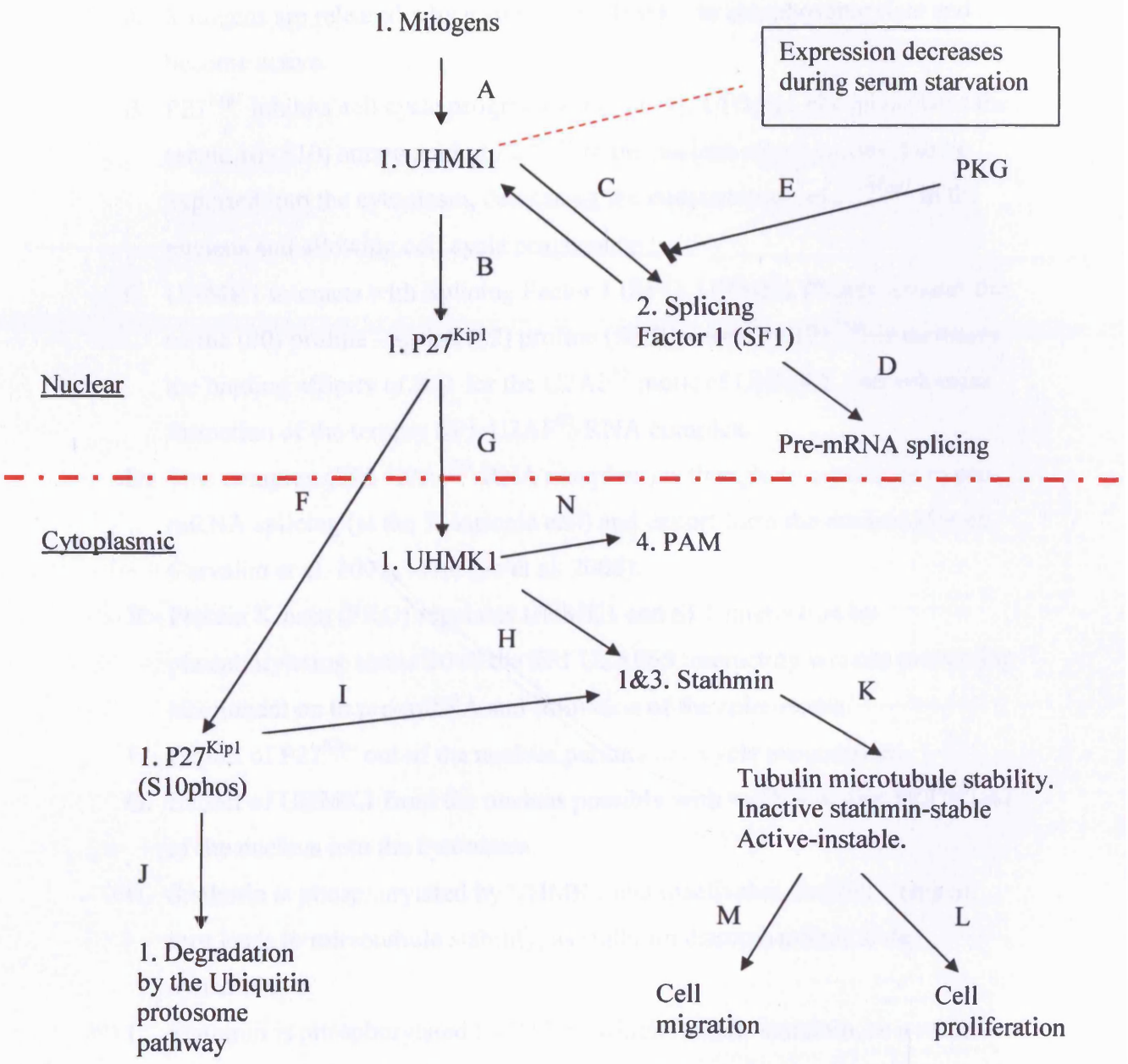
UHMK1 has also been found to interact with the cytoplasmic domain of the peptidyl-glycine α-amidating mono-oxygenase (PAM) suggesting an additional function for the protein (Alam et al. 1996; Maucuer et al. 1997; Maucuer et al. 2000). PAM also known as P-CIP2 (PAM COOH-terminal interactor protein) and a second interacting protein P-CIP10 which is similar to Huntingtin-associated protein-interacting protein (Caldwell et al. 1999), was found to directly interact with PAM by means of the yeast two hybrid system in a rat hippocampal cDNA library, all three proteins are highly expressed in the brain in neurons (Alam et al. 1996).

PAM has two enzymatically active domains with catalytic activities - peptidylglycine alpha-hydroxylating monooxygenase (PHM) and peptidyl-alpha-hydroxyglycine alpha-amidating lyase (PAL). These catalytic domains work sequentially to catalyze neuroendocrine peptides to active alpha-amidated products.

PAM is also involved in the regulated secretory pathway in neurons and endocrine cells in which biologically active peptides are stored in large dense core vesicles (LDCVs) and undergo regulated release (Alam et al. 1996). PAM is thought to be involved in activating peptides and routing them to LDCVs. PAM itself can be tethered to the vesicles and does so almost exclusively in the nervous system (Alam et al. 1996) PAM is also likely to be involved in vesicle routing and trafficking. UHMK1 and PAM are both found to be expressed in neurons (Caldwell et al. 1999), and UHMK1 was found to bind to the C-terminal domain of integral membrane PAM. UHMK1 is thought to be an important part in the routing and distribution of

the PAM protein by the process of phosphorylation of serine (949) residues. The expression of PAM may signal UHMK1 to become more diffuse in the cytoplasm.

**Figure 9:5 Diagram showing protein interactions with UHMK1 deduced from a number of sources.**




1. Boehm et al (2002)
2. Manceau et al (2006)
3. Iancu-Rubin et al (2005)
4. Alam et al (1996)

## 9.4 DESCRIPTION OF PATHWAYS OCCURRING IN Figure 9:5

A. Mitogens are released which stimulate UHMK1 to autophosphorylate and become active.

B. $P27^{Kip1}$ inhibits cell cycle progression G0 to G1, UHMK1 phosphorylates the serine 10 (S10) amino acid of $P27^{Kip1}$ in the nucleus which causes it to be exported into the cytoplasm, decreasing the concentration of $P27^{Kip1}$ in the nucleus and allowing cell cycle progression

C. UHMK1 interacts with Splicing Factor 1 (SF1). UHMK1 Phosphorylates the serine (80) proline – serine (82) proline (SPSP) motif of SF1. This increases the binding affinity of SF1 for the $U2AF^{65}$ motif of UHMK1, and enhances formation of the ternary SF1-$U2AF^{65}$-RNA complex.

D. This complex (SF1-$U2AF^{65}$-RNA complex ) is thought to contribute to pre-mRNA splicing (at the 3' intronic end) and export form the nucleus (Gama-Carvalho et al. 2001; Manceau et al. 2006).

E. Protein Kinase (PKG) regulates UHMK1 and SF1 interaction by phosphorylating serine 20 of the SF1 U2AF65 interacting domain preventing attachment on to pre-mRNA and formation of the splicosome.

F. Export of $P27^{Kip1}$ out of the nucleus permits cell cycle progression.

G. Export of UHMK1 from the nucleus possibly with mRNA and/or $P27^{Kip1}$ out of the nucleus into the cytoplasm.

H. Stathmin is phosphorylated by UHMK1 and inactivates stathmin. This in turn leads to microtubule stability, as stathmin disrupts microtubule dynamics.

I. Stathmin is phosphorylated by $P27^{Kip1}$ which inhibits stathmin, as a result stabilises microtubule formation, which decreases cell migration.

J. $P27^{Kip1}$ is degraded by a number of pathways, one being the ubiquitin proteosome pathway.

K. Dependent on the Stathmin phosphorylation state (whether phosphorylated or not, in addition to the location of phosphorylation) Stathmin will effect tubulin stability and in turn microtubule formation.
L. Inactivation of stathmin will lead to Cell proliferation.
M. Inactivation of stathmin will lead to reduced cell migration.
N. PAM interacts with UHMK1 by yeast 2 hybrid system, PAM is phosphorylated by UHMK1 at serine 949, it is suggested that this is required to localise PAM to large dense core vesicles which aid in activating bioactive peptides in neurons. PAM is seen to be involved in the regulated secretory pathway in neurons and endocrine cells.

Because it is known that the stathmin mouse knockout exhibits behavioural abnormalities, it is of interest to know if pharmacological interventions might effect UHMK1 gene expression.

This was investigated with the antipsychotic drug clozapine in an independent microarray study comparing the effect of clozapine and haloperidol antipsychotic therapy in mice Rizig et al (2009). It was found that UHMK1 was significantly down regulated by clozapine but not haloperidol. This does not prove involvement of UHMK1 in schizophrenia but does hint at a possible reason why clozapine is a more effective drug than haloperidol.

With the known effect of UHMK1 on neuronal growth and differentiation and its high level of expression in the hippocampus, it is posible to envisage the effect of an abnormally expressed UHMK1 would have on an individual. Currently there are no animal models for altered UHMK1 expression, and whether neuronal loss would be a result of such an abnormality. However, studies have shown the effect and possible causes of neuronal loss around the hippocampal formation, and the importance of growing neurons in this area. One such study showed that neural stem cells, located in the subventricular zone (svz) and the subgranular zone (SGZ) of the dentate gyrus

(DG), produce new neuronal and glial cells in the hippocampus of adult mammals (neural stem cell proliferation; NSP). Some of these neuronal cells differentiate into neurons that intergate functionally and structurally into existing neural networks a process known as adult neurogenesis. Thereby maintaining neuronal plasticity and is probably involved in memory formation or stress responses. The hippocampus is thought to be involved in the etiopathology of depression and adult neurogenesis has been suggested to be involved in the pathophysiology and treatment of mood disorders with both tricyclic and serotonin reuptake inhibitor antidepressants (Reif et al. 2006).

