The Search for a Schizophrenia Susceptibility Locus on Chromosome 17

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy at Cardiff University

Liam Stuart Carroll

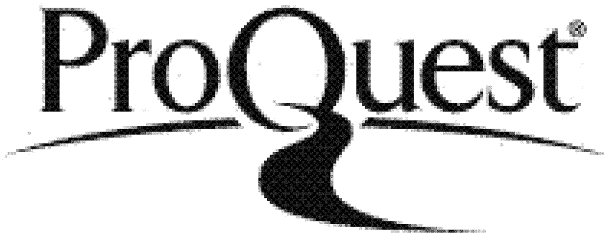
2007

Department of Psychological Medicine, School of Medicine, Cardiff University

Supervisors:

Professor Michael J. Owen

Dr. Nigel M. Williams



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

Summary

The search for genetic variants that alter the risk of developing schizophrenia has met with little success (Owen et al., 2005a); with the best example coming from a large multiply affected pedigree (Blackwood et al., 2001, Millar et al., 2000, St Clair et al., 1990). In this study a genome-wide significant schizophrenia linkage region in a single pedigree (Williams et al., 2003a) has been refined to an 11.7Mb region at 17q23-q24 where all 6 affected males share 21 consecutive microsatellite marker genotypes. Analysis of this region by oligonucleotide-array comparative genome hybridisation shows that no large deletions explain the linkage signal. High-density genotyping identified a region of homozygosity present in C702 affecteds that was not identified in 2709 individuals from the UK. This rare diplotype encompasses the 3' of the gene encoding Protein Kinase C Alpha (*PRKCA*), implicated in the pathogenesis of schizophrenia and related disorders (Mirnics et al., 2001, Hahn and Friedman, 1999, Birnbaum et al., 2004). Mutation screening of *PRKCA* in the linked pedigree C702 identified an exonic haplotype with a minor allele frequency of 0.003, which is homozygous in affected individuals. The haplotype is associated in a UK case-control sample with major mental illness (p=0.05. OR=1.8) due to risk in males (p=0.005. OR=3.9). Also, the pedigree C702 genotype was not observed in ~9000 Europeans. Association mapping of *PRKCA* using schizophrenia and psychosis case-control samples from Europe identified a common associated allele (rs3803821, meta-analysis p=0.02, OR=1.1) that shows significant overtransmission in a trios sample (p=0.03). Therefore, *PRKCA* may represent the locus causing the pedigree C702 linkage signal and contains genetic variation associated with schizophrenia and related disorders.

Declaration and Statements

DECLARATION

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ..... (candidate)

Date ..... 11.02.08 .....

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD

Signed ..... (candidate)

Date ..... 11.02.08 .....

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed ..... (candidate)

Date ..... 11.02.08 .....

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..... (candidate)

Date ..11.\02.\08.....

STATEMENT 4 - BAR ON ACCESS APPROVED

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans after expiry of a bar on access approved by the Graduate Development Committee.

Signed ..... (candidate)

Date .....

Acknowledgements

The completion of this thesis owes a great deal to the guidance and help I have received from many people and I will try to remember them all now.

An enormous amount of thanks is directed to Nigel, who has never failed to offer advice, criticism and praise and without whom my research and thesis would have suffered. A huge debt of gratitude also goes to Mike, who has been as interested in this project as myself and despite his overflowing diary has always both challenged and guided my research. There are also those who are just as busy but have provided me with help and suggestions whenever they have been asked for, and my thanks go to Mick and George.

When I began this research I had little lab experience, however I have been lucky enough to have colleagues who have gone out of their way to help me, and so special thanks go to Hywel and Nadine for this. My thanks also extends to those who have helped me throughout my time in the department; Tim, Nick B, Lyn, Lucy, Sarah, Amy, Angela, Didi, Dobril, Elaine, Darko, Dragana, Beate, Richard, Nick C, Ian, Becky, Jenny W, Jenny T, Anna and Elisa and anyone who I have forgotten.

To my family I owe so much, as without them I would never have been given the opportunity to get this far and who have always given me all the support and help they can; a special thanks goes to Mum and Dad, James, Rory, Siobhân and to Dom, Sue and Sarah. Thanks also to my friends, who make me realise that there is more to life than what I do. Finally, I would like to thank Joanne, who has been my constant companion and source of strength throughout this work.

Table of Contents

Chapter 1: General Introduction

1.1.1. Schizophrenia

Schizophrenia is a severe psychiatric disorder that cannot be clearly defined at the psychological or biological levels of scientific understanding (Owen, O'Donovan *et al.*, 2002). The disorder is at best inadequately treated by current therapies, where affected individuals are rarely completely repatriated to society and at worst it robs people of their personality and condemns them and others to lifelong suffering. Sufferers may experience a plethora of symptoms ranging from delusions and language problems to motor dysfunction. The phenotype is common and affects ~1% of the population worldwide, placing a huge monetary burden on society and disrupting millions of lives (Owen, O'Donovan *et al.*, 2002). Understanding schizophrenia at the level of genetics will eventually lead to the full understanding and treatment of this illness. The purpose of the following research is to identify and understand genetic variants that influence the expression of schizophrenia and related phenotypes.

1.1.2. The History of the Phenotype

To understand the term "schizophrenia" we must first acknowledge that the phenotype was largely classified before it was named. Work by Emil Kraepelin attempted to demarcate broad terminologies such as dementia and insanity, by clustering a group of symptoms commonly co-expressed into a disease which he termed dementia praecox

(Wing and Agrawal, 2003). He described symptoms that co-occurred and appeared as one disorder; delusions, hallucinations, thought disorder, blunted affect, depression and lack of insight. However, what is now seen as his major contribution is a distinction between a phasic form of mania and a regular, sometimes progressive and deteriorating disorder (Wing and Agrawal, 2003). Later, Eugen Bleuler coined the term schizophrenia to describe the disorder, separate from the manic-depressive phenotypes we know as the bipolar disorders. Although he also recognised that affective symptoms could appear in schizophrenia he believed that a unifying thought disruption was common to all schizophrenics (Wing and Agrawal, 2003). However, since the time of Kraepelin and Bleuler, there have been and still are those who question the absolute segregation of the disorders (Craddock and Owen, 2005).

1.1.3. Modern Classification of Schizophrenia

Despite much research into the aetiology, biological and behavioural aspects of schizophrenia, the disorder is still diagnosed based solely upon the symptoms presented and reported (Wing and Agrawal, 2003). There have been two independent attempts to catalogue the symptoms of schizophrenia and other mental disorders, to allow for the reliable and reproducible diagnosis of the syndrome. These are the ICD (International Statistical Classification of Diseases and Related Health Problems) and DSM (Diagnostic and Statistical Manual of Mental Disorders) classifications of psychiatric disorders, as defined by consortia of psychiatric health workers worldwide (Wing and Agrawal, 2003).

However, it must be noted, that between these independently produced criteria there are differences.

Schizophrenia as a disorder can be generally summarised into both positive and negative symptoms (Fuller, Schultz *et al.*, 2003, Wing and Agrawal, 2003). The positive symptoms consist of a thought disorder, as well as delusions and hallucinations. Negative symptoms include a flatness of affect, psychomotor retardation and social withdrawal. However, a diagnosis of schizophrenia does not require all of these behaviours to be apparent, and rather it is the combination of factors that warrants the identification of schizophrenia. The result of such a classification system is that individuals who are classified as schizophrenic may differ from each other phenotypically, but are more similar to each other than they are to those with other disorders or the common person. There are also those who do not meet the criteria for schizophrenia; but who have a certain complement of symptoms very similar to schizophrenia but not quite fitting the diagnosis, further supporting the view that schizophrenia represents a dimension of a continuum of related disorders (Craddock and Owen, 2005, Cutting, 2003, Fuller, Schultz *et al.*, 2003, Wing and Agrawal, 2003). Therefore, the phenotype of schizophrenia can be described as heterogeneous, and may represent part of a spectrum of psychiatric illnesses that may extend into other psychotic and mood disorders and into the general population (Craddock, O'Donovan *et al.*, 2005, Craddock and Owen, 2005).

1.2.1. The Genetic Epidemiology of Schizophrenia

The lifetime morbid risk of developing schizophrenia is remarkably uniform across populations, being ~1% worldwide (Jablensky, 2003, Owen, O'Donovan *et al.*, 2002) . A proportion of the risk of developing schizophrenia is attributable to environmental factors (Hyde and Lewis, 2003, Maki, Veijola *et al.*, 2005, McGrath and Murray, 2003) and the many suggested risk factors include gestation and delivery complications, season of birth, maternal viral infection during pregnancy, low IQ, urban upbringing, migrant parents, social class, certain personality traits, gender and possibly the use of cannabis and other illegal drugs (Hyde and Lewis, 2003, Maki, Veijola *et al.*, 2005, McGrath and Murray, 2003). Such risk factors, particularly those that occur before the child enters society, support the view that schizophrenia has an organic origin and is not purely a psychological manifestation.

There is now overwhelming evidence that the risk of developing schizophrenia is largely attributable to genetic factors. The risk of a relative of a schizophrenic developing the disorder has been shown to be proportional to the amount of genetic material shared with that individual (Jablensky, 2003, McGrath and Murray, 2003, Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003, Shih, Belmonte *et al.*, 2004). Epidemiological reviews of family and twin studies unequivocally confirm that the risk of developing schizophrenia is increased in relatives of affecteds, as compared to the general population (Jablensky, 2003, McGrath and Murray, 2003, Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003, Shih, Belmonte *et al.*, 2004). The worldwide population risk of acquiring schizophrenia is low (1%), as compared to first cousins (1.5-2%), siblings (3-

9%) and dizygotic and monozygotic twins (4-17% and 48-53% respectively). Stringently controlled family studies have concluded that the risk for schizophrenia in first degree relatives of affecteds is 5-15 times greater than in the general population (Owen, O'Donovan *et al.*, 2002). Other evidence from family studies includes the risk to offspring of parents who are both schizophrenic, being similar to that of monozygotic twins at around 46% (Owen, O'Donovan *et al.*, 2002).

Family based studies do not prove unequivocally that shared genes are responsible for the disorder, as families also share a very similar environment, and common environmental factors may influence the risk of developing schizophrenia (Jablensky, 2003, Maki, Veijola *et al.*, 2005, McGrath and Murray, 2003). Adoption and twin studies attempt to discern and demarcate the roles of gene and environment. Adoption based studies assess whether the risk of adopted children developing schizophrenia is related to the affection status of their biological parents and siblings and not to their non-biological relatives (Jablensky, 2003, Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003, Shih, Belmonte *et al.*, 2004). A large series of Danish adoption studies found that 8% of biological relatives of schizophrenic adoptees also had the disorder, compared to only 1% with non-schizophrenic adoptees (Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003, Shih, Belmonte *et al.*, 2004). A Finnish study also found increased rates of schizophrenia in adopted children that had biological mothers with schizophrenia, as compared to those adopted from unaffected mother (Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003, Shih, Belmonte *et al.*, 2004).

Although adoption studies strongly suggest that genetic factors are the primary vector of risk of developing schizophrenia, they do not exclude a role for environmental risk factors that may be at work pre- and peri-natally (Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003, Shih, Belmonte *et al.*, 2004). Analysis of disease concordance rates in dizygotic (DZ) and monozygotic (MZ) twins should however show if there is any increase risk to those twins who are genetically identical, controlling for the possible confounding factors of pregnancy and birth (although it should be noted that environmental influences still differ between DZ and MZ twins (Riley, Asherson *et al.*, 2003)). Several twin studies of populations from Europe and a study from Japan all conclude that disease concordance rates for MZ twins are 41-65% compared to 0-28% for DZ twins (Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003). Consequently, the heritability of schizophrenia, that is the proportion of variability in the phenotype that is due to genetic factors, has been estimated to be >80% (Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003). Therefore, the proportion of risk of developing schizophrenia due to genetic factors is extremely large, being comparable to type I diabetes (~80%) but much greater than for breast cancer (~30%) and type II diabetes (26%) (Kirov, O'Donovan *et al.*, 2005).

In conclusion, family, twin and adoption studies collectively display that schizophrenia is caused by variation between individuals primarily at the genetic level. However, the nature of this genetic component is unknown and a proportion of the heritability is likely to be due to gene-environment interactions.

The frequency of discordance of schizophrenia among MZ twins suggests that what is shared is a liability or predisposition towards presenting with schizophrenia (Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003). Corroborating evidence comes from the studies of the offspring of discordant MZ twins; both sets of offspring have the same genetic risk of developing schizophrenia, again indicating that carrying "schizophrenic" genetic material only increases risk of developing the disorder (Jablensky, 2003, Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003, Shih, Belmonte *et al.*, 2004). Furthermore, the decrement in risk as genetic distance increases among relatives is indicative of a multigenic model (Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003, Shih, Belmonte *et al.*, 2004). The likely explanation is that the risk of developing schizophrenia is dependent upon multiple variants possibly in multiple genes.

1.2.2. Genetic Models of Schizophrenia

Epidemiological studies of schizophrenia have shown a decrement in risk of developing schizophrenia as genetic distance increases among relatives (Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003, Shih, Belmonte *et al.*, 2004), which is inconsistent with a single major genetic locus causing the majority of instances of schizophrenia. Therefore, the mode of transmission model for schizophrenia is consistent with other common disorders, being non-Mendelian and complex and involving more than one locus (Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003). In an important series of papers, Risch (Risch, 1990b, Risch, 1990a) calculated that the

recurrence risk (the risk of a family member of an affected acquiring the same phenotype) among relatives of probands is incongruous with any single genetic locus conferring a relative risk (to what extent a particular risk factor influences an outcome, designated as λ) of >3; models of two or three loci each bestowing a λs of less than two were more probable. Therefore, the mode of transmission of schizophrenia will probably be oligogenic with loci of modest effect, or polygenic, where many loci have weak effects, or a combination of the two (Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003). Additionally, environmental factors must play a role in imparting risk and may not act independently of the genetic factors (Maki, Veijola *et al.*, 2005, McGrath and Murray, 2003, Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003). The number of genetic and environmental risk factors and the interaction between those risk factors remains unknown.

Schizophrenia is a human disease syndrome which may consist of multiple related phenotypes that are likely to be continuous in their distribution, however the disorder itself is categorised as either present or absent. Also, the aetiology of the disorder is likely to involve multiple genetic and environmental factors in any given individual. The resolution between schizophrenia being a clinically defined disorder and the recognition that schizophrenia may represent an extreme of a continuum whereby many factors combined to produce an outcome can be explained by the multifactorial liability-threshold (MFTM) model of complex disorders (Cardno and McGuffin, 2002).

The MFTM model (figure 1.1) proposes that the risk or liability of developing diseases such as schizophrenia follows a continuous distribution in a population (i.e. everyone is exposed to a certain amount of genetic and environmental risk). Individuals

below a certain threshold on this distribution are unaffected (or not brought to medical attention), while individuals who have achieved a certain degree of liability are diagnosed with schizophrenia. The threshold for schizophrenia is determined by the lifetime morbid risk of developing the disorder; therefore ~1% of the population will achieve the threshold for being diagnosed schizophrenic. The liability threshold has been shown to differ between populations for some disorders but not for schizophrenia (Jablensky, 2003, Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003, Shih, Belmonte *et al.*, 2004), although for example males may have a lower threshold of liability than females (Jablensky, 2003, Maki, Veijola *et al.*, 2005) implying that males and females have different or additional risk factors.

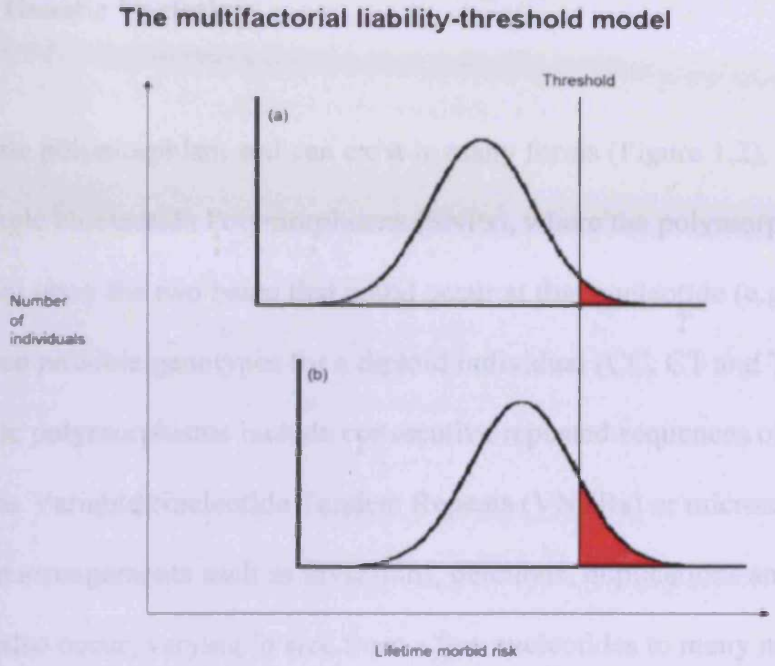


Figure 1.1: The Multifactorial-Liability Threshold Model of complex disorders, with schizophrenia as an example. The disorder has a continuous liability distribution in the population; that is multiple genetic and environmental factors contribute to risk of developing the disorder. In the case of schizophrenia in the general population (a), the lifetime morbid risk of developing schizophrenia is 1%, therefore in 1% of individuals (red) risk factors have combined to cause schizophrenia while in the remainder of unaffected individuals (white), there is no clinical diagnosis. Considering siblings of schizophrenics (b) these individuals have, on average, a lifetime morbid risk of ~10%, meaning their risk threshold is lower than for the general population. Figure adapted from (Cardno and McGuffin, 2002).

1.3.1. Human Genetic Variation

A genetic polymorphism and can exist in many forms (Figure 1.2), the simplest of these being Single Nucleotide Polymorphisms (SNPs), where the polymorphism has two forms dependent upon the two bases that could occur at that nucleotide (e.g. C and T), resulting in three possible genotypes for a diploid individual (CC, CT and TT). More complex genetic polymorphisms include consecutive repeated sequences of variable length known as Variable Nucleotide Tandem Repeats (VNTRs) or microsatellites. Major chromosomal rearrangements such as inversions, deletions, duplications and translocations also occur, varying in size from a few nucleotides to many megabases of sequence. The nature of a polymorphism on any given chromosome is termed an allele, the arrangement of alleles on a single chromosome gives an individuals' haplotype and a combination of haplotypes over a set of polymorphisms can be termed an individuals "diplotype" (O'Donovan and Owen, 2002, Strachan and Read, 2003a).

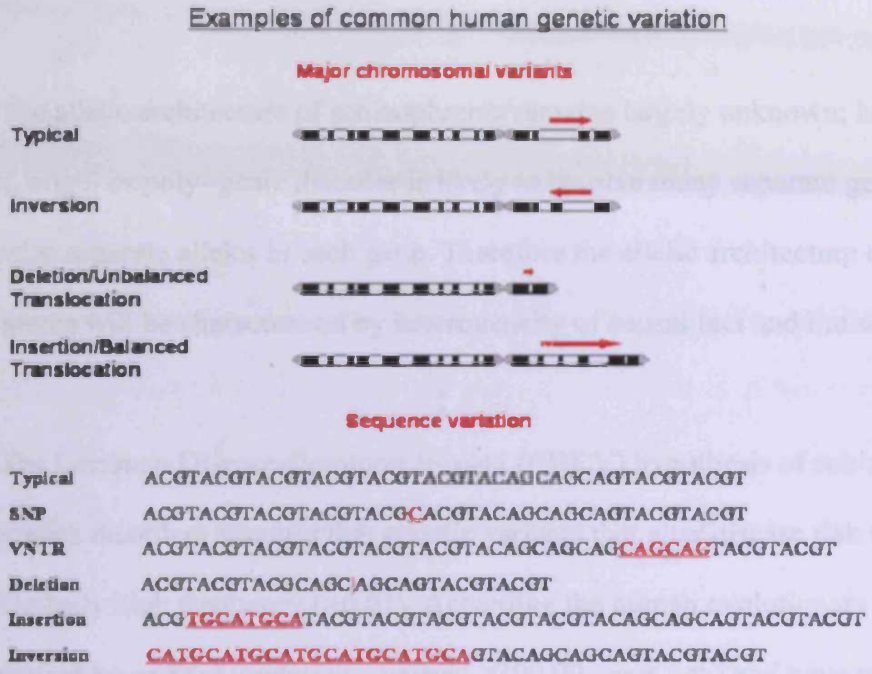


Figure 1.2: Common examples of human genetic variation, as will be discussed throughout these studies. Red arrows or sequence indicate the site of a genetic variant (Or | for small sequence deletion). Large chromosomal alterations are shown, displayed are: a typical chromosome, an inverted segment from the same chromosome, a deletion (can also be an unbalanced translocation of this locus to another chromosome without the reciprocate insertion of the dislocated chromosomal segment) and an insertion (can also be a balanced translocation, where the dislocated locus is inserted into the donor chromosomal region). Also shown are smaller sequence variants: a SNP, a variable nucleotide tandem repeat (in this case a tri-nucleotide repeat), a deletion, insertion and an inversion occurring in the same sequence.

1.3.2. Allelic Architecture of Schizophrenia

The allelic architecture of schizophrenia remains largely unknown; however a complex, oligo- or poly- genic disorder is likely to involve many separate genes and also may involve separate alleles in each gene. Therefore the allelic architecture of schizophrenia will be characterised by heterogeneity of causal loci and individual risk alleles.

The Common Disease-Common Variant (CDCV) hypothesis of schizophrenia and other complex disorders assumes that genetic variants that alter disease risk will occur with a relatively high frequency (>0.01). According the human evolutionary theory these must therefore be ancient (mutation occurred >100,00 years ago) and have undergone little negative evolutionary selection (Lohmueller, Pearce *et al.*, 2003, Reich and Lander, 2001, The International HapMap Consortium, 2003). Supportive arguments include the rapid expansion of the human population from a small founder pool (Tishkoff and Verrelli, 2003) and the similar risk of developing schizophrenia between populations (Jablensky, 2003, Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003, Shih, Belmonte *et al.*, 2004). However, there is evidence to suggest schizophrenics produce fewer offspring than normal individuals which is not consistent with a common risk allele hypothesis (Keller and Miller, 2006). Such concerns have been explained by two main hypotheses: current risk alleles were not detrimental to the fitness of ancestral carriers and carriers of risk alleles that do not develop schizophrenia have a reproductive advantage (balancing selection) (Keller and Miller, 2006).

In contrast, the Common Disease Rare Variant (CDRV) model of complex disease (Liu, Zhang *et al.*, 2005, McClellan, Susser *et al.*, 2007, Pritchard, 2001, Terwilliger and Hiekkalinna, 2006, Zhu, Fejerman *et al.*, 2005) proposes that the common diseases reflect the action of multiple rare variants. Proponents of this model utilize evidence that the majority schizophrenics have no close affected relatives, that paternal age is associated with an increased risk of schizophrenia (paternal age is associated with an increased mutation rate), the decreased fertility of patients as being consistent with many independent and recently occurring mutations causing increased susceptibility to schizophrenia (McClellan, Susser *et al.*, 2007) and also that the genes underlying human behaviour are proposed to have a high mutation rate (Keller and Miller, 2006). Also, there are model predictions of complex disease that demonstrate that common alleles are unlikely to explain the majority of complex disease as rare alleles (Pritchard, 2001). Furthermore, there are many late-onset Mendelian disorders, in which functional alleles are unlikely to be under evolutionary pressure, that show both common and rare disease allele frequencies (Zondervan and Cardon, 2004).

There are now examples from complex diseases that show a wide scale of allelic diversity. Alleles that are reasonably rare have been found to explain instances of phenotypes such as Deep-Vein-Thrombosis, Crohn's disease and low high-density lipoprotein cholesterol levels (Cohen, Kiss *et al.*, 2004, Zondervan and Cardon, 2004). There have in addition been discoveries supportive of the CDCV hypothesis, such for late-onset Alzheimer's Disease and Type-II diabetes (Lohmueller, Pearce *et al.*, 2003,

Zondervan and Cardon, 2004). Therefore, an objective view is that both common and rare variants are likely to explain most instances of schizophrenia.

1.4. The Neurobiology of Schizophrenia

A familiarity with the neurobiological research findings within schizophrenia is essential to the molecular biologist attempting to generate a hypothesis on the genetic basis of schizophrenia. The phenotype of schizophrenia is defined purely on the basis of traditional psychiatric rationale, however the disorder is one of thought and behaviour and therefore must have a neural basis as the CNS is the vector of human thought and behaviour. However, the nervous system representing the primary locus of disease generation is not unequivocal, as shown for example in demyelinating diseases of the nervous system where external factors are prominent (e.g. myasthenia gravis and MS).

Importantly, the compounds that ameliorate and induce schizophrenia like symptoms have their action at neural receptors (Moghaddam and Krystal, 2003, Waddington, Kapur *et al.*, 2003), and the lack of autoimmune or inflammatory pathology indicate the aetiology of schizophrenia to lie within the genetics of the neurons or glia that form the CNS. However, once this premise is accepted, how are the cells of the CNS disrupted to create schizophrenia? As opposed to a few decades ago where the search for schizophrenia pathology was termed a "graveyard" for those foolish enough to try (Harrison and Lewis, 2003), there is now strong and widespread evidence for neuropathologies of schizophrenia (Harrison and Weinberger, 2005).

1.4.1. The Micropathology or Molecular-Pathology of Schizophrenia

Human behaviour reflects information conducted by the electrochemistry of the hundreds of millions of neurones, communicating via billions of synapses. There are countless factors that influence neurotransmission, which all affect the number, structure and chemistry of the synapse, or the conduction of the electrical impulse to the synapse. Therefore, regardless of the primary pathology, the synapse will be entailed in the mechanism of schizophrenia. Concurrently, the evidence for synaptic disruption in schizophrenia is probably the best characterised of the potential schizophrenia neuropathologies (Harrison and Lewis, 2003, Harrison and Weinberger, 2005, Owen, O'Donovan *et al.*, 2005). One of the putative schizophrenia macropathologies is an increase in ventricular volume and concomitant reduction of the whole brain volume due to loss of grey matter in several brain regions including the prefrontal cortex (PFC), Hippocampal formation (HF) and the medio-dorsal and anterior thalamic nuclei of the thalamus (Harrison and Lewis, 2003, Harrison and Weinberger, 2005). Such findings are corroborated by a reduction in synaptic markers in such regions (Harrison and Lewis, 2003, Harrison and Weinberger, 2005), from cytoarchitectural studies show a reduction of dendritic arborisation rather than any other neuronal pathology (Black, Kodish *et al.*, 2004) and microarray studies contending dysregulation of synaptic transcripts in schizophrenia (Mirnics, Middleton *et al.*, 2001).

Alternative or complementary explanations describe a purely glial or myelin related disruption as the primary aetiopathological site in schizophrenia, which also has documented neuropathological evidence similar to that outlined above (Hakak, Walker *et*

al., 2001, Moises, Zoega *et al.*, 2002). However, the volume of supportive evidence is not equivalent to that for synaptic pathology in schizophrenia (Harrison and Weinberger, 2005, Owen, O'Donovan *et al.*, 2005), although microarray studies report both synaptic (Mirnics, Middleton *et al.*, 2001) and myelin (Hakak, Walker *et al.*, 2001) related transcript dysregulation and pathological studies similarly show evidence for white (Davis, Stewart *et al.*, 2003) and grey matter (Harrison and Lewis, 2003, Harrison and Weinberger, 2005) disruption . Furthermore, destabilisation of the synapse will have repercussions throughout the cytoarchitecture of the neuron: a loss of synaptic input and the trophism that it supplies will result in atrophy of the neuron and the loss of axonal white matter and a proportion of the supportive glial population. Therefore the white matter and glial decreases reported in schizophrenia may represent concomitant secondary pathologies; alternatively the synaptic pathology may also be secondary to white matter pathology.

1.4.2.The Macropathology or Brain Region Pathology of Schizophrenia

None of the pathologies reported in schizophrenia CNS encompass the entire brain. The synaptic pathology apparent in schizophrenia is not ubiquitous throughout the CNS and there are certain regions that consistently show disruption in schizophrenics. Neuropsychological testing (Braff, Geyer *et al.*, 2001, Molina, Sanz *et al.*, 2005), functional neuroimaging (Liddle and Pantelis, 2003), structural and molecular studies (Harrison and Lewis, 2003, Liddle and Pantelis, 2003) of the schizophrenic CNS converge on the PFC as a neural correlate of the disease (Harrison and Lewis, 2003,

Harrison and Weinberger, 2005, Liddle and Pantelis, 2003). The evidence is just as compelling for the temporal lobe, particularly the hippocampal formation (HF), consisting of the hippocampus and heteromodal association cortices (Harrison and Lewis, 2003, Harrison and Weinberger, 2005, Liddle and Pantelis, 2003). Finally, there is less numerous but just as coercing data for thalamic disturbances in schizophrenia, particularly the medio-dorsal and anterior thalamic nuclei which project to the PFC (Harrison and Lewis, 2003, Harrison and Weinberger, 2005, Liddle and Pantelis, 2003). These disparate neuroanatomical loci possibly dysregulated in schizophrenia exist on a direct neural pathway between the HF, medio-dorsal an anterior thalamus and PFC, which has been implicated in the formation of emotional episodic memory (Aggleton and Brown, 1999).

However, it must be noted that studies attempting to replicate functional neuropathological findings often fail to find the same effect or find the opposite result (Liddle and Pantelis, 2003). Therefore, although functional studies consistently find differences between affected and unaffected individuals and there is no unifying neural or anomaly in schizophrenics. It may be the case therefore, that the disorder is generated by many heterogeneous processes even at the neural systems level, where for example the control of activity in a particular region (e.g. the PFC) is more important than the overall apparent functioning of that region (Winterer and Weinberger, 2004).

1.4.3. Other Clues to the Pathology of Schizophrenia

The behavioural and physical data implicate the spatial locality of schizophrenia pathology; however the temporal qualities of schizophrenogenesis may also inform us of the neural mechanisms involved. Pre-natal environmental risk factors for schizophrenia are well established (Jablensky, 2003, Maki, Veijola *et al.*, 2005, McGrath and Murray, 2003), however schizophrenia does not typically become apparent enough to generate a psychiatric diagnosis until early adulthood, despite some pre-morbid behavioural and physical predictors. The development of the nervous system is not finalised at birth, but continues into early adulthood, and there is even regenerative potential in the mature CNS, which may be involved in the pathogenesis of schizophrenia (Reif, Fritzen *et al.*, 2006). Therefore, schizophrenia is likely to be induced during the development of the nervous system (Lewis and Levitt, 2002, Weinberger and Marenco, 2003). The major risk factors occur during early life and adolescence; corresponding temporally with neural induction, migration and synaptogenesis in the embryo and infant, and then with the selective synaptic loss or "pruning" that occurs during early adulthood (Lewis and Levitt, 2002, Weinberger and Marenco, 2003). The mechanisms by which the "two-hit" hypothesis acts, and the inter-dependence and independence of the two stages of disease induction are unknown (Lewis and Levitt, 2002, Weinberger and Marenco, 2003) however the development of the neural system is expected to be the temporal location of disease aetiology.

1.4.4. A Unifying Hypothesis of the Neuropathology of Schizophrenia?

Reasoning that schizophrenia pathology occurs primarily within association cortices, hippocampal formation and medial-anterior thalamic synapses is a large but justifiable hypothesis (see above). However, our limited knowledge of the sub-genome responsible for the correct construction of this system would still identify thousands of genes potentially involved in the generation, maintenance and plasticity of synapses in these regions. However, the neuroleptics and psychomimmetics that act to ameliorate and induce the multiple symptoms of schizophrenia incriminate several neurotransmitter systems, notably dopamine, glutamate, GABA and serotonin (Moghaddam and Krystal, 2003, Rang, Dale *et al.*, 1999, Waddington, Kapur *et al.*, 2003). The neuropathological literature also points to a glutamatergic hypofunction in schizophrenics (Collier and Li, 2003, Harrison and Lewis, 2003, Harrison and Weinberger, 2005, Owen, O'Donovan *et al.*, 2005) and finally the current list of putative schizophrenia susceptibility genes also converge on a glutamatergic disruption in schizophrenia (Harrison and Weinberger, 2005, Owen, O'Donovan *et al.*, 2005) (see section 1.12).