Findings by Reif et al (2006) suggest that reduced neural stem cell proliferation is present in the hippocampus of schizophrenia patients and may contribute to the pathogenesis of schizophrenia, due to a reduction in the amount of newly formed cells (Reif et al. 2006). These findings fit with the known functions of UHMK1 such as being involved in cell cycle progression, neuronal migration and outgrowth in addition to its high expression in the hippocampal formation.

## 9.5 CONCLUSION

With the evidence shown, it can be seen that the actions of UHMK1 protein are complex and multifunctional. Functions of UHMK1 involve activation of the cell cycle, pre-mRNA processing involving intronic splicing and export into the nucleus. As discussed earlier UHMK1 is also involved in interacting with PAM, which is involved in the neuronal secretory pathway of bioactive peptides. UHMK1 also interacts with stathmin which is involved in microtubule destabilisation and neuronal integrity. It seems that UHMK1 is part of a delicate and complex network of neuronal control, regulating neuronal growth and possibly neuronal signalling. UHMK1 gene is highly expressed in the hippocampus, which is known to be part of the brain where neurons still divide and regenerate. UHMK1 is an attractive candidate gene for schizophrenia susceptibility as shown through the genetic linkage and association studies as well as from neurobiological pespective.

# 10.0 THESIS DISCUSSION

In this thesis I have tested for association of previously implicated genes RGS4 and CAPON to schizophrenia (section 4.0 & 5.0). We found no allelic association with schizophrenia with either of these genes using the UCL case control sample. As a result further fine mapping had been carried in the chromosomal region between the two genes as described in section 6.0

Section 6.0 and 7.0 had shown significant allelic and haplotypic association had been detected implicating the U2AF homology motif kinase 1 (UHMK1) as a schizophrenia susceptibility gene.

Sequencing of UHMK1 exons, 3' and 5' UTR, had produced three mutations in the promoter region. Genotyping of these genetic variants in the case control samples showed no evidence of association with schizophrenia (section 8.1.3). Sequencing of the greater promoter region and the introns are suggested, which may hold further interesting and potential aetiological mutations.

It was also seen in this thesis that UHMK1 is highly expressed in the nervous system and brain, and has a role in its development, the interactions of UHMK1 with other proteins such as stathmin and PAM added further weight to the possibility of UHMK1s role in behavioural characteristics and neuronal development and signalling as discussed in section 9.0.

UHMK1 is known to be highly expressed in regions of the brain implicated in schizophrenia, it has been found to be significantly down regulated in mice treated with antipsychotic drug clozapine (Rizig et al (2009)). Further confirmation of the involvement of this gene in schizophrenia is needed followed by further efforts to detect genetic variation in or next to the gene (which may even effect neighbouring genes) .

## 10.1 FUTURE WORK

### 10.1.1 FURTHER REPLICATIONS

Further fine mapping around the gene should be carried out to exclude neighbouring genes more convincingly and to show that the association was not attributed to linkage to a close neighbouring gene. This, as a result would give further support to UHMK1 as the candidate susceptibility gene. Clearly, for the UHMK1 gene to become a credible contender as an important component in the susceptibility of schizophrenia the association findings need to be repeated in several more populations. However, before any replications are attempted, attention should be paid to the methodology. Replication of association studies have become a universal cause for concern, in the field of genetics as a whole. Several journals have proposed guidelines or state in their editorial "Policies and Practices" that genetic association studies related to complex disorders are unlikely to be accepted (Saito et al. 2006).
A recent meta-analysis concluded that popultaion heterogeneity was in evidene for association studies (Ioannidis et al. 2001).
False-negative results may come from studying genes of modest effect that may only have odds ratios of between 1.1-1.5, explaining 1-8% of the overall disease risk in a population. Locus heterogeneity may further weaken the genetic signal. In reality locus heterogeneity has been proven for all common complex disease.
False-positive results may also be due to population stratification; this can result from hidden allele frequency differences detected due to ethnic and ancestral differences between cases and controls. This is observed when cases and controls are drawn from two or more racial groups where the disease prevalence varies by race along with the genetic variant frequency.
However, despite the above, over 50% of allelic associations in case control samples are replicated and confirmed Lohmueller (2003).
Therefore, any future investigations of UHMK1 and schizophrenia must involve attempts at replicating the positive associations reported here. Such replications should aim to use the same markers of the initial study and sufficient numbers of cases and controls so as to obtain sufficient power.

### 10.1.2 RE-SEQUENCING

In the re-sequencing of the UHMK1 gene or any gene thought to be involved in the susceptibility to schizophrenia it is of utmost importance either to discover any aetiological base-pair changes or to confirm that no such abnormalities are present. The bi-directional method of re-sequencing is employed in this thesis is not without faults. It would be appropriate to suggest that any areas which have been re-sequenced be repeated with an alternative method e.g. capillary sequencing. In addition a larger region of the 5' and 3' UTR should be sequenced to detect long range control regions or enhancer elements which could affect the expression efficiency of the gene. These locus control regions could be further than 14 kilobases away and situated within another gene (Ho et al. 2006). Indeed UHMK1 could be containing a long distance control region that could effect another gene, hence complete sequencing of the entire genomic extent of the gene is important.

Furthermore, as stated above, although UHMK1 is a highly plausible candidate for the susceptibility to schizophrenia, it may simply be that there is an alternative gene or control region within the region of UHMK1 responsible for the association.
It is not therefore unreasonable to suggest that any future work may involve the re-sequencing of introns of the UHMK1 gene. Indeed several diseases have been found to have a mutation within the introns as noted above. Myotonic dystrophy (type 2) has been found to be caused by a CCTG expansion (mean~5000 repeats) located in intron 1 of the zinc finger protein 9 (ZNF9) gene (Liquori et al. 2001). Also closely linked or even causative abnormalities can be found in the introns, such as the three SNP haplotype in intron 1 of OCA2 (Oculocutaneous Albinism, type 2) which is now thought to influence eye colour variation (Duffy et al. 2007).

### 10.1.3 GENOTYPING

In any future work, the non-dbSNPs discovered during re-sequencing of the introns and UTRs should be genotyped in the whole case-control sample. In addition, more

data-base SNPs including those found by re-sequencing and further SNPs chosen from the HapMap should be genotyped across the region so that the importance of UHMK1 as a susceptibility gene in schizophrenia can be investigated further and also, further fine mapping around the gene to help exclude neighbouring genes that may be the true susceptibility gene linked to the positive UHMK1 markers. It should also be pointed out that HapMap does not contain all the known SNPs covering all regions, and that typing of tagged SNPs within UHMK1 has been carried out to the best of our knowledge. The gene should be fully sequenced in order to find additional SNPs.

This study has failed to replicate association with previously implicated genes CAPON (NOS1AP) and RGS4, although this reduces evidence for association to schizophrenia with these genes we can not exclude it. NOS1AP is a large gene that according to the data in HapMap would require over 18 tagged SNPs to thoroughly evaluate the association with schizophrenia. Therefore both of the implicated genes could be more vigorously genotyped with tagged SNPs to increase confidance to exclude the genes as candidates to schizophrenia susceptibility. In addition the neighbouring genes to UHMK1 (UAP1 and DDR2) should also be typed with tagged SNPs to prove that they are not implicated in schizophrenia susceptibility and UHMK1 is not picking up LD with these genes.

### 10.1.4 GENOME SCANS

Whole genome scan association studies using large numbers of SNPs have been carried out looking at several other diseases including myocardial infarction, osteoarthritis (Abel et al. 2006) and breast cancer. Recently, Mah et al (2006) carried out a genome-wide scan using over 25,000 SNPs located within approximately 14,000 genes. They repeated the scan within several populations. They found a marker on 1q32 within a novel candidate gene for schizophrenia (PLXNA2).
Work in the future will include high density genome wide association studies aiming to reinforce current findings and/or identify further candidate genes. The

identification of genes by markers associated with specific loci may not be the final route by which complex disorders are completely understood. The loci may simply be regions for interaction or control. It is the understanding of these interactions and control mechanisms that will possibly be the focus of the attention of scientists over the next decade.

### 10.1.5 EXPRESSION VECTORS AND KNOCK-OUT/KNOCK-DOWN MOUSE MODELS

Once association with UHMK1 is beyond doubt and aetiological base pair changes have been identified then expression vectors and "knock down" experiments can be conducted. Transgenic knock out experiments are also highly desirable and would be of further interest to investigate what the effect it would have on UHMK1 expression and the organisms brain pathology and also behaviour.

### 10.1.6 METHYLATION STATUS AT CpG

DNA methylation involves the addition of a methyl group to the number 5 carbon of the cytosine pyrimidine ring. Methylation of the CpG island suppresses gene expression of the upstream gene. This activity is important for developmental roles and specific tissue expression. We have found a rare non-database SNP in the CpG island as described in section 8.1.3, and also two other mutations further upstream. It would be interesting to investigate how these mutations could affect the methylation status in the promoter region, especially in the case where we have the rare SNP.

Measuring DNA methylation can be carried out by sodium bisulfite treatment, which creates sequence differences by converting unmethylated cytosines to uracils, but leaving methylated cytosine unchanged. The differences can then be detected quantitatively by several techniques, such as sequencing of subclones or PCR products, restriction-digestion or pyrosequencing (Clark et al. 1994; Xiong et al. 1997; Colella et al. 2003).