The many molecular/cellular phenotypes that may increase the risk of acquiring schizophrenia may converge on a common biological pathway; however the current neurofunctional and neuropathological studies suggest that a unifying biological system is unlikely (Hakak, Walker *et al.*, 2001, Harrison and Lewis, 2003, Mirnics, Middleton *et al.*, 2001). Rather then, if it is accepted that schizophrenia is brought about by a regional disruption of the neural signalling routes that forge human behaviour (language, logical thought, social interaction), i.e. systems entailing the PFC and limbic structures, then

such a disruption can be brought about in multiple fashions: the neurodevelopmental positioning of a neurone and connections, the number, type and maintenance of synapses, neurotransmission and trophism at the synapse, the disruption of the action potential and the nutritive, structural and signalling facilitation provided by neuro-glia interrelations. It is not suggested that all these pathways are commonly involved in schizophrenia, but that more than one of these pathways explains the heterogeneity of schizophrenia. Therefore, a glutamatergic-synapse hypothesis of schizophrenia pathology is not in itself a unifying theory and is unlikely to explain all instances of schizophrenia, but is the model that best fits the current neuropathological findings (Collier and Li, 2003, Harrison and Lewis, 2003, Harrison and Weinberger, 2005, Owen, O'Donovan *et al.*, 2005).

1.5. Genetic Linkage and Linkage Analysis

1.5.1 Genetic Linkage

Genetic linkage refers to a departure from the independent assortment of genetic material; that is the co-inheritance of genetic alleles in a non-random fashion. To understand this phenomenon requires an understanding of the transmission of genetic material between generations.

The generation of haploid cells during parental gametogenesis involves the pairing of homologous chromosomes in the diploid cell before cell division. When the homologous chromosomes of the diploid cell anneal during meiosis they exchange chromosomal segments at cross-over points called chiasma. The operation is termed

meiotic homologous recombination or meiotic cross-over and generates gametic chromosomes consisting of alternating chromosomal segments from the paternal and maternal chromosomes (Figure 1.2). Recombination therefore generates increased diversity among the human species and restructures genes and their alleles (Riley, Asherson *et al.*, 2003, Sham and McGuffin, 2002, Strachan and Read, 2003b). There are approximately 35 recombinations or crossovers per gamete, and therefore each chromosomal segment inherited from the father or mother will likely contain many genes/loci from that parent. Two loci on the same chromosome will be transmitted together unless they are separated by a crossover point. Therefore, the closer the two loci are together the larger the likelihood they will be co-transmitted in a gamete. This is the phenomena of linkage and is a departure from the law of independent assortment (Riley, Asherson *et al.*, 2003, Sham and McGuffin, 2002, Strachan and Read, 2003b).

1.5.2. Linkage Analysis

Linkage analysis is a molecular genetic technique for the identification of genomic regions influencing expressed traits or phenotypes. Essentially, linkage analysis aims to discern if a genetic marker co-segregates with a phenotype (Riley, Asherson *et al.*, 2003, Sham and McGuffin, 2002, Strachan and Read, 2003b). The marker itself need not be causing the phenotype, but its coinheritance with the phenotype is indicative that a susceptibility locus causing the phenotype resides on the same chromosomal region as the linked marker (Figure 1.3) (Riley, Asherson *et al.*, 2003, Sham and McGuffin, 2002, Strachan and Read, 2003b).

Linkage analysis in human disease or trait genetic mapping begins with the ascertainment of pedigrees expressing the phenotype of interest. The individuals in these pedigrees are then genotyped for polymorphic markers. Following this, there are two general types of linkage analysis; model based or parametric linkage analysis and model free or non-parametric linkage analysis. Both involve statistical methods for inferring the likelihood that the data indicates a locus is linked, demonstrated as a logarithm of the odds (lod) score. Parametric methods involve stipulating a precise model for the data (e.g. population prevalence of disease, penetrance and phenocopy rate); however this is often difficult for studies of schizophrenia pedigrees where the precise mode-of-inheritance and disease model is unknown.

In model-free methods the evidence for linkage is dependent solely upon the degree of allele sharing within disease afflicted families; allele sharing being higher for a locus containing a disease allele between related individuals concordant for the disorder than between unaffected and affected relations. Individuals are said to be Identical-By-State (IBS) if they share they same copy of an allele at a locus, and Identical-By-Descent (IBD) if the allele is demonstrably an allele from a common ancestor. Therefore IBD is the more informative measure, as IBS can occur by chance and mask the true sharing of the same chromosomal segment. (Riley, Asherson *et al.*, 2003, Sham and McGuffin, 2002).

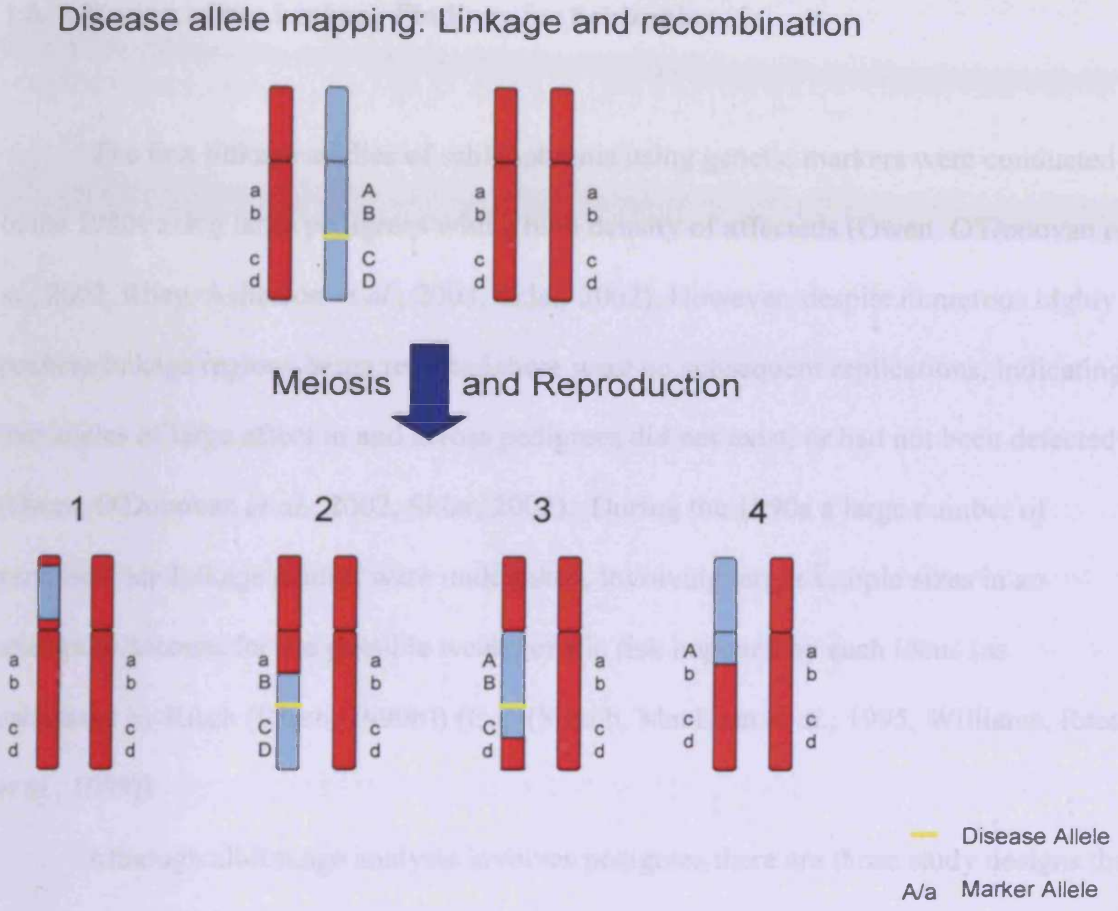


Figure 1.3: Meiotic recombination and Linkage Mapping: A sexual reproduction resulting in four offspring. Four markers have been genotyped across the long arm of the chromosome for all individuals and the alleles and therefore haplotypes for each individual are shown. Recombinations can be identified by the genotypes of the markers: Individual 2 is recombinant between markers A and B, individual 3 is recombinant at the C and D position and individual 4 is a recombinant showing the haplotype A-b-c-d which is not present in any parent. A disease allele may be present in this pedigree (yellow bar). If the disease allele is dominant and fully penetrant then individuals 2 and 3 will co-segregate the phenotype and also the haplotype B-C, while the unaffecteds will never have this haplotype.

1.6. A Review of the Linkage Findings for Schizophrenia

The first linkage studies of schizophrenia using genetic markers were conducted in the 1980s using large pedigrees with a high density of affecteds (Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003, Sklar, 2002). However, despite numerous highly positive linkage regions being reported there were no subsequent replications, indicating that alleles of large effect in and across pedigrees did not exist, or had not been detected (Owen, O'Donovan *et al.*, 2002, Sklar, 2002). During the 1990s a large number of genome-wide linkage studies were undertaken, involving larger sample sizes in an attempt to account for the possible weak genetic risk imposed by each locus (as calculated by Risch (Risch, 1990b)) (E.g. (Straub, MacLean *et al.*, 1995, Williams, Rees *et al.*, 1999))

Although all linkage analysis involves pedigrees there are those study designs that incorporate extended and multiply affected pedigrees, and those that examine affected siblings and their affected/unaffected sibling pairs. It is apparent from subsequent linkage replication findings and follow up association studies that studies of multiply affected families have been the most fruitful (Riley, Asherson *et al.*, 2003, Sklar, 2002). Studies of samples of under 100 multiply affected pedigrees in a wide selection of populations has identified significant evidence of linkage to 1q, 6p, 6q, 8p, 13q, 22q (Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003, Sklar, 2002) with regions such as 1q, 6p, 6q and 13q amongst others showing evidence for linkage in further studies. It has also become apparent from the linkage analysis of large, multiply affected pedigrees that genes of large effect can be discovered by linkage analysis of even a single pedigree (Blackwood,

Fordyce *et al.*, 2001). On the other hand, large collaborative family based studies have identified many linkages, such as to 6p, 6q, 8p and 22q (Owen, O'Donovan *et al.*, 2002). Some of these regions have subsequently shown genetic association signals that have replicating support (see section 1.12).

A less common approach in schizophrenia genetics is to use ASPs, which are often easier to obtain than multiply affected, multi-generational families. Studies of equivalent numbers of ASPs to the extended family results mentioned above (e.g. less than 100) has identified some regions of linkage, such as 6q and 10p (Riley, Asherson *et al.*, 2003, Sklar, 2002). Directed chromosomal region studies using many hundreds of ASPs has identified linkage signals; 13q being a good example (Sklar, 2002). However the largest genome-wide Affected Sib Pair (ASP) linkage studies to date were performed using sufficiently powered sample sizes to discover relative risks of >2 across the majority of the genome. Williams *et al* (Williams, Norton *et al.*, 2003) used 353 ASPs of western European ancestry and found evidence for suggestive evidence for linkage to 10q25.3-q26.3 and 17p11.2-q25.1, however the latter linkage was due to genome-wide significant linkage in a single pedigree. A slightly larger study of 382 sibs (DeLisi, Shaw *et al.*, 2002) analysing a population of similar ancestry claimed genome-wide significant linkage at 10p and suggestive evidence of linkage to 2q and 3q with weaker evidence for 22q. Therefore, apparently between two homogenous populations there may be loci conferring relative risks of ~2, however these loci do not replicate indicating both a degree of aetiological heterogeneity and probably small (and therefore not detected) effects.

1.6.1. Linkage Meta-Analyses of Schizophrenia

There have been a few attempts to meta-analyse the linkage findings across studies to account for weak effects, the most recent and comprehensive of which reanalysed many sets of genome scan data (Lewis, Levinson *et al.*, 2003) The meta-analysis involved selecting 20 genome-wide scans involving 2945 affected individuals from 1208 pedigrees of predominantly European ancestry. Each selected study involved a high density of markers genotyped in families diagnosed with schizophrenia and/or schizoaffective disorder. Analysis of each study was performed using linkage scores (e.g. Maximum multipoint LOD scores) or corresponding p values. The genome was divided into ~30cM bins containing all markers and each bin was assigned a rank in that study and also each ranked bin in each study was weighted according to the number of affecteds analysed for that study. The average values across all studies for each bin were calculated and nominal and permuted p values generated. The results of this analysis highlighted one genome-wide significant region at 2q (according to Lander and Kruglyak criteria (Lander and Kruglyak, 1995)) and suggestive loci at 5q, 3p, 11q, 6p, 1q, 22q, 8p, 20q, 14p, 16q, 18q, 10p, 15q, 6q, and 17q (Lewis, Levinson *et al.*, 2003).

Another study by Badner and Gershon (Badner and Gershon, 2002) combined significance values across studies using a Multiple Scan Probability method. The design relied on analysing 18 studies, totaling 1787 affecteds from 681 pedigrees of primarily European descent. Each study was analysed for linkage regions that showed a significance of $p \leq 0.01$ under an *a priori* defined model, after correction for the genetic size of the linked region. Once achieved the significance of the same region across other

studies also underwent the same analysis and the p values across all studies combined (with and without the original study). Also, the same type of analysis was performed using the best p value for a region (regardless of model fitted) and the combined p value corrected for the number of comparisons. All the above analyses identified 3 regions showing statistically significant evidence for linkage across all replication studies excluding the original study; at 8p, 13q and 22q. 12 other regions showed evidence for significant linkage across studies when the original study was included (Badner and Gershon, 2002).

Although drawn from predominantly the same datasets the two meta-analyses performed different analyses and subsequently the best results do not overlap, although some chromosomal regions do appear in both analyses; such as 1q, 2q, 6p, 6q, 8p, 10p, 11q, 14p and 22q (Badner and Gershon, 2002, Lewis, Levinson *et al.*, 2003). The analyses differed at several points that may explain the results, for example the Badner & Gershon analysis (Badner and Gershon, 2002) only looked at regions achieving nominal significant linkage, while the Lewis *et al* study (Lewis, Levinson *et al.*, 2003) ranked all the linkage regions for each study.

If each of the linkage studies reporting evidence for linkage to a region, either in individual or combined samples, was to be believed then much of the genome would be implicated in schizophrenia susceptibility. The Lewis *et al* (Lewis, Levinson *et al.*, 2003) meta-analysis regions alone encompass ~18% of known human genes (Sullivan, 2005) which still represents a major obstacle to mapping a schizophrenia disease gene. Therefore, as allelic and genetic aetiological heterogeneity may exist between samples a reasonable approach to the data should be to define the best linked regions in each

population and study those further. Such approaches have thus far yielded the most successful results in terms of an end product (see section 1.12). A representation of the replicating or highly significant linkage regions for schizophrenia is shown in figure 1.4.

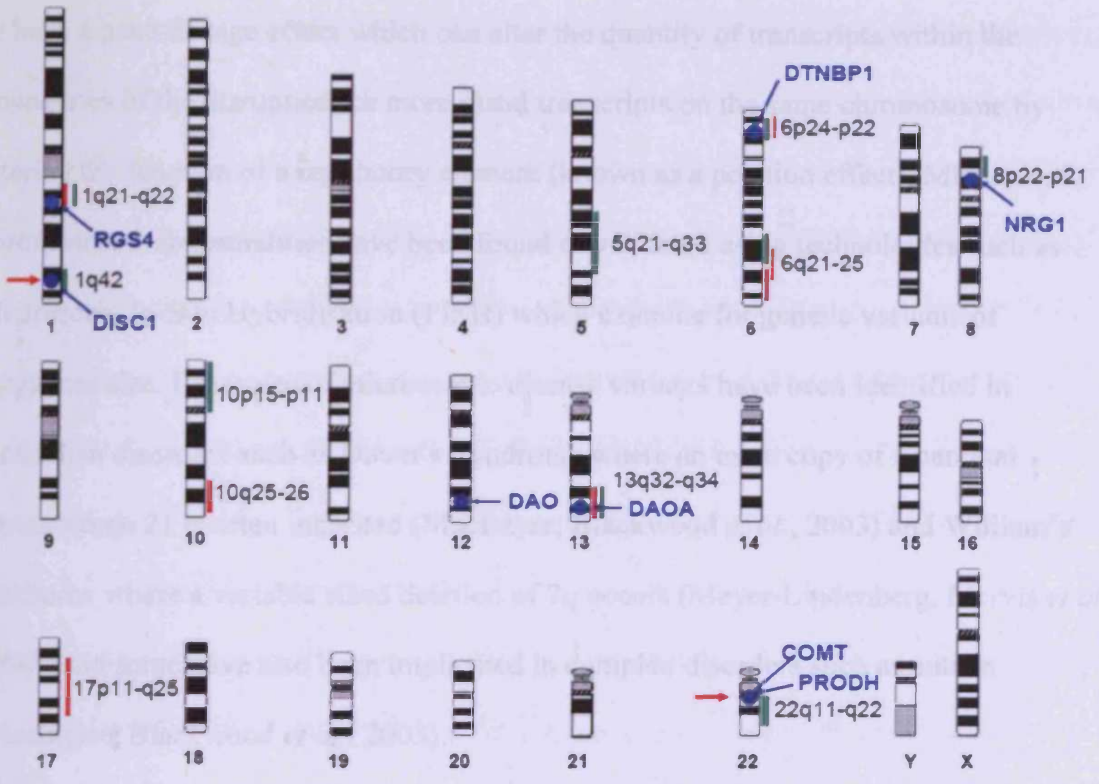


Figure 1.4: Ideogram showing chromosomal regions of linkage to schizophrenia and candidate genes that may be associated (see section 1.12). Vertical bars denote regions where evidence for linkage has been found in more than one study (green) and where linkage has reached genome-wide significance (red) according to the criteria of Lander and Kruglyak (Lander and Kruglyak, 1995). Arrows (red) identify regions where chromosomal abnormalities are involved with schizophrenia. The chromosomal positions of the major candidate genes discussed in this chapter are also displayed.

1.7. Chromosomal Abnormalities and Complex Disorders

A major chromosomal abnormality segregating with a phenotype can indicate the presence of a disease relevant gene. Large deletions, duplications, translocations and inversions may disrupt the qualitative nature of a gene (i.e. alter protein coding sequence) or have a gene-dosage effect which can alter the quantity of transcripts within the boundaries of the disruption, or more distal transcripts on the same chromosome by altering the function of a regulatory element (known as a position effect). Microscopic chromosomal abnormalities have been found or validated using technologies such as Fluorescent In-Situ Hybridisation (FISH) which examine for genetic variants of megabase size. Examples of microscopic disease variants have been identified in mendelian disorders such as Down's Syndrome where an extra copy of a parental chromosome 21 is often inherited (MacIntyre, Blackwood *et al.*, 2003) and William's syndrome where a variable sized deletion of 7q occurs (Meyer-Lindenberg, Mervis *et al.*, 2006), and some have also been implicated in complex disorders such as autism (MacIntyre, Blackwood *et al.*, 2003).

Currently detected large chromosomal abnormalities range from those that are rare and *de novo*, to those that are widespread in populations (Stefansson, Helgason *et al.*, 2005). Inherited, but more frequently spontaneous deletions occur commonly in Velo-Cardio-Facial-Syndrome (VCFS) where size-variable hemizygous deletion of 22q11 is the primary cause this relatively common disorder of chromosomal abnormality (1 in 4000 live births) (Murphy and Owen, 2001, Williams, O'Donovan M *et al.*, 2006). More population prevalent rearrangements have been shown to exist (Stefansson, Helgason *et*

al., 2005), however their relevance to disease requires thorough validation by family or population based analysis, however robust genotyping methods for CNVs comparable to those for more simple variants has yet to be made widely available (Myers and McCarroll, 2006).

With the advent of high resolution technologies to detect structural variation, such as oligonucleotide-array comparative genome hybridisation (e.g. (Locke, Sharp *et al.*, 2006)), SNP-arrays looking for loss-of-heterozygosity (LOH) (e.g. (Conrad, Andrews *et al.*, 2006)) and large-scale re-sequencing, sub-microscopic structural variants (SSVs) even down to a single base deletion can potentially be discovered. There are a growing number of publications that have attempted to catalogue copy-number variants (CNVs – deletions and duplications) in a genome-wide manner by such means, using unaffected individuals (e.g. (Conrad, Andrews *et al.*, 2006, Hinds, Kloek *et al.*, 2006, Locke, Sharp *et al.*, 2006, McCarroll, Hadnott *et al.*, 2006, Redon, Ishikawa *et al.*, 2006)) which have identified thousands of potential CNVs across the genome and validated hundreds (Eichler, 2006, Myers and McCarroll, 2006). A database exists attempting to catalogue all the potential CNVs from published studies into a "CNV map" of the genome (http://projects.tcag.ca/variation/project.html) , which at the time of writing has catalogued 3643 CNV loci and 77 inversions using 40 research articles (Iafrate, Feuk *et al.*, 2004). The general conclusion from these studies is that the between individual differences in genetic sequence due to SSVs are ten-times that due to SNPs (Redon, Ishikawa *et al.*, 2006), while the vast majority of CNVs identified are less than 5kb in size (Eichler, 2006). Furthermore, as the power of current technologies to accurately

identify small CNVs is less than for larger ones, it is conceivable that many smaller CNVs remain undiscovered.

Therefore, it would seem that a large amount of genomic variation has been recently revealed, however just as for SNPs and other simple genetic variants many of the identified SSVs are likely to by innocuous polymorphisms (Eichler, 2006, Feuk, Carson *et al.*, 2006, Myers and McCarroll, 2006) while others could impart a disease risk or be alleles for mendelian phenotypes. However, their potential impact on genetic disease studies is huge as these variants are just as likely as SNPs to influence disease risk and also may not be detected by conventional genotyping technologies (which they may even confound) (Feuk, Carson *et al.*, 2006, Myers and McCarroll, 2006).

Recent publications have identified CNVs present in small samples used for linkage disequilibrium structure analysis by the International Haplotype Mapping Consortium (Conrad, Andrews *et al.*, 2006, McCarroll, Hadnott *et al.*, 2006, Redon, Ishikawa *et al.*, 2006). The aim of these studies was to identify CNVs and also determine if they are correlated (in linkage disequilibrium, see section 1.9) with SNPs. The results of these studies produced mixed findings however, with some concluding that many CNVs identified were ancestral and so correlated with SNPs (Hinds, Kloek *et al.*, 2006, McCarroll, Hadnott *et al.*, 2006, Redon, Ishikawa *et al.*, 2006) which has been demonstrated for other SSVs (Stefansson, Helgason *et al.*, 2005), while another study pointed out that many will not be uncovered by current association mapping strategies (Conrad, Andrews *et al.*, 2006), either due to lack of correlation with SNPs or lack of SNPs in regions containing CNVs. Therefore, as a proportion of SSVs will not be identified via current association methods, the technology to identify and validate (i.e. to

genotype) such variants must become robust and standardised before high-throughput, genome-wide disease studies can begin in populations. However, the directed use of these technologies as a screen for SSVs, i.e. in disease pedigrees or sub-samples, may be followed by simpler validation procedure such as sequencing or quantitative PCR.

1.8. Chromosomal Abnormalities and Schizophrenia

There are numerous reports of chromosomal rearrangements that are associated with neurological phenotypes, such as mental retardation (MacIntyre, Blackwood *et al.*, 2003), and there have been many reports of large genomic alterations that either cause or increase risk of developing schizophrenia (MacIntyre, Blackwood *et al.*, 2003). Unbalance translocations and deletions involving 5q have been reported to co-segregate with schizophrenia in independent, but small pedigrees (MacIntyre, Blackwood *et al.*, 2003), a *de novo* inversion and deletion at 9q have been reported in unrelated patients with schizophrenia and schizoaffective disorder respectively (MacIntyre, Blackwood *et al.*, 2003) and the study of a fragile site (a region prone to breakpoints) at 9q identified evidence of disruption in ~80% of patients compared to ~10% of controls (MacIntyre, Blackwood *et al.*, 2003). Furthermore, there are many studies that find pedigrees with many psychiatric phenotypes co-segregating with a chromosomal abnormality, such as a translocation involving 18p co-segregating with schizophrenia, paranoid traits and psychotic episodes in three separate individuals (MacIntyre, Blackwood *et al.*, 2003). There are many more instances of partial co-segregation of schizophrenia with a

chromosomal abnormality, however none involve large sample sizes or coincide with widely replicated linkage signals (MacIntyre, Blackwood *et al.*, 2003).

However, there are two prominent chromosomal abnormalities that show robust evidence for involvement in schizophrenia disease pathology. There are several reports that adults with deletions of 22q11 have an increased risk of developing psychosis such as schizophrenia which may be as high as 30% (Murphy and Owen, 2001, Williams, O'Donovan M *et al.*, 2006). The rarity of such identified deletions in the general population and schizophrenic cases means that they cannot contribute to the common form of the disorder, however linkage to 22q11-13 is one of the most replicated linkage findings (Badner and Gershon, 2002, Lewis, Levinson *et al.*, 2003) suggesting that a gene or genes involved in schizophrenia reside within the deleted and/or linked region.

There has been extensive documentation of an extended Scottish pedigree where a balanced translocation (1;11)(q42;q14.3) co-segregates with major mental illness' including schizophrenia (Blackwood, Fordyce *et al.*, 2001, Millar, Wilson-Annan *et al.*, 2000, St Clair, Blackwood *et al.*, 1990). This rearrangement directly disrupts 2 known genes at 1q42, named Disrupted In Schizophrenia 1 and 2 (*DISC1* and *DISC2*), which may be particularly relevant to other European populations where linkage has been reported to 1q (Owen, O'Donovan *et al.*, 2002, Sklar, 2002). A more detailed review of this data is given in section 1.12.6.

Future genome-wide scans for chromosomal abnormalities, particularly copy-number variations (CNVs) may reveal that linkage regions and associations are, in part, due to common structural polymorphisms (Moon, Yim *et al.*, 2006), however it is

important that association studies of such CNVs involve assays with low type I and II error rates.

1.9. Linkage Equilibrium and Linkage Disequilibrium

An understanding of the phenomena of Linkage Equilibrium and Linkage Disequilibrium is fundamental to population genetics. Linkage equilibrium (LE) refers to the complete independence of genotypes at two loci, for example loci on separate chromosomes as measured across individuals from a given population. Linkage disequilibrium (LD) occurs when there is a co-occurrence or correlation between genotypes at two loci on the same chromosome, as measured in a population. Therefore, two adjacent alleles on the same chromosome may be completely independent (LE), or show some degree of correlation (LD). The calculation and measures of LE and LD are performed using data obtained from a population; however measures of inter-marker LD are applicable to any individual of the same ancestry.

Gauging the extent of LD between two loci is done through statistical dimensions, of which there are two commonly used measures; D' and r^2 (Devlin and Risch, 1995, Mueller, 2004). Both measures are based on the pairwise disequilibrium coefficient D, which is the co-variance between two loci; that is a measurement of observing the alleles of each marker on the same haplotype or independently of each other in a given population. To calculate a value for D between two marker alleles (i.e. A and B), the allele frequency of the two alleles (p_A and q_B) and the haplotype frequency (α_{AB}) are required (Devlin and Risch, 1995, Mueller, 2004).

$$D_{AB} = a_{AB} - p_A q_B$$

A preferred measurement of this value is to normalise D to give a value of D' which lies on a scale of 0-1, where a value of 0 implies complete un-relatedness while 1 indicates complete LD (but not necessarily perfect LD). *D'* is calculated by dividing the observed value for *D* by the theoretical maximal value for *D* (D_{max}) (Devlin and Risch, 1995, Mueller, 2004):

$$D' = \frac{D}{D_{max}}$$

A more informative measure still is r^2, where a value of 0 again shows independence of the two markers' alleles, but a value of 1 indicates the perfect correlation of alleles of the two markers, where the alleles of both markers always co-occur and therefore are the same frequency (perfect LD) (Zondervan and Cardon, 2004):

$$r^2 = \frac{D^2}{f(A_1)\, f(A_2)\, f(B_1)\, f(B_1)}$$

Typically, the closer two alleles are together on a chromosome the longer it will take for a recombination to separate them, therefore alleles that lie close together are

often in high LD in any population. However, the extent of LD within a region is dependent upon multiple factors including the recombinational history of the locus and the mutation rate across the region (Sham and McGuffin, 2002, Tishkoff and Verrelli, 2003, Zondervan and Cardon, 2004). Therefore, the extent of LD can vary dramatically between regions, with LD being complete, i.e. $D'=1$ and $r^2<1$ (where the allele of one polymorphism always occurs with the allele of another polymorphism, but the frequencies of the two alleles differ), perfect, i.e. $D'=1$ and $r^2=1$ (where there is perfect correlation or co-occurrence between the alleles of two polymorphisms), incomplete, i.e. $D'<1$ and $r^2<1$ or absent, i.e. $D'=0$ and $r^2=0$, for loci separated by the same distance.

The primary factor that diminishes LD between markers is recombination. New mutations will occur on a haplotype and so should be in complete LD (not perfect) with the surrounding ancestral alleles of that haplotype, however this will be altered by recombination. It is known that recombination rates differ across the genome (Strachan and Read, 2003b) and so there should be regions where the D' and possibly r^2 between all markers is high, as recombination does not often occur. These should be broken-up in areas where the recombination rate is high and so the haplotypes are constantly re-shuffled. This rationale has since been shown to be the case across the genome (Daly, Rioux *et al.*, 2001, The International HapMap Consortium, 2003) and regions of high LD between markers where there is little haplotype diversity have become known as "haplotype blocks", which are interspersed by regions of low LD with higher recombination rates (Daly, Rioux *et al.*, 2001).

1.10. Association

1.10.1. Association and Association Mapping

Association in genetic studies refers to the co-occurrence of an allele or genotype with a trait more often than would be expected by chance. Such a finding indicates that the allele is either a disease causing variant or is serving as a proxy for the true disease causing variant due to linkage disequilibrium (Riley, Asherson *et al.*, 2003, Sham and McGuffin, 2002, Zondervan and Cardon, 2004). Association studies generally take on two forms; the indirect association study and the direct association study.

1.10.1.i. Direct Association

The purpose of an association study is to find the disease risk altering allele(s) or genotype, which should exhibit a direct association. A direct association study tests markers with a putative biological function for association with a disease. However, the design assumes that the disease causing allele or genotype is being directly tested by the analysis performed, which may not be the case for example if complex genetic mechanisms of risk are at work (e.g. diplotypes). Direct association studies typically follow the directed mutation screening of a gene to identify all putative functional variants. However, this process involves the assumption that a true functional sequence has not been overlooked by the strategy. Although many protein coding, transcribed, and putative regulatory motifs have been catalogued (e.g. http://genome.ucsc.edu/) there still

remains a large proportion of the genome that either has no function, or a function not yet characterised (ENCODE Project Consortium, 2007). Direct association studies are therefore vulnerable to false negative findings because of these assumptions.

1.10.1.ii. Indirect Association

A SNP can be associated with disease by virtue of being in LD with a true causal variant. This is termed indirect association and allows one to examine a region for association without actually examining the true mutation of interest (Zondervan and Cardon, 2004). Consequently, an indirect association study does not set out to identify a disease allele, but attempts to demarcate a region where a disease allele is present. Further re-sequencing of the causative region will be required to find and test putative disease alleles. Possibly a more informative method of indirect association analysis involves association analysis of haplotypes, as a disease allele occurs on a haplotype background and so forms a disease haplotype that should exhibit the strongest evidence for indirect association when directly tested.