### 10.1.7 MRI SCANS OF UHMK1 ASSOCIATED CASES.

Magnetic Resonance Imaging (MRI) is a powerful scanning devise which is able to build up detailed pictures of internal tissue organs, using safer form of radio waves, unlike the ionising waves of x-rays.

Upon further implication and conformation that UHMK1 is a schizophrenia susceptibility gene. One may want to investigate how the abnormally expressed gene would affect the brain morphology of schizophrenic patients which have the abnormal UHMK1 gene. This can be carried out by scanning the patients using an MRI machine; it would be interesting to see which areas of the brain (if any) are affected, compared to other schizophrenics and to normal controls. The results could be compared to what we have already discovered by MRI scanning of schizophrenic patients to see if they overlap, and potentially be implicated as a main cause factor to abnormal brain morphology.

### 10.1.8 INVESTIGATING EXPRESSION IN SCHIZOPHRENIC PATIENTS.

An important implication of a gene causing schizophrenia is whether it is abnormally expressed in schizophrenic patients compared to controls. That can be carried out by using quantitative RT-PCR, unfortunately obtaining suitable brain tissue not easy. One would need to use human brain tissue to extract mRNA, and this could only be obtained from external organisations. As a result it would not be possible to screen before hand which patients will be the most likely to have an abnormal UHMK1 gene. However, it would be of great interest if mRNA can be extract from the hippocampal formation where UHMK1 was found to be highly expressed, from a number of patients and compare it to the same number of matched controls. In addition the tissue samples could be used to extract the UHMK1 protein and investigate if there are different isoforms present in higher concentrations in the

schizophrenics or the controls when compared to one another. Caution is needed in evaluating post motem mRNA levels because drug effects can influence UHMK1 mRNA levels as has been shown with clozapine.

### 10.1.9 BETTER UNDERSTANDING OF PROTEIN-PROTEIN INTERACTIONS

In order to fully understand UHMK1's role in the pathology of schizophrenia and to look to find potential therapeutic targets to control the disease, one needs to fully understand the protein interactions associated with UHMK1. For future work, it would be worth investigating the protein interactions further. Biochemical techniques such as the yeast two hybrid screening should be employed to help build up the proein network picture. Coprecipitation and 2D gel studies also looking at glycosylation and phosphorylation changes may also aid in the study of the protein function and control.

### 10.1.10 GENE-GENE INTERACTIONS

It is important to consider gene-gene interactions, as well as recessive and dominant gene effects. The associated gene may not be directly responsible for schizophrenia susceptibility but its product may interact with other abnormal genes effecting their expression which may have an even greater effect on the pathology of schizophrenia. Specific gene interactions may occur in specific sub-types of schizophrenia. The gene interactions will not only help to confirm known candidate genes, but help to implicate other genes in schizophrenia susceptibility.

### 10.1.11 GENE SUSCEPTIBILITY AND SUB-TYPES OF SCHIZOPHRENIA

Sequencing and screening of the gene was carried out as previously stated (section 2.8.4 & 8.1) by using the individuals with the most significantly associated haplotypes. Screening of these individuals lead to the detection of the rare SNP in

the CpG island and also the two INDELs. Further genotyping of these changes in the entire population showed them not to be associated with schizophrenia. However, they could nevertheless be responsible for very rare subtypes of schizophrenia. Further work should be done to investigate whether these mutations are in fact a rare cause of schizophrenia. In the future even larger sample size would be justified by increasing the power to detect polymorphisms with high frequency SNPs as well to screen for additional rare susceptibility mutations.

## 10.2 CONCLUSION

UHMK1 is attractive as a schizophrenia susceptibility gene because it contains both a kinase and a RNA recognition motif which is widely expressed in the brain. The gene has been implicated in the British UCL case control sample and also replicated in a larger independent Scottish Aberdeen case control sample. UHMK1 is an attractive gene highly expressed in the developing nervous system and in the mature brain. UHMK1 expression is greatest in the hippocampal formation which links it to behaviour and memory. The protein is involved in cell cycle progression, neuronal out growth and possibly neuronal signalling. Further association studies in independent populations are still required, to fully confirm UHMK1 as a susceptibility gene for schizophrenia.

# 11.0 ELECTRONIC REFERENCES:


Allen Brain atlas, Allen Institute for Brain Science. http://www.brainatlas.org/

Ensembl http://www.ensembl.org/index.html

GDB http://www.gdb.org/

Haploview http://www.broad.mit.edu/mpg/haploview/

Human Brain Informatics:

http://www.hubin.org/facts/history/history_schizophrenia_en.html

International HapMap Project http://www.HapMap.org/

KBiosciences http://www.Kbioscience.co.uk

Laboratory of Neuro imaging, UCLA history of schizophrenia:

http://www.loni.ucla.edu/Research/Projects/Schizophrenia

LocusView http://www.broad.mit.edu/mpg/locusview/

MWG http://www.mwg-biotech.com/html/all/index.php

NCBI http://www.ncbi.nih.gov/

NetPrimer http://www.premierbiosoft.com/netprimer/netprlaunch/netprlaunch.html

OMIM SCZD [MIM 181599]:

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM

Primer3 http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi

Schizophrenia.com http://www.schizophrenia.com

Sigma http://www.sigmaaldrich.com

UCSC http://genome.ucsc.edu/

# 12.0 REFERENCES:


Abel, K., R. Reneland, et al. (2006). "Genome-wide SNP association: identification of susceptibility alleles for osteoarthritis." Autoimmun Rev **5**(4): 258-63.

Akbarian, S., J. J. Kim, et al. (1996). "Maldistribution of interstitial neurons in prefrontal white matter of the brains of schizophrenic patients." Arch Gen Psychiatry **53**(5): 425-36.

Alam, M. R., B. D. Caldwell, et al. (1996). "Novel proteins that interact with the COOH-terminal cytosolic routing determinants of an integral membrane peptide-processing enzyme." J Biol Chem **271**(45): 28636-40.

Andreasen, N. C. (1995). "Symptoms, signs, and diagnosis of schizophrenia." Lancet **346**(8973): 477-81.

APA (1994). APA 1994 Diagnostic and statistical manual of mental disorders. Washington, DC, American Psychiatric Association.

Badner, J. A. and E. S. Gershon (2002). "Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia." Molecular Psychiatry **7**: 405-411.

Baldassarre, G., B. Belletti, et al. (2005). "p27(Kip1)-stathmin interaction influences sarcoma cell migration and invasion." Cancer Cell **7**(1): 51-63.

Baron, M. (1986a). "Genetics of schizophrenia: I. Familial patterns and mode of inheritance." Biol Psychiatry **21**(11): 1051-66.

Baron, M. (1986b). "Genetics of schizophrenia: II. Vulnerability traits and gene markers." Biol Psychiatry **21**(12): 1189-211.

Baron, M., R. Gruen, et al. (1982). "Schizoaffective illness, schizophrenia and affective disorders: morbidity risk and genetic transmission." Acta Psychiatr Scand **65**(4): 253-62.

Barrett, J. C., B. Fry, et al. (2005). "Haploview: analysis and visualization of LD and haplotype maps." Bioinformatics **21**(2): 263-5.

Bengra, C., T. E. Mifflin, et al. (2002). "Genotyping of essential hypertension single-nucleotide polymorphisms by a homogeneous PCR method with universal energy transfer primers." Clin Chem **48**(12): 2131-40.

Benjamin J. Sadock, M. D. and M. D. Virginia A. Sadock (2003). Kaplan & Sadock's Synopsis of Psychiatry Behavioral Sciences/ Clinical Psychiatry. New York, Lippincott Williams & Wilkins.

Bieche, I., V. Manceau, et al. (2003). "Quantitative RT-PCR reveals a ubiquitous but preferentially neural expression of the KIS gene in rat and human." Brain Res Mol Brain Res **114**(1): 55-64.

Bishop, Y., S. E. Fienberg, et al. (1975). Discreate Multivariate Analysis: Theory and Practice, Cambridge, Mass., MIT Press.

Blackwood, D., A. Fordyce, et al. (1998). "Thirty year follow up of a family showing association of schizophrenia with a balanced translocation t(1:11)(q42.1,q14.3)." American Journal of Medical Genetics **81**: 532.