1.10.2. Association Study Design: Case-Control and Family Based

When performing association mapping across a region to identify a disease locus there are many factors to consider: The type of population studied, the selection of markers to genotype, the methods of analysis of markers genotyped the frequency and number of disease allele(s) and how they manifest risk, the effect size of allele(s) on the

phenotype, the power of the sample size studied to detect an association if true and reject association if none is found, possible environmental or epistatic influences on the risk imposed by the genotype and once the disease locus has been refined, the probable re-sequencing efforts to find putative functional variants. There are two main approaches to association studies, these are known as case-control and family based studies.

1.10.2.i. Case-Control Association Studies

The most straightforward association study involves assaying the frequency of alleles or genotypes in a sample of affecteds (cases), and comparing this with the frequency in unaffecteds (controls) from the same population. If a genetic variant can be shown to be present at significantly different levels between cases and controls then the genetic variant is said to be associated with disease. There are many advantages of this study design, such as the relative ease of construction of a large sample of cases and controls which can therefore have power to detect weak effects.

When constructing a case-control sample several factors must be considered to minimise confounding factors and allow the correct interpretation of findings. Firstly, the size of the sample to ascertain must be considered because the power of a case-control study to detect true association and exclude association is dependent on the effect size of the disease allele. The calculation that the relative risk (λ) imposed by each schizophrenia risk allele is likely to be small (≤ 2) gives us a guideline for design of a case-control study (Risch, 1990b, Risch, 1990a). Therefore, the best case scenario would be a common disease allele of large effect (i.e. Allele frequency of 50% and relative risk of 1.5 for a

heterozygote and 2 for a homozygote risk allele). A sample of 275 cases and an equal sized sample of controls will be required for 80% power to detect an effect at a $p \leq 0.05$ significance level (http://pngu.mgh.harvard.edu/~purcell/gpc/) or exclude with a reasonable degree of certainty any large effect association with such a variant. A more pragmatic scenario however, would involve an allele frequency of 0.2 and a relative risk of 1.2 for a heterozygote and 1.5 for a homozygote, which would require 1236 cases and 1236 controls for similar power (http://pngu.mgh.harvard.edu/~purcell/gpc/).

The major detractor of case-control association studies is the phenomenon of population stratification (Hirschhorn and Daly, 2005, Newton-Cheh and Hirschhorn, 2005, Sham and McGuffin, 2002). Population stratification refers to differences between allele frequencies between cases and controls in an association study that is unrelated to affection status (Cardon and Palmer, 2003). The phenomenon can lead to reports of an association where none exists, so called "spurious" associations. Alternatively, a negative association can be claimed where the association is real in that population. Population stratification occurs when cases and controls are not well matched. This can be a result of unforeseen population admixture, where an allele with different frequencies across populations is tested in a sample made up of individuals from different populations or where another ascertainment bias exists for cases or controls (Cardon and Palmer, 2003).

To protect against population stratification a homogeneous population of similar ancestry is desirable (Cardon and Palmer, 2003). Furthermore, a process of matching cases to controls for possible confounding factors such as age and sex is advantageous, as is obtaining the sample from the same geographical area in an attempt to control for other risk or unrelated but confounding factors. Another worry for association studies is the

selection of the control sample. Biases can also be introduced using controls selected from blood repositories, and as such may be selectively sampling "super-healthy" individuals that may generate associations with health altering alleles. Ideally, controls should be screened for psychiatric phenotypes, however this is impractical. Nonetheless, it has been argue that a random sample of unscreened unaffected individuals from the same population is desirable. (Clark, Boerwinkle *et al.*, 2005, Hirschhorn and Daly, 2005, Newton-Cheh and Hirschhorn, 2005, Sham and McGuffin, 2002, Zondervan and Cardon, 2004). The use of unscreened controls has been shown to be suitable for disease with a reasonably uncommon population prevalence (Moskvina, Holmans *et al.*, 2005).

A method that attempts to control for population stratification is the genotyping of unlinked and highly polymorphic markers throughout the genome of samples. Statistical methods can then be employed to account for any stratification observed in the sample when performing association analysis. This method, called genomic control (Sham and McGuffin, 2002), does not prove or disprove population substructure but studies using the technique have implicated that population stratification is unlikely in well matched case-control studies (Hirschhorn and Daly, 2005). There are other forms of analysis, such as where individuals are grouped into populations based upon their multilocus genotype data using a program called STRUCURE (Pritchard, Stephens *et al.*, 2000).

The measures discussed to combat population stratification, such as matching cases and controls and genomic control are tools for addressing the problems of population substructure in a case-control association sample. However, these measures can only minimise the risk of population stratification having an impact on an association

study. An alternative to the case-control association study that is robust to this confounder is the family based study.

1.10.2.ii. Family-Based Association Studies

The principle of family-based association studies is that if the co-segregation of a particular allele occurs with a phenotype in multiple pedigrees above the levels expected by chance then an association can be claimed between the allele and disease (Laird and Lange, 2006, Sham and McGuffin, 2002). As all that is required is a distortion of transmission of an allele from heterozygous parents (who can be of any affection status) to affected offspring the design is robust to population stratification.

The most popular family-based association study design is the use of affected proband-parent trios, where both parents and their affected offspring are genotyped (Laird and Lange, 2006, Sham and McGuffin, 2002). If there is a distortion in the number of times an allele is transmitted to the affected offspring from a heterozygous parent (as measured across many trios) that is greater than would be expected by chance, then the allele can be claimed to be associated. There are a numerous statistical methods to test for such an association, however the Transmission Disequilibrium Test (TDT) is now the most widely used (Laird and Lange, 2006, Sham and McGuffin, 2002).

Family based association studies are comparable to case-control studies in terms of their power to detect association (Laird and Lange, 2006). However, a major detractor of the family based study as compared to population methods is the difficulty in acquisition of the sample. This is to the need for three samples for each trio and also the

possible difficulty in obtaining co-operation from relatives of an individual affected with a socially disruptive and stigmatic disorder such as schizophrenia and late onset disorders. A further drawback is the genotyping of the sample; with at least two genotypes being required for each trio for a given variant. It is also noteworthy that while a suitably powered family based study is resistant to population stratification, it remains open to both false positive and false negative result due to locus and allelic heterogeneity.

Both population and family studies have their benefits and pitfalls and a researcher must decide which type of sample suits the needs of the project. A wholistic view realises that the two methods are mutually beneficial and either study can be replicated by a similar finding in the alternative sample from the same population.

1.10.3. Marker Selection for Association Studies

1.10.3.i. Which markers and why?

The human genome contains approximately 2,858,034,764 sequenced bases (http://genome.ucsc.edu/goldenPath/stats.html#hg18) and 11,170,620 SNPs are listed in the most recent assembly of SNPs (http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=9606). Furthermore there could be more than 300,000 polymorphic repeats such as VNTRs (http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=9606), and hundreds of large structural rearrangements (http://projects.tcag.ca/variation/) are

catalogued. This data represents the current knowledge of the human genome, and it is estimated that the majority of polymorphisms are rare and have not been documented (The International HapMap Consortium, 2003). Therefore the scope for variation within an association sample is huge and marker discovery and/or selection is in itself a hugely complex and controversial topic, but crucial to the success of any study. As any direct association study is often governed by indirect study findings, the indirect association study marker selection strategy is considered in detail.

The ideal association study design would involve the genotyping of all polymorphisms in a given sample across the whole genome. However, this situation will only be available once the complete re-sequencing of all samples has been performed. Furthermore, many alleles will have a high correlation with other alleles, in fact the LD may occur to such an extent as to make the genotyping of many alleles or genome re-sequencing a redundant exercise. Therefore, an alternative approach is to first assay a proportion of the variants in a region in a small subset sample and estimate the degree of LD between marker alleles. Then a selection of the informative markers and removal of redundant markers can be performed, leaving a researcher with a smaller number of "tag SNPs" to genotype that should capture the same or an acceptable amount of variation within the region and whole sample (Newton-Cheh and Hirschhorn, 2005, Stram, 2004, Zondervan and Cardon, 2004). Such a premise was the basis for the entire International Haplotype Map Consortium project or HapMap project, which has become the primary tool for the contemporary indirect association study (The International HapMap Consortium, 2003, The International HapMap Consortium, 2005).

1.10.3.ii. The HapMap as a Tool for Indirect Association Studies

The HapMap project was initially launched under the premise that due to inter-marker LD in any given population only a relatively small amount of common genetic variation would need to be genotyped to capture a high percentage of the total common genetic variation of the human genome (The International HapMap Consortium, 2003). Therefore by assaying a small sample of individuals from a given population the majority of the common genetic variation and the LD structure in the whole population can be inferred (The International HapMap Consortium, 2003, The International HapMap Consortium, 2005). A method of tag SNP selection can then be performed and these analysed in an association sample that shares the same ancestry as the population of LD inference. The HapMap project has catalogued over 1 million variants as part of a first phase and this number will increase to over 5.5 million SNPs as the end of phase 2; all genotyped in 45-90 individuals from each of their populations of inference (individuals of Western European, Western African, Han Chinese and Japanese ancestry), and all the data is free and publicly available (The International HapMap Consortium, 2003, The International HapMap Consortium, 2005) (http://www.hapmap.org/).

As expected, some variants have a high degree of pairwise correlation and therefore occur in regions where multiple SNP alleles show a low amount of haplotype diversity ("haplotype blocks") interspersed by regions of SNPs showing little LD (The International HapMap Consortium, 2003, The International HapMap Consortium, 2005).Furthermore, the project has also seen the re-sequencing of smaller regions of the

genome in the some of the same inference individuals (The ENCODE project) to gauge the success of the SNP genotyping studies in accurately constructing the LD structure, which have concluded that a small amount (~10%) of power to detect the effects of genetic variation is lost when only tag SNPs are considered (de Bakker, Yelensky *et al.*, 2005, The International HapMap Consortium, 2003, The International HapMap Consortium, 2005).

The major detractors of the HapMap project argue that rare variants may be more important for common disease than is currently accepted (Cohen, Kiss *et al.*, 2004, Hirschhorn and Daly, 2005, Liu, Zhang *et al.*, 2005, Newton-Cheh and Hirschhorn, 2005, Pritchard, 2001, Zhu, Fejerman *et al.*, 2005, Zondervan and Cardon, 2004), that the populations of inference may not accurately reflect the LD structure of the populations they are meant to (Sawyer, Mukherjee *et al.*, 2005), values of LD statistics are overestimated in small samples and some have even gone as far as to claim the HapMap theory that a genotyped "proxy" SNP in apparently complete LD with the causal variant will not show any association on the majority of occasions, even in the largest samples (Terwilliger and Hiekkalinna, 2006). A theory recently supported from simulated genotype data where a proxy allele may need to be the near perfect correlate of an ungenotyped allele for association to be detected and replicated independently (Lin, Vance *et al.*, 2007).

1.10.3.iii. Tagging approaches to LD/association mapping

Tagging in genetic studies refers to the selection of markers that act as surrogates or proxies for unassayed alleles. There are many methods of tagging, but broadly speaking two methods that have become popular both of which involve the use of haplotype tagging SNPs (here termed htSNPs) and pairwise tag SNPs (here called tag SNPs) (de Bakker, Yelensky *et al.*, 2005, Pe'er, de Bakker *et al.*, 2006, Stram, 2004, Zeggini, Rayner *et al.*, 2005).

Pairwise SNP based tagging

The pairwise tag SNP method involves the selection of markers based upon their independence, as measured by pairwise r^2 measurement of LD (de Bakker, Yelensky *et al.*, 2005, Stram, 2004, Zeggini, Rayner *et al.*, 2005). In short, the perfect tagSNP selection strategy would genotype all markers where alleles were not in an r^2 of 1 with any other marker, as defined in the sample of LD inference. The disease causing allele may of course be one of the genotyped markers, but alternatively the power to detect association with an un-genotyped marker in LD with a genotyped proxy will be dependent upon the formula $1/r^2$ (where 1 is the sample size that enables detection at $p<0.05$ if the allele is tested directly) (Terwilliger and Hiekkalinna, 2006). Relaxing the r^2 threshold will reduce the number of proxies to be genotyped, and a commonly used value is to remove SNPs showing $r^2>0.8$ with a genotyped SNP, which only marginally reduces power (especially for loci closer together) but can substantially reduce cost of genotyping (de Bakker, Yelensky *et al.*, 2005). However the performance of such a threshold is

questionable (Lin, Vance *et al.*, 2007, Terwilliger and Hiekkalinna, 2006) as obviously the degree of correlation with an ungenotyped variant may also decrease with the removal of such SNPs, although a potential benefit is the removal of another assay and test to correct for (see section 1.10.5).

Software enabling the analysis of LD structure in a given population (such as a HapMap inference population) and selection of pairwise tagSNPs as defined by user setting is provided by the implementation of Tagger in the package Haploview (Barrett, Fry *et al.*, 2005). The analysis of pairwise defined tagSNPs should simply involve the assessment of each variant for association, with adequate corrections for any multiple tests. However, the proponents of the method advocate exhaustive haplotype analysis (i.e. all possible haplotypes) of the pairwise tags, which can raise power to detect causative alleles of $\leq$5% minor alleles frequency fro ~20% to ~60% in the CEU sample, at the cost of correcting for multiple tests (de Bakker, Yelensky *et al.*, 2005).

Haplotype based tagging

The alternative method of association mapping using tagging SNPs involves htSNPs, where rather than single-marker analysis the haplotype diversity in a population is taken into account (Clark, 2004, de Bakker, Yelensky *et al.*, 2005, Newton-Cheh and Hirschhorn, 2005, Pe'er, de Bakker *et al.*, 2006, Schaid, 2004, Stram, 2004). The result is that far less genotyping is required and so the study design is more efficient, however problems arise from interpreting the results as the addition of markers exponentially increases the amount of haplotypes to analyse.

There are currently widely available programs for the reconstruction of haplotypes from unphased (i.e. not known what chromosome alleles lie on) genotype data. Such software can reliably predict haplotypes and haplotype frequencies and assess any association therein for both family-based and case-control association samples (EH+, UNPHASED, Haploview). Many methods of haplotype reconstruction from genotype data rely on implementations of the Expectation-Maximisation algorithm (Clark, 2004), which involves estimating the maximum likelihood probability of observing a haplotype in an individual or group of individuals, then using this as evidence that the haplotype exists.

The consideration of a multi-marker haplotype as a proxy for an un-genotyped marker may increase the economy and efficiency of an association study as compared to the pairwise tagSNP approach. This could be useful in analysing data from whole genome association study platforms, meaning that ungenotyped tagSNPs are inferred from the genotypes of the fixed panel of markers using a multi-marker approach (Marchini, Howie *et al.*, 2007, Pe'er, de Bakker *et al.*, 2006). However, the theory can be applied to any genomic region and methods based upon entropy as a measure of haplotype diversity (Georgieva, Moskvina *et al.*, 2006, Hampe, Schreiber *et al.*, 2003, Peirce, Bray *et al.*, 2006) as well as forms of htSNP selection software (Ding, Zhang *et al.*, 2005, Forton, Kwiatkowski *et al.*, 2005) have been published. However, a current problem with these methods is deciding upon which haplotypes to test from the large number that could be examined, the computing power required to do so, and the interpretation of positive and negative findings (Schaid, 2004).

It must be noted that both pairwise tagSNP and htSNP tagging should identify associations with disease alleles, be they SNP alleles or haplotypes. The conundrum when choosing which method is only in the economy and necessity of genotyping many SNPs when only a few may be needed versus the complexity of any analyses to be performed

1.10.4. Power of Association Studies: The Frequency and Effect Size of Disease Alleles

Further major factors affecting the efficacy of both direct and indirect association analysis are the frequency and also the effect size of the disease alleles (Clark, Boerwinkle *et al.*, 2005, Hirschhorn and Daly, 2005, Zondervan and Cardon, 2004). Typically the more common a risk allele the greater is the power to detect an association. Of equal importance is the size of the risk of developing the phenotype that the allele imbues; the larger the effect size the greater the power to detect an association.

Therefore, the best scenario in the association mapping of any disease loci is a common allele with a large phenotypic effect, such as with the Apoε(4) allele for late-onset Alzheimer's Disease (Zondervan and Cardon, 2004). Conversely, a rare allele (i.e. >1% MAF) with a small relative risk will be difficult to identify by current association methods. The situation with schizophrenia is that the allelic architecture is completely unknown, with common but probably non-functional haplotypes and also rare chromosomal abnormalities (Millar, Wilson-Annan *et al.*, 2000, Williams, O'Donovan *et al.*, 2005) both showing promising evidence for involvement in the aetiology of

schizophrenia. Probably the best case scenario for schizophrenia has already been described above, with a mere 275 cases and 275 controls being required to detect an association with 80% power at $p \leq 0.05$ with a directly tested allele with a relative risk of 2. However, a more unattractive scenario would be a rare allele of small effect: To have 80% power to detect an association ($p \leq 0.05$) with a RISK allele of 1% frequency providing a relative risk of 1.5 would require 3930 cases and an equal number of controls (http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html).

1.10.5. Assessing the Statistical Evidence for Association.

Statistical evidence for association in case-control studies comes from the estimation of the odds ratio or chi square statistic from a contingency table of allele or genotype counts in cases and controls. Usually, the evidence is converted into a p or probability value (probability that the null/alternative hypothesis is false/true) and typically a p value of 0.05 or less indicates that the null hypothesis of no association can be rejected (Owen, Craddock *et al.*, 2005, Petrie and Sabin, 2005, Sham and McGuffin, 2002). A p value of 0.05 means that such a value will occur on only 5% of occasions by chance (assuming all SNPs are independent), however many association studies test >20 SNPs for association and so such a p value can be expected for one of the variants.

There are several methods that attempt to adjust a probability estimate for multiple comparisons, the most prominent of which are the Bonferonni method, experiment or gene-wide adjustments and permutation methods (Hirschhorn and Daly, 2005). The Bonferonni method requires the obtained p value to be multiplied by the

number of independent tests performed. However, such an adjustment of findings for multiple comparisons may be over-conservative as because of inter-marker LD many markers tested are not completely independent and also, a weak true positive may be missed when many hundreds or thousands of SNPs are analysed.

There are some many alternative methods to the Bonferonni correction for assessing the evidence for association across a gene, such as those that involve "gene-based" or "experiment-wide" tests for association. For example, accounting for all the variation or independent tests (at some pre-defined LD threshold) in and around a gene of interest (regardless of whether this is tested) and then correcting for this (Neale and Sham, 2004, Nyholt, 2004). However, such methods are often shown to be over-conservative in the presence of weak, but true, genetic effects (Salyakina, Seaman *et al.*, 2005). The alternative and preferred method is the permutation approach which involves the random assortment of affection status to samples, each time the evidence for association is calculated and the amount of times a p value matching any initially positive p values is determined. The permutation procedure then corrects the tabulated p value by the amount of times the p value occurs by chance (Hirschhorn and Daly, 2005). However, true positives may again be missed as the more data analysed (i.e. the more SNPs) the more highly significant p values will be observed during permutation (just by chance) and so again the vast amounts of SNPs needed to localise an association means that the importance of the detected effect is diminished. Therefore, statistical attempts to correct for multiple testing, although justified due to the amount of false positives that association studies of schizophrenia have generated (Owen, Craddock *et al.*, 2005), may be overlooking true disease associations.

The gold standard in assessing the evidence for an association is the replication study (Clark, Boerwinkle *et al.*, 2005, Hirschhorn and Daly, 2005, Owen, Craddock *et al.*, 2005). In short, if more than one separate type of association study conjointly displays an association at a locus this is seen as an independent replication and is supportive evidence that a causal locus exists. The weight of evidence supporting the association can be evaluated by the amount of studies showing the same effect, with more studies increasing the confidence that an association is true. Replication can be confounded by different LD structures between samples (especially if differing tag SNPs are selected), or different aetiological factors (genetic or environmental heterogeneity), or differing effect sizes. In such a case the evidence must be summarised across many studies, culminating in a meta-analysis where all findings are considered (Talkowski, Seltman *et al.*, 2006).

1.11. Consideration of More Complex Genetic Mechanisms

1.11.1. Gene-Environment Interactions

Until recently, research into the aetiology of schizophrenia research has taken a "main-effect" approach where a particular genetic allele or the exposure to an environmental influence has been tested for association with the disorder (Clark, Boerwinkle *et al.*, 2005). This division between those seeking to establish that nature or nurture is responsible for schizophrenia seems counterintuitive considering the archetypal hypothesis for the disorder is that neither genetic variants nor environmental features

alone are likely to cause most instances of schizophrenia. Therefore, genetic association studies and environmental studies that have thus far found no relevance of their feature to the disease may be missing the fact that the expression of an environmental or genetic effect may be (and the likelihood is that it is) conditional on other environmental or genetic effects (Clark, Boerwinkle *et al.*, 2005, Newton-Cheh and Hirschhorn, 2005)

1.11.2. Gene-Gene interactions

Epistasis in general terms denotes the interaction between genes or gene products and the term has historically been used to describe genetic effects such as the "masking" of the effect of one genotype by the genotype at a separate locus (Clark, 2004, Cordell, 2002, Hirschhorn and Daly, 2005). However, true epistasis in terms of genetic studies refers to the multiplicative or synergistic actions of >1 allele or gene.

Accounting for possible epistatic effects is a burgeoning field in complex genetics, and methods testing epistasis using association data have been developed that essentially involve logistic regression to fit multiple genotypes to affection status (Cordell, 2002, North, Curtis *et al.*, 2005). A successful example of this approach has been applied to genes that feasibly interact at the biological level (as show by correlation of expression levels of gene involved in CNS myelination) in research that identified an epistatic effect on risk that was independent of main effects at two loci that also show main effects (Georgieva, Moskvina *et al.*, 2006, Peirce, Bray *et al.*, 2006)

1.12. A Review of Association Analyses of the Most Promising Candidate Genes for Schizophrenia.

Many of the genes currently implicated in the genetic aetiology of schizophrenia are discussed below and an ideogram (figure 1.4) displays their chromosomal positions in relation to interesting linkage findings.

1.12.1. Dysbindin-1 (DTNBP1)

One of the most significant and widely replicated linkage regions for schizophrenia lies on chromosome 6q21-q22.3 (Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003, Sklar, 2002), a region in which Straub *et al* (Straub, Jiang *et al.*, 2002) performed association mapping using Irish-descent families with a high incidence of schizophrenia. Inside this locus significant single marker and haplotypic association was obtained. Some corroboration of these findings came from another family based study (Schwab, Knapp *et al.*, 2003) where some of the same individual SNPs were also associated, as were numerous 2-6 marker haplotypes including some of the 3 marker haplotypes identified by Straub *et al,* (Straub, Jiang *et al.*, 2002). A large case-control study analysing two independent samples from the UK and Ireland (Williams, Preece *et al.*, 2004a) provided no evidence for single marker association of previously identified and novel variants in *DTNBP1* with schizophrenia. Regardless of this many haplotypes spanning the gene were found to have significantly different frequencies between affecteds and controls. However while these were not the same as identified previously,

they all containing a novel putative promoter variant (rs2619538 or SNP-A). Remarkably, the Irish sample in this study had earlier been studied and no association found, however the novel SNPs and subsequently haplotypes identified in the Williams *et al* (Williams, Preece *et al.*, 2004a) study yielded positive association that replicated between the samples. A family study consisting of 488 Bulgarian parent-proband trios (Kirov, Ivanov *et al.*, 2004) has shown support for 2 individual marker associations with schizophrenia previously identified (Straub, Jiang *et al.*, 2002), as well as multiple 2, 3 and 4 marker haplotype associations, many of which included the putative promoter SNP identified by Williams *et al* (Williams, Preece *et al.*, 2004a)

Further support for the worldwide role of *DTNBP1* in conferring susceptibility to schizophrenia resulted from a large case-control study of 670 Japanese cases and 588 Japanese controls (Numakawa, Yagasaki *et al.*, 2004). This study provided single marker positive associations for many markers tested by Straub *et al,* and also for SNP-A, and also numerous haplotypic associations of combinations of those SNPs. A further study (van den Oord, Sullivan *et al.*, 2003) re-examined the sample initially (Straub, Jiang *et al.*, 2002) tested by Straub *et al,* and refined the original associated haplotype in this sample, identifying a specific haplotype containing some novel SNPs as affecting the risk of developing schizophrenia. Funke *et al* (Funke, Finn *et al.*, 2004) performed a case control study analysing 3 sub-sets of the U.S. populations and found associations between individual SNPs tested by previous studies in white and Hispanic populations, but not those of African descent.

Significant associations between haplotypes composed of previously identified SNPs and schizophrenia in a study of families from China have also been demonstrated

(Tang, Zhou *et al.*, 2003). Van den Bogaert *et al* (Van Den Bogaert, Schumacher *et al.*, 2003) found little evidence for association in their study of German, Polish and Swedish populations, with single marker association for one SNP and only in the small Swedish sample. Some haplotypes also showed positive association in the Swedes only, but interestingly associations were made more significant when families with a high incidence of schizophrenia were included. Such a finding is interesting as it mimics earlier findings in multiply affected pedigrees (Straub, Jiang *et al.*, 2002, Van Den Bogaert, Schumacher *et al.*, 2003).

Therefore, the evidence is substantial that genetic variation in the locus encoding Dysbindin-1 affects the risk of developing schizophrenia. The many studies declaring association between genetic variants and combinations of those variants do not always concur on the nature of the associated alleles, however the only large study to show no evidence in favour of *DTNBP1* being a schizophrenia susceptibility locus is that by Hall and colleagues (Hall, Gogos *et al.*, 2004) which examined 5 markers in small samples of US and South African descent. Interestingly, the analysis of Caucasians from the US also reported nominal associations in a schizophrenia case-control sample but these did not meet their criteria for statistical association when all markers analysed were considered (Wood, Pickering *et al.*, 2006).

Recently, an article examining the multiple *DTNBP1* associations and the HapMap CEU sample across *DTNBP1* concluded that there is little haplotype diversity at the *DTNBP1* locus and the associated alleles of all studies of European populations lie on multiple backgrounds (Mutsuddi, Morris *et al.*, In press). They conclude that there cannot therefore be one common causal allele to all studies. However it is more the case that

there may be a common causal allele, however allelic heterogeneity and LD structural difference are also likely to exist that could mask effects as one locus or unveil the effects of another (Lin, Vance *et al.*, 2007).

Therefore, as the weight of association evidence is considerable it seems aetiological and allelic heterogeneity as well as distinct LD structures between populations may explain the observed results.

1.12.2. Neuregulin (NRG1)

The genetic locus encoding Neuregulin-1 was first implicated in affecting schizophrenia susceptibility in an Icelandic case-control and family based study in which association mapping was performed within the 8p21-22 linkage region (Stefansson, Sigurdsson *et al.*, 2002). Evidence for single marker association and strong evidence for association of several multi-marker haplotypes comprised of microsatellites and SNPs was presented, but also for a shorter core haplotype spanning the 5' region of *NRG1*. The same variants and haplotypes were then tested for association in an independent Scottish sample (Stefansson, Sarginson *et al.*, 2003). 3 SNPs achieved highly significant association, as did the seven marker at-risk core haplotype implicated in the Icelandic study and the same SNP allele implicated in the Icelandic study (SNP8NRG221533).

Corroborative data has arisen since, with a large case control study of the UK population displaying a significant increase of the risk haplotype in UK affecteds which appeared to originate from cases with a family history of schizophrenia, although no single markers approached statistical significance (Williams, Preece *et al.*, 2003). Further

replication from examination of an Irish case-control sample found positive association for 2 markers not previously examined by Stefansson *et al,* (Stefansson, Sarginson *et al.*, 2003, Stefansson, Sigurdsson *et al.*, 2002), and claimed to have refined the at-risk within the *NRG1* locus (Corvin, Morris *et al.*, 2004). Supporting evidence for association of *NRG1* haplotypes in populations of European ancestry has also been noted in a second Scottish sample (Thomson, Christoforou *et al.*, 2007), Portuguese familial sample (Petryshen, Middleton *et al.*, 2005) Bulgarian trios sample (Kirov *et al,* unpublished data) and similarly in South Africans (Hall, Gogos *et al.*, 2004) and Hispanics (Walss-Bass, Raventos *et al.*, 2006). Additionally, recent evidence implicates a missense polymorphism may increase the risk of schizophrenia in a familial Hispanic sample (Walss-Bass, Liu *et al.*, 2006), although such isolated findings require replication.

Yang *et al*, (Yang, Si *et al.*, 2003) examined Han Chinese schizophrenic family trios and found single marker association for 3 variants, as well as multiple haplotypes from these three markers that achieved significance. Examination of a Chinese population also found numerous single marker associations and haplotypic associations spanning two separate regions within the 5' region of *NRG1* (Tang, Chen *et al.*, 2004). In a huge sample of Han Chinese patients and controls evidence for haplotype association was again displayed (Zhao, Shi *et al.*, 2004), however within the Chinese samples there are differences between the markers and alleles carrying the association as well as when compared to the European studies, which could be expected as the associated haplotypes and markers are not the true pathogenic variants. These differences were further supplemented in a separate examination of a Han Chinese population using family based and case-control methods (Li, Stefansson *et al.*, 2004) which reported numerous positive

results for single markers and haplotypes spanning the 5' and central *NRG1* region. Further support for *NRG1* risk variants in non-European samples comes from case-control analysis of the Japanese (Fukui, Muratake *et al.*, 2006) and African-Americans (Lachman, Pedrosa *et al.*, 2006) populations, both claiming variants at the 5' of *NRG1* influence disease risk. It is apparent therefore that differences in LD structure and/or allelic heterogeneity may explain the wide diversity of association signals generated at the 5' region of *NRG1* in populations of varying ancestry. The former explanation may be true given the complex patterns of LD in the region (Gardner, Gonzalez-Neira *et al.*, 2006).

There have been some studies where no association has been found between any single markers or haplotypes within the gene encoding Neuregulin-1. A detailed study of the regions polymorphisms in an Irish population with a high density of schizophrenics yielded no evidence in support of *NRG1* being a susceptibility locus for schizophrenia (Thiselton, Webb *et al.*, 2004) neither did examinations of case-control and familial US samples (Duan, Martinez *et al.*, 2005, Hall, Gogos *et al.*, 2004) and a small Danish case-control study (Ingason, Soeby *et al.*, 2006). Further corroborative evidence dismissing a worldwide role for *NRG1* in schizophrenia comes from a Japanese case-control study where no evidence for association was gained fro the Icelandic at-risk haplotype or in single markers (Iwata, Suzuki *et al.*, 2004). However, the samples thus far showing evidence for association may be showing allelic and also structural LD heterogeneity, with such factors and sample size influencing the power of these studies to detect association. In agreement, two meta-analyses and a review of the literature data have both concluded that the NRG1 locus is associated with schizophrenia across populations

(Munafo, Thiselton *et al.*, 2006, Talkowski, Seltman *et al.*, 2006, Tosato, Dazzan *et al.*, 2005), although care must be taken to ensure the sample size is not only large enough to detect any effect, but homogeneous in its ancestry to prevent the effects of stratification that may occur population admixture has occurred unknowingly into sub-groups.

There is fairly strong evidence implicating the locus encoding NRG1 as a region harbouring genetic risk to developing schizophrenia. However, the evidence against this eventuality cannot be discounted; neither can the heterogeneity between studies of differing populations. It may be the case that the Asian and European samples that show association do so because they share the same causative variant, carried on different genetic backgrounds. Another eventuality is genetic heterogeneity between the groups; different causative variants within the same gene are affecting the risk of developing schizophrenia in the disparate populations.