Blackwood, D. H., A. Fordyce, et al. (2001). "Schizophrenia and affective disorders--cosegregation with a translocation at chromosome 1q42 that directly

disrupts brain-expressed genes: clinical and P300 findings in a family." Am J Hum Genet **69**(2): 428-33.
Bleuler, M. (1978). The schizophrenia disorders: long term patient and family studies. Translated by S. Clements. New Haven, Yale University Press.
Boehm, M., T. Yoshimoto, et al. (2002). "A growth factor-dependent nuclear kinase phosphorylates p27(Kip1) and regulates cell cycle progression." Embo J **21**(13): 3390-401.
Botstein, D. and N. Risch (2003). "Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease." Nat Genet **33 Suppl**: 228-37.
Breen, G., D. Prata, et al. (2006). "Association of the dysbindin gene with bipolar affective disorder." Am J Psychiatry **163**(9): 1636-8.
Bromet, E. J. and S. Fennig (1999). "Epidemiology and natural history of schizophrenia." Biol Psychiatry **46**(7): 871-81.
Brzustowicz, L. M., K. A. Hodgkinson, et al. (2000). "Location of a major susceptibility locus for familial schizophrenia on chromosome 1q21-q22." Science **288**(5466): 678-82.
Brzustowicz, L. M., J. Simone, et al. (2004). "Linkage disequilibrium mapping of schizophrenia susceptibility to the CAPON region of chromosome 1q22." Am J Hum Genet **74**(5): 1057-63.
Cai, G., X. Wu, et al. (2002). "[Linkage analysis of susceptibility genes for familial schizophrenia on chromosome 1 in Chinese population]." Zhonghua Yi Xue Yi Chuan Xue Za Zhi **19**(6): 491-4.
Caldwell, B. D., D. N. Darlington, et al. (1999). "The novel kinase peptidylglycine alpha-amidating monooxygenase cytosolic interactor protein 2 interacts with the cytosolic routing determinants of the peptide processing enzyme peptidylglycine alpha-amidating monooxygenase." J Biol Chem **274**(49): 34646-56.
Cardinaux, J. R., P. J. Magistretti, et al. (1997). "Brain-derived neurotrophic factor stimulates phosphorylation of stathmin in cortical neurons." Brain Res Mol Brain Res **51**(1-2): 220-8.
Cardno, A. G. and Gottesman, II (2000). "Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics." Am J Med Genet **97**(1): 12-7.
Chen, X., C. Dunham, et al. (2004). "Regulator of G-protein signaling 4 (RGS4) gene is associated with schizophrenia in Irish high density families." Am J Med Genet **129B**(1): 23-6.
Chen, Y. S., N. Akula, et al. (2004). "Findings in an independent sample support an association between bipolar affective disorder and the G72/G30 locus on chromosome 13q33." Mol Psychiatry **9**(1): 87-92; image 5.
Chowdari, K. V., K. Mirnics, et al. (2002). "Association and linkage analyses of RGS4 polymorphisms in schizophrenia." Hum Mol Genet **11**(12): 1373-80.
Chumakov, I., M. Blumenfeld, et al. (2002). "Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia." Proc Natl Acad Sci U S A **99**(21): 13675-80.

Clark, D., I. Dedova, et al. (2006). "A proteome analysis of the anterior cingulate cortex gray matter in schizophrenia." Mol Psychiatry **11**(5): 459-70, 423.
Clark, S. J., J. Harrison, et al. (1994). "High sensitivity mapping of methylated cytosines." Nucleic Acids Res **22**(15): 2990-7.
Colella, S., L. Shen, et al. (2003). "Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites." Biotechniques **35**(1): 146-50.
Collins, F. S. (1992). "Positional cloning: let's not call it reverse anymore." Nat Genet **1**(1): 3-6.
Collins, F. S., M. S. Guyer, et al. (1997). "Variations on a theme: cataloging human DNA sequence variation." Science **278**(5343): 1580-1.
Cordell, H. J. (2002). "Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans." Hum Mol Genet **11**(20): 2463-8.
Curtis, D., J. Knight, et al. (2006). "Program report: GENECOUNTING support programs." Ann Hum Genet **70**(Pt 2): 277-9.
Curtis, D., B. V. North, et al. (2002). "A quick and simple method for detecting subjects with abnormal genetic background in case-control samples." Ann Hum Genet **66**(Pt3): 235-44.
de Bakker, P. I., R. Yelensky, et al. (2005). "Efficiency and power in genetic association studies." Nat Genet **37**(11): 1217-23.
Denicourt, C. and S. F. Dowdy (2004). "Cip/Kip proteins: more than just CDKs inhibitors." Genes Dev **18**(8): 851-5.
Devlin, B., S. A. Bacanu, et al. (2002). "Genome-wide multipoint linkage analyses of multiplex schizophrenia pedigrees from the oceanic nation of Palau." Mol Psychiatry **7**(7): 689-94.
Devlin, B. and N. Risch (1995). "A comparison of linkage disequilibrium measures for fine-scale mapping." Genomics **29**(2): 311-22.
Done, D. J., E. C. Johnstone, et al. (1991). "Complications of pregnancy and delivery in relation to psychosis in adult life: data from the British perinatal mortality survey sample." Bmj **302**(6792): 1576-80.
Duffy, D. L., G. W. Montgomery, et al. (2007). "A Three-Single-Nucleotide Polymorphism Haplotype in Intron 1 of OCA2 Explains Most Human Eye-Color Variation." Am J Hum Genet **80**.
Eastwood, S. L. (2005). "Does the CAPON gene confer susceptibility to schizophrenia?" PLoS Med **2**(10): e348.
Ekelund, J., W. Hennah, et al. (2004). "Replication of 1q42 linkage in Finnish schizophrenia pedigrees." Mol Psychiatry **9**(11): 1037-41.
Ekelund, J., D. Lichtermann, et al. (2000). "Genome-wide scan for schizophrenia in the Finnish population: evidence for a locus on chromosome 7q22." Hum Mol Genet **9**(7): 1049-57.
Ekelund, J., D. Lichtermann, et al. (1997). "A genome scan for schizophrenia loci in affected sib-pairs collected in Finland." American Journal of Medical Genetics **74**: 559.
Elston, R. C. and M. A. Campbell (1970). "Schizophrenia: evidence for the major gene hypothesis." Behav Genet **1**(1): 3-10.

Eve C. Johnstone, C.P.L. Freeman, et al. (1998). Companion To Psychiatric Studies Sixth Edition, Churchill Livingstone.

Falconer, D. S. (1965). Maternal effects and selection response. Genetics today: proceedings of the XI international congress of genetics. S. J. Greet. Oxford, Pergamon. **3:** 763-764.

Farmer, A. E., P. McGuffin, et al. (1984). "Searching for the split in schizophrenia: a twin study perspective." Psychiatry Res **13**(2): 109-18.

Frankle, W. G., J. Lerma, et al. (2003). "The synaptic hypothesis of schizophrenia." Neuron **39**(2): 205-16.

Freking, B. A., S. K. Murphy, et al. (2002). "Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals." Genome Res **12**(10): 1496-506.

Fried, M. and D. M. Crothers (1981). "Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis." Nucleic Acids Res **9**(23): 6505-25.

Gama-Carvalho, M., M. P. Carvalho, et al. (2001). "Nucleocytoplasmic shuttling of heterodimeric splicing factor U2AF." J Biol Chem **276**(16): 13104-12.

Gerber, D. J., D. Hall, et al. (2003). "Evidence for association of schizophrenia with genetic variation in the 8p21.3 gene, PPP3CC, encoding the calcineurin gamma subunit." Proc Natl Acad Sci U S A **100**(15): 8993-8.

Giampietro, C., F. Luzzati, et al. (2005). "Stathmin expression modulates migratory properties of GN-11 neurons in vitro." Endocrinology **146**(4): 1825-34.

Gomes, I., A. Collins, et al. (1999). "Hardy-Weinberg quality control." Ann Hum Genet **63**(Pt 6): 535-8.

Goring, H. H. and J. D. Terwilliger (2000). "Linkage analysis in the presence of errors IV: joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified." Am J Hum Genet **66**: 1310-1327.

Gottesman, II and J. Shields (1967). "A polygenic theory of schizophrenia." Proc Natl Acad Sci U S A **58**(1): 199-205.

Gottesman, II and J. Shields (1976). "A critical review of recent adoption, twin, and family studies of schizophrenia: behavioral genetics perspectives." Schizophr Bull **2**(3): 360-401.

Gottesman, I. I. (1991). Schiziophrenia Genesis. New York, New York: W.H.Freeman.

Green, E. K., R. Raybould, et al. (2005). "Operation of the schizophrenia susceptibility gene, neuregulin 1, across traditional diagnostic boundaries to increase risk for bipolar disorder." Arch Gen Psychiatry **62**(6): 642-8.

Guo, S., W. Tang, et al. (2006). "RGS4 polymorphisms and risk of schizophrenia: an association study in Han Chinese plus meta-analysis." Neurosci Lett **406**(1-2): 122-7.

Gurling, H. M., H. Critchley, et al. (2006). "Genetic association and brain morphology studies and the chromosome 8p22 pericentriolar material 1

(PCM1) gene in susceptibility to schizophrenia." Arch Gen Psychiatry **63**(8): 844-54.
Gurling, H. M., G. Kalsi, et al. (2001). "Genomewide genetic linkage analysis confirms the presence of susceptibility loci for schizophrenia, on chromosomes 1q32.2, 5q33.2, and 8p21-22 and provides support for linkage to schizophrenia, on chromosomes 11q23.3-24 and 20q12.1-11.23." Am J Hum Genet **68**(3): 661-73.
Gusella, J. F., N. S. Wexler, et al. (1983). "A polymorphic DNA marker genetically linked to Huntington's disease." Nature **306**(5940): 234-8.
Hacia, J. G., L. C. Brody, et al. (1996). "Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis." Nat Genet **14**(4): 441-7.
HapMap (2003). "The International HapMap Project." Nature **426**(6968): 789-96.
Hardy, G. H. (1908). "Mendelian proportions in a mixed population." Science **18**: 49-50.
Harrison, P. J. and D. R. Weinberger (2005). "Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence." Mol Psychiatry **10**(1): 40-68; image 5.
Hashimoto, R., T. Suzuki, et al. (2005). "Association study of the frizzled-3 (FZD3) gene with schizophrenia and mood disorders." J Neural Transm **112**(2): 303-7.
Hattori, E., C. Liu, et al. (2003). "Polymorphisms at the G72/G30 gene locus, on 13q33, are associated with bipolar disorder in two independent pedigree series." Am J Hum Genet **72**(5): 1131-40.
Hedrick, P. W. (1987). "Gametic disequilibrium measures: proceed with caution." Genetics **117**(2): 331-41.
Ho, Y., F. Elefant, et al. (2006). "Locus control region transcription plays an active role in long-range gene activation." Mol Cell **23**(3): 365-75.
Hovatta, I., T. Varilo, et al. (1999). "A genomewide screen for schizophrenia genes in an isolated Finnish subpopulation suggesting multiple susceptibility loci." Am J Hum Genet **65**: 1114-1124.
Hwu, H. G., C. M. Liu, et al. (2003). "Linkage of schizophrenia with chromosome 1q loci in Taiwanese families." Mol Psychiatry **8**(4): 445-52.
Iancu-Rubin, C. and G. F. Atweh (2005). "p27(Kip1) and stathmin share the stage for the first time." Trends Cell Biol **15**(7): 346-8.
Ikeda, M., N. Iwata, et al. (2005). "Association analysis of chromosome 5 GABAA receptor cluster in Japanese schizophrenia patients." Biol Psychiatry **58**(6): 440-5.
Ioannidis, J. P., E. E. Ntzani, et al. (2001). "Replication validity of genetic association studies." Nat Genet **29**(3): 306-9.
Ishiguro, H., Y. Horiuchi, et al. (2006). "RGS4 is not a susceptibility gene for schizophrenia in Japanese: Association study in a large case-control population." Schizophr Res.
Jakob, H. and H. Beckmann (1986). "Prenatal developmental disturbances in the limbic allocortex in schizophrenics." J Neural Transm **65**(3-4): 303-26.

Jin, K., X. O. Mao, et al. (2004). "Proteomic and immunochemical characterization of a role for stathmin in adult neurogenesis." Faseb J **18**(2): 287-99.
Johnstone, E. C., T. J. Crow, et al. (1976). "Cerebral ventricular size and cognitive impairment in chronic schizophrenia." Lancet **2**(7992): 924-6.
Johnstone, E. C., J. F. Macmillan, et al. (1987). "The occurrence of organic disease of possible or probable aetiological significance in a population of 268 cases of first episode schizophrenia." Psychol Med **17**(2): 371-9.
Jorde, L. B. (1995). "Linkage disequilibrium as a gene-mapping tool." Am J Hum Genet **56**(1): 11-4.
Kan, Y. W., Dozy, A.M. (1978). "Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation." Proc. Natl. Acad. Sci. **75(11)**: 5631-5.
Kety, S. S. (1983). "Mental illness in the biological and adoptive relatives of schizophrenic adoptees: findings relevant to genetic and environmental factors in etiology." Am J Psychiatry **140**(6): 720-7.
Kidd, K. K., Cavalli-Sforza, L.L. (1973). "An Analysis of the genetics of schizophrenia." Soc. Biol. **20**: 254-265.
Kimura-Yoshida, C., K. Kitajima, et al. (2004). "Characterization of the pufferfish Otx2 cis-regulators reveals evolutionarily conserved genetic mechanisms for vertebrate head specification." Development **131**(1): 57-71.
Kinoshita, Y., T. Suzuki, et al. (2005). "No association with the calcineurin A gamma subunit gene (PPP3CC) haplotype to Japanese schizophrenia." J Neural Transm **112**(9): 1255-62.
Kirov, G., D. Ivanov, et al. (2004). "Strong evidence for association between the dystrobrevin binding protein 1 gene (DTNBP1) and schizophrenia in 488 parent-offspring trios from Bulgaria." Biol Psychiatry **55**(10): 971-5.
Kleinjan, D. A., A. Seawright, et al. (2001). "Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6." Hum Mol Genet **10**(19): 2049-59.
Kleinjan, D. A. and V. van Heyningen (2005). "Long-range control of gene expression: emerging mechanisms and disruption in disease." Am J Hum Genet **76**(1): 8-32.
Kosower, N. S., L. Gerad, et al. (1995). "Constitutive heterochromatin of chromosome 1 and Duffy blood group alleles in schizophrenia." Am J Med Genet **60**(2): 133-8.
Kovelman, J. A. and A. B. Scheibel (1984). "A neurohistological correlate of schizophrenia." Biol Psychiatry **19**(12): 1601-21.
Krazeisen, A., R. Breitling, et al. (1999). "Determination of cDNA, gene structure and chromosomal localization of the novel human 17beta-hydroxysteroid dehydrogenase type 7(1)." FEBS Lett **460**(2): 373-9.
Lalouel, J. M., D. C. Rao, et al. (1983). "A unified model for complex segregation analysis." Am J Hum Genet **35**(5): 816-26.
Lander, E. S. and N. J. Schork (1994). "Genetic dissection of complex traits." Science **265**(5181): 2037-48.

Law, A. J., B. K. Lipska, et al. (2006). "Neuregulin 1 transcripts are differentially expressed in schizophrenia and regulated by 5' SNPs associated with the disease." Proc Natl Acad Sci U S A **103**(17): 6747-52.

Lehman, A. F. and J. A. Lieberman (1997). "Practice guideline for the treatment of patients with schizophrenia. American Psychiatric Association." Am J Psychiatry **154**(4 Suppl): 1-63.

Lettice, L. A., S. J. Heaney, et al. (2003). "A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly." Hum Mol Genet **12**(14): 1725-35.

Lettice, L. A., T. Horikoshi, et al. (2002). "Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly." Proc Natl Acad Sci U S A **99**(11): 7548-53.

Levine, M. and R. Tjian (2003). "Transcription regulation and animal diversity." Nature **424**(6945): 147-51.

Levinson, D. F., M. D. Levinson, et al. (2003). "Genome scan meta-analysis of schizophrenia and bipolar disorder, part I: Methods and power analysis." Am J Hum Genet **73**(1): 17-33.

Lewis, C. M., D. F. Levinson, et al. (2003). "Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia." Am J Hum Genet **73**(1): 34-48.

Lewontin, R. C. (1964)a. "The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models." Genetics **49**: 49-67.

Lewontin, R. C. (1964)b. "The Interaction of Selection and Linkage. Ii. Optimum Models." Genetics **50**: 757-82.

Lewontin, R. C. (1988). "On measures of gametic disequilibrium." Genetics **120**(3): 849-52.

Li, T., X. Ma, et al. (2004). "Evidence for association between novel polymorphisms in the PRODH gene and schizophrenia in a Chinese population." Am J Med Genet B Neuropsychiatr Genet **129**(1): 13-5.

Liddle, P. F., K. J. Friston, et al. (1992). "Patterns of cerebral blood flow in schizophrenia." Br J Psychiatry **160**: 179-86.

Liedtke, W., E. E. Leman, et al. (2002). "Stathmin-deficient mice develop an age-dependent axonopathy of the central and peripheral nervous systems." Am J Pathol **160**(2): 469-80.

Liou, Y. J., I. C. Lai, et al. (2006). "Genetic analysis of the human ENTH (Epsin 4) gene and schizophrenia." Schizophr Res **84**(2-3): 236-243.

Liquori, C. L., K. Ricker, et al. (2001). "Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9." Science **293**(5531): 864-7.

Litt, M. and J. A. Luty (1989). "A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene." Am J Hum Genet **44**(3): 397-401.

Lohmueller, K. E., C. L. Pearce, et al. (2003). "Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease." Nat Genet **33**(2): 177-82.

Mah, S., M. R. Nelson, et al. (2006). "Identification of the semaphorin receptor PLXNA2 as a candidate for susceptibility to schizophrenia." Mol Psychiatry.

Manceau, V., M. Swenson, et al. (2006). "Major phosphorylation of SF1 on adjacent Ser-Pro motifs enhances interaction with U2AF65." Febs J **273**(3): 577-87.
Matthysse, S., K. Lange, et al. (1979). "Continuous variation caused by genes with graduated effects." Proc Natl Acad Sci U S A **76**(6): 2862-5.
Matthysse, S. W. and K. K. Kidd (1976). "Estimating the genetic contribution to schizophrenia." Am J Psychiatry **133**(2): 185-91.
Maucuer, A., J. H. Camonis, et al. (1995). "Stathmin interaction with a putative kinase and coiled-coil-forming protein domains." Proc Natl Acad Sci U S A **92**(8): 3100-4.
Maucuer, A., J. P. Le Caer, et al. (2000). "Specific Ser-Pro phosphorylation by the RNA-recognition motif containing kinase KIS." Eur J Biochem **267**(14): 4456-64.
Maucuer, A., S. Ozon, et al. (1997). "KIS is a protein kinase with an RNA recognition motif." J Biol Chem **272**(37): 23151-6.
McGlashan, T. H. and R. E. Hoffman (2000). "Schizophrenia as a disorder of developmentally reduced synaptic connectivity." Arch Gen Psychiatry **57**(7): 637-48.
McGue, M. and Gottesman, II (1989a). "A single dominant gene still cannot account for the transmission of schizophrenia." Arch Gen Psychiatry **46**(5): 478-80.
McGue, M. and Gottesman, II (1989b). "Genetic linkage in schizophrenia: perspectives from genetic epidemiology." Schizophr Bull **15**(3): 453-64.
McGue, M., Gottesman, II, et al. (1985). "Resolving genetic models for the transmission of schizophrenia." Genet Epidemiol **2**(1): 99-110.
McGuffin, P., A. E. Farmer, et al. (1984). "Twin concordance for operationally defined schizophrenia. Confirmation of familiality and heritability." Arch Gen Psychiatry **41**(6): 541-5.
Millar, J. K., S. Christie, et al. (2001). "Genomic structure and localisation within a linkage hotspot of Disrupted In Schizophrenia 1, a gene disrupted by a translocation segregating with schizophrenia." Mol Psychiatry **6**(2): 173-8.
Millar, J. K., J. C. Wilson-Annan, et al. (2000). "Disruption of two novel genes by a translocation co-segregating with schizophrenia." Hum Mol Genet **9**: 1415-1423.
Miranda, A., J. Garcia, et al. (2006). "Putative association of the carboxy-terminal PDZ ligand of neuronal nitric oxide synthase gene (CAPON) with schizophrenia in a Colombian population." Schizophr Res **82**(2-3): 283-5.
Mirnics, K., F. A. Middleton, et al. (2001). "Disease-specific changes in regulator of G-protein signaling 4 (RGS4) expression in schizophrenia." Mol Psychiatry **6**(3): 293-301.
Morra, M., J. Lu, et al. (2001). "Structural basis for the interaction of the free SH2 domain EAT-2 with SLAM receptors in hematopoietic cells." Embo J **20**(21): 5840-52.
Morris, D. W., A. Rodgers, et al. (2004). "Confirming RGS4 as a susceptibility gene for schizophrenia." Am J Med Genet **125B**(1): 50-3.
Morton, N. E. (1955). "Sequential tests for the detection of linkage." Am J Hum Genet **7**(3): 277-318.