1.12.3. Catechol-O-Methyl-Transferase (COMT)

The evidence examining the gene encoding Catechol-O-Methyl-Transferase in schizophrenia is extensive and contradictory, although the *a priori* evidence is encouraging. The gene lies within one of the best characterised linkage regions on 22q11, also within regions deleted in the various 22q11 deletion syndromes such as VCFS (Velo-Cardio-Facial-Syndrome), where there is evidence of association of these disorders with schizophrenia (Williams, O'Donovan M *et al.*, 2006). Within this region, the gene that stands out as a functional candidate is *COMT*, the isoforms of which catalyse the methylation of catecholamines such as dopamine and are particularly expressed in

prefrontal cortical regions and the hippocampus, localities alleged to be erroneous in schizophrenia.

Peripheral variation in *COMT* activity has been consistently reported and the SNP responsible was identified as a non-synonymous mutation; Val158Met (Grossman, Szumlanski *et al.*, 1992), the Val allele having significantly less activity than the Met. This polymorphism and the candidate gene have been comprehensively examined for association, but no conclusion been drawn. There have been numerous reports of association and lack of, between the Val allele and schizophrenia and a recent meta-analysis was inconclusive (Glatt, Faraone *et al.*, 2003). There is a wealth of conflicting evidence discussing this polymorphism representing a risk factor for schizophrenia, therefore the polymorphism either has no effect on disease risk or the effect is very weak and/or dependent upon interacting factors. However, many reported association studies of COMT have just taken a direct analysis of the Val58Met polymorphism and so fewer indirect studies have been attempted.

Indirect association studies examining the *COMT* gene have been just as equivocal. Original studies using only a few markers failed to find any association in Chinese patients (Chen, Lee *et al.*, 1999) or French families (de Chaldee, Corbex *et al.*, 2001). A substantial study of Ashkenazi Jews reported strong evidence for association between haplotypes including the Val Allele and flanking intronic SNPs (Shifman, Bronstein *et al.*, 2002), further corroborating that although *COMT* may be a susceptibility locus for schizophrenia this is not attributable to the Val158Met polymorphism. Such a realization is supported by two recent studies, which implicate different haplotypes in US and Irish samples respectively, spanning the *COMT* locus as being associated with

schizophrenia (Chen, Wang *et al.*, 2004, Sanders, Rusu *et al.*, 2005). Intriguingly, Bray *et al* (Bray, Buckland *et al.*, 2003b) found a haplotype including the Val/Met and two intronic SNPs comprising an at risk haplotype identified by Shifman *et al* (Shifman, Bronstein *et al.*, 2002) caused a reduction in prefrontal cortex *COMT* mRNA. However, a large collaborative study of almost 1200 cases and ~1600 controls provides no evidence for association of any variants or haplotypes of the *COMT* gene with schizophrenia (Williams, Glaser *et al.*, 2005). Although, it is worth mentioning that schizophrenia disease risk may be altered by epigenetic mechanisms that act at the *COMT* locus that have yet to be investigated by large-scale studies (Abdolmaleky, Cheng *et al.*, 2006).

1.12.4. D-Amino Acid-Oxidase and D-Amino Acid-Oxidase Activator (DAO and DAOA)

An association map was built across the 13q34 linkage region by Chumakov *et al* (Chumakov, Blumenfeld *et al.*, 2002), in case-control populations from Canada and Russia. Positive associations with schizophrenia were gained for markers spanning two novel genes, *G72* and *G30*, which overlap but are transcribed on opposite strands. Taking the positive finding within *G72* further to elucidate possible functions of this gene, a search for proteins that interact with G72 was performed using Yeast-2-hybrid technology yielding a strong binding partner; D-amino Acid Oxidase. DAO is expressed in the brain, primarily in the cerebellum, and is capable of oxidizing D-serine, an NMDA receptor activator, but resides on chromosomal region 12q24 where significant linkage has not been reported (Chumakov, Blumenfeld *et al.*, 2002). Further *in vitro* study

revealed that G72 enhances the activation of DAO, and G72 is now known as D-amino Acid Oxidase Activator (DAOA). The authors then performed mutation screening of *DAO* and found 4 single markers to be associated in the Canadian sample only and subsequently tested the interaction of genotypes within both *DAO* and *DAOA* loci and found modest evidence of a statistical epistatic interaction. The authors construe that both genes influence disease risk by altering NMDA receptor, possibly identifying a common pathophysiological pathway (Chumakov, Blumenfeld *et al.*, 2002).

Attempts to replicate these findings have met some success; Schumacher *et al,* (Schumacher, Jamra *et al.*, 2004) genotyped the same SNPs identified by Chumakov *et al* as being associated in a German sample and found single marker associations with schizophrenia in both *DAO* and *DAOA*, as well as numerous positive haplotype associations within the two loci. A study of the Han Chinese also gave support for single marker and haplotypic association at the *DAOA* locus, although *DAO* was not examined (Wang, He *et al.*, 2004). Similarly, a case-control examination of a Chinese population yielded highly significant SNP and haplotype associations (Liu, He *et al.*, 2004). Further corroboration has stemmed from a study of South African trios (Hall, Gogos *et al.*, 2004), which identified several multi-marker haplotypic associations but no single marker significance in *DAOA* and also a case-control analysis of Ashkenazi Jews yielded single variant and haplotypic association (*DAO* not examined in either study). Another study analysed a US sample of trios comprised of early onset psychosis subjects and found positive SNP and 2-marker haplotype associations (Addington, Gornick *et al.*, 2004). Fascinatingly, studies of mood disorders, which can include subjects showing schizophrenia symptomology, have found stronger evidence for association , indicating a

role for genotypes at this locus helping to refine psychiatric phenotypes (Hattori, Liu *et al.*, 2003, Williams, Green *et al.*, 2006). Therefore many studies have examined and found association at the *DAOA* locus, but few have attempted to study the *DAO* locus for association, although association has been noted (Wood, Pickering *et al.*, 2006).

As is the case for the aforementioned genes, there is no unifying haplotype or associated functional SNP allele for either *DAOA* or *DAO*. Again raising the possibilities of allelic and/or LD heterogeneity between populations, inadequate samples sizes, population stratification issues and many more possible explanations for heterogeneous findings at the two loci. In a large case-control study of the UK population there is weak evidence for association of *DAO* with schizophrenia, but strangely not with *DAOA*. The issue is further confused as there is evidence for an interactive effect between the two loci that increase significance, indicating epistasis may be required to increase risk in some populations as originally proposed by Chumakov *et al* (Chumakov, Blumenfeld *et al.*, 2002).

1.12.5. Regulator of G-protein Signalling 4 (RGS4)

The gene encoding Regulator of G-protein Signaling 4 is located within a reported linkage region at 1q22 (Owen, O'Donovan *et al.*, 2002). However, unlike the previously discussed genes this was not the major reason for assessment, but that of decreased expression of *RGS4* found in schizophrenic prefrontal cortex by microarray analysis (Mirnics, Middleton *et al.*, 2001). Genetic evidence for association of the *RGS4* locus with schizophrenia was found via family oriented analysis within two US samples

(Chowdari, Mirnics *et al.*, 2002). Although no significant association was discovered in a separate Indian sample, the combination of the 3 samples yielded increased significance of association of abundant SNPs and haplotypes situated at the 5' region of *RGS4* with a diagnosis of schizophrenia, although some incongruencies exist between the associated alleles and haplotypes even for the US samples (Chowdari, Mirnics *et al.*, 2002).

Support that *RGS4* is a candidate gene for schizophrenia susceptibility has surfaced from studies of UK and Irish populations. A large case-control sample from the UK was inspected, yielding modest single marker and 2-marker haplotype evidence for association (Williams, Preece *et al.*, 2004b). Morris *et al* (Morris, Rodgers *et al.*, 2004) provide more substantial evidence, replicating the exact same at risk haplotype reported by Chowdari *et al* (Chowdari, Mirnics *et al.*, 2002) in a case-control study of an Irish population. Further validation of these findings emerged from studying Irish families with a high incidence of schizophrenia, where modest single marker evidence and strong haplotypic association was presented (Chen, Dunham *et al.*, 2004).

A meta-analysis of thirteen studies, consisting of >2000 pedigrees and a >7000 individual case-control sample revealed weak overall haplotype evidence in the familial sample and a faintly positive SNP in the case-control sample (Talkowski, Seltman *et al.*, 2006). However, such a finding is, at least, promising especially if allelic heterogeneity within and between populations exists at this locus. An interesting observation is the reported association between a neighbouring gene, UHMK1 and schizophrenia (Puri, McQuillin *et al.*, 2006), discovered by association mapping of the linkage region and so the possibility remains that associations at RGS4 may be capturing an association with neighbouring genes.

1.12.6. Disrupted In Schizophrenia 1 (DISC1)

A study of a Scottish population for psychiatric illness identified an extended pedigree where there was a high incidence of major mental illness including bipolar disorder, major depression and schizophrenia (St Clair, Blackwood *et al.*, 1990). Cytogenetic analysis of individuals within the pedigree resulted in the discovery of a balanced autosomal translocation between chromosomes 1 and 11, t(1;11) (q43;q21) (St Clair, Blackwood *et al.*, 1990). A psychiatric diagnosis was verified for 16 of the 34 members harbouring the translocation contrasting with only 5 of the 43 without it. Therefore the chromosomal rearrangement was co-segregating with diagnosable mental illness more often than would be expected by chance.

Further evaluation of the pedigree (Millar, Wilson-Annan *et al.*, 2000) and disrupted regions revealed that of the core disturbed regions, the chromosome 11 region contained very few protein coding sequences, whereas the chromosome 1 region contained two novel genes termed Disrupted In Schizophrenia 1 and 2 (*DISC1* and *DISC2*) the genomic structure of which were subsequently determined (Millar, Wilson-Annan *et al.*, 2000). Remarkably this region has also been implicated from linkage studies of Finnish schizophrenics (Ekelund, Hennah *et al.*, 2004) and also intriguing, given the repertoire of mental illnesses co-segregating with disruption of this region, are reports of linkage of 1q42 to bipolar disorder (Macgregor, Visscher *et al.*, 2004). Independent analysis of the pedigree revealed robust evidence for linkage of the translocation to schizophrenia, bipolar and affective disorders, and also captivating was the observation of an altered P300 waveform in seemingly unaffected carriers of the

translocation as this is believed by some to be a possible marker or endophenotype for schizophrenia related disorders (Blackwood, Fordyce *et al.*, 2001).

Of the few reported studies of the *DISC1/2* locus the results are mixed, the first study attempted to gain association in the Scottish population using four SNPs and a microsatellite genotyped in a small case-control sample of schizophrenics (Devon, Anderson *et al.*, 2001). A resounding lack of association was achieved for individual markers and haplotypes, although some variants were present only in affecteds but were too rare to be tested via their methods. Interestingly, an independent Scottish case-control study found strong evidence for association at the DISC1 locus in a large case-control sample (Zhang, Sarginson *et al.*, 2006) and a further study of the Scottish population, which took in to account the LD structure across this locus in the same population, also noted evidence in favour of association at DISC1 (Thomson, Wray *et al.*, 2005).

A comprehensive family study of the Finnish population which analysed 28 SNPs covering the three candidate genes in the region (*TRAX, DISC1* and *DISC2*) found a significant under-transmission of some *DISC1* haplotypes to schizophrenics, an effect which was more pronounced for females (Hennah, Varilo *et al.*, 2003). This result is made more interesting by the substantial linkage this group have to 1q42 (Ekelund, Hennah *et al.*, 2004) and has also been substantiated by a replication of association signal at this locus in another Finnish sample (Cannon, Hennah *et al.*, 2005). Corroborative data has also emerged from a case-control studies of a US population, (Hodgkinson, Goldman *et al.*, 2004), as has a mapping study of 1q42 in a Taiwanese familial sample (Liu, Fann *et al.*, 2006). However, of note are two Japanese case-control studies which examined regions of the *DISC1* gene and found no evidence for association (Kockelkorn, Arai *et*

al., 2004, Zhang, Tochigi *et al.*, 2005), a Chinese familial study that only showed a weak association in females (Chen, Chen *et al.*, 2006) and a reasonably sized US Caucasian sample showing no support for the role of *DISC1* in schizophrenia susceptibility (Wood, Pickering *et al.*, 2006).

The above studies have analysed the *DISC1* locus looking for relatively common variants of a reasonable effect size, however this locus was not discovered by association mapping and so may not harbour such alleles. Studies taking a different approach have yielded interesting findings, with a rare frameshift mutation of *DISC1* found in a multiply affected pedigree (Sachs, Sawa *et al.*, 2005), however this was not found to be associated with disease in a large case-control study (Green, Norton *et al.*, 2006) although even a sample of this size may not be suitably powered. Therefore, the best course of action for putative functional, rare alleles in *DISC1* is to demonstrate some biological consequence of carrying the given allele, a consequence that can be directly linked to schizophrenia aetiology. Such studies are underway and yielding results implicating DISC1 mutations in neurite growth (Kamiya, Kubo *et al.*, 2005) for example. However, such results can only be confirmed by large scale studies of many schizophrenics to either test for association of the allele, or alleles with the same biological outcome (for example in terms of the Scottish pedigree, who may be the only carriers of that mutation in the human population).

1.13. The Search for a Schizophrenia Susceptibility Locus on Chromosome 17

As part of a whole-genome linkage study of schizophrenia in a UK and Irish population, Williams *et al* (Williams, Norton *et al.*, 2003) noted a large multiply affected pedigree (C702) that explained almost the entirety of their suggestive linkage signal to chromosome 17. On closer inspection of pedigree C702 in isolation, all 6 affected male siblings shared a large paternal and maternal haplotype across the majority of the chromosome. Linkage analysis of this pedigree in isolation resulted in genome-wide significant evidence for linkage to chromosome 17, with no evidence for linkage elsewhere in the genome (Williams, Norton *et al.*, 2003). Therefore, the pedigree is co-segregating a chromosomal region with schizophrenia in a Mendelian fashion and thus presents the possibility of a highly penetrant schizophrenia disease variant (in this pedigree).

The initial aim of this thesis is to refine the genome-wide significant linkage region in pedigree C702 to make the linked region amenable to more detailed molecular genetic analysis. Refinement of the linkage region will allow other techniques to be applied to the linked region, such as high-density CNV mapping, directed re-sequencing and eventually full re-sequencing. Furthermore, directed re-sequencing and association analysis of candidate genes within the C702 linkage region will be conducted in an attempt to identify the allele(s) causing schizophrenia in family C702. Association analysis of any interesting loci will be conducted in large association samples with schizophrenia and related phenotypes in an attempt to verify that the locus is a psychiatric disease susceptibility locus. If possible, any putative disease alleles will be analysed for a

molecular function that would indicate the biological mechanism of schizophrenia in this exceptional pedigree.

Chapter 2. Materials and Methods

2.1. DNA Samples

2.1.1. Case-control and Familial DNA Samples

All samples used in the studies reported here provided written, informed consent to participate in genetic studies. Protocols and procedures for sample ascertainment and further scientific research were all approved by relevant ethics committees.

UK Schizophrenia Cases

All schizophrenia cases were unrelated Caucasians born in the UK or Ireland. All 709 cases (483 males and 226 females) met DSM-IV criteria for schizophrenia with consensus diagnosis being determined following a semi-structured interview by trained psychiatrists and psychologists using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) (Wing, Babor *et al.*, 1990) and examination of case notes. Pedigree C702 was collected as part of a linkage study (Williams, Norton *et al.*, 2003) and an affected sibling is included in the above association sample. This case sample was used for the construction of DNA pools (see section 2.1.4).

The cases used in a whole-genome association study of schizophrenia (O'Donovan *et al*, submitted) were 476 individuals from the above 709 schizophrenia case sample (321 males and 155 females).

UK Bipolar Disorders Cases

All bipolar disorder cases were unrelated Caucasians born in the UK or Ireland. All 702 bipolar disorder I cases met DSM-IV criteria (BPI. 257 males and 441 females) with consensus diagnosis being determined following a semi-structured interview, analysis according to SCAN (Wing, Babor *et al.*, 1990) and examination of case notes. Also analysed in this study were individuals with diagnoses of schizoaffective bipolar disorder (SABP. N=78) and individuals with bipolar disorder II (BPII. N=20), psychosis not-otherwise-specified (PNOS. N=24) and a further 28 individuals with bipolar related phenotypes.

UK Unipolar Depression Cases

All unipolar depression cases were unrelated Caucasians born in the UK or Ireland, and represents a collaborative collection effort of research groups in Cardiff and London (Institute of Psychiatry). All 992 unipolar depression cases (298 males and 656 females) met DSM-IV criteria for unipolar depression with consensus diagnosis being determined following a semi-structured interview, analysis according to SCAN (Wing, Babor *et al.*, 1990) and examination of case notes.

UK Post-Natal Depression Cases

All post-natal depression cases were unrelated Caucasians born in the UK or Ireland. All 98 female cases met DSM-IV criteria for post-natal depression with consensus diagnosis being determined following a semi-structured interview, analysis using SCAN (Wing, Babor *et al.*, 1990) and examination of case notes.

UK Blood Donor Controls

Blood-donor controls (N=2220. 971 males, 965 females, 284 with no gender information) were all Caucasian, from the United Kingdom and the majority (N=1470) were obtained directly from the UK National Blood Service. This sample is made of three control samples matched for age, sex and ethnicity to the schizophrenia case samples (control N=716), bipolar disorders sample (control N=754) and unipolar depression sample (control N=750). The controls matched to the schizophrenia and bipolar disorders samples were not specifically screened for psychiatric illness, but individuals were not receiving regular prescribed medications for psychiatric illness. Controls matched to the unipolar depression sample were screened for evidence of psychiatric illness and obtained through collaboration (Institute of Psychiatry, London). In the United Kingdom, blood donors are not paid and therefore the sample is not over-representative for impoverished or socially disadvantaged persons. Samples were collected in Wales, the

Midlands (England) and controls used for the unipolar depression sample were obtained from London through collaboration (Institute of Psychiatry, London).

UK 1958 Control Birth Cohort

A further sample of 1058 UK control subjects was obtained from the Wellcome Trust; a population control cohort (529 males, 529 females) using a sub-set from 17,000 individuals born in England, Scotland and Wales on one day in 1958. For these individuals, their sex and their vague ethnic origin we the only details disclosed.

Wellcome Trust Case-Control Consortium (WTCCC) control samples

Whole genome-association data for a sample of 3000 controls from the WTCCC study of common diseases was utilised as part of this research (Wellcome Trust Case Control Consortium, 2007). This sample included 1500 individuals from the 1958 Control Birth Cohort study outlined above (750 males and 750 females). Another set of 1500 controls from the UK was obtained from the UK National Blood Service, with 750 male and 750 female samples. After quality control procedures to identify degraded samples, duplicated samples, samples with a large amount of missing genotype data and samples with non-Caucasian ancestry (Wellcome Trust Case Control Consortium, 2007) 2938 control samples had analysable data (1446 males and 1492 females)

Bulgarian Trios Sample

A sample of parent-proband trios was recruited in Bulgaria. The recruiting clinicians performed a semi-structured interview with an abbreviated version of the SCAN (Wing, Babor *et al.*, 1990) instrument. Consensus best-estimate diagnoses were made according to DSM-IV criteria by two psychiatrists, based on all available information, which included in each case discharge summaries from hospital admissions, an interview using SCAN (Wing, Babor *et al.*, 1990), and a summary prepared by the doctor in charge. The Bulgarian proband-parent trio sample consists of a total of 957 patients with a psychiatric disorder diagnosis and their parents from 952 families. The majority of probands had diagnoses of DSM-IV schizophrenia (N=598), however a proportion have bipolar I disorder (N=170) or schizoaffective disorder (N=105).

US National Institute for Mental Health (NIMH) Autism Trios

The NIMH autism proband-parent trio sample consisted of 95 patients and their parents. The probands met a diagnosis of ICD-10 autism and the sample was obtained with permission from the NIMH (United States of America).

Irish Schizophrenia Case-Control Sample

A schizophrenia and schizoaffective disorder case-control sample from the Republic of Ireland was made available for direct DNA analysis through collaboration

(Derek Morris and Mike Gill. Trinity College, Dublin, Ireland). All schizophrenia cases were unrelated Caucasians of Irish nationality. All 76 schizoaffective, 296 schizophrenia and 1 schizophreniform case (223 males and 127 females) met DSM-IV criteria. The sample also included 812 blood donor controls from the Republic of Ireland health services, consisting of 518 males and 288 females.

German Case-Control Sample

A schizophrenia case-control sample from Germany was made available for analysis of data only, through collaboration (Dan Rujescu. Department of Psychiatry, Ludwig Maximilians University, Germany). All schizophrenia cases were unrelated Caucasians of German nationality. All 513 schizophrenia cases (334 males and 179 females) met DSM-IV and ICD-10 criteria for schizophrenia. The sample also included 1332 controls randomly selected from the Munich area of Germany, consisting of 609 males and 721 females.

Pooled DNA samples

The UK schizophrenia case-control sample has 709 cases and 716 controls. Of this sample, 709 cases and 710 controls were used for construction of DNA pools (see section 2.1.4). The case and control sample were divided into four matched stages; the first three stages consist of 184 cases and 184 controls, while stage four has 157 cases and 158 controls. For details on pool construction see section 2.1.4.

2.1.2. DNA Extraction and Storage

All DNA samples ascertained in the UK or Bulgaria were obtained as high molecular weight fractions from lymphocytes as acquired from blood intravenously, or from buccal cavity epithelial cells via mouthwash or from Epstein-Barr-Virus (EBV) transformed β-lymphoblasts, originally obtained from intravenous blood. In each case DNA was extracted using standard phenol-chloroform procedures followed by ethanol precipitation. All DNA samples used in this study were extracted by other members of the Department of Psychological Medicine research group (Cardiff University) or by members of our collaborative research groups (except EBV transformed lymphoblastoid samples from pedigree C702, extracted by myself). All stock and diluted samples are stored at -20°C in water or TE buffer.

2.1.3. DNA/RNA Quantification and Assessment

The quality and quantity of DNA was measured using a spectrophotometer (Beckman DU 640B Spectrophotometer). An aliquot of a DNA sample was first diluted to 5%, and then the absorbance (A) of UV light at wavelengths (λ) of 260nm and 280nm was calculated. Assuming that an A (260nm) of 1 is equivalent to 50μg of DNA, a ratio of A (260nm) to A (280nm) of above 1.8 indicated a suitable amount of clean DNA without contaminating RNA and protein. The spectrophotometer calculates a concentration (in ng/μl) for the sample based on the UV absorbance.

For more accurate quantification of DNA samples were analysed using Pico Green (Invitrogen) and a fluorimeter (Fluoroskan Ascent. Thermo Labsystems). The Pico Green reagent specifically interacts with double-stranded DNA only and is therefore more accurate than DNA quantification by spectrophotometry. Samples were first diluted to less than 50ng/µl (typically 4ng/µl) in sterile water, based on the concentration calculated using the spectrophotometer. Then the sample is diluted to 1% in a 1x TE buffer in a white 96 well labsystems cliniplate (Thermo Labsystems). A working Pico Green solution is produced by adding 5µl Pico Green at 200x to 995µl of 1x TE buffer. The fluorimeter dispenses 100µl of Pico Green working dilution into each sample and measures the DNA concentration using a UV excitation wavelength of 485nm and an emission wavelength of 538nm. The DNA is quantified by comparison to the gradient of a standard curve, which can be generated each time by the user using the calibrant standard provided, or compared to a standard curve generated recently (Dr. Nadine Norton).

For measuring RNA concentration an Amersham Ultrospec 2100 pro UV/Visible spectrophotometer was used.

2.1.4. DNA Pool Construction

Each DNA sample was quantified using Pico Green using the same standard curve (as section 2.1.3) and samples diluted to 4ng/µl ±0.5ng/µl. The construction of separated case and control pools was performed using a Microlab S (Hamilton) automated sample processor. Samples for DNA pools were quantified by multiple

personnel (including myself) and pools were constructed by Anna Preece and Nadine Norton (Norton, Williams *et al.*, 2002).

2.2. Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR) is a means of amplification of a specific DNA sequence that lies between two regions of know sequence. The technique is performed *in vitro* and is dependent upon the thermostable enzyme *Thermus Aquatics* (Taq). Taq polymerase performs the synthesis of the complementary DNA strand from a denatured DNA template in the presence of a suitable buffer containing $MgCl_2$, deoxy nucleotide triphosphates (dNTPs) and two oligonucleotide primers.

The oligonucleotide primers represent the known sequence and are designed to flank the region to be amplified. The Taq polymerase requires the primers as a double stranded initiation point for 5' to 3' synthesis of DNA. The PCR reaction involves the denaturation of the double-stranded template DNA at a high temperature to produce single strands, the annealing of the oligonucleotide primers to their complimentary and the enzymatic extension of the DNA product. The process of denaturation, annealing and extension is repeated in a cyclical fashion where the template and synthesised product is successively re-amplified for approximately 30-40 cycles.

Oligonucleotide PCR primers used in this study were typically designed *in silico* using the online program Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Oligonucleotides used in this thesis were synthesised by

Invitrogen, Sigma-Aldrich and Eurogentec. As with the design of any oligonucleotide primer, the sequence must ideally not overlap with any polymorphic.

2.2.1. PCR Optimisation

Three types of Taq polymerases were used in this study for PCR purposes: HotStarTaq (Qiagen), GC Rich System (Roche) and Titanium Taq (BD Biosciences). The latter was only used for Amplifluor based genotyping (see section 2.5.5).

The Qiagen HotStarTaq based PCR was typically performed in a 12µl volume using 3µl of genomic DNA (4ng/µl), 0.28µl of each primer (5pmol), 0.96µl dNTPs (5mM) each, 1.2µl buffer (10x) provided with the kit, 6.22µl of water and 0.06µl of Taq (10U/µl). The advantage of the HotStarTaq is that it has a chemically modified site that prevents activity until it has been heated for 15 minutes at 95°C. This prevents non-specific elongation by primer binding as the temperature increases, therefore annealing will only occur at the correct temperature as the reaction cools.

The PCR cycling conditions used comprise of either a standard three step method as outlined above, or a "touch-down" PCR:

Standard PCR.

1. 94°C-96°C for 15 minutes
2. 94°C-96°C for 20 seconds
3. Tm°C for 20 seconds

4. 72°C for 30-45 seconds
5. Repeat steps 2-4 for 35-45 cycles
6. 72°C for 10 minutes
7. 15°C for 10 minutes

Where Tm°C is the primer set annealing temperature.

Touch-down PCR

1. 94°C-96°C for 15 minutes
2. 94°C-96°C for 5 seconds
3. Tm°C + 6°C for 5seconds (-0.5°C per cycle)
4. 72°C for 10 seconds
5. Repeat steps 2-4 for 11-15 cycles
6. 94°C for 5 seconds
7. Tm°C for 5 seconds
8. 72°C for 10 seconds
9. Repeat steps 6-8 for 19-25 cycles
10. 72°C for 10 minutes
11. 15°C for 10 minutes

There are DNA sequences that have GC content to the extent that the HotStarTaq has difficulties amplifying the sequence due to the difficulty in denaturing the strong GC secondary structures that form. The GC Rich PCR system (Roche) involves a Taq

polymerase and proofreading enzyme plus a number of additives in the buffer including DMSO (Di-methyl Sulphoxide), 7-deaza dGTP, formamide and glycerol that facilitate the PCR amplification of GC rich DNA sequences. PCR reactions were performed in 12µl volumes with 3µl of genomic DNA (4ng/µl), 0.28µl of each primer (5pmol), 0.96µl dNTPs (5mM each), 2.4µl of provided reaction buffer, 1.2µl of resolution buffer, 0.8µl water and 0.06µl of provided Taq mix. The PCR conditions were as follows:

GC Rich PCR System

1. 95°C for 3 minutes
2. 95°C for 20 seconds
3. Tm°C for 20 seconds
4. 72°C for 30 seconds
5. Repeat steps 2-4 for 35-45 cycles
6. 72°C for 10 minutes
7. 15°C for 10 minutes

The PCR reactions for genotyping methods such as the Sequenom hME and iPlex chemistries and Amplifluor will be described in sections 2.5.5 and 2.5.7.

2.3. Agarose Gel Electrophoresis

DNA fragments can be fractionated according to their size when a potential difference is applied through a porous substance such as an agarose gel, as the negative

phosphate groups allow the movement of DNA towards the anode. Analysis of pre- or post-PCR samples was performed using 0.5-2% agarose gels, dependent upon fragment size and resolution required. To construct a 1% agarose gel for electrophoresis 1g of agarose (Sigma-Aldrich) was dissolved in 100ml 0.5x TBE buffer (Ultra pure electrophoresis sequencing grade, National Diagnostics). The mixture was heated until the solution was clear, and then cooled and 1µl Ethidium Bromide solution (10mg/ml) added. The solution was then poured into a gel-former and appropriate well formers added before allowing the gel to cool and form a solid.

To run samples on the gel an appropriate volume of PCR product was mixed with a loading buffer (6x loading buffer: 15% Ficoll, 0.25% Bromophenol blue, 0.25% Xylene cyanol, in water) and placed in a formed well. An aliquot of size-standard (e.g. 1kb plus DNA ladder, Invitrogen) was also run alongside samples to allow size delineation. The gel and sample are then run at between 100-120V in an electrophoresis tank for an appropriate amount of time to see the DNA size expected.

Samples assayed by agarose gel electrophoresis were visualised using a UV transilluminator (UVP) and photographs taken using an attached Kodak Electrophoresis Gel documentation and analysis system.

2.4. Sample Processing

For large scale PCR and post-PCR reactions involving many DNA samples, reagents and samples were aliquoted using robotic liquid handling systems. DNA samples are typically present at intermediate (40ng/µl) and working (4ng/µl) dilutions in

shallow-well DNA boxes (ABgene). These are amenable to processing using the Beckman-Coulter FX and Hydra microdispensers. DNA samples were aliquoted into suitable microtitre plates (ABgene). The Beckman-Coulter FX liquid handler can also dispense PCR reagents and was used to this effect. All programs for use with the Beckman-Coulter FX were written by Sarah Dwyer, and programs for the Hydra microdispenser were constructed by myself.

2.5. Genotyping and Sequencing Methods

2.5.1. Microsatellites

Microsatellite genotyping was performed using a 5' fluor-labelled oligonucleotide PCR primer set (where one primer is labelled and the other unlabelled). The fluorescent labels used were FAM, HEX and NED, which fluoresce at different wavelengths. The fluorescent labelled PCR products were analysed using an automated ABI3100 PRISM Genetic Analyser (Applied Biosystems). As with all samples analysed using the ABI3100 PRISM Genetic Analyser, samples are loaded by electrokinetic injection. Then the samples (in this case fluorescent labelled PCR) are electrophoretically moved through a polyacrylamide filled capillary where the fluorophore is excited by a laser; the resultant fluorescence is measured by a Photo-Multiplier-Tube. The PCR samples were run through a 36cm long capillary using POP4 polyacrylamide. Fluorescent size standards were added to each of reactions post-PCR. The size standards used were ROX-400HD or ROX-500. The raw data and analysis of the fluorescent fragments was performed

automatically using Genescan Analysis v3.7 v3.7(Applied Biosystems) software. This data was then analysed using Genotyper software (Applied Biosystems), which allows the user to delineate the size of the fluorescently labelled product and judge whether alleles are present for size-polymorphic markers.

Panels of microsatellites were designed; each panel consisted of microsatellite markers where the size of the marker (and all allelic variants) did not overlap with the size of any other marker (and that markers alleles) labelled with the same fluorophore. All microsatellite marker PCRs used in this study had a size greater than 100bp and smaller than 400bp. Each fluorescent PCR was performed separately and the independent products pooled at an appropriate ratio to allow genotyping to be performed. Post-PCR samples were often diluted in water before 4μl of each individual product or the pool of individual diluted PCRs was added to 8μl HiDi formamide (Applied Biosystems).

Genotype results were exported to a Microsoft Excel file for further analysis. If alleles were genotyped in samples that contained water instead of DNA then a PCR contamination had occurred and the experiment was re-performed. Experiments were also excluded for markers where more than two alleles were present, and the experiment re-performed. Markers showing >2 alleles and/or non-mendelian errors in the families studies were excluded.