Munafo, M. R., L. Bowes, et al. (2005). "Lack of association of the COMT (Val158/108 Met) gene and schizophrenia: a meta-analysis of case-control studies." Mol Psychiatry **10**(8): 765-70.

Murray, R. M., S. W. Lewis, et al. (1985). "Towards an aetiological classification of schizophrenia." Lancet **1**(8436): 1023-6.

Myakishev, M. V., Y. Khripin, et al. (2001). "High-throughput SNP genotyping by allele-specific PCR with universal energy-transfer-labeled primers." Genome Res **11**(1): 163-9.

Nakamura, Y., M. Leppert, et al. (1987). "Variable number of tandem repeat (VNTR) markers for human gene mapping." Science **235**(4796): 1616-22.

Numakawa, T., Y. Yagasaki, et al. (2004). "Evidence of novel neuronal functions of dysbindin, a susceptibility gene for schizophrenia." Hum Mol Genet **13**(21): 2699-708.

Okasha, A. (1999). "Mental health in the Middle East: an Egyptian perspective." Clin Psychol Rev **19**(8): 917-33.

Ott, J. (1991). Analysis of Human Genetic Linkage. Baltimore, John Hopkins University Press.

Petryshen, T., A. Kirby, et al. (2003). "LocusView2.0 Copyright (c) 2003, 2004 - The Broad Institute, unpublished software."

Petryshen, T. L., F. A. Middleton, et al. (2005). "Genetic investigation of chromosome 5q GABAA receptor subunit genes in schizophrenia." Mol Psychiatry **10**(12): 1074-88, 1057.

Pfeifer, D., R. Kist, et al. (1999). "Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to SOX9: evidence for an extended control region." Am J Hum Genet **65**(1): 111-24.

Pierri, J. N., C. L. Volk, et al. (2001). "Decreased somal size of deep layer 3 pyramidal neurons in the prefrontal cortex of subjects with schizophrenia." Arch Gen Psychiatry **58**(5): 466-73.

Pimm, J., A. McQuillin, et al. (2005). "The Epsin 4 gene on chromosome 5q, which encodes the clathrin-associated protein enthoprotin, is involved in the genetic susceptibility to schizophrenia." Am J Hum Genet **76**(5): 902-7.

Pritchard, J. K. and M. Przeworski (2001). "Linkage disequilibrium in humans: models and data." Am J Hum Genet **69**(1): 1-14.

Purcell, S., S. S. Cherny, et al. (2003). "Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits." Bioinformatics **19**(1): 149-50.

Puri, V., A. McQuillin, et al. (2008). "Confirmation of the genetic association between the U2AF homology motif (UHM) kinase 1 (UHMK1) gene and schizophrenia on chromosome 1q23.3." Eur J Hum Genet 16Apr.

Puri, V., A. McQuillin, et al. (2007). "Fine Mapping by Genetic Association Implicates the Chromosome 1q23.3 Gene UHMK1, Encoding a Serine/Threonine Protein Kinase, as a Novel Schizophrenia Susceptibility Gene." Biol Psychiatry **61**(7): 873-879.

Puri, V., A. McQuillin, et al. (2006). "Failure to Confirm Allelic Association Between Markers at the CAPON Gene Locus and Schizophrenia in a British Sample." Biol Psychiatry **59**: 195-197.

Rao, D. C., N. E. Morton, et al. (1981). "Path analysis of qualitative data on pairs of relatives: application to schizophrenia." Hum Hered **31**(6): 325-33.
Raybould, R., E. K. Green, et al. (2005). "Bipolar disorder and polymorphisms in the dysbindin gene (DTNBP1)." Biol Psychiatry **57**(7): 696-701.
Reich, D. E. and E. S. Lander (2001). "On the allelic spectrum of human disease." Trends Genet **17**(9): 502-10.
Reif, A., S. Fritzen, et al. (2006). "Neural stem cell proliferation is decreased in schizophrenia, but not in depression." Mol Psychiatry **11**(5): 514-22.
Risch, N. (1990a). "Genetic linkage and complex diseases, with special reference to psychiatric disorders." Genet Epidemiol **7**(1): 3-16; discussion 17-45.
Risch, N. (1990b). "Linkage strategies for genetically complex traits. I. Multilocus models." Am J Hum Genet **46**(2): 222-8.
Risch, N. (1990c). "Linkage strategies for genetically complex traits. II. The power of affected relative pairs." Am J Hum Genet **46**(2): 229-41.
Risch, N. (1990d). "Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs." Am J Hum Genet **46**(2): 242-53.
Risch, N. and M. Baron (1984). "Segregation analysis of schizophrenia and related disorders." Am J Hum Genet **36**(5): 1039-59.
Risch, N. and K. Merikangas (1996). "The future of genetic studies of complex human diseases." Science **273**(5281): 1516-7.
Rizig, M. A. (2009). "Deciphering the cellular mechanism of antipsychotics' action, a
microarray gene expression study and pathways analysis of
antipsychotic drugs Clozapine and haloperidol." In Press.
Rizig, M. A., A. McQuillin, et al. (2006). "Failure to confirm genetic association between schizophrenia and markers on chromosome 1q23.3 in the region of the gene encoding the Regulator of G-protein Signalling 4 protein (RGS4)." Am J Med Genet B Neuropsychiatr Genet **141**(3): 296-300.
Robert Cancro, M. D. M. D. S. (2005). Comprehensive Textbook of Psychiatry.
Rosa, A., L. Fananas, et al. (2002). "1q21-q22 locus is associated with susceptibility to the reality-distortion syndrome of schizophrenia spectrum disorders." Am J Med Genet **114**(5): 516-8.
Rozen, S. and H. Skaletsky (2000). "Primer3 on the WWW for general users and for biologist programmers." Methods Mol Biol **132**: 365-86.
Saito, Y. A., N. J. Talley, et al. (2006). "Case-control genetic association studies in gastrointestinal disease: review and recommendations." Am J Gastroenterol **101**(6): 1379-89.
Saugstad, J. A., M. J. Marino, et al. (1998). "RGS4 inhibits signaling by group I metabotropic glutamate receptors." J Neurosci **18**(3): 905-13.
Schumacher, J., R. A. Jamra, et al. (2004). "Examination of G72 and D-amino-acid oxidase as genetic risk factors for schizophrenia and bipolar affective disorder." Mol Psychiatry **9**(2): 203-7.
Schwab, S. G., G. N. Eckstein, et al. (1997). "Evidence suggestive of a locus on chromosome 5q31 contributing to susceptibility for schizophrenia in German

and Israeli families by multipoint affected sib-pair linkage analysis." Mol Psychiatry **2**(2): 156-60.
Sham, P. C. and D. Curtis (1995). "An extended transmission/disequilibrium test (TDT) for multi-allele marker loci." Ann Hum Genet **59 ( Pt 3)**: 323-36.
Sham, P. C. and D. Curtis (1995). "Monte Carlo tests for associations between disease and alleles at highly polymorphic loci." Ann Hum Genet **59**(Pt 1): 97-105.
Shaw, S. H., M. Kelly, et al. (1998). "A genome-wide search for schizophrenia susceptibility genes." Am J Med Genet **81**(5): 364-76.
Shenton, M. E., C. C. Dickey, et al. (2001). "A review of MRI findings in schizophrenia." Schizophr Res **49**(1-2): 1-52.
Shenton, M. E., R. Kikinis, et al. (1992). "Abnormalities of the left temporal lobe and thought disorder in schizophrenia. A quantitative magnetic resonance imaging study." N Engl J Med **327**(9): 604-12.
Shifman, S., M. Bronstein, et al. (2002). "A highly significant association between a COMT haplotype and schizophrenia." Am J Hum Genet **71**(6): 1296-302.
Shumyatsky, G. P., G. Malleret, et al. (2005). "stathmin, a gene enriched in the amygdala, controls both learned and innate fear." Cell **123**(4): 697-709.
Slater, E., Cowie, V. (1971). The Genetics of Mental Disorders. London, Oxford University Press,.
Sobell, J. L., C. Richard, et al. (2005). "Failure to confirm association between RGS4 haplotypes and schizophrenia in Caucasians." Am J Med Genet B Neuropsychiatr Genet **139**(1): 23-7.
Spitzer, R. L. and J. Endicott (1977). The Schedule for Affective Disorders and Schizophrenia, Lifetime Version, 3rd edition. New York, New York State Psychiatric Institute.
St Clair, D., D. Blackwood, et al. (1990). "Association within a family of a balanced autosomal translocation with major mental illness." Lancet **336**: 13-16.
Stefansson, H., E. Sigurdsson, et al. (2002). "Neuregulin 1 and susceptibility to schizophrenia." Am J Hum Genet **71**(4): 877-92.
Straub, R. E., Y. Jiang, et al. (2002). "Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia." Am J Hum Genet **71**(2): 337-48.
Straub, R. E., C. J. MacLean, et al. (1997). "Support for a possible schizophrenia vulnerability locus in region 5q22-31 in Irish families." Mol Psychiatry **2**(2): 148-55.
Sullivan, P. F., K. S. Kendler, et al. (2003). "Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies." Arch Gen Psychiatry **60**(12): 1187-92.
T Bedirhan Ustun, Jurgen Rehm, et al. (1999). "Multiple-informant ranking of the disabling effects of different helth conditions in 14 countries." The Lancet **354**.
Talkowski, M. E., H. Seltman, et al. (2006). "Evaluation of a Susceptibility Gene for Schizophrenia: Genotype Based Meta-Analysis of RGS4 Polymorphisms from Thirteen Independent Samples." Biol Psychiatry.

Tang, R. Q., X. Z. Zhao, et al. (2006). "Family-based association study of Epsin 4 and Schizophrenia." Mol Psychiatry **11**(4): 395-9.
Taymans, J. M., H. K. Kia, et al. (2004). "Dopamine receptor-mediated regulation of RGS2 and RGS4 mRNA differentially depends on ascending dopamine projections and time." Eur J Neurosci **19**(8): 2249-60.
Taymans, J. M., J. E. Leysen, et al. (2003). "Striatal gene expression of RGS2 and RGS4 is specifically mediated by dopamine D1 and D2 receptors: clues for RGS2 and RGS4 functions." J Neurochem **84**(5): 1118-27.
Thompson, P. M., C. Vidal, et al. (2001). "Mapping adolescent brain change reveals dynamic wave of accelerated gray matter loss in very early-onset schizophrenia." Proc Natl Acad Sci U S A **98**(20): 11650-5.
Tsuang, M. T., K. D. Bucher, et al. (1983). "A search for 'schizophrenia spectrum disorders'. An application of a multiple threshold model to blind family study data." Br J Psychiatry **143**: 572-7.
Vieland, V. J. and J. Huang (2003). "Two-locus heterogeneity cannot be distinguished from two-locus epistasis on the basis of affected-sib-pair data." Am J Hum Genet **73**(2): 223-32.
Wang, D. G., J. B. Fan, et al. (1998). "Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome." Science **280**(5366): 1077-82.
Weber, J. L. and P. E. May (1989). "Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction." Am J Hum Genet **44**(3): 388-96.
Wei, J. and G. P. Hemmings (2004). "Lack of a genetic association between the frizzled-3 gene and schizophrenia in a British population." Neurosci Lett **366**(3): 336-8.
Weinberg, W. (1908). "On the demonstration of heredity in man."
Weinberger, D. R. (1995). "From neuropathology to neurodevelopment." Lancet **346**(8974): 552-7.
Wender, P. H., Rosenthal, D., Kety, S.S., Schulsinger, F., Welner, J. (1974). "Crossfostering. A research strategy for clarifying the role of genetic and experiential factors in the etiology of schizophrenia." Arch Gen Psychiatry **30**: 121-128.
WHO (1992). The ICD-10 classification of mental and behavioural disorders: clinical descriptions and dignostic guidelines. Geneva, World Health Organisation.
Williams, N. M., A. Preece, et al. (2004). "Identification in 2 independent samples of a novel schizophrenia risk haplotype of the dystrobrevin binding protein gene (DTNBP1)." Arch Gen Psychiatry **61**(4): 336-44.
Williams, N. M., A. Preece, et al. (2004). "Support for RGS4 as a susceptibility gene for schizophrenia." Biol Psychiatry **55**(2): 192-5.
Winokur, G., M. T. Tsuang, et al. (1982). "The Iowa 500: affective disorder in relatives of manic and depressed patients." Am J Psychiatry **139**(2): 209-12.
Wright, A., B. Charlesworth, et al. (2003). "A polygenic basis for late-onset disease." Trends Genet **19**(2): 97-106.

Wright, I. C., S. Rabe-Hesketh, et al. (2000). "Meta-analysis of regional brain volumes in schizophrenia." Am J Psychiatry **157**(1): 16-25.
Xiong, Z. and P. W. Laird (1997). "COBRA: a sensitive and quantitative DNA methylation assay." Nucleic Acids Res **25**(12): 2532-4.
Xu, B., N. Wratten, et al. (2005). "Increased expression in dorsolateral prefrontal cortex of CAPON in schizophrenia and bipolar disorder." PLoS Med **2**(10): e263.
Yang, J., T. Si, et al. (2003). "Association study of the human FZD3 locus with schizophrenia." Biol Psychiatry **54**(11): 1298-301.
Zaidel, D. W., M. M. Esiri, et al. (1997). "Size, shape, and orientation of neurons in the left and right hippocampus: investigation of normal asymmetries and alterations in schizophrenia." Am J Psychiatry **154**(6): 812-8.
Zhao, J. H., D. Curtis, et al. (2000). "Model-free analysis and permutation tests for allelic associations." Hum Hered **50**(2): 133-9.
Zhao, J. H., S. Lissarrague, et al. (2002). "GENECOUNTING: haplotype analysis with missing genotypes." BIOINFORMATICS **18**: 1694-1695.
Zheng, Y., H. Li, et al. (2005). "Association of the carboxyl-terminal PDZ ligand of neuronal nitric oxide synthase gene with schizophrenia in the Chinese Han population." Biochem Biophys Res Commun **328**(4): 809-15.

# 13.0 APPENDIX

## 13.1 RAW RESULTS OF LD BETWEEN ALL 55 MARKERS USED.

**Figure 13:1 LD plot of all 55 markers used in the UCL sample. Above is Cramer's V. Below are P-values.**

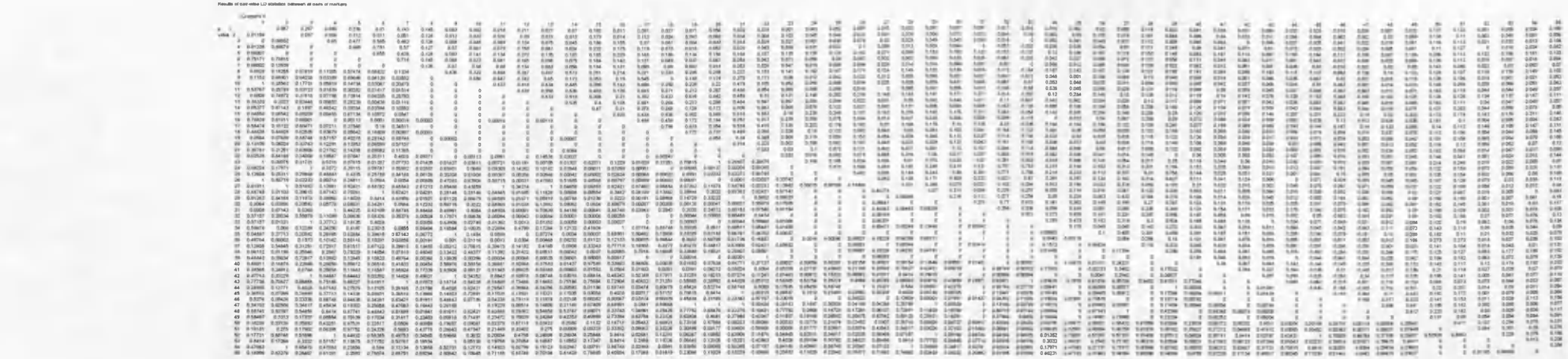
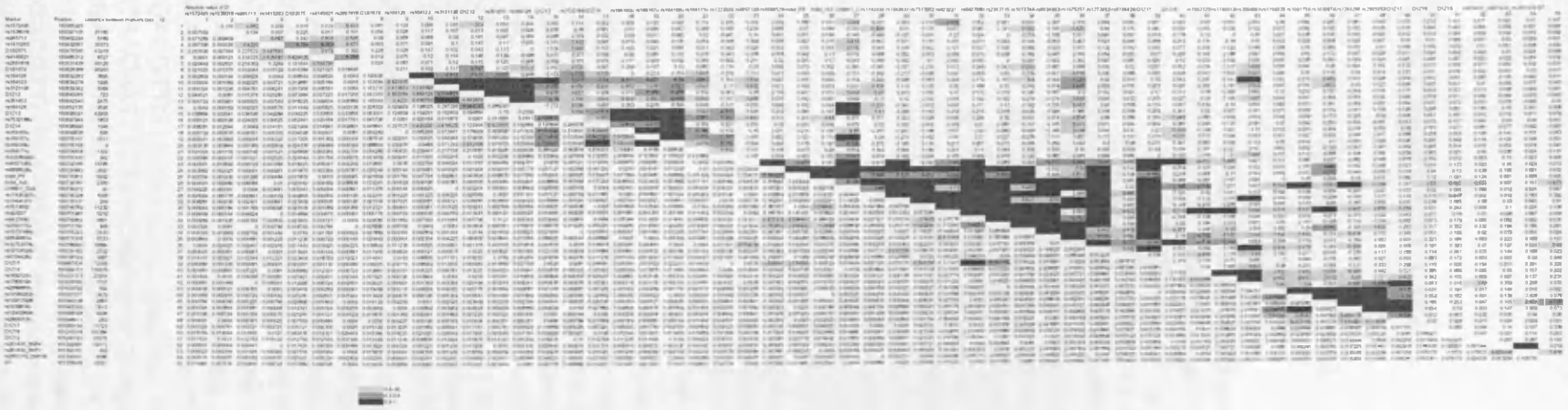