Microsatellites analysed in this study were selected from the following sources: Marshfield genetic map (http://research.marshfieldclinic.org/genetics/home/index.asp) , deCODE genetic map (Kong, Gudbjartsson *et al.*, 2002), and the UCSC genome browser track "STS markers". Markers were primarily selected based on their

polymorphism using the Marshfield and deCODE measurements of heterozygosity and number of alleles.

2.5.2. Sequencing

All sequencing was performed using the fluorescent Sanger sequencing method via Big Dye termination chemistry (Applied Biosystems) and analysed using the ABI3100 PRISM Genetic Analyser. The fluorescent sequencing reaction involves obtaining large amounts of template DNA (i.e. by PCR) and then the random incorporation of four fluorescently labelled ddNTPs (di-deoxy-dinucleotide-triphosphates. ddATP, ddCTP, ddGTP and ddTTP) that terminate after extending one base during primer extension. This produces a series of DNA fragments where the chain growth has been terminated at each successive position. When electrophoresed in a capillary sequencer and detected by a laser the each base of the sequence will be fractionated by size and fluoresce according to the base at that size.

There were two different protocols employed in this study; the majority of sequencing was performed by a user intense manual process using Millipore consumables, the other method was the more automated Agencourt protocol. Both stages involved PCR of the sample to be sequenced in a 12μl reaction.

Millipore Based Sequencing

PCR clean-up

Prior to sequencing, PCR products were cleaned of excess primers, dNTPs and Taq. The PCR clean-up stage entailed the use of a vacuum manifold and a Multiscreen 96 well PCR clean-up plate (Millipore). The PCR product was transferred to a well in the plate and the plate placed upon the vacuum manifold. After 10 minutes at full vacuum pressure (~1000mbar) all contaminants are removed. The clean PCR product is removed from the membrane via the addition of 25µl of water and vigorous aspirating with a pipette.

Sequencing reaction

Big Dye version 3.2 sequencing (Applied Biosystems) involves four individually fluorescently labelled dideoxynucleotides (ddNTP), collectively called BigDye terminators, a sequenase enzyme and one oligonucleotide primer. The sequencing reaction involves 5µl of cleaned PCR product, added to a reagent mix consisting of 2µl of BigDye termination mix, 2µl of BigDye sequencing buffer and 1µl of either the forward or reverse orientation oligonucleotide primer (usually the PCR primer) at 3pmol/µl, to initiate primer extension. The cyclic sequencing reactions were performed on an MJ thermocycler using the following conditions:

1. 96°C for 2 minutes
2. 96°C for 30 seconds
3. 55°C for 15 seconds
4. 60°C for 4 minutes
5. Repeat steps 2-4 for 24 cycles
6. 4°C for 4 minutes
7. 15°C forever

Post-sequencing clean-up

Removal of excess ddNTPs and other sequencing contaminants was performed using a Multiscreen 96-well filtration plate (Millipore) and Sephadex 500 (Sigma). A small amount of Sephadex was added to each of the filtration plate wells using the adaptor provided. 300µl of water was added to each well and the plate left for >3 hours. The plate is then centrifuged at 2500xg for 5 minutes to remove excess water. The sequencing reaction product is made up to 20µl with water and added to a well of the Sephadex laden filtration plate. The plate is then spun for 10 minutes at 2500xg and the cleaned product collected in a plate placed underneath the filtration plate. The cleaned product is then dried down at 65°C and 10µl HiDi formamide added before analysis on the ABI3100 PRISM Genetic Analyser.

Agencourt Based Sequencing

Agencourt based sequencing was employed as more automated method of sequencing to improve consistency, reduce the possibility of error and reduces user involvement time. The method was automated via a Beckman-Coulter NX liquid handler (programs written by Sarah Dwyer).

PCR clean-up

The Agencourt method removes unincorporated dNTPs, primers, DNA polymerase and salts. The PCR product (12µl) is mixed with AMPure (Agencourt) reagent (21.6µl), containing metallic beads which adhere to the amplimers. The solution and products not adhered to the magnetic beads are washed off and removed using successive 85% ethanol wash steps. The PCR amplimers are then eluted in 195µl of pure water into a new plate.

Sequencing reaction

The cleaned PCR product was added in a 5µl volume to a 5µl reagent mix consisting of: 0.166µl of BigDye termination mix, 1.917µl water, 1.917µl BigDye sequencing buffer and 1µl of primer (3pmol/µl). The sequencing reaction was performed as for the Millipore based sequencing protocol.

Post-sequencing clean-up

Post –sequencing clean up employs a CleanSEQ chemistry protocol (Agencourt) that is semi-automated. The process removes unincorporated dye terminators and further contaminants. Sequencing reaction product (10μl) is added to CleanSEQ reagent (10μl) and 85% ethanol (41.59μl) and aspirated to mix. The sequencing product binds to the magnetic beads contained in the CleanSEQ reagent. The non-bound sequencing reaction contaminants are washed off and removed in successive 85% ethanol wash steps. The cleaned sequencing product can then be eluted in pure water (75μl) and is ready for analysis using a capillary sequencer.

2.5.3. Sequencing Analysis

The raw data generated by the ABI3100 PRISM Genetic Analyser is automatically analysed by Sequence Analysis Software (Applied Biosystems). The software calls the fluorescence at each nucleotide as the corresponding base. Sequencher software (Gene Codes) was then used, which aligns multiple sequencing traces and allows comparison with a reference sequence. The software then highlights where differences occur between the traces, allowing the user to manually inspect these differences and judge whether a polymorphism exists.

2.5.4. Allele Frequency Estimation and Genotyping Using Fluorescent Single Base Primer Extension

A polymorphism which varies at one particular nucleotide (for example a SNP) can be genotyped in an individual or pooled DNA sample (Norton, Williams *et al.*, 2002) via oligonucleotide primer mediated extension of a single fluorescently labelled ddNTP using SNaPshot chemistry (Applied Biosystems). The SNaPshot reaction entails the PCR of the samples of interest, which are then cleaned before primer extension by a single fluorescent ddNTP corresponding to the next 3' base (fluorescent ddATP, ddCTP, ddGTP and ddTTP). This is followed by another clean up step to remove excess ddNTPs and then analysis using an ABI3100 PRISM Genetic Analyser (Applied Biosystems) followed by manual inspection using Genotyper software (Applied Biosystems). Examples of SNaPshot genotypes can be found in appendices 10.4. The degree of fluorescence of each fluorescent base is directly proportional to the amount of the allele present in each sample, and therefore the procedure is amenable to allele frequency estimation in a DNA pool, allowing case-control studies to be performed in a few samples rather than genotyping every sample individually.

Oligonucleotide extension primers were designed using the internet based algorithm FP Primer (http://m034.pc.uwcm.ac.uk/FP_Primer.html. Dobril Ivanov).

SNaPshot individual genotyping.

The sample to be genotyped undergoes a standard 12μl PCR reaction. The PCR is then cleaned to degrade unincorporated dNTPs and unextended primers by addition of Shrimp Alkaline Phosphatase (SAP) (Amersham) and exonuclease I (Amersham) respectively. The reaction involves the addition of: 0.5μl SAP (1U/μl), 0.1μl exonuclease I (10U/μl) and 4.5μl water to each 12μl PCR product. The reaction conditions were as follows:

1. 37°C for 1 hour
2. 80°C for 15 minutes

The samples are then ready to undergo SNaPshot primer extension. An 8μl reaction mix consisting of 1.25μl SNaPshot reagent (containing fluor-labelled ddNTPs and a sequenase), 3.75μl reaction buffer, 2μl water and 1μl of extension primer diluted to an appropriate concentration was added to 2μl of the cleaned PCR product. Typically extension primers were used at 0.5pmol/μl although this can be altered to gain optimum peak heights using the equation c=Y'/(YX), where Y' is the required peak height (typically 3000 fluorescence intensity units as displayed by Genotyper software (Applied Biosystems)), Y is the initial peak height and X is the initial primer concentration (Norton, Kirov *et al.*, 2002). The following reaction was then executed:

1. 96°C for 2 minutes
2. 96°C for 5 seconds
3. 43°C for 5 seconds

4. 60°C for 5 seconds
5. Repeat steps 2-4 for 24 cycles

A further stage of SAP clean-up is performed to degrade the unincorporated ddNTPs. A 5µl reaction mix comprised of 0.5µl SAP and 4.5µl water is added to the SNaPshot reaction product and the same conditions as for the SAP PCR clean-up performed. After the reaction is finished, 3µl of the product is added to 10µl of HiDi formamide and the sample is ready for analysis. The samples are run through a 36cm capillary using POP4 polyacrylamide (Applied Biosystems) without any internal size standard. The raw data is analysed using the Genescan Analysis v3.7 software (Applied Biosystems) and subsequently imported into Genotyper software (Applied Biosystems).

The Genotyper software allows the discrimination of the correct alleles and the amount of that allele present in a sample (as indirectly measured via the "peak height" of the fluorescence). The amount of each allele present is given a numerical value (arbitrary absorbance units) and can be placed into a Microsoft Excel file for further analysis. Individual samples can then be genotyped as homozygous or heterozygous based on the presence of the fluorescence of the corresponding nucleotide. Where alleles were detected in water samples the experiment was re-performed as a contamination had occurred.

SNaPshot DNA pool case-control analysis

Allele frequencies are estimated in a DNA pool using the SNaPshot chemistry just as for an individual sample (Norton, Williams *et al.*, 2002); however the fluorescence now corresponds to the total fluorescence of many different individuals' alleles. DNA Pools were constructed from our large case-control sample of DSM-IV diagnosed schizophrenics and controls as section 2.1.4.

Once the pools have been prepared the procedure involves the duplicate PCR of each case and control pool as for SNaPshot individual genotyping. The remaining steps are also the same as for the SNaPshot individual genotyping method above. The resultant project files from the ABI3000 Genetic Analyser (Applied Biosystems) are analysed using Genescan software (Applied Biosystems) and input into the software Genotyper (Applied Biosystems). Fluorescence intensity for each allele can then be recorded for each pool and the duplicates and these results imported into an excel file for further analysis. Allele frequencies are then estimated using the following formula:

$$\text{Allele Frequency} = A / (A + (k \times B))$$

Where: A = Peak height of Allele 1 (allele frequency to be estimated)

B = Peak height of Allele 2

K = Average ratio of alleles from a heterozygote

(Peak height Allele 1/Peak Height Allele 2)

The value of k can also be estimated when actual allele frequency is known using the formula:

$$k = (F-F*f)/f-F*f)$$

Where F is the observed frequency of allele A as calculated from formula below, and f is the actual frequency of allele A (e.g. CEPH individuals genotyped in the HapMap project or other individual genotyping data):

Peak Height Allele A/ (Peak Height Allele A + Peak Height Allele B)

Once the value of k is known it can be applied to the original formula for allele frequency estimation using k.

The calculation is applied to each pools sample, then both case and control pool frequencies are combined and a mean taken. The allele frequencies are then converted to allele counts by multiplying by the number of chromosomes present in the case and control groups. The result allele counts are placed into a 2 x 2 contingency table for the derivation of a chi square and corresponding p value for the association of that allele with schizophrenia.

Allele frequencies can also be artificially estimated over a range of values for k (0.25, 0.32, 0.47, 0.88, 1, 1.14, 2.13, 3.1, 4) to account for any problems in estimating k from heterozygotes or if heterozygotes were unavailable. All pooling allele frequency estimation and case-control analysis was performed using Microsoft Excel macros designed by Nigel Williams and George Kirov. When analysing peak heights, values were excluded for heights below 100 and above 6000, where the experiment was

repeated. Experiments were also repeated when alleles were detected in samples that should have no DNA, as a contamination had occurred. DNA pooling experiments that resulted in an allelic association p value of <0.1 were repeated.

2.5.5. Individual Genotyping Using Amplifluor UniPrimer Chemistry

The Amplifluor UniPrimer assay involves the use of 5 separate primers. The reaction involves an allele specific PCR where there is a universal anti-sense reverse PCR oligonucleotide primer and two sense forward primers which differ at their 3' ends corresponding to the complementary base of the SNP alleles. The forward primers also differ at the 5' end of the oligonucleotide, where each allele specific primer has a ~20bp stretch of nucleotides complimentary to nucleotides of two universally fluorescently labelled primers (UniPrimers). Each of the uniprimers is labelled either with a green or red fluorophore, however they do not fluoresce due to a hairpin structure that brings a quencher into contact with the fluorophore.

The unlabelled, allele-specific primers initiate a competitive allele-specific PCR and these allele specific amplimers serve as templates for the binding of the universal fluorescent labelled primers (UniPrimers). The incorporation of the uniprimer into the allele-specific amplimer displaces the hairpin structure of the uniprimer, releasing the fluorophore from its quencher and generating fluorescence. The resultant levels of red or green fluorescence can distinguish the levels of each allele.

The assays are designed using the internet based software Amplifluor AssayArchitect (https://apps.serologicals.com/AAA/). The design process entails the

input of the SNP details and the flanking DNA sequence, or the specific rs number for the polymorphism in question into the design software. The software then designs the two forward and one reverse primers.

The polymorphism specific pcr primer mix for the reaction is as follows (where primers were obtained at 100pmol/µl):

Forward primer (Allele 1)	2.5µl
Forward primer (Allele 2)	2.5µl
Reverse primer	25µl
Water	470µl

The reagent mix for the reaction is as follows (using the polymorphism specific primer mix made above):

10x Titanium Taq Buffer (BD Biosciences)	0.5µl
dNTPs (2.5mM each)	0.4µl
Primer mix	0.07µl
SR labeled primer (BD Bioscience)	0.07µl
FAM (BD Bioscience)	0.07µl
Titanium Taq Polymerase (BD Biosciences)	0.05µl
Reaction Mix S (BD Biosciences))*	0.0625µl*
Water	3.84µl (3.215µl with Mix S)

*Reaction Mix is an optional additive which can improve the assay e.g. genotype clusters.

.

The total volume of the reaction mix is 5µl. This is added to 12ng of dried sample DNA in a black 96 well or 384 well microtitre plates (ABGene). For large scale genotyping the reaction ix was added using a Beckman-Coulter FX liquid handler. The cycling conditions for the reaction are as follows (using an MJ thermocycler):

1. 96°C for 4 minutes
2. 96°C for 10 seconds
3. 58°C* for 5 seconds
4. 72°C for 10 seconds
5. Repeat steps 2-4 for 20 cycles
6. 96°C for 10 seconds
7. 55°C for 20 seconds
8. 72°C for 40 seconds
9. Repeat steps 6-8 for 15-20 seconds*
10. 68°C for 7 minutes
11. 15°C for 10 minutes

*These steps require optimisation for each specific polymorphism reaction

The first stage of the reaction conditions (steps 1-5) involves the denaturation of the target DNA sample, the annealing of the allele specific primers and the elongation of the fragments which include the complimentary tails for the universal UniPrimer. The second stage (steps 6-11) involves the denaturation of the PCR product, the annealing of the fluorescently labeled primers to the complimentary sequence in the PCR product. The fluorescently labeled primers form a hairpin structure in the unbound state and the fluorescence prevented. When the allele specific PCR product binds to the complimentary sequence the fluorescence is allowed.

The fluorescence of each sample is analysed using an Analyst HTS Assay Detection Platform (LJL Biosystems) at the wavelengths shown below for each fluorophore:

FAM	Excitation at 485nm	Emission at 520nm
SR	Excitation at 580nm	Emission at 620nm

The results are given in signal intensity for the two fluorophores which can be plotted on a graph program (https://apps.serologicals.com). The clusters of sample fluorescent points correspond to three genotype classes which can be assigned manually be the user. The output of the graph program is in the form of 11, 12 or 22 where 1 corresponds to the FAM allele and 2 the SR allele.

2.5.6. Sequenom MassARRAY: Homogenous MassEXTEND and iPlex Genotyping Platforms

The Sequenom MassARRAY genotyping system allows the highly accurate genotyping of simple polymorphisms by combining primer extension chemistry with MALDI-ToF (Matrix Assisted Laser Desorption Ionisation – Time of Flight) Mass Spectrometry (MS). There are two different types of chemistry used for genotyping by MassARRAY in this study: homogenous MassEXTEND (hME) and iPlex. Both involve primer extension over the polymorphism of interest and the examination of the mass of the extended product to discern the genotype of a sample. Results are collated and can be analysed using the software Typer (Sequenom). The main advantages of the MassARRAY genotyping systems are their accuracy and the high assay multiplexing level.

hME MassARRAY Genotyping

The initial step of MassARRAY genotyping involves the design of the multiplex assay; this is performed using the Sequenom Assay Design software. For each polymorphism to be genotyped, the flanking DNA sequence is obtained and DNA sequences or variants that may confound any assay are highlighted (e.g. known SNPs and repetitive sequence) to prevent assay design over these regions. The highest multiplex assay possible is then designed from this information using Sequenom MassARRAY Assay Design software, which details PCR primers and an appropriate MassEXTEND

extension primer in an output file (.trs) on design completion. The design software adds a 10bp non-specific "tag" sequence to the 5' end of PCR primers where necessary, to ensure they are detected later in the MALDI-ToF mass spectrum (and therefore prevent the need for an exonuclease step). The software designs the PCR oligonucleotides to create the shortest amplimer possible to allow efficient PCR, and the primers are to have an annealing temperature of as close to 56°C to match the universal PCR conditions (below).

MassEXTEND primers are specifically designed to extend over the polymorphic base, and with the correct reagent mix to extend one base past the SNP before termination. This allows un-extended and extended primer to be separated by sufficient mass for accurate genotyping of alleles. All SNPs within a multiplex must display the same alleles at the SNP and extra extended base, as each type of SNP sequence requires its own reagent which has the correct two dNTPs and ddNTP, to allow the extension of the dNTP or ddNTP over the polymorphic base then the extension of the other ddNTP. The MALDI-ToF system can resolve masses that differ by as little as 3 Daltons. hME Assay Design software can design multiplex assays where up to ten SNPs can be assayed for each sample. Regardless of the number of assays, the reaction conditions are universal for all stages, the only optimisation required is the adjustment of extension primer concentration.

Each PCR reaction is performed with 8ng of dried DNA in a 384 well microtitre plate (ABgene) with the addition of a 5ul PCR mix as follows:

hME PCR mix (1 reaction)

Water	3.39µl
PCR buffer (10x)	0.625µl
$MgCl_2$ (25mM)	0.325µl
dNTPs (25mM)	0.1µl
HotStarTaq	0.06µl
Forward and Reverse primers (@1pmol/µl)	0.5µl

The following PCR is then performed:

1. 95°C for 15 min
2. 94°C for 20 sec
3. 56°C for 30 sec
4. 72°C for 1 min
5. Repeat steps 2-6 for 44 cycles
6. 72°C for 3 min
7. 15°C for 10 minutes

A number of genomic DNA positive and negative samples are electrophoresed on a 2% agarose gel to check for PCR contamination. If none is found a 2µl SAP mix is added to the 5µl PCR reaction, consisting of:

hME SAP mix (1 reaction)

SAP	0.3µl
SAP buffer	0.17µl
Water	1.53µl

And the reaction undergoes the following thermocycling conditions;

1. 37°C for 30 minutes
2. 85°C for 10 minutes
3. 95°C for 5 minutes
4. 15°C for 10 minutes

An extension primer mix is constructed consisting of optimized concentrations of the MassEXTEND primers. The extension primer mix is defined by an optimization procedure involving a small number of DNA samples where the extension primers are split into four groups dependent upon their mass, which are diluted initially to final concentrations of 0.625µM, 0.78µM, 0.95µM and 1.25µM (lowest mass to highest mass). Extension primers are divided into mass groups because lower mass products generate a lower signal-to-noise ratio when detected by MALDI-ToF MS. After the initial test run, extension primers are diluted according to the following equation:

Primer dilution Factor = Optimum peak height / (Actual peak height / 10)

The optimum peak height (user decided, ~30-50) and actual peak height obtained from the raw MassARRAY spectra. At the optimisation stage, failed or anomalous assays (e.g. self-priming) are removed.

The extension reaction involves the addition of the unextended hME primer, the specific 3 base termination mix (two ddNTPs and one dNTP corresponding to the polymorphic base plus the extra non-polymorphic base) and thermosequenase. With the optimised extension primer mix the following 2µl hME reaction mix is added to the 7µl PCR and SAP reaction product:

hME Extension mix (1 reaction)

Water	0.76µl
Appropriate hME extend mix	0.2µl
Adjusted extension primer mix	1µl
Thermosequenase	0.04µl

Then the following reaction is performed:

1. 94°C for 2 minutes
2. 94°C for 5 seconds
3. 52°C for 5 seconds
4. 72°C for 5 seconds
5. Repeat steps 2-4 for 100 cycles

6. 15°C for 10 minutes

To remove residual salts in the reaction that may interfere with MALDI-ToF analysis, 6mg of clean-up resin (added using a Sequenom dimple plate) and 16µl of water is added to the reaction mix before incubation and mixture on a rotor for >15 minutes. The clean-up resin removes all ions that may alter the resulting spectra and analysis. The samples are spun in a centrifuge at 3000rpm for 15 minutes to remove the resin from the solution and the samples are ready for analysis.

Samples are spotted onto a Sequenom MassARRAY SpectroCHIP automatically using a nanodispenser liquid handler (Sequenom). Each chip holds 384 spots; each spot is composed of a combustible matrix that allows ionisation of the product when excited by a laser. Each ionised extended and unextended MassEXTEND primer product differs in mass and is therefore amenable to MALDI-ToF MS analysis using MassARRAY RT software (SpectroAcquire, Sequenom). The software estimates genotypes for each sample based upon the assay design output file (.trs) and certain parameters such as the peak heights (intensity of mass signal) of each allele and also the extension primer yield (successful extension of the unextended primer compared to residual unextended primer). These genotypes can then be viewed and manually revised by the user using the Typer (Sequenom) software. An approach adopted during the studies reported here was "double-genotyping", where another experienced user of the Sequenom genotyping system and Typer software checked the genotypes for every assay.

iPlex MassARRAY Genotyping

The iPlex method of genotyping relies on primer extension over the polymorphic nucleotide of interest and the discernment of alleles by MassARRAY, as for hME, but the advantage is a higher multiplex (up to a 29-plex). However there are differences, critically in the extension reaction where the extension is terminated after a single base during the iPlex reaction. The mass is resolved either by the extended primer or by the addition of a non-specific sequence to the extension primer by the modified MassARRAY Assay Design software, allowing for better size discrimination between allele masses of different assays. There are also modifications to the PCR stage, where more DNA (12ng) and DNA polymerase is used per reaction to accommodate for the higher multiplex. However, the PCR and SAP reaction and conditions remain otherwise unchanged from hME.

iPlex PCR (1 reaction)

Water	3.35µl
PCR buffer	0.625µl
$MgCl_2$ (25mM)	0.325µl
dNTPs (25mM)	0.1µl
PCR primer mix	0.5µl
HotStarTaq	0.1µl

See the hME genotyping procedure for PCR thermocycling conditions, which are unchanged for iPlex. The same is also true for the iPlex SAP reagents and conditions (see above hME section).

As for hME, extension primers were divided into four groups of lower mass to higher mass for optimisation (0.9375μM, 1.17μM, 1.425μM and 1.875μM final concentration). The thermocycling conditions for the extension reaction are altered shortening the cycle length but increasing the number of cycles.

iPlex Extension (1 reaction)

iPlex reaction buffer	0.2μl
iPlex termination mix	0.2μl
iPlex enzyme	0.041μl
Adjusted un-extended primer mix	1.559μl

After the extension reaction desalting of the solution using Clean Resin (Sequenom) is performed by the addition of 6mg of the resin using the supplied dimple plate followed by the 25μl pure water. The remaining steps are identical to hME for analysis using the Sequenom MassARRAY, however hME and iPlex require different run parameters applied to the SpectroAcquire software (Sequenom).

2.6. Data Storage and Analysis

2.6.1. Department of Psychological Medicine Database

To permanently store data and reduce the chances of user-error, a departmental database cataloguing genotypic and phenotypic information for was created by Ivan Nikolov (http://w011.pc.uwcm.ac.uk/db/login.asp), which has since been modified (https://u004.pc.uwcm.ac.uk/new_ver/default.asp). Another database for the Bulgarian trios is also available (Microsoft Access. Assembled by Dobril Ivanov, Ivan Nikolov, Lyudmila Georgieva and George Kirov), however this is fully complementary to the most recent departmental database. The majority of variants individually genotyped and their corresponding samples are present on one of these databases. Information such as how the variant was genotyped, the assay details and the allele coding is available. Phenotypic details on each of the samples are accessible, such as the sex and affection status of an individual. The data is easily and remotely available using the internet, and the data can be output in a pre-makeped format, or as pre-makeped Haploview input files.
A pre-makeped file is a standard input file for many programs. The columns are as follows:

Column (a): Sample identifier (case-control) or pedigree identifier (families/trios)

Column (b): Individual identifier (only informative in family/trio analysis).

Column (c): Father's identifier (only informative in family/trio analysis).

Column (d): Mother's identifier (only informative in family/trio analysis).

Column (e): Sex identifier (0=UNKNOWN, 1=MALE, 2=FEMALE)

Column (f): Affection status (0=UNKNOWN, 1=UNAFFECTED, 2=AFFECTED)

Column (g): Marker 1, allele 1 (either 0, 1 or 2. Note, for haploview input alleles A, C, G, T are coded 1,2,3,4 respectively).

Column (h): Marker 1, allele 2

Column (i): Marker 2, allele 1

2.6.2. Haploview

Haploview is a software program designed for genetic association studies (Barrett et al, 2005). The program allows the user to:

- Import marker genotype data, such as case-control or familial samples (such as generated by the HapMap)
- Assess the quality of the genotype data by displaying the number of individuals in the sample, the percentage of successful genotyping, Hardy-Weinberg Equilibrium p values and non-Mendelianisms.
- View LD between markers (measures of D' and r^2)
- Select "tag" markers for association testing (Tagger) based upon user defined LD, allele frequency and type of analysis criteria.
- Test markers and haplotypes for association.
- Markers can be exclude/included based upon their genotype quality or frequency for example.

- Furthermore, human genomic position and known genes can also be viewed in relation to this data.

Haploview was used in this study to examine LD between genotyped markers and test for single marker association. The association results were always compared to manually analysed data for case-control analysis.

2.6.3. Statistical Analysis

Several types of statistical analysis were performed in this thesis:

The testing of consanguinity in pedigree C702 was performed by Prof. Peter Holmans (see chapter 3).

Tests of single-marker association in cases and controls was performed using either Minitab software by construction of a 2x2 (allelic) or 2x3 (genotypic) contingency table before chi square (χ^2) analysis, or alternatively by Haploview (http://www.broad.mit.edu/mpg/haploview/). Fisher's exact tests of allelic association were performed using an online tool (http://www.matforsk.no/ola/fisher.htm). The χ^2 test examines the null hypothesis that the proportions of a characteristic in two groups are the same. The resulting χ^2 value can be compared to values from a known probability distribution to gain a p value.

Inverse variance meta analysis for cases and controls was performed using a Microsoft Excel formula sheet designed by Dr. Valentina Moskvina, where each study is weighted according to the effect size and a mean effect size across all studies computed.

The output is an association p value and odds ratio, but also a heterogeneity p value, which tests the null hypothesis that the effect size in all studies goes in the same direction.

For case and control studies, odds ratios (ORs) were calculated using the formula AD/BC, where A is allele 1 in cases, B is allele 2 in cases, C is allele 1 in controls and D is allele 2 in controls, using a Microsoft Excel formula sheet provided by Dr. Nadine Norton.

Tests of single-marker association in familial samples was performed by a Transmission Disequilibrium Test (TDT) (Spielman, McGinnis *et al.*, 1993) using Haploview. The method tests the null hypothesis that the transmission of either allele from a heterozygous parent to affected offspring will occur half of the time. Therefore, if an allele is over- or under- transmitted to an affected offspring to the extent that it is unlikely to have occurred by chance, an association can be claimed.

Linkage Disequilibrium measurements of D' and r^2 were performed using Haploview. Analysis of variants for deviations from Hardy-Weinberg Equilibrium was also performed using Haploview.

Tests of data normality, comparisons of group variance and means and Pearson correlation analysis were performed using SPSS version 14.

Sample size power calculations for association samples were performed using an online Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/).

Haplotype frequency estimations in case-control samples and subsequent global haplotype association analysis was performed using Fast EH+ software (Zhao, Curtis *et al.*, 2000). For individual haplotype association testing, haplotypes were converted into

alleles to form a 2x2 contingency table (the counts of haplotype to be tested versus the counts of all other haplotypes, in cases and controls separately) and a χ^2 test performed (Minitab). Haplotype frequency estimations and association analysis in familial samples was performed using Haploview.

2.7. Post-Genomic Materials and Methods

2.7.1. Cell Culture

During this study it was required for cell lines to be generated and studied for pedigree C702. During the DNA extraction procedure from whole blood, β-Lymphoblasts can be isolated by centrifugation sedimentation (Dr. Nadine Norton). These cells were sent to ECACC (European Centre for Analysis of Cell Culture) where they were immortalized by transformation with an Epstein-Barr virus (EBV). The EBV transformed β-Lymphoblasts can then be stored and re-cultured when necessary. Four of the six affected siblings of C702 were successfully transformed and used in further analysis.

Cell culture was performed in a Microflow Advanced Biosafety Cabinet – Class 2 (BioQuell). Surfaces and containers/utensils were cleaned with ethanol prior to use and effervescent chlorine tablets (Haz-Tabs. Guest Medical) were also used to disinfect the fume hood and waste reagents and consumables.

When EBV transformed β-Lymphoblast cell lines are ordered or removed from storage (Liquid Nitrogen. BOC) they are present in a cell "pellet", which is an adhesion of many thousands of frozen cells. For *in vitro* cell culture the media used was

GlutaMAX (GIBCO Invitrogen cell culture) which contains a stabilized form of L-glutamine. The media also requires nutrients found in Fetal-Bovine-Serum (FBS - GIBCO Invitrogen cell culture), an ampicillin and streptomycin antibiotic mixture to prevent microbial contamination of the culture, and a further supplement of glutamine (as glutamine has a tendency to degrade over time). The media is constructed as follows for 100ml:

GlutaMAX	68ml
FBS	15ml
Glutamine (200mM)	1ml
Ampicillin/Streptomycin in RPMI-1640	1ml

The media is warmed to 36°C in a water bath (Grant Instruments) , then the cell pellet is removed from nitrogen and 1-2ml of media added to the cell pellet before aspiration into a serological pipette tip and placement into a T-flask that can hold 15cm^2 (15ml) of solution (Nunc). Further media is added to a volume of 15ml; the T-flask is manually shaken and placed into an incubator (36°C, 5%CO_2) for culture. The initial culture is left until the mixture changes colour from red to yellow (indicating that the glutamine has been used up and that the cells are growing), which usually takes 2-3 days.

The integrity of the cell culture can be viewed using a light microscope after labeling with Trypan Blue (Invitrogen). Any signs of unusual levels of apoptosis/necrosis may indicate that the culture is not receiving the necessary amount of media and this can

be adjusted or the cell line re-grown from liquid Nitrogen or re-ordered from the Coriell cell repository.

All cell culturing of EBV transformed β-Lymphoblasts used in this study were performed by myself and/or Lyn A Elliston.

2.7.2. DNA extraction from cultured transformed β-Lymphocytes

An appropriate volume of actively growing cell culture was pippetted into a Falcon tube (15ml, corresponding to approximately 10^{-7} cells). The culture was centrifuged at 375xg for 10 minutes at 4°C. The resulting solution was removed from the cell pellet and an equal volume of Phosphate Buffered Saline (PBS) added and the centrifugation step repeated. The wash step is repeated twice, and then all solution removed.

To the pellet is added 500µl STE buffer (10mM Tris-Cl, pH7.5 10mM NaCl 1mM EDTA), 50µl sodium dodecyl sulphate and 50µl proteinase K. The sample is then vortexed vigorously and left overnight at 55°C. The next day the whole sample is added to a 2ml Eppendorf tube. 1 ml of tris-saturated phenol is then added to this solution and vortexed to gain a precipitate. The sample is then spun in a microcentrifuge at 3500rpm for 4 minutes and the resultant upper visible layer removed via pipette (discard remaining solution) and this is transferred to a new Eppendorf tube. Then, 400µl chloroform and 400µl tris-saturated phenol is added and the sample, vortexed and then spun for 4 minutes at 3500rpm in a microcentrifuge. The resultant upper layer is again removed and transferred to a new Eppendorf tube and 1ml of chloroform added before vortexing and

spinning again for 4 minutes at 3500rpm in a microcentrifuge. The top layer is again removed and added to 1ml of 100% ethanol. The DNA can now be visualized as a white precipitate and can be removed using a pipette tip and added to 100-1000µl of 1x TE buffer.

2.7.3. RNA Extraction

In this study, RNA was isolated from both EBV transformed β-Lymphoblast cultures and also from neural tissue. All neural tissue RNA and gDNA extraction performed by Dr. Nicholas Bray, while lymphoblast culture and RNA extraction was performed by myself under the supervision of Lyn A. Elliston and Dr. Nicholas Bray, and the RT-PCR carried out by myself.

All RNA extractions were performed using the Ambion RNAqueous-Midi kit. For RNA extraction from lymphoblasts the following procedure was adhered to: An appropriate volume of actively growing cell culture was pippetted into a Falcon tube (15ml, corresponding to approximately 10^{-7} cells). The culture was then placed in a centrifuge and spun at 375xg for 10 minutes at 4°C at the minimum settings of acceleration and deceleration. The resulting solution was removed from the cell pellet and an equal volume of Phosphate Buffered Saline (PBS) added and the centrifugation step repeated. The wash step is repeated twice, and then all solution removed.

To the pellet is added 400µl of lysis/binding solution and the pellet is vortexed until cells have disrupted. 400µl of the 64% ethanol is then added and the whole contents spun in the supplied filter cartridge (placed into a 1.5ml Eppendorf) at 11500 rpm for 1

minute using a microcentrifuge, discarding the flow-through. 700µl of wash solution 1 is then added to the cartridge membrane and the same microcentrifuge step performed (flow-through discarded). Then, 450µl of wash solution 2/3 is added and spun identically with the flow through discarded (this step is repeated). The empty cartridge is then spun at 11500 rpm with nothing added. The cartridge is then placed into a new collecting tube and 70µl of heated elution buffer (80°C) added and spun at 11500 for 1 minute. DNAase is then added to the resultant total RNA, which is then stored at -80°C.

2.7.4. Reverse Transcription/cDNA Synthesis from Total RNA

All reverse transcriptions (RTs) used in this study were done using a RETROscript kit (Ambion) and total RNA (DNAase treated) and were performed by myself or Dr. Nick Bray. RT reactions were performed in a PCR-free environment and RNAzol was constantly applied to work surfaces, gloves etc. Reagents were kept on ice at all times unless required. The RT protocol was as according to RETROscript guidelines as follows:

Mastermix A (1 reaction)

Water	9µl
Random decamers	2µl
Total RNA	1µl (at 1µg/ml)

The solution is heated to 85°C for 3 minutes and then placed on ice.

Mastermix B (1 reaction)

10x RT buffer	2µl
dNTPs	4µl
RNAase inhibitor	1µl
MMLV-RT	1µl

Add 8µl of mix B to 12µl of mix A and perform RT reaction (in microtitre plate or PCR tubes using an MJ thermocycler):

1. 44°C for 40 minutes
2. 55°C for 20 minutes
3. 92°C for 10 minutes
4. 4°C for 15 minutes

A volume of 120µl of water is then added to each sample and each 140µl sample is stored at -80°C.

Note, to control for contamination, control RT steps were performed in the absence of template RNA.

2.7.5. Protein Extraction from Transformed β-Lymphocytes

Cell lines were obtained from ECACC as in section 2.7.1. Cells were all cultured using the same reagents and conditions as section 2.7.1. 20ml of growing cell media solution is extracted into a 150cm^2 Falcon tube and centrifuged at 250xg for 10 minutes to pellet cells. The supernatant was removed and replaced with 20ml D-PBS (Dulbecco's-Phosphate Buffered Saline. Invitrogen) to wash, and the centrifugation repeated. The PBS wash step is then repeated. A suitable extraction buffer was obtained (to perform cell lysis while preventing protein degradation or interference with the proteins' biological activity or immunoreactivity), either RIPA buffer (Sigma) or IP buffer (Sigma). The extraction buffer was supplemented with protease inhibitors (Roche). Note that all comparative experiments were performed using the same extraction buffer.

A 400ul volume of extraction buffer was added to the cell pellet (after removal of the PBS supernatant), and the cell/buffer mixture gently aspirated to successfully lyse the cells. The mix is then left to stand for 5 minutes (on dry ice) before further aspiration. Aliquots are then made of the cell extract and stored at -80°C.

Note that protein extraction from the cerebellar sample used in this study was performed by Dr. Angela Hodges using a Thiourea based extraction buffer.

2.7.6. Bradford Assay to Quantify the Protein Content of a Solution

To estimate the amount of total protein present in a cellular extract a Bradford Assay was performed. The assay exploits the fact that the absorbance maximum for an acidic dye

(Coomassie Brilliant Blue G-250. Peirce) shifts from 465nm to 595nm when protein is present. Basic amino acids stabilize the anionic dye and cause a visible colour change. Accurate quantification of a test sample is gained once a standard curve based upon known concentrations of albumin is generated.

The Bradford Assay protocol used in this study was based upon the Bio-Rad Protein Assay system. To generate a standard curve a solution of Bovine-Serum-Albumin (BSA) (Sigma) at a known concentration was diluted over a 7 point range of concentrations from 0-2mg/ml, in duplicate and in the same extraction buffer as the test samples. Test samples were mixed gently and diluted 1 in 20, again in the same extraction buffer. A 1 in 5 dilution of the BioRad Protein Assay Bradford reagent was made at a suitable volume. Each of the standard curve and test samples is then mixed and 10µl added to 990µl of the Bradford reagent and gently mixed (all in an identical fashion). The mixture is then left for 15 minutes then added to a plastic cuvette for analysis on a spectrophotometer (Amersham Ultrospec 2100 pro). The mean of each duplicate standard curve concentration datapoint is plotted on a graph of absorbance (nm) versus protein concentration (µg/µl). The absorbance values for the test sample therefore allow the inference of the protein concentration.

2.7.7. Relative Allelic Expression Assay

To assay transcripts for factors regulating steady-state mRNA levels in an allele-specific manner (such as *cis* or haplotype acting factors) a quantitative assay of allelic expression has been developed in-house (Bray, Buckland *et al.*, 2003a). The principle of

the assay is that in any given tissue at any given time the relative amount of steady state mRNA levels (as transcribed from each paternal chromosome) should have a 1:1 ratio unless factors regulating the stead-state levels of the mRNA differentially affect the two autosomal transcripts. To measure for such differences, a transcribed heterozygote marker can be used where the two alleles should have a 1:1 ratio in a quantitative assay. The assay is performed by a PCR within an exon that encompasses the variant in question, which will amplify genomic DNA (gDNA) and complementary DNA (cDNA) equally. Therefore, any divergence of the 1:1 allelic ratio of gDNA observed in cDNA samples is indicative of an allele-specific regulation of steady state mRNA levels in the tissue examined, such as a *cis* acting polymorphism regulating transcription.

All post-mortem samples used in this study were obtained from four sources: The MRC, London Neurodegenerative Diseases Brain Bank, London, UK. The Department of Clinical Neuroscience, Karolinska Institute, Sweden. The Mount Sinai School of Medicine, New York, USA and the Stanley Medical Research Institute Brain Bank, Bethesda, USA. Samples are from multiple brain regions including frontal, temporal and parietal lobes. All samples either had no known neurological or psychiatric disorder at time of death (controls) or details of their phenotype were given (e.g. Alzheimer's disease, schizophrenia, bipolar I disorder, unipolar depression). Furthermore, details such as the cause of death and post-mortem interval were also given.

Polymorphisms within predicted mRNA sequences that had the highest known minor allele frequencies were selected as markers for genes studied so as to maximise the number of heterozygotes studied. Alternatively, an interesting polymorphism (i.e. associated) may be chosen. All markers are assayed by using "universal" primers based

on single exonic sequence capable of amplifying genomic or complementary DNA. The same analytical conditions were used for genomic DNA and cDNA to enable us to employ the average of the ratios observed from genomic DNA (representing a 1:1 ratio of the two alleles) in order to correct allelic ratios obtained from cDNA analyses for any inequalities in allelic representation specific to each assay (Bray, Buckland *et al.*, 2003a). RNA samples also underwent PCR prior to an RT step to examine for detectable levels of product that would indicate a genomic DNA contamination.

Each cDNA sample was assayed alongside the corresponding heterozygous genomic DNA to ensure uniformity of conditions. PCR amplification was carried out as either standard or touchdown (see section 2.2) however 6µl of cDNA was used. Primer extension was carried out with SNaPshot chemistry (Applied Biosystems) as described in section 2.5.4. Aliquots of 3µl SNaPshot reaction product were combined with 8µl HiDi formamide and loaded onto an ABI3100 PRISM Genetic Analyser (Applied Biosystems). Products were electrophoresed on a 36-cm capillary array filled with POP4 and data were processed by using Genescan Analysis v3.7 software (Applied Biosystems). Peak heights representing allele-specific extended primers were determined by using Genotyper software (Applied Biosystems) and were used to give a ratio of allelic representation for each sample. All peak heights were between 500-6000 fluorescence intensity units and assays showing alleles present in water or blank samples were excluded.

Analysis for relative differential allelic mRNA expression the following protocol was adapted. Heterozygote samples were analysed in parallel for gDNA and duplicate cDNA samples (independent RT reactions for the same sample). Raw data from Genotyper software was collated for each sample gDNA and cDNA. A ratio of alleles

was then obtained from the fluorescence intensity values for each allele. The mean of all the gDNA ratios was then taken. All gDNA and cDNA ratios were corrected by the mean gDNA value (therefore forcing the gDNA data to be near to 1). Samples showing a standard deviation between duplicate RTs of ±0.2 were removed. The average cDNA ratio of the two duplicate RTs was taken as the corrected cDNA ratio for that sample. The average standard deviation over all corrected duplicate RT standard deviations was taken as a quality score for the assay.

All ratios (corrected genomic or cDNA) were transformed to the natural log of the ratio datapoint and absolute values then taken for each measurement. The distribution of the gDNA and cDNA data was analysed for normality by a Kolmogorov-Smirnov test of normality. If the data was normal, then group means (either for gDNA versus cDNA, or cDNA of haplotype carriers versus cDNA of non-carriers) were compared by a 2 independent sample *t*-test. If the data was not normal then a Mann-Whitney U test was employed. Multiple group comparisons were done firstly by a one-way ANOVA (assuming normality of the data) or a Kruskal-Wallis test (not assuming normality of the data), before the appropriate two-group comparison, as above. A difference between the mean of gDNA and cDNA for an assay indicates differential allelic expression, while differences between cDNA groups could indicate a genotype or phenotype effect on those groups. Individual samples showing a corrected cDNA ratio >20% (outside a corrected ratio of 0.83-1.2) were also said to show evidence for differential allelic expression.

2.7.8. Taqman Gene Expression Assay

Real-time quantitative PCR (QPCR), performed using the TaqMan Gene Expression Assay system (Applied Biosystems), was used to measure total *PRKCA* levels in EBV-transformed lymphoblast cell lines from 1 C702 sibling and 1 unrelated control that did not carry the rare E19A allele. This method was also used to compare *PRKCA* expression between 10 DSM-IV schizophrenic cases and 11 control individuals using post-mortem cerebral cortex tissue (obtained from the Stanley Medical Research Institute Brain Bank, Bethesda, USA).

The Taqman Gene Expression Assay involves the use of a PCR primer set and a 5' FAM labelled probe (oligonucleotide) that is also 3' labelled with a quencher moiety. The PCR primer set and probe are specific to the transcript(s) of interest and so PCR will only occur over cDNA (not genomic DNA). The process involves the binding of the Taqman probe to the cDNA of interest. While bound, the fluorophore will be quenched, however during PCR of the fragment the 5' nuclease activity of DNA polymerase releases the fluorophore from the quencher and fluorescence is generated that is proportional to the amount of PCR product (and so the amount of cDNA or transcript).

Total *PRKCA* mRNA levels were assayed using TaqMan Gene Expression Assay (Hs00176973_m1), that assays 5 putative *PRKCA* isoforms including AB209475 and X52479. In order to control for inter-sample variability in RNA input or reverse transcription efficiency, samples were simultaneously assayed for a 'housekeeping' gene (assumed to be unrelated to disease state) that could be used to normalise *PRKCA* measures. Lymphoblast-based assays were normalised using measures of *ACTB* mRNA

(TaqMan Gene Expression Assay hs99999903_m1), whereas post-mortem brain experiments were normalised using measures of *18S* (TaqMan Gene Expression Assay hs99999901_s1).

cDNA samples used for analysis were obtained as described in section 2.7.7. Real-time QPCR reactions were performed in a total volume of 20µl, containing 9µl cDNA (of sample cDNA diluted 1 in 2), 10µl TaqMan Universal PCR mix (Applied Biosystems) and 1µl TaqMan Gene Expression Assay reagent (Applied Biosystems). Individual cDNA samples were assayed in triplicate. Individual reactions were quantified against a standard curve constructed by serial dilution of a cDNA sample for both *PRKCA* and the normaliser gene. Negative controls, containing water rather than cDNA, were included for each gene assay and data only analysed if these showed no evidence of contamination.

A typical plate layout is shown in figure 2.1. PCR was carried out on an ABI 7000 Sequence Detection System (Applied Biosystems) using an appropriate microtitre plate (Applied Biosystems) covered with an optical adhesive cover. The PCR reaction conditions were as standard as follows: A denaturation step of 95°C for 10 min, followed by 40 cycles of 95°C for 15s and 60°C for 1 min. Raw cycle threshold (C_T) values for the test or normaliser transcript were recorded using SDS 2.0 software (Applied Biosystems) during the linear phase of PCR. Standard curve dilutions were given arbitrary values based on their relative concentration (i.e. '100', '50', '25', '12.5' and '6.25') and these were used by the software to calculate relative expression values for each sample. For each individual sample, a mean expression value was calculated for *PRKCA* and the normaliser transcript based on the values derived from the 3 replicate reactions. PRKCA

expression values were then divided by the corresponding values for the selected housekeeping gene to give a normalised PRKCA expression value.

	1	2	3	4	5	6	7	8	9	10	11	12
A	NTC	100	100	100	50	50	50	25	25	25	NTC	NTC
B	X	12.5	12.5	12.5	6.25	6.25	6.25	Sample1	Sample1	Sample1	X	X
C	X	Sample2	Sample2	Sample2	Sample3	Sample3	Sample3	Sample4	Sample4	Sample4	X	X
D	X	Sample5	Sample5	Sample5	Sample6	Sample6	Sample6	Sample7	Sample7	Sample7	X	X
E	NTC	100	100	100	50	50	50	25	25	25	NTC	NTC
F	X	12.5	12.5	12.5	6.25	6.25	6.25	Sample1	Sample1	Sample1	X	X
G	X	Sample2	Sample2	Sample2	Sample3	Sample3	Sample3	Sample4	Sample4	Sample4	X	X
H	X	Sample5	Sample5	Sample5	Sample6	Sample6	Sample6	Sample7	Sample7	Sample7	X	X

Figure 2.1: Plate layout for TaqMan QPCR analysis. Shown are the rows (A-H) and columns (1-12) of a 96 well plate (Applied Biosystems). Green wells denote sample set-up for one TaqMan probe-set, which is duplicated for the normalising probe-set (yellow wells). NTC = Non-template control, X = blank well. "SampleN" denotes a 1 in 2 dilution of a test sample; each sample is replicated in triplicate for each probe. Also, a range of cDNA concentrations starting at an undiluted sample (100) and then serial dilutions of this sample (to 6.25, or 1 in 8 dilution) are shown in triplicate.

Tests for statistically significant differences between case and control samples were performed using an unpaired two sample t test.

2.7.9. Western Blot

Western Blotting is a method to detect a specific protein or small number of proteins in a mixture, while simultaneously displaying the size of that protein(s). The method involves the gel electrophoresis of a protein mixture to separate by size, then the electrophoretic movement of the size electrophoresed proteins onto a PVDF membrane

followed by incubation of the membrane with a highly specific primary antibody (Ab) and subsequent secondary antibody that allows visual and semi-quantitative detection of the protein (e.g. by fluorescence or radiation). The gel systems used in this study were NuPAGE Bis-Tris Pre-Cast Gels (Invitrogen), for small to mid-size molecular weight proteins and NuPAGE Tris-Acetate Gels (Invitrogen) for larger proteins. Secondary Ab/protein detection was done via chemiluminescence using horse-radish peroxidase conjugated secondary antibodies (Pierce Biotechnology) detected using LabWorks software.

Total protein/cell extract (see section 2.7.5) was defrosted and centrifuged gently to remove any cellular debris. Up to 100µg of total protein (as delineated by Bradford assay, see section 2.7.6) was combined in a PCR tube with 7.5µl NuPAGE LDS buffer (Invitrogen) and 3µl NuPAGE reducing agent (Invitrogen). The volume was made up to 20 or 30µl with the same extraction buffer used to obtain the protein. The samples were then heated to 70°C for 10 minutes using an MJ thermocycler. The appropriate gel and running buffer was then constructed in the supplied gel tank (Invitrogen). 12 or 17 well NuPAGE 12% Bis-Tris gels (Invitrogen) are used with 1x MOPS running buffer (Invitrogen). 12 or 17 well NuPAGE 3-8% Tris-Acetate gels were run with 1x Tris-Acetate running buffer (Invitrogen). In all cases the reduced samples were run with 500µl antioxidant (Invitrogen) added to the running buffer in the watertight upper buffer chamber. To allow size discrimination directly on the western blot MagicMark XP western protein standard was used (Invitrogen). Samples were added to the wells of the gel and run at 200V for 50 minutes (Bis-Tris) or 150V for 1 hour (Tris-Acetate) using a power pack (Bio-Rad power Pac 3000)

The PVDF membrane (Millipore) and filter paper (x2) was cut as the same size as the gel. The membrane was immersed in methanol (15 seconds), followed by immersion in pure water (2 minutes) then left in transfer buffer (1x transfer buffer with 10% methanol) with the filter paper cuttings. The post-electrophoresis gel is then combined with the membrane and sandwiched between the filter paper before insertion into the supplied transfer tank (Invitrogen). The upper chamber of the transfer tank is filled with transfer buffer and the lower chamber is filled with water. The transfer stage involves a 30V for 1 hour step (Tris-Acetate) or a 30V for 40 minutes step (Bis-Tris) using a power pack (Bio-Rad). Meanwhile, a solution of 1x Phosphate buffered saline (NaCl 81.8g; KCl 2g; Na_2HPO_4 14.3g; KH_2PO_4 2.4g. Add to 1 litre pure water) is made, to which is added 1ml of Tween 20 (Sigma-Aldrich) to make PBS-T. 100ml of PBS-T is added to 5g of Marvel milk powder and mixed to make a blocking agent.

The gel tank is dismantled; the gel is stained with Coomassie Blue for 1 hour before de-staining with 100% methanol. A photograph was then taken of the gel to make sure no anomalies had occurred during gel electrophoresis or transfer steps. The PVDF membrane was transferred to the blocking agent and gently mixed on a rotator for 1 hour. The membrane was then added to a falcon tube containing 10ml of the blocking agent solution with an appropriate concentration of primary Ab (determined during optimization of antibody using the same protocol but applying multiple dilutions of the primary antibody to the same set of samples) and left to rotate for 1 hour. The membrane is then washed in 3 separate quantities of PBS-T for a total of 15 minutes. The membrane was then added to a falcon tube containing 10ml of the blocking agent solution with an appropriate concentration of secondary Ab (Pierce Biotechnology anti-mouse HRP or

anti-goat HRP) and rotated for 1 hour. The membrane is then washed in PBS-T as before. A solution composed of a 1:1 ratio of SuperSignal West Dura Luminol/Enhancer Solution (Pierce Biotechnology) and SuperSignal West Dura Stable peroxide Buffer was then constructed and the membrane immersed in this for 5 minutes. The membrane is then placed onto a transparent plastic tray (AutoChemi systems) and covered with cling-film before insertion into the AutoChemi systems gel doc (UVP Bioimaging Systems). The chemiluminescence was captured by a dynamic integration procedure, recording 99 frames in 50 minutes and combining the images.

Quantification using LabWorks Image Acquisition and Analysis software, version 4.6, which allows a user to identify and quantify any bands present in arbitrary chemiluminescence units.

The concentration of total protein in a mixture was estimated using the Bradford assay; however to more accurately quantify the levels of a test protein in a sample of interest a normaliser was used to correct for possible differences in concentration or errors in gel loading etc. A normaliser is a stably expressed/transcribed gene/protein that is not disease/state related, such as α-tubulin. Both the test and normalizing proteins underwent western blots based upon the same gel and transfer steps. Quantification is taken as the image optical density (OD) for each band, and then a ratio of the test and normaliser band ODs is taken as an arbitrary measure of test protein levels for that sample.

2.8. Bioinformatic Resources

2.8.1. The UCSC Genome Browser

http://genome.ucsc.edu/

The UCSC genome browser lists biological information for the genomes of many species, most extensively for *Homo sapiens*. The database allows the user to quickly localise an area of interest in the human genome (be it a chromosomal region, an mRNA transcript or gene, complex polymorphisms such as CNVs or simple variants such as SNPs) armed only with the approximate location of the item, the raw DNA sequence or the identifying name of the point of interest (e.g. NM accession prefixes for human mRNA transcripts or rs prefixes for identified human SNPs). Therefore, the whole human genome is annotated with the recent state of knowledge of genetic variants (duplications, microsatellites, SNPs), genes and the DNA sequence of the entire genome.

An entire disease locus mapping project can be successfully choreographed using the resources available through the UCSC genome browser website. Other analogous databases include NCBI Entrez (http://www.ncbi.nlm.nih.gov/Entrez/) which similarly catalogues human genome information.

2.8.2. A List of Bioinformatic Resources Used in this Study

Scientific literature searches at PubMed and ScienceDirect

http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed

http://www.sciencedirect.com/

UCSC Genome Browser

http://genome.ucsc.edu/

International HapMap Project

http://www.hapmap.org/

National Center for Biotechnology Information

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

Standard PCR primer design at Primer3

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

Database of Genomic Variants. A catalogue of structural variants in the human genome.

http://projects.tcag.ca/variation/

Department of Psychological Medicine Databases, hosted by the Biostatistics and Bioinformatics Unit t Cardiff University.

http://x001.psycm.uwcm.ac.uk/

Regulatory motif search using ClusterBuster.

http://zlab.bu.edu/cluster-buster/

Regulatory motif search in UTR using UTResource

http://www.ba.itb.cnr.it/UTR/

micro-RNA binding site motif search using TargetScan (UCSC Genome Browser)

http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=102889095&c=chrX&g=targetScanS

Genetic Power Calculator

http://pngu.mgh.harvard.edu/~purcell/gpc/

Nyholt correction

http://gump.qimr.edu.au/general/daleN/SNPSpD/

Chapter 3. Refinement of a Schizophrenia Linkage Region on Chromosome 17 in a Single Pedigree

3.1. Introduction

Recent linkage studies of schizophrenia and related psychiatric illnesses have implicated the long arm of chromosome 17 as linked to major mental illnesses such as schizophrenia and bipolar disorders (Dick, Foroud *et al.*, 2003, Ewald, Wikman *et al.*, 2005, Klei, Bacanu *et al.*, 2005, Williams, Norton *et al.*, 2003). The most compelling evidence indicative of a psychiatric disease locus at 17q stems from analysis of a large schizophrenia sample obtained in the UK over the last decade (Williams, Norton *et al.*, 2003). Williams and colleagues (Williams, Norton *et al.*, 2003) performed a genome-wide linkage scan on a large sample of 353 affected sibling pairs (schizophrenia and/or schizoaffective disorder) from the UK, Ireland, Sweden and the USA. The study incorporated a high density of markers, spaced at ~10cM throughout the genome, and identified a near (Williams, Norton *et al.*, 2003) genome-wide significant linkage peak at 17p11.2-q25.1 (LOD = 3.35, *suggestive* evidence of linkage (Lander and Kruglyak, 1995)). This spanned much of the chromosome, with the maximum lod score -1 region spanning 57–74 cM (26.6–58.7 Mb UCSC Genome Browser. May 2004). The authors identified that much of the linkage evidence was due to allele sharing within a single pedigree (C702); specifically the sharing of maternal and paternal haplotypes among all affecteds forming a possible region IBD2 (Williams, Norton *et al.*, 2003).

A number of positional cloning attempts to identify susceptibility loci for the related bipolar disorders have implicated 17q as a potential linkage region. A high density genome scan of 154 bipolar I disorder ASPs (Affected Sibling Pairs) from the UK and Ireland detected a multipoint MLS of 1.38 at 17q23-q24 (Bennett, Segurado *et al.*, 2002). An independent high resolution study employed 250 ASPs from the US and of their two most significant findings one was a detection of suggestive evidence for linkage at 17q in bipolar I and schizoaffective patients, with both single and multipoint measures at the microsatellite marker D17S928 (Dick, Foroud *et al.*, 2003). Finally, another independent US study of 151 siblings primarily of Caucasian ancestry as in the Dick *et al* study (Dick, Foroud *et al.*, 2003), identifies evidence for linkage in the same region using the same genetic model and phenotype (non-parametric lod score = 2.07, at D17S1531) (McInnis, Dick *et al.*, 2003).

Ewald and colleagues (Ewald, Wikman *et al.*, 2005) used homozygosity mapping to analyse the genotypes of SNPs and microsatellites in 22 consanguineous families from Cuba exhibiting bipolar affective disorder. The best evidence for linkage was across the region of 17q24-q25 with a maximum parametric multipoint lod score of 2.08.

Further corroborative evidence for a psychiatric disease locus on 17q comes from linkage analysis of large, multiply affected pedigrees such as from the Balearic islands (Tomas, Canellas *et al.*, 2006), the population isolate of Palau (Klei, Bacanu *et al.*, 2005) and also using pedigrees from Japan and the UK (Rees, Fenton *et al.*, 1999). Tomas *et al* analysed 12 large pedigrees with multiple instances of Bipolar I disorder and found markers at 17q24-qter showed favourable evidence for non-parametric linkage

(LOD=1.55 at D17S949), however the signal was strongest at 17q24 where the best evidence of linkage was under a recessive mode of inheritance (LOD=2.1, D17S949). Klei and colleagues (Klei, Bacanu *et al.*, 2005) studied families segregating both bipolar and schizophrenia, with schizophrenics alone considered as a narrow phenotype. Both the narrow and broad disease models show evidence for linkage at 17q23.2 (LOD=1.79 and 2.8 for narrow and broad diagnosis at D17S1290), under a recessive disease model only. Finally, a linkage study using multiply affected pedigrees (expressing primarily schizophrenia and bipolar disorders) from the UK and Japan (Rees, Fenton *et al.*, 1999)identified evidence in favour of linkage at D17S784 in the UK pedigrees (LOD = 1.15), but this failed to reach evidence for suggestive or significant linkage. Two point analysis using a recessive model incorporating a broad diagnosis for the UK and Japanese pedigrees combined yielded a more significant result (LOD = 1.65). When multipoint evidence was considered, markers from D17S798-D17S784 displayed a maximum MLS of 1.67 in the UK and 1.58 in the combined sample, again under a broad-recessive disease type. None of the above linkages were enough to warrant a claim of suggestive linkage (Rees, Fenton *et al.*, 1999).

Support for a schizophrenia susceptibility locus at 17q also comes from meta-analysis of 20 genome-scans of schizophrenia pedigree samples from around the world (Lewis, Levinson *et al.*, 2003) evidence for linkage to 17q21.33-q24.3 was detected, with much of the evidence coming from samples with a large preponderance of west European ancestry (e.g. (Blouin, Dombroski *et al.*, 1998, Levinson, Mahtani *et al.*, 1998, Straub, MacLean *et al.*, 2002)). However, although the majority of linkage genome-wide scans have been performed on individuals of European descent, the evidence from both the Klei

et al study (Klei, Bacanu *et al.*, 2005) of Polynesian pedigrees and the Rees *et al* study of Japanese pedigrees (Rees, Fenton *et al.*, 1999) suggests that other populations may too exhibit risk conferred by this locus.

The suggestive evidence of linkage reported to chromosome 17 gained from the analysis of 353 ASPs in the Williams *et al* study (Williams, Norton *et al.*, 2003) was due to a multiply affected family with schizophrenia, from Ireland. The remaining evidence for linkage in this region when C702 are removed is weak (multipoint lod score or MLS=0.89) but does not provide evidence against linkage to this region implying that the evidence for linkage to chromosome 17 in the Williams *et al* ASP study (Williams, Norton *et al.*, 2003) stems largely from pedigree C702. Pedigree C702 consists of six affected male siblings diagnosed with DSM-IV schizophrenia who also display elements of disturbed mood (Prof. Nick Craddock, personal communication) and can also be diagnosed according to Karl Leonhard's criteria with "Affect-Laden-Paraphrenia" (Dr. George Kirov, personal communication), a sub-phenotype of the schizophrenias that may show a recessive mode-of-inheritance (Ban, 2004). Separate linkage analysis of only pedigree C702 achieved an MLS of 8.38 at 77cM (D17S787-D17S944), that reaches the Lander & Kruglyak (Lander and Kruglyak, 1995) criteria for genome wide significance (p=0.02). The linked region (LOD-1) spans a 53cM region covering 17p11.2-q25.1 and secondary analysis distinguished maternal and paternal haplotypes spanning this region which cosegregate with each affected member of C702 (figure 3.1). (Williams, Norton *et al.*, 2003). The pedigree showed no evidence for linkage elsewhere in the genome and there was no indication of greater than average homozygosity across the genome that would indicate a consanguineous pedigree (Williams, Norton *et al.*, 2003).

Pedigree C702 represents an excellent opportunity for molecular genetic analysis. However the major obstacle to elucidating the genetic factors that cause the C702 phenotype is the size of the C702 linkage region. Figure 3.1 displays the paternal and maternal haplotypes of the affected individuals (Williams, Norton *et al.*, 2003), spanning 76.2Mb and 58.8Mb respectively. The evidence for linkage stems from the affecteds all sharing both alleles for 6 markers apparently Identical By Descent (IBD2). The IBD2 status of the affected siblings may reflect the convergence of two copies of the same disease allele(s) from a common ancestor (an example of coalescence) (Rosenberg and Nordborg, 2002). The unaffected status of the parents also supports the view that a convergence of disease alleles has occurred to result in a recessive phenotype, dependent upon the affected siblings sharing two copies of the disease allele(s).

Maternal DNA was not available and therefore the large regions where the affected members of C702 apparently share paternal and maternal haplotypes may be Identical By State (IBS). It is noteworthy that falsely inferring that IBS regions are IBD may inflate the linkage signal gained for this pedigree (Williams, Norton *et al.*, 2003). The genetic distance between the two markers defining the reported IBD2 region (D17S799-D17S785) is 78.34cM, 115.08cM and 41.6cM (sex-averaged, female and male genetic distance respectively) (Kong, Gudbjartsson *et al.*, 2002). This means that a recombination event can be expected for a transmitted chromosome to occur on 78.34% of occasions (115.08% and 41.6% of occasions for female and male chromosomes). It is therefore highly unlikely that a recombination event has not occurred within this region in one of the transmitted parental chromosomes. High density mapping of the reported IBD2 region should establish this and therefore refine the region of linkage.

Therefore, reported here is an attempt to define the shared paternal and maternal chromosomal segments in affected members of pedigree C702 by genotyping further markers across the putative IBD2 region reported by Williams *et al* (Williams, Norton *et al.*, 2003).

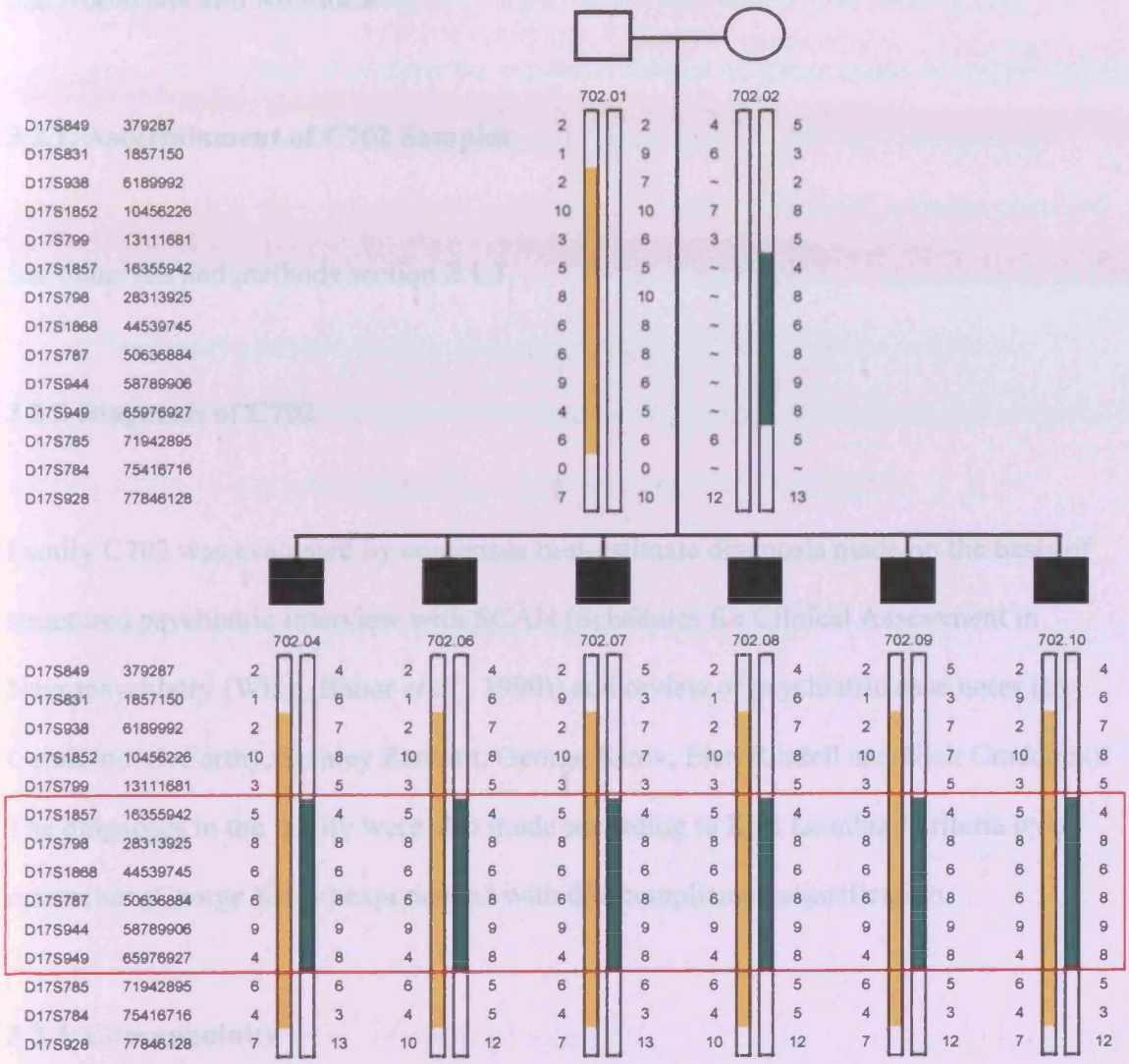


Figure 3.1: A pedigree diagram displaying markers genotyped within family 702 (Williams, Norton *et al.*, 2003). The pedigree is shown as the unaffected father (square), unaffected mother for which DNA was absent (circle) and six affected siblings (black-fill square). Given for each individual are marker names (prefix D17S), position (UCSC May 2004) and genotype. Haplotypes are inferred where possible and the maximum regions where the siblings share one (IBD1) and two (IBD2) chromosomal segments are shown in orange and green respectively. The putative maximum IBD2 region shown by a red rectangle spans 58.8Mb (D17S799-D17S785). Missing parental genotypes were inferred where possible.

3.2. Materials and Methods

3.2.1. Ascertainment of C702 Samples

See materials and methods section 2.1.1.

3.2.2 Diagnosis of C702

Family C702 was evaluated by consensus best-estimate diagnosis made on the basis of structured psychiatric interview with SCAN (Schedules for Clinical Assessment in Neuropsychiatry (Wing, Babor *et al.*, 1990)) and review of psychiatric case notes (by Geraldine McCarthy, Stanley Zammit, George Kirov, Elen Russell and Nick Craddock). The diagnoses in the family were also made according to Karl Leonhard criteria by a researcher (George Kirov) experienced with this complicated classification.

3.2.3. Consanguinity

Within-family relationships were confirmed by examination of the degree of homozygosity across the genome using two related methods, both developed in-house by Peter Holmans.

Firstly, the probability of homozygosity in the absence of inbreeding was evaluated for each genetic marker analysed as the expected probability of a homozygous genotype being the sum of the squares of the allele frequencies. The allele frequencies

were obtained from the Williams *et al* study (Williams, Norton *et al.*, 2003). The probabilties were summed to show the expected number of homozygous genotypes for an individual. The number of homozygous genotypes for each C702 test individual is calculated and then a chi square test (1 degree of freedom) performed between observed values and expected values.

The second method takes in to account the frequency of alleles in the test subject(s). The method involved the calculation of an inbreeding coefficient (F), which is the probability that an individual's two alleles at a given locus are IBD:

The probability (P) of a homozygous genotype (for allele *i*) is:

$$P(i/i) = ((F \times Pi) + (1 - F))\ Pi,$$

(where P*i* is the population frequency of allele *i*)

Whereas the probability of a heterozygous genotype (alleles *i* and *j*) is:

$$P(i/j) = (1 - F)\ 2PiPj$$

The calculation was performed for each locus and the overall likelihood obtained by forming the product of the contributions from each locus. Twice the natural logarithm of the ratio of the likelihood maximised with respect to F to its value when F=0 is the test

statistic of the null hypothesis of no inbreeding (i.e. F=0), and has a chi-squared distribution on one degree of freedom when the null hypothesis is true.

3.2.4. Marker Selection

Previously, 372 microsatellite markers were genotyped across the genome in pedigree C702 (Williams, Norton *et al.*, 2003), including 14 markers across the chromosome 17. Included in this study were 36 microsatellite markers in addition to those already analysed in C702 (Williams, Norton *et al.*, 2003). Markers were selected for maximum heterozygosity according to the Marshfield genetic map (http://research.marshfieldclinic.org/genetics/) and also to give an even physical marker spacing across the original Williams *et al* (Williams, Norton *et al.*, 2003) C702 IBD2 region (figure 3.1). The resulting dataset provides a 50 marker map across chromosome 17. The range of marker heterozygosity for all 50 markers was 0.19-0.9 (mean = 0.69), not including D17S2182. The average inter-marker physical distance for all 50 markers was ~1.6Mb or ~2.5cM in genetic distance (sex-averaged, minus D17S2182). However, as the linkage region was refined in stages the marker density increases around the linkage region as it was refined.

All markers are present on the UCSC Genome Browser (May 2004 freeze). All markers but one (D17S2182) are found with details of their genetic map position on the Marshfield Center for Medical Genetics website: (http://research.marshfieldclinic.org/genetics/). All but 10 markers are found on the deCODE genetic and physical map (Kong, Gudbjartsson *et al.*, 2002). The deCODE map

was used as the spatial template for further analysis, but was supplemented by a more detailed 17q map that reverses the order of consecutive markers D17S1809, D17S1825 and D17S1792 (Chen, Saarela *et al.*, 2004). All markers were genotyped in the six affected siblings and father except markers D17S1820 and D17S784 (missing paternal genotype only).

Details of microsatellite markers studied, their chromosomal position and the corresponding PCR primers used for genotyping are given in table 3.1. Example genotypes for some of the markers are given in appendices 10.1.

3.2.5. Genotyping

Protocols for the genotyping of microsatellites can be found in the materials and methods chapter, sections 2.2 and 2.5.1.

Examples of microsatellite marker genotyping in pedigree C702 are shown in appendices 10.1.

Marker Name	PCR primer 1	PCR primer 2	Start position†
D17S849*	Williams et al. 2003	Williams et al. 2003	379287
D17S831*	Williams et al. 2003	Williams et al. 2003	1857160
D17S938*	Williams et al. 2003	Williams et al. 2003	6189992
D17S1852*	Williams et al. 2003	Williams et al. 2003	10456226
D17S799*	Williams et al. 2003	Williams et al. 2003	13111681
D17S1857*	Williams et al. 2003	Williams et al. 2003	16355942
D17S1871	AAGGCCTCTGGATTTGGACT	TCTGATTTAGGTAGGGTTCTCCA	20702466
D17S798*	Williams et al. 2003	Williams et al. 2003	28313925
D17S1818	TCTTGGCACATCTGAAAGCA	AGCCTGGGCAACAAGAGTAA	34416617
D17S1787	CCTTCACGCTTTGTCTCACA	GGTTCCCTTCGTCCTTGAG	36978463
D17S806	GTGGTGGTGTGTGTGTACGG	GGCATTTGTGCCTGGAAG	43166788
D17S1868*	Williams et al. 2003	Williams et al. 2003	44539745
D17S1795	GCATTTCAGACAGTGCCAGAG	AGTGCAGAGGCAGAAGATCC	45279978
D17S1820	TGGTAGGAAATGGCAGAGTC	CCAGTTTTTCTATAAAACCCAAACC	47139854
D17S787*	Williams et al. 2003	Williams et al. 2003	50636864
D17S1807	TTCCTCTCTTCTCCAAATTTTCTT	GACCTCAGATGAGGGGGAAT	51164578
D17S1806	GGTAACTGGTATTCAATCCTGGAG	CTGGCCTGTGTGTTCATGTT	52958631
D17S1160	AGGGATGAAAGGAAAATGCA	TCCAGCATTTAACATACCCA	52996889
D17S1804	CAGCCTGGGTGACAGAAA	GCAATCATGTCAAAACTTTAGAGTAG	56332261
D17S916	CAACAGTGCAACAGTGTGAGACC	AGCCCCCACCCCTGCTAGTCC	56576231
D17S792	CCTGGCTTCTGAGACAGTTC	CTACAGTGAAGGAGTGAAGGC	56594527
D17S923	ACCCAACATTGTAAATGGG	GCTGAGACCATATCACTTAAATC	56723603
D17S1838	CTTTTTACAGTCTTAAAATGAAAGA	GATCACGGGATGGTTT	56807191
D17S808	ACCCTAGACAGGATGCCA	TGTGGGTTTTCTCAGGTTAT	58026831
D17S794	CTGGGGGACAGAAGAAGTCC	CCTGTCTATTCCCAGCCCTA	58170964
D17S924	GATTTCCTCTGGCTTTGTGG	CCACGATGACCAAAATGTGA	58290509
D17S944*	Williams et al. 2003	Williams et al. 2003	58789906
D17S1297	GGTGGCAGATGCCTTTAGTA	GTTTATTCCCTCCCTCTTGC	59567899
D17S1792	GTTTTCGGTTATTAACAACAACAA	CAGGTGGCTGATGGAGA	60127864
D17S1825	GACTTTATCAAGATGTTTCCAGG	TATTGAGCCCAGGAGTTCTA	60458408
D17S1809	TTGCCAAGATACAGGAATG	AAGCCCACCAGGTCAT	60489935
D17S1874	TAGGAGGCAGCAATGG	TGTGGTATGTGTGGCA	61627275
D17S1816	ATGGCACCACTGCACTAT	TCACAGCAAGACTCTGTCAC	61851032
D17S942	CTGATGCCAAGATTTTTTAC	TGATTCTATTTCCCTAATATACTG	61919561
D17S1821	GAAACCAGGGCAGTGCTAT	CCAAAACGGACAGGCTATT	62715931
D17S1813	TGACCCAGGTCTCCTCATTC	GTAGAGACGGGGTTTCACCA	62916451
D17S2193	TCTCTAATCTGCCCAAGCAC	ACACTTAGGTCATTGTGTGGC	64059280
D17S789	AGCCTGGGTGACAGAGTGAG	TGTCTTCAGCCGAATTTTCC	64140110
D17S940	CCTGTTTTTTGATGTACTTATCTC	TCTTTAATGGAATTTTCATACC	64589957
D17S2182	TCTGGACAGGTTATTATACTGGG	CAGGGTGAGCAAGCTGTATT	64851294
D17S1295	TCTAGCCGTTGTGTGTGGTA	TTCTCAGACTCTCTCAGTCTTGG	65525565
D17S949*	Williams et al. 2003	Williams et al. 2003	65976927
D17S1350	TCACAAATGTCATATATCCA	TGGCAATCTAACAGATGAGA	66600357
D17S1797	CTAAGGCATTCATTTGCG	GGACTTATCTTTAACTGGGCTG	68037333
D17S617	CAGCTTTCAGTTTCCTGAGTGTTG	CCTTGACATCAGACTAGGATGAGTT	68828387
D17S1352	CACCTTCATGGGGCTTAC	CTAGCACTCTGCCTTCCAAC	69509436
D17S1301	AAAGAAGATGAAATTGCCATG	TAAAAAGAATGAAGGTAAAAATGTG	70192381
D17S785*	Williams et al. 2003	Williams et al. 2003	71942895
D17S784*	Williams et al. 2003	Williams et al. 2003	75416716
D17S928*	Williams et al. 2003	Williams et al. 2003	77846128

* Microsatellite markers analysed by Williams *et al* (2003)
† Chromosome 17 start position (JCSC Build 35 May 2004)

Table 3.1: Microsatellite markers studied. PCR primers are given for those markers analysed in this study. Position of polymorphic repeat start site is given for all markers (Chromosome 17. NCBI Build 35).

3.3. Results

3.3.1. Pedigree C702

Figure 3.2 shows a pedigree diagram of family C702.

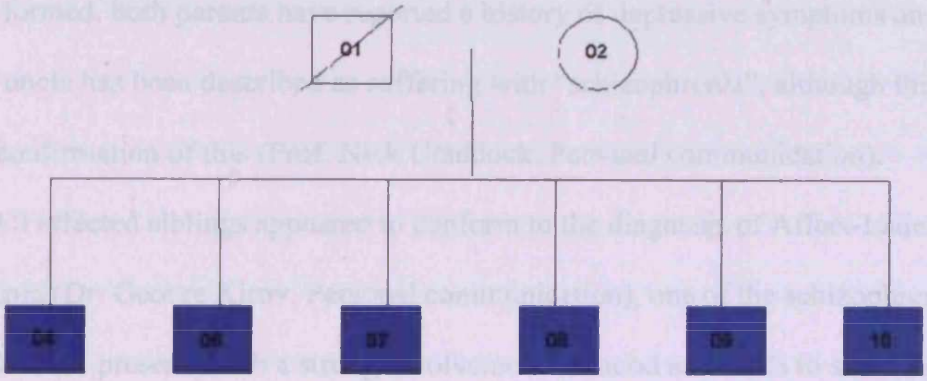


Figure 3.2: Pedigree C702. Shown are the six male siblings diagnosed with DSM-IV Schizophrenia (shaded blue) and the parents (white) for whom no formal psychiatric evaluation has been performed, however no severe psychiatric illness has been reported. Sample identifiers are shown within each box/circle. DNA is available for the 6 male affected siblings and the father only.

The affected siblings of pedigree C702 were diagnosed with DSM-IV and ICD-10 schizophrenia by several psychiatrists. The siblings all show an age-at-onset of 17-22 and a chronic course of illness, with $\geq$ 4 psychiatric hospital admissions for each sibling. Of the 6 affected siblings, three have documented attempts at suicide while the others have reported suicidal ideation. Their illness is characterized by prominent features of

pathological mood disturbance including both manic and depressive symptoms, but not of sufficient extent to warrant a diagnosis of DSM-IV schizoaffective disorder (Prof. Nick Craddock, personal communication). Both parents have a history of depressive symptoms and a paternal uncle is described by the family as suffering with "schizophrenia", however there is no clinical confirmation of this observation (Prof. Nick Craddock. Personal communication). While no formal psychiatric assessment of the parents has been performed, both parents have reported a history of depressive symptoms and a paternal uncle has been described as suffering with "schizophrenia", although there is no clinical confirmation of this (Prof. Nick Craddock. Personal communication).

All affected siblings appeared to conform to the diagnosis of Affect-Laden Paraphrenia (Dr. George Kirov. Personal communication), one of the schizophrenia subtypes which presents with a strong involvement of mood and tends to show bipolar features (Dr. George Kirov. Personal communication). Outcome of this subtype can vary, but typically patients develop severe deterioration of personality after successive acute episodes (Dr. George Kirov. Personal communication).

3.3.2. Consanguinity

Table 3.2 shows the results of two methods of testing for consanguinity of pedigree C702 (Analysis performed by Peter Holmans). The table shows the 6 affected siblings of pedigree C702 (Samples 4, 6, 7, 8, 9 and 10), the number of homozygous and heterozygous genotypes for each sample (from a genome-scan of 372 autosomal

microsatellite markers) and an expected value for the number of homozygous genotypes (assuming no inbreeding).

The chi square statistic for the two methods of analysis is then shown: CHI1 corresponds to the measure of inbreeding based purely on the expected number of homozygous genotypes assuming no consanguinity (E(HOM)). CHI2 takes into account the frequency of alleles that an individual is homozygous for. A chi square of 2.706 gives a p value of 0.05 and none of the siblings achieve this value indicating that the affected siblings of pedigree C702 are not the result of a consanguineous mating. The power of CHI2 (the more informative and therefore powerful measure) based upon the 372 genome scan markers to detect consanguinity in an individual ranges from a power of 20% to detect a second cousin mating, 87% to detect a first cousin mating and 100% to detect a sibling mating (at $p \leq .05$) when F (the maximised inbreeding coefficient) =0.015 corresponds to second-cousin mating, F=0.0625 to first-cousin mating and F=0.25 to sib mating. The pedigree C702 values for F are consistent with a measure of parent relatedness that is approximately equal to or less than a second cousin mating.

FAMILY	SAMPLE	HOM	HET	E(HOM)	CHI1	CHI2	F	F-LO	F-HI
702	4	85	261	75 1	1.305	2.319	0.033	0	0.083
702	6	72	274	74 99	0	0 193	0 008	0	0 052
702	7	76	272	75 3	0.006	0.022	0.003	0	0 051
702	8	72	275	75.32	0	0	0	0	0.047
702	9	87	262	75.64	1.705	2.003	0.03	0	0.082
702	10	81	268	75.45	0.408	0	0	0	0.048

HOM= number of homozygous genotypes
HET = number of heterozygous genotypes
E(HOM) = expected value of HOM assuming no inbreeding
CHI1= chi-square based on HOM, E(HOM) - i e not taking account of frequency of homozygosity for alleles
CHI2 = chi-square based on genotype likelihoods - i e taking into account the frequency of homozygous alleles (rare alleles give more evidence for inbreeding).
F = max likelihood estimate of inbreeding coefficient
F-LO, F-HI upper and lower 95% confidence interval for F

Table 3.2: Results of consanguinity testing using microsatellite marker genotypes and allele frequencies from a previous study (Williams, Norton *et al.*, 2003) and algorithms developed and implemented by Peter Holmans.

3.3.3. Marker and Haplotype Analysis

A total of 50 microsatellite markers spanning chromosome 17 were analysed in pedigree C702, consisting of the 14 markers studied previously (Williams, Norton *et al.*, 2003) supplemented by a further 36 microsatellites.

The original Williams *et al* study (Williams, Norton *et al.*, 2003) identified maternal and paternal haplotypes constructed from the genotype data from the pedigree C702 offspring and father (figure 3.1), with maternal genotypes inferred from these where possible. The transmitted paternal haplotype could be refined only at the telomeric end where marker D17S1301 showed a recombination of the inherited paternal haplotype in sibling 702-08. The shared paternal haplotype is defined at either end by markers

D17S831 and D17S1301, giving a maximal region of chr17:1857150-70192381 (68.3Mb).

In this study an additional 36 markers were genotyped and added to the 14 markers from the original study (Williams, Norton *et al.*, 2003). The pedigree diagram displaying all microsatellite markers genotyped in pedigree C702 and the maximum shared haplotype regions they define is shown in figure 3.3. Lod score analysis was not performed using the high-density map as the markers are likely to be in some degree of LD and accurately gauging allele frequencies for 36 microsatellites was not economical. The siblings share chromosomal segments apparently Identical By Descent 2 (IBD2), shown in figure 3.3 spanning a genetic distance of 12.57cM (Kong, Gudbjartsson *et al.*, 2002) corresponding to the chromosomal region 17q23.2-q24.3.

The refined IBD2 region contains 21 consecutive markers (including marker D17S944 from the original study (Williams, Norton *et al.*, 2003)) and is defined at the proximal end by two consecutive recombinant markers in individual 702-10 (D17S1160), and at the distal end by another two consecutive recombinant markers in 702-04 (D17S2182) (see figure 3.3). This refined IBD2 region corresponds to chr17:52,996,269-64,651,294 (UCSC May 2004 freeze). The IBD2 region contains approximately 140 known genes (NCBI Map Viewer) and is 11,655,026 bp in size. Therefore, a genome-wide significant region of linkage for schizophrenia has been refined from a 69.2Mb region to an 11.7 Mb region at 17q23.2-q24.3.

Analysis of the refined IBD2 for stretches of homozygosity that could indicate the convergence of haplotypes bearing a founder mutation shows a contiguous stretch of 5

microsatellite markers from D17S944 to D17S1809, while several markers homozygous in isolation (D17S808, D17S942 and D17S940). These and other regions are examined further in chapter 4.

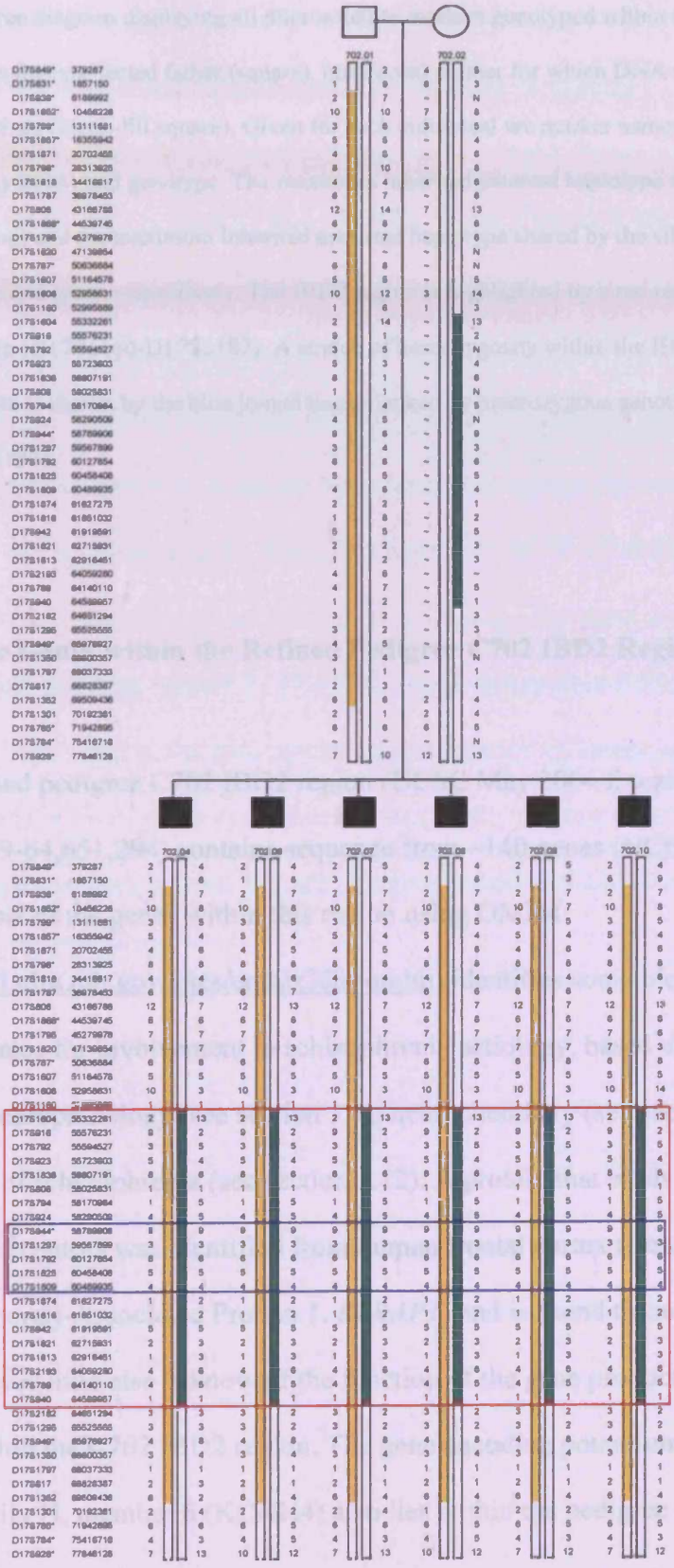


Figure 3.3: A pedigree diagram displaying all microsatellite markers genotyped within family 702. The pedigree is shown as the unaffected father (square), unaffected mother for which DNA was absent (circle) and six affected siblings (black-fill square). Given for each individual are marker names (prefix D17S), position (UCSC May 2004) and genotype. The maximum inherited paternal haplotype shared by the siblings (IBD1 region) and the maximum inherited maternal haplotype shared by the siblings (IBD2 region) are shown in orange and green respectively. The IBD2 region is highlighted by a red rectangle and spans a maximum of 11.7Mb (D17S1160-D17S2182). A stretch of homozygosity within the IBD2 region spanning a maximum of 3.3Mb is shown by the blue joined lines, flanked by heterozygous genotypes at markers D17S924 and D17S1874.

3.3.4. Candidate Genes within the Refined Pedigree C702 IBD2 Region

The refined pedigree C702 IBD2 region (UCSC May 2004 freeze, chr17:52,996,269-64,651,294) contains sequence from ~140 genes (NCBI Map Viewer). Literature analysis of the genes within this region using OMIM (http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim) identifies some biologically plausible candidates for involvement in schizophrenia aetiology, based upon a knowledge of the putative neuropathology (see section 1.4), neurochemistry (see section 1.4) and current genetics of schizophrenia (see section 1.12). A protein that binds to benzodiazepine receptors was identified from human frontal cortex (Benzodiazapine Receptor (Peripheral)-Associated Protein 1. *BZRAP1*) and is found throughout limbic structures, however little else is know of the function of the gene product that is encoded by sequence within the C702 IBD2 region. The gene encoding potassium voltage-gated channel, subfamily H, member 6 (KCNH4) also lies within the pedigree C702 refined

IBD2 region and has expression primarily in telencephalic regions of the brain and is likely to be involved with controlling neuronal excitability. Regulator of G-protein Signalling 9 (*RGS9*) is expressed in forebrain regions receiving dopaminergic innervation and is therefore has strong plausibility as a candidate gene. The gene encoding Protein Kinase C Alpha (*PRKCA*) has been implicated in the prefrontal cortical regulation of behaviour and thought and also in episodic memory and is also therefore a strong biological candidate. A further candidate is the *FALZ* gene, encoding Fetal Alzheimer Antigen, which shows expression changes from being throughout the neuron to being nucleus specific during development. Also, the expression of the encoded protein is high in neurodegenerative diseases such as Alzheimer's Disease. The gene encoding Amyloid β Precursor Protein Binding protein 2 (*APPBP2*) binds to Amyloid β Precursor Protein (*APP*), which is implicated in the pathogenesis of Alzheimer's Disease and may have some relevance to the C702 phenotype. Finally, Clathrin is a major protein component of intracellular organelles such as vesicles, and the gene encoding Clathrin (*CLTC*) lies within the refined IBD2 region.

3.4. Discussion

Original genotyping of pedigree C702 (Williams, Norton *et al.*, 2003) showed that all affected siblings shared the same genotype at 6 consecutive markers on chromosome 17 and shared the same allele over a further 5 flanking markers, indicating a shared inherited maternal and larger paternal haplotype. The genome-wide significant linkage generated from these haplotypes over the putative IBD2 region is compelling. However the linked region covered much of the chromosome, which presented a considerable hurdle to further molecular genetic analysis.

The Williams *et al* study (Williams, Norton *et al.*, 2003) concluded "Secondary analysis has allowed us to identify maternal and paternal haplotypes spanning markers D17S1852 to D17S785 and D17S1857 to D17S949, respectively,". However this can only be affirmed where the haplotypes are demonstrably shared copies from each parent. Nevertheless, some marker alleles are unquestionably IBD (Williams, Norton *et al.*, 2003): Markers D17S799, D17S798, D17S1868, D17S944 and D17S949 form the haplotype 3-8-6-9-4, which can only be transmitted from the father and is transmitted to all the siblings forming a large IBD1 region. Therefore, it is likely that the siblings share a large paternally transmitted haplotype, and what is shown in figure 3.1 is the possible extent of that IBD1 region (orange) spanning markers D17S938 to D17S784 (chr17:6190259-75416716, UCSC May 2004).

The transmitted maternal haplotype and sharing status is also a debatable feature of the original C702 analysis (Williams, Norton *et al.*, 2003). The siblings share identical genotypes for 6 markers (figure 3.1) and are likely to be IBD1 across this region.

However, the question of whether this IBS region shows the same inherited paternal and maternal haplotype in each sibling (an IBD2 region) is unanswerable without maternal genotypes or further genotyping. This study attempted to irrefutably narrow the putative IBD2 linkage region by fine mapping to identify recombinant siblings that would refine the region to a size amenable to directed re-sequencing.

The genotyping of an additional 36 microsatellites to the 14 originally genotyped (Williams, Norton *et al.*, 2003) elucidated multiple recombinations within the original IBD2 region of C702 (figure 3.3), as expected from the estimates of the recombination rates across the region (Kong, Gudbjartsson *et al.*, 2002). The marker defining the centromeric boundary IBD2 region is D17S1160, where sibling 702-10 was recombinant. The adjacent marker (D17S1606) is also recombinant in the same sibling supporting the view that a different maternal chromosomal segment has been inherited by this sibling. The telomeric end of the original IBD2 region (Williams, Norton *et al.*, 2003) was refined by marker D17S2182, showing evidence of a recombination in sibling 702.04. The refined C702 IBD2 linkage region includes 21 microsatellite markers and spans 11655025bp (UCSC May 2004. D17S1160-D17S2182) and incorporates 153 genes (NCBI Map Viewer). Therefore, the region of linkage is now 57,571,432 bases smaller than reported in Williams *et al,* (Williams, Norton *et al.*, 2003).

It is important to note that a marker from the original study (Williams, Norton *et al.*, 2003) still remains within the current IBD2 region (D17S944) which therefore supports the view that the original linkage was not due to chance.

The refined maternal and paternal haplotypes still span a substantial proportion of the chromosome (11.7 and 69.2 Mb respectively), however the current dataset may be

able to refine the region further. As both parents are unaffected it is plausible that the IBD2 linkage signal favours a founder disease allele residing on both the paternal and maternal haplotypes and interacting to cause disease. Such a recessive mechanism is also supported by the available schizophrenia and bipolar disorders linkage data (Bennett, Segurado *et al.*, 2002, Dick, Foroud *et al.*, 2003, Ewald, Wikman *et al.*, 2005, Klei, Bacanu *et al.*, 2005, McInnis, Dick *et al.*, 2003, Rees, Fenton *et al.*, 1999, Tomas, Canellas *et al.*, 2006) and anecdotally by the possible recessive mode-of-inheritance of "Affect-laden paraphrenia" (Ban, 2004). If true, then it is most likely the C702 linkage is explained by the convergence of alleles from a common founder ancestor. However, an alternative supposition is that the allele sharing at chromosome 17 is a consequence of the pedigree being inbred.

To disprove that consanguinity explains the C702 linkage data the genome-wide linkage scan data of the Williams *et al* study (Williams, Norton *et al.*, 2003) was used. The dataset provides genotypes for 372 microsatellite markers spaced throughout the genome in 353 ASPs from the UK and Ireland (including C702). Each of the 6 family C702 affected siblings was first analysed for an excess volume of homozygous marker genotypes as compared to the expected value obtained from the whole sample. No evidence for inbreeding was found in any of the siblings by this method. However homozygosity for several rare alleles could indicate the pedigree is inbred. Therefore, a second method analysed the expected frequency of homozygous genotypes and compared that to those observed for the six C702 siblings. No evidence was obtained to suggest the mother and father are first or second degree relatives and therefore all the current data would suggest that the linkage obtained for pedigree C702 is not due to consanguinity.

Indirect supportive evidence comes from the whole genome C702 linkage data as consanguinity should generate multiple regions of strong linkage, however the next most significant multipoint lod score (MLS) was a value of 2.93 on chromosome 20p12, which does not approach genome-wide significance (Dr. Marian Hamshere. Personal communication). In summary, this data strongly suggest that the linkage at family C702 is not solely attributable to inbreeding.

The mapping of a recessive allele within the C702 IBD2 region may be helped by autozygosity or homozygosity mapping, which has been successful in other disorders (Saba, Montpetit *et al.*, 2005, Strachan and Read, 2003b, Valente, Abou-Sleiman *et al.*, 2004). Dependent upon the age of the allele and the rates of recombination and mutation, the location of the disease allele will be surrounded by the same alleles resulting in homozygous genotypes and a homozygous diplotype. Therefore, the large region of homozygosity (the diplotype of markers D17S944-D17S1809) and homozygous markers D17S808, D17S942 and D17S940 could identify a locus IBD2 for the disease allele. However, the linkage disequilibrium is not known between these markers and the 5 marker haplotype may be common. Furthermore, the density of markers leaves many megabase gaps which could harbour a homozygous region and so an even higher density genetic map is required (see chapter 4).

The same contentions of the Williams *et al* (Williams, Norton *et al.*, 2003) findings apply to this study. Again, without maternal DNA the exact regions IBD1 and IBD2 cannot be definite. It therefore remains plausible that some of the peripheral markers included in the maximum possible IBD1 and IBD2 regions will prove to be IBS. However, it is unlikely that all 21 markers apparently IBD2 are all falsely so, just as it is

unlikely that the mother is homozygous for all 21 markers, unless extreme LD exists between these markers (i.e. the diplotype is common).

The lack of widely replicated linkage to the 17q region in schizophrenia and bipolar disorder suggests that the disease allele at this locus is unlikely to exhibit a strong dominant effect and/or be common (i.e. $\geq$ 1-5% disease allele frequency). The 17q region is an infrequently reported region of linkage to both schizophrenia and bipolar illnesses, however strikingly all studies overlap with the C702 IBD2 region (figure 3.4). Also, many of the linkage studies exhibiting corroborative evidence for linkage to the same 17q region do so under a recessive mode of inheritance (figure 3.4).

Considering the current data, a reasonable proposition is that the affected C702 siblings represent the convergence of a highly penetrant disease allele from a coalescent founder ancestor resulting in a phenotype that appears recessive in this pedigree. A further supposition is that the mood symptoms exhibited by the C702 siblings and the recessive nature of both the C702 and literature linkage to overlying regions supports the view that the disease relevant gene and biological mechanism conferring risk is common to all populations. Finally and controversially it could be argued that the disease allele itself may be common to all studies, however if this is the case the allele must be extremely old, rare and recessively acting to explain the dearth of linkage findings to this region.

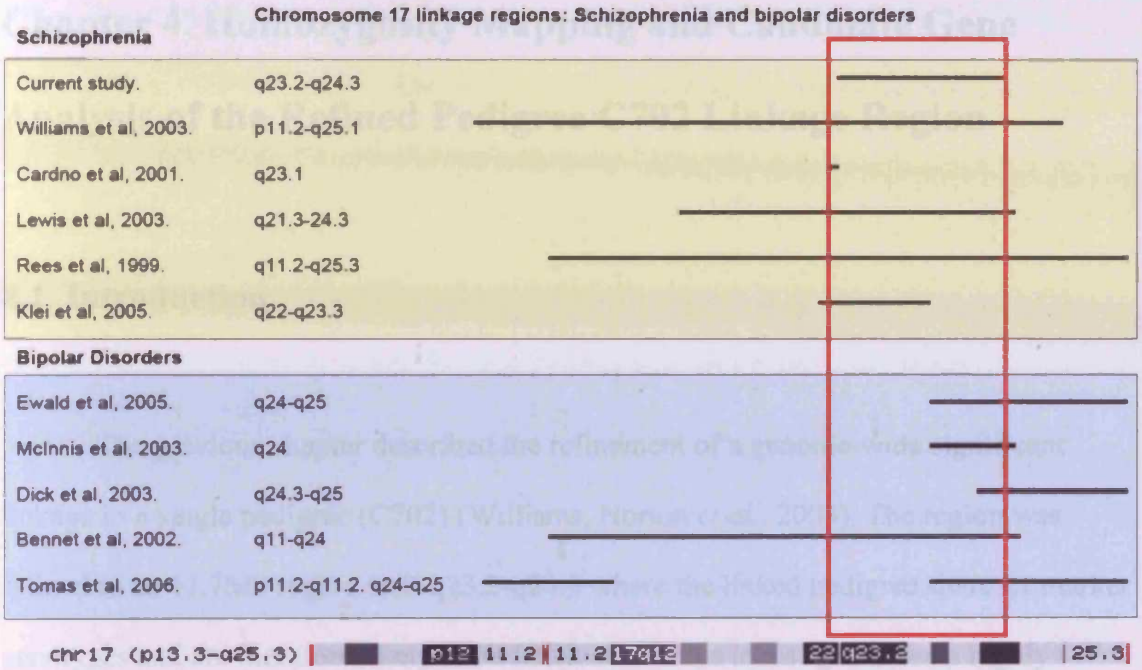


Figure 3.4: Literature evidence suggests chromosome 17 harbours a disease locus for schizophrenia, bipolar disorders and related illnesses. Shown are regions showing evidence for linkage in each of the studies and the novel C702 linkage region reported in this study. Studies are grouped either as "Schizophrenia" or "Bipolar Disorder" according to the principal phenotype analysed by the study. The red box indicates a region of linkage common to a large proportion of reports including this study. Note that the Current study and several others (Cardno, Holmans *et al.*, 2001, Lewis, Levinson *et al.*, 2003, Williams, Norton *et al.*, 2003) have overlapping samples.

Chapter 4. Homozygosity Mapping and Candidate Gene Analysis of the Refined Pedigree C702 Linkage Region

4.1. Introduction

The previous chapter described the refinement of a genome-wide significant linkage in a single pedigree (C702) (Williams, Norton *et al.*, 2003). The region was refined to an 11.7Mb region at 17q23.2-q24.3 where the linked pedigree share 21 marker genotypes and are therefore likely to be IBD2 across this locus. Therefore a highly penetrant allele(s) or genotype for schizophrenia may exist within the refined region.

The IBD2 status of all the affected individuals of family C702 combined with the lack of schizophrenia in the parents is suggestive of recessive genetic mechanism acting at this locus to induce schizophrenia. This is concordant with the growing body of literature implicating 17q as a linkage region for schizophrenia and related phenotypes that indicate a recessive genetic mechanism as the most likely explanation of linkage to this region (Bennett, Segurado *et al.*, 2002, Dick, Foroud *et al.*, 2003, Ewald, Wikman *et al.*, 2005, Klei, Bacanu *et al.*, 2005, McInnis, Dick *et al.*, 2003, Rees, Fenton *et al.*, 1999, Tomas, Canellas *et al.*, 2006).

The pattern of linkage observed in C702 is consistent with the convergence of disease alleles from a common founder ancestor. Founder disease alleles have been identified in consanguineous and non-consanguineous pedigrees by autozygosity mapping where stretches of homozygosity in linked regions are used to identify a disease

locus (Moynihan, Bundey *et al.*, 1998, Saba, Montpetit *et al.*, 2005, Valente, Abou-Sleiman *et al.*, 2004, Vazza, Zortea *et al.*, 2000). The C702 recessive acting disease allele would have occurred on a specific haplotype and depending upon the age of the disease allele the haplotype may have been broken up by recombination or altered by new mutations. Consistent with a founder model, the convergence of the disease haplotype in C702 affected individuals would be expected to result in a contiguous stretch of homozygous markers of the risk haplotype (i.e. a homozygous diplotype). Therefore, regions of homozygosity within the refined C702 IBD2 region may distinguish the location of the genetic variant responsible for the C702 phenotype.

This chapter describes the analysis of microsatellite genotypes in pedigree C702 to identify several regions of homozygosity. Directed resequencing of candidate genes in these regions was then undertaken and association analysis performed in a schizophrenia case-control sample from the UK by use of DNA pools. Concurrently, one of the C702 siblings was genotyped as part of a whole-genome association study of mood and psychotic disorders (Wellcome Trust Case Control Consortium, 2007) and schizophrenia (O'Donovan *et al*, submitted). This SNP map was used to further characterise the homozygous regions across the refined pedigree C702 IBD2. Homozygous SNP diplotypes in C702 were analysed in a sample of 2938 UK controls and 476 UK cases (O'Donovan *et al*, submitted) looking for a rare or associated diplotype that could indicate the presence of the C702 disease variant.

4.2. Materials and Methods

4.2.1. DNA Samples

Family C702 was acquired as part of a whole-genome linkage scan of schizophrenic pedigrees from the UK and Ireland (Williams, Norton *et al.*, 2003, Williams, Rees *et al.*, 1999).

DNA pools were constructed from a UK schizophrenia case and blood donor control sample consisting of 709 cases and 710 controls (see section 2.1.4). The sample was divided into four matched stages each containing 184 cases and an equal number of controls, except for the fourth stage, which contains 157 cases and 158 controls. Each DNA sample was quantified using Pico Green (Invitrogen) and a Fluoroskan Ascent fluorimeter (Thermo Labsystems) and samples diluted to approximately 4ng/µl (±0.5ng/µl), as detailed in section 2.1.3. The construction of separate case and control pools was performed using a Microlab S (Hamilton) automated sample processor. Samples for DNA pools were quantified by multiple personnel (including myself) and pools were constructed by Anna Preece and Nadine Norton (Norton, Williams *et al.*, 2002).

The association sample used in this chapter for homozygosity mapping using the Affymetrix GeneChip platform consists of 476 schizophrenics from the UK and Ireland (O'Donovan et al, submitted) and 2938 controls from the UK (O'Donovan et al, submitted (Wellcome Trust Case Control Consortium, 2007). For a full description of these association samples, see section 2.1.1.

4.2.1. Affymetrix GeneChip Human Mapping 500K Array Set Genotyping

Genotyping of a C702 affected sibling (702-06), 2938 UK controls and 476 schizophrenics from the UK was performed by Affymetrix using the Affymetrix GeneChip Human Mapping 500K Array Set according to their standard protocols. The genotyping was performed conjointly as part of The Wellcome Trust Case Control Consortium (WTCCC) project to identify genetic variants for complex diseases, in this case for mood-psychosis spectrum psychiatric disorders ((Wellcome Trust Case Control Consortium, 2007) and O'Donovan *et al*, submitted). Quality control protocols have been performed by Affymetrix and the WTCCC, where dubious samples and assays were removed (e.g. due to low call rate, errors in Hardy-Weinberg Equilibrium, internal control genotyping errors). The data are then stored on a departmental database (http://x001.psycm.uwcm.ac.uk/).

4.2.2. Homozygosity Mapping

Homozygosity Mapping of the C702 IBD2 region (chr17:52995889-64651294. UCSC May 2004 freeze) followed a three step approach:

First, all regions of homozygosity in the C702 affecteds were defined using the microsatellite data (chapter 3) as being until the nearest flanking heterozygote genotyped marker. The second approach looked specifically at the largest maximal homozygous section in the C702 IBD2 region identified in chapter 3, consisting of 5 contiguous homozygous genotypes and spanning a maximum of 3,336,402bp (D17S924-D17S1874),

which was the only homozygous region to have contiguous homozygous markers. To validate or disprove this region, 41 known SNPs (dbSNP) and flanking sequence spanning the region were bi-directionally sequenced in C702 sibling and ≥ 1 unrelated control (see materials and methods section 2.5.2). The SNPs assayed and corresponding PCR and sequencing primers are shown in appendices 10.2. The homozygous regions obtained from this set of markers (table 4.2) were used to define candidate genes for mutation screening.

A third approach was performed at a later date using the Affymetrix Gene-Chip technology. The study involved the genotyping of 476 UK schizophrenia cases, a pedigree C702 sibling (702.06) and 2938 UK controls (O'Donovan et al, submitted . The Affymetrix SNP genotypes generated and unbiased map of where 702.06 showed contiguous homozygous marker genotypes. A homozygous region was defined as ≥ 2 homozygous markers (SNP or microsatellite). The C702 SNP diplotypes were analysed for their presence and frequency in the case-control sample, using a Microsoft Excel macro designed by Dr. Valentina Moskvina and implemented by myself. The macro firstly excluded individuals with a failed genotype and then summed the number of cases and controls that had an identical diplotype to the genotyped C702 sibling. If present at high enough frequency, the diplotypes were analysed for association by a standard chi-square test and rare diplotypes were also noted. Note that only Affymetrix SNPs were included in this analysis as the case-control sample has not been genotyped for the other markers described in this chapter.

Note, that a homozygous loss of DNA could explain the C702 IBD2 linkage region and therefore stretches of marker fails (for the C702 sibling) were also noted.

4.2.3. Mutation Screening

The UCSC genome browser (May 2004 freeze) was used to gain the complete DNA sequence for identified candidate genes. Exonic sequences were defined based upon the "Known Genes" and "Human mRNAs from GenBank" tracks from the genome browser. Potential regulatory regions were classified as 1kb 5' of the first base of the first exon (putative core promoter region) and all regulatory motif sequences ascertained using the algorithm ClusterBuster (http://zlab.bu.edu/cluster-buster/). PCR and bi-directional sequencing (see materials and methods chapter 2.2.2) was performed over all of the potentially functional sequences, in a C702 sibling and $\geq$ 1 unrelated control(s) as a reference. Any polymorphisms identified were examined for association by allele frequency estimation using case and control DNA pools (see materials and methods chapter, sections 2.1.3 and 2.1.4).

PCR/Sequencing primers used for mutation screening are shown in appendices tables 4.A2.

4.2.4. Association Analysis Using Pooled Case and Control DNA samples.

Each case and control pool and was simultaneously assayed using PCR followed by SNaPshot chemistry (Applied Biosystems) (see sections 2.2 and 2.5.4) and analysed by capillary electrophoresis using an ABI3100 Genetic Analyser (Applied Biosystems) (Norton, Williams *et al.*, 2002). Fluorescent peak heights for each allele were ascertained through Genescan and Genotyper software (Applied Biosystems).

Primer extension products may be un-equally represented by the fluorescence values obtained (e.g. possibly due to differential PCR efficiency for different alleles, or differential incorporation of ddNTPs). To account for this unequal representation of alleles, the estimated allele frequencies from pools were corrected using the mean of the ratios obtained from analysis of $\geq$ 2 heterozygotes. Allele frequencies in case and control pools are then estimated using the formula:

$$\text{Allele Frequency} = A/ (A + (k \times B))$$

Where: A = Peak height of Allele 1 (allele frequency to be estimated)

B = Peak height of Allele 2

k = Average ratio of alleles, e.g. from a heterozygote

(Peak Height Allele 1/Peak Height Allele 2)

The calculation is applied to each sample pool, then both case and control pool frequencies are combined and a mean taken. The allele frequencies are then converted to allele counts by multiplying by the number of chromosomes present in the case and control groups. The result allele counts are placed into a 2 x 2 contingency for chi square analysis. All pooling allele frequency estimation and case-control analysis was performed using Microsoft Excel macros designed by Dr. Nigel Williams and Dr. George Kirov.

Extension primers used for single base primer extension and allele frequency estimation in DNA pools by SNaPshot chemistry are also shown in appendices tables 4.A2.

4.3. Results

4.3.1. Homozygosity Mapping using Microsatellite Markers and Sequencing

Analysis of the 21 microsatellite markers in family C702 that define the refined IBD2 region (chapter 3), identified 4 regions that contained homozygous markers. Only one region contained contiguous homozygous marker genotypes, chr17:582908730-61627275 (flanking markers D17S924-D17S1874). The maximal extent of these regions (i.e. until the next flanking heterozygous genotyped marker) and the microsatellite markers that define those regions are shown in table 4.1.

Microsatellite Marker Homozygous Regions	
Maximal Region	**Flanking Heterozygous Markers**
chr17:56807439-58170964	D17S1838-D17S794
chr17:58290873-61627275	D17S924-D17S1874
chr17:61851420-62715931	D17S1816-D17S1821
chr17:64140423-64651294	D17S789-D17S2182

Table 4.1: Maximum possible homozygous regions in the C702 IBD2 region based upon the microsatellite data shown in chapter 3. Positions were gained from UCSC May 2004 freeze. The region defined by markers D17S924 and D17S1874 spans 5 consecutive homozygous microsatellite markers. The other regions contain one homozygous marker genotype.

To further characterise the largest homozygous region defined by the microsatellite data containing 5 contiguous markers and spanning 3,336,402bp, 41 known SNPs and flanking sequence were sequenced in a C702 sibling and unrelated control. Table 4.2 shows the results of this genotyping and displays that the 5 contiguous microsatellites are not a stretch of homozygosity. There are therefore 6 resulting regions within the IBD2 segment where the microsatellite data supplemented by the SNP sequencing results indicated that the C702 siblings are homozygous across that locus (where a locus contains $\geq$ 1 microsatellite). These regions are shown in figure 4.1.

Marker	C702 Genotype	UCSC May 2004 position
D17S924	HET	58290509
rs241062	HET	58383684
rs958334	HET	58557996
rs7214345	HOM	58590764
rs1268007	HOM	58627283
rs3744279	HOM	58699106
rs9913301	HOM	58707078
rs2429426	HOM	58746777
D17S944	HOM	58789906
rs10491168	HOM	58795195
rs4968648	HOM	58895111
rs4295	HOM	58910030
rs4311	HET	58914495
rs4267385	HOM	58937488
rs6504165	HET	58983808
rs3760252	HOM	59205918
rs2058194	HOM	59374080
rs6504204	HET	59566252
D17S1297	HOM	59567899
rs2047153	HOM	59623901
rs4968716	HET	59719211
rs1550792	HOM	59757060
rs4968723	HOM	59770666
rs2070782	HET	59785865
rs6504227	HET	59831227
rs3744409	HET	59922875
rs1991402	HET	60043792
rs6504248	HET	60101961
D17S1792	HOM	60127854
D17S1825	HOM	60458408
D17S1809	HOM	60489935
rs1476171	HOM	60493874
rs7218921	HOM	60499438
rs299842	HOM	60551691
rs938299	HOM	60578935
rs908089	HOM	60610404
rs8077696	HOM	60623023
rs6504297	HOM	60783427
rs3926150	HOM	60835391
rs9908987	HOM	60878953
rs11868547	HET	60954065
rs1468329	HET	61003783
rs876843	HOM	61081459
rs4377195	HET	61148391
rs2052153	HOM	61230844
rs972952	HET	61368935
rs11079641	HET	61500109
D17S1874	HET	61627275

Table 4.2: The large microsatellite defined homozygous region spanning 5 homozygous markers (maximal region D17S924-D17S1874) underwent sequencing encompassing known SNPs. Shown is each marker and the corresponding C702 genotype (either HET=heterozygous, or HOM=homozygous).

4.3.2. Candidate Gene Analysis

The homozygous regions defined by microsatellite and SNP sequencing data (section 4.3.1) are shown in figure 4.1 and table 4.2. The regions overlap with a total of 41 known and hypothetical genes (UCSC May 2004). The prioritisation of candidate gene mutation screening focused upon microsatellite defined homozygous regions that had been refined by SNP sequencing (table 2). Although this approach is biased as the whole IBD2 region was not assayed for SNPs, these regions were considered more reliable than homozygous regions defined by only one microsatellite. Furthermore, any gene in any homozygous region that was considered a strong functional candidate was also screened.

According to the above criteria, 5 genes were chosen for mutation screening. The *PRKCA* gene encoding Protein Kinase C Alpha is a strong functional candidate and is considered in chapter 6. The gene *RGS9* encodes the protein Regulator of G-protein Signalling 9, which belongs to a family of proteins that when activated restrict G-protein signalling, and are therefore key mediators of chemical transmission within cells (Ishii and Kurachi, 2003). *RGS9* is primarily expressed within neurones, particularly dopaminergic neurones projecting from the striatum to the prefrontal cortices, and also their post-synaptic targets (Song, Waataja *et al.*, 2006). The enzyme is also thought to interact with Dopamine D2 receptors (Kovoor, Seyffarth *et al.*, 2005), implicated in both the pathogenesis of schizophrenia and in the action of antipsychotics (Moghaddam and Krystal, 2003, Waddington, Kapur *et al.*, 2003). Furthermore, another family member, *RGS4*, has been implicated as a schizophrenia susceptibility gene (Chowdari, Mirnics *et*

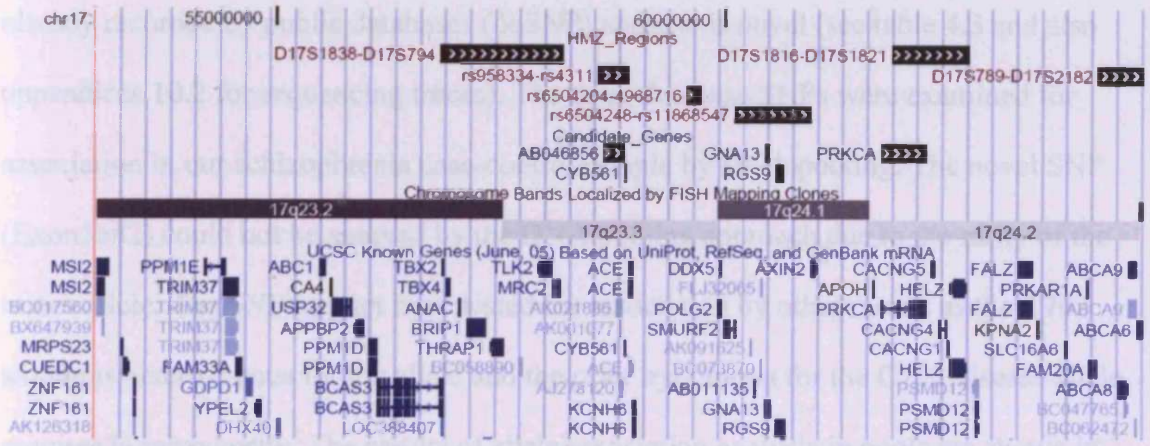
al., 2002, Mirnics, Middleton *et al.*, 2001, Talkowski, Seltman *et al.*, 2006), making RGS9 a plausible candidate susceptibility gene for schizophrenia.

Guanine nucleotide binding proteins regulate the transduction of cell signalling at G-protein coupled receptors (Hill, 2006). As a member of this family, the *GNA13* gene encodes Guanine Nucleotide binding protein Alpha 13, which has a role in neurite motility (Kuner, Swiercz *et al.*, 2002) and also embryonic and neuronal development in C. Elegans (Yau, Yokoyama *et al.*, 2003). Therefore, the limited physiological study of this protein suggests involvement with neural signalling which may be disrupted in schizophrenia (Owen, O'Donovan *et al.*, 2005). Furthermore, evidence suggests that the wider family of G-protein coupled neurotransmitter receptors and G-proteins may be involved in the pathogenesis of schizophrenia (Hill, 2006, Mirnics, Middleton *et al.*, 2001, Talkowski, Seltman *et al.*, 2006).

Cytochrome b561 (CYB561) is a transmembrane protein that is found primarily in catecholamine and neuropeptide secretory vesicles of neuroendocrine tissues, the hippocampus, thalamus and also at moderate levels in the cerebral cortex of mice (Srivastava, Pollard *et al.*, 1998). Therefore, *CYB561* is expressed in neural tissue that may show dysregulation in schizophrenia (Harrison and Lewis, 2003) and at levels which increase during neurodevelopment (Srivastava, Pollard *et al.*, 1998). Also the spatial expression of the transcript is consistent with a role in dopaminergic neurotransmission (Srivastava, Pollard *et al.*, 1998), making it a reasonable candidate for involvement in neural signalling processes that could be anomalous in schizophrenia.

AB046856 hypothetically encodes protein called KIAA1636, although only transcript evidence has been submitted to the public databases (Nagase, Kikuno *et al.*,

2000). Although the biological candidacy for this transcript is the weakest for the genes examined, the transcript was identified from human brain (Nagase, Kikuno *et al.*, 2000) and spans a large proportion of a homozygous region confirmed by SNP genotyping (figure 4.1), making it a positional candidate.



Region (UCSC May 2004)	Defining Markers	Size (bp)	Candidate Genes
chr17:56807191-58171285	D17S1838-D17S794	1364094	None recognised
chr17:58557996-58914495	rs958334-rs4311	356499	*CYB561, AB046856*
chr17:59566252-59719211	rs6504204-4968716	152959	None recognised
chr17:60101961-60954065	rs6504248-rs11868547	852104	*GNA13, RGS9*
chr17:61851420-62715931	D17S1816-D17S1821	864511	*PRKCA*
chr17:64140423-64651294	D17S789-D17S2182	510871	None recognised

Figure 4.1: A UCSC May 2004 track displaying the homozygous regions within the C702 IBD2 region as defined by microsatellite and SNP sequencing data, designated as "HMZ_Regions". The candidate genes which have undergone mutation screening and association analysis in this chapter and chapter 6 are shown by the track "Candidate_Genes". Also given are base positions, chromosomal bands and some of the Known Genes in the region. The accompanying table details the maximal possible homozygous regions in IBD2 region of C702 (based upon available genotype data), corresponding heterozygous genotype marker names defining the homozygous region, the size of each region and recognised candidate genes prioritised for mutation screening and association analysis.

***CYB561* - Cytochrome b-561**

All putative functional sequence at *CYB561* were successfully sequenced except 112 of the first two exons and the 1kb putative core promoter (chr17:58,877,227-58,878,398. UCSC May 2004). In total 5885 bases were sequenced in a C702 sibling and unrelated controls at *CYB561*. A total of 4 SNPs were discovered at this locus, 3 are already recorded by public databases (dbSNP) and one is novel (see table 4.3 and also appendices 10.2 for sequencing traces). The three database SNPs were examined for association in our schizophrenia case-control sample by DNA pooling. The novel SNP (Exon5b#2) could not be assayed by the DNA pooling approach due to the rarity of the minor allele. The SNP has not been tested for association by other means as the C702 sibling is heterozygous for the allele and the prior hypothesis for the C702 disease allele requires homozygosity. The results of allele association analysis in pools are shown in table 4.3.

***AB046856* - tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 2 (TANC2)**

All exons were successfully sequenced except for a 232 base exon (chr17:58771194-58771425) and an amplimer in the putative promoter region (chr17:58630795-58631213). In total 21,277 bases were bidirectionally sequenced and 6 SNPs discovered, all of which were present in publicly available databases (table 4.3). The 6 variants were tested for association in a pooled DNA case-control sample, none

showed any evidence for association with schizophrenia (see table 4.3). Furthermore, none of the C702 sibling genotypes were rare and homozygous.

***GNA13* - guanine nucleotide binding protein (G protein), alpha 13**

Exons and regulatory regions totalling 8342bp were sequenced in C702 and an unrelated control. The majority of the putative core promoter region (chr17:60,483,306-60,484,279. UCSC May 2004), 177bp of the last exon (chr17:60,437,296-60,437,472. UCSC May 2004) and a possible regulatory region identified by ClusterBuster (chr17:60485919-60486915. UCSC May 2004) were not successfully sequenced. Only one SNP was identified (table 4.3), however this was an intronic SNP where a control individual was heterozygous (C/T) for an allele that was too rare for analysis in the DNA pools (SNP Reg6C. Estimated MAF $\leq$0.03) and so was not analysed further as the C702 affected was homozygous for the common allele (C).

***RGS9* – Regulator of G-protein signalling 9**

The putative promoter, regulatory motifs identified by ClusterBuster and exons of *RGS9* were bidirectionally sequenced in a C702 sibling and $\geq$ 1 control, except for exon 7 positioned at chr17:60594780-60594861 (UCSC May 2004 freeze). Approximately 10,185 putative functional bases were sequenced and 9 polymorphisms identified. Three of these SNPs were too rare for analysis in the case-control DNA pools (estimated MAF $\leq$0.03) and as the C702 sibling did not display the minor allele these SNPs were not

analysed further (table 4.3). SNP rs11658673 failed to optimise for analysis in DNA pools, however as the affected C702 sibling was homozygous for the common allele in the HapMap CEU sample (MAF=0.79) it was not analysed further. The other 5 variants were tested successfully, of these rs11658773 showed evidence for association that just failed to reach nominal significance (p=0.08). This is however unlikely to explain the C702 phenotype as the siblings are homozygous for the common (MAF~0.8) allele.

***PRKCA* – Protein Kinase C – Alpha**

A detailed investigation of PRKCA is presented in chapter 6.

Gene ID	SNP ID	UCSC position†	Alleles$^{\Sigma}$	702 sib genotype	Type	Case*	Control*	CEU*	P value"
AB046856	rs2460111	chr17:58669004	G/A	AA	Synonymous	0.5	0.5	0.45	0.97
AB046856	rs4968765	chr1758786506	A/G	GG	Intronic	0.5	0.51	0.5	0.64
AB046856	rs4968768	chr17:58799652	T/C	CC	Intronic	0.51	0.5	0.51	0.67
AB046856	rs8065950	chr17:58810949	C/G	CC	Intronic	0.5	0.5	-	0.96
AB046856	rs4968644	chr17:58821784	G/C	CC	Intronic	0.52	0.52	0.52	1
AB046856	rs2270133	chr17:58827057	T/G	AA	Intronic	0.47	0.47	0.5	0.91
CYB561	rs3087776	chr17:58864010	T/C	TT	3' UTR	0.49	0.48	0.48	0.48
CYB561	rs4968773	chr17:58868990	G/A	AA	Intronic	0.5	0.5	-	0.7
CYB561	Exon5b#2	chr17:58869002	A/C	AC	Intronic	A rare‡	A rare‡	-	-
CYB561	rs4968646	chr17:58869053	C/T	CC	Intronic	0.49	0.48	-	0.65
RGS9	rs402137	chr17:60563744	A/G	AA	5'	0.75	0.75	-	0.87
RGS9	rs11658673	chr17:60615579	T/C	CC	Intronic	FAIL	FAIL	0.79	FAIL
RGS9	rs11654243	chr17:60615579	G/A	AA	Intronic	0.84	0.84	0.79	0.55
RGS9	Exon10a#3	chr17:60615646	G/A	AA