**Figure 13:2 Lod Score Values**

**Figure 13:3 LD Plot of all 55 markers. Above is D' values, below is r2. Distances between markers is on the left.**



## 13.2 PRIMER SEQUENCES OF UHMK1 SEQUENCING

**Table 13:1 Primer sequences used to sequence UHMK1 as described in section 8.1.2**

| Primer Name | Location | Sequence |
|---|---|---|
| KIS_P9_L_m13r | 5' UTR | GGATAACAATTTCACACAGGcaacaaagcaggggtgatg |
| KIS_P9_R_m13F | 5' UTR | CACGACGTTGTAAAACGACgggggaaatttgcatctgta |
| KIS_P8_L_m13r | 5' UTR | GGATAACAATTTCACACAGGtctcatttaatatcttcagggttgg |
| KIS_P8_R_m13F | 5' UTR | CACGACGTTGTAAAACGACtacaggcatccaccaccac |
| KIS_P7_L_m13F | 5' UTR | CACGACGTTGTAAAACGACtgggtgaaattccatctcaaa |
| KIS_P7_R_m13r | 5' UTR | GGATAACAATTTCACACAGGcaccatgcccggttaaat |
| KIS_P6_L_m13F | 5' UTR | CACGACGTTGTAAAACGACtggggttaaacaaaaccccatc |
| KIS_P6_R_m13r | 5' UTR | GGATAACAATTTCACACAGGagcctaagctgttgctctgc |
| P5L_r_M13R | 5' UTR | GGATAACAATTTCACACAGGggacctcacgcaagaaagaa |
| P5R_r_M13F | 5' UTR | CACGACGTTGTAAAACGACttcaaaccagagcaactcca |
| P4aL_r_M13F | 5' UTR | CACGACGTTGTAAAACGACcatggtggcaaattcctgta |
| P4aR_r_M13R | 5' UTR | GGATAACAATTTCACACAGGggaaagtttgaggagaggttga |
| P4bL_r_M13F | 5' UTR | CACGACGTTGTAAAACGACctggaaatgcagcccagt |
| P4bR_r_M13R | 5' UTR | GGATAACAATTTCACACAGGggcatgatctcggcatct |
| P3L_r_M13R | 5' UTR | GGATAACAATTTCACACAGGgcaagagagaagacaggtcaga |
| P3R_r_M13F | 5' UTR | CACGACGTTGTAAAACGACtttctgggccaaatcaatgt |
| KIS_P2_L_m13R | 5' UTR | GGATAACAATTTCACACAGGtgagccaactatgagtgacca |
| KIS_P2_R_m13F | 5' UTR | CACGACGTTGTAAAACGAC cccatgattctggaagccta |
| Ex1a0_L_M13R | CpG | GGATAACAATTTCACACAGGgctgtggcaggtacagagc |
| Ex1a0_R_M13F | CpG | CACGACGTTGTAAAACGACactgtgacctgggcaagaag |
| KIS_EX1a3_Lm13R | CpG | GGATAACAATTTCACACAGGaaggaccagttttggcttca |
| KIS_EX1a3_Rm13F | CpG | CACGACGTTGTAAAACGACgttaagggcacggacacg |
| KIS_EX1a2_Rm13F | CpG | CACGACGTTGTAAAACGACcagcgaacccgatacacc |
| KIS_EX1a2_Lm13R | CpG | GGATAACAATTTCACACAGGtgtcatggctgcttgaagtc |
| KIS_EX1a1_Lm13R | Exon 1 | GGATAACAATTTCACACAGGcgtgcccttaacccacac |
| KIS_EX1a1_Rm13F | Exon 1 | CACGACGTTGTAAAACGACctttgcggaaaccatactcg |
| UHMK_EX2_L_M13R | Exon 2 | GGATAACAATTTCACACAGGcagtggaagcttgctcatca |
| UHMK_EX2_R_M13F | Exon 2 | CACGACGTTGTAAAACGACcggtggcactgaaattcttt |
| UHMK_EX3_L_M13R | Exon 3 | GGATAACAATTTCACACAGGtgcccagttaatgaaccaaa |
| UHMK_EX3_R_M13F | Exon 3 | CACGACGTTGTAAAACGACtgaaagtaaggaccccaaagc |
| KIST_EX4r_L_M13R | Exon 4 | GGATAACAATTTCACACAGGtgttttctgttgcattttactctca |
| KIST_EX4r_R_M13R | Exon 4 | CACGACGTTGTAAAACGACgcaaaagagcattcctttgaa |
| KIS_EX5_M13R | Exon 5 | GGATAACAATTTCACACAGGcaacaatcatcccaccgata |
| KIS_EX5_M13F | Exon 5 | CACGACGTTGTAAAACGACagtgggaagcatgaccagat |
| KIS_EX6_M13R | Exon 6 | GGATAACAATTTCACACAGGgcttcatgatgatccaagca |
| KIS_EX6_M13F | Exon 6 | CACGACGTTGTAAAACGACtgaggcttgaacccagga |
| UHMK_EX7_L_M13R | Exon 7 | GGATAACAATTTCACACAGGttcaaggagagtagaaggtggaa |
| UHMK_EX7_R_M13F | Exon 7 | CACGACGTTGTAAAACGACtgcaaaactcataatccttagca |
| UHMK_EX8A_L_M13F | Exon 8 | CACGACGTTGTAAAACGACttggcatcacctggaagttt |
| UHMK_EX8A_R_M13R | Exon 8 | GGATAACAATTTCACACAGGccaaatgtatgcaacgcagt |
| UHMK_EX8B_L_M13F | Exon 8 | CACGACGTTGTAAAACGACgcaggactacccccttacca |

Continued.

| | | |
|---|---|---|
| UHMK_EX8B_R_M13R | Exon 8 | GGATAACAATTTCACACAGGgtgcaagatcaagcatcagc |
| UHMK_EX8C_L_M13R | Exon 8 | GGATAACAATTTCACACAGGataagctggcactggatgct |
| UHMK_EX8C_R_M13F | Exon 8 | CACGACGTTGTAAAACGACggcacaacttagtctttgttcca |
| UHMK_EX8D_L_M13F | Exon 8 3'UTR | CACGACGTTGTAAAACGACgagaggcccagaacaaact |
| UHMK_EX8D_R_M13R | Exon 8 3'UTR | GGATAACAATTTCACACAGGtgggatgctgctcctctagt |
| UHMK_EX8E_L_M13R | Exon 8 3'UTR | GGATAACAATTTCACACAGGtatccctgctccctttttcc |
| UHMK_EX8E_R_M13F | Exon 8 3'UTR | CACGACGTTGTAAAACGACaaaggagagtcctgaatttgaca |

M13F/R tails are indicated in uppercase of the primers.

# Appendix 2

## 13.3 Published Papers

American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 141B:296–300 (2006)

# 13.3.1 Failure to Confirm Genetic Association Between Schizophrenia and Markers on Chromosome 1q23.3 in the Region of the Gene Encoding the Regulator of G-Protein Signaling 4 Protein (*RGS4*)

Mie A. Rizig,[1] Andrew McQuillin,[1] Vinay Puri,[1] Khalid Choudhury,[1] Susmita Datta,[1] Srinivasa Thirumalai,[2] Jacob Lawrence,[1] Digby Quested,[4] Jonathan Pimm,[1] Nicholas Bass,[1] Graham Lamb,[3] Helen Moorey,[3] Allison Badacsonyi,[5] Katie Kelly,[6] Jenny Morgan,[2] Bhaskar Punukollu,[4] Gomathinayagam Kandasami,[7] David Curtis,[5] and Hugh Gurling[1]*

## 13.3.2 Failure to Confirm Allelic Association Between Markers at the CAPON Gene Locus and Schizophrenia in a British Sample

Vinay Puri, Andrew McQuillin, Srinivasa Thirumalai, Jacob Lawrence, Robert Krasucki, Khalid Choudhury, Susmita Datta, Simon Kerwin, Digby Quested, Nicholas Bass, Jonathan Pimm, Graham Lamb, Helen Moorey, Gomathinayagam Kandasami, Allison Badacsonyi, Katie Kelly, Jenny Morgan, Bhaskar Punukollu, Haitham Nadeem, David Curtis, and Hugh M.D. Gurling

BIOL PSYCHIATRY 2006;59:195–197

## 13.3.3 Fine Mapping by Genetic Association Implicates the Chromosome 1q23.3 Gene UHMK1, Encoding a Serine/Threonine Protein Kinase, as a Novel Schizophrenia Susceptibility Gene.

Vinay Puri, Andrew McQuillin, Khalid Choudhury, Susmita Datta, Jonathan Pimm, Srinivasa Thirumalai, Robert Krasucki, Jacob Lawrence, Digby Quested, Nicholas Bass, Helen Moorey, Jenny Morgan, Bhaskar Punukollu, Gomathinayagam Kandasami, David Curtis, and Hugh Gurling

BIOL PSYCHIATRY 2007;61:873–879

## 13.3.4 Confirmation of Genetic association between the U2AF homology motif (UHM) Kinase 1 (UHMK1) gene and schizophrenia on chromosome 1q23.3.

European Journal of Human Genetics (2008), 1–8

www.nature.com/ejhg

ARTICLE

Vinay Puri[1], Andrew McQuillin[1], Susmita Datta[1], Khalid Choudhury[1], Jonathan Pimm[1], Srinivasa Thirumalai[2], Robert Krasucki[1], Jacob Lawrence[1], Digby Quested[3,9], Nicholas Bass[1], Caroline Crombie[4], Gillian Fraser[4], Nicholas Walker[5], Helen Moorey[6], Manaan Kar Ray[8], Akeem Sule[8], David Curtis[7], David St. Clair[4] and Hugh Gurling*[,1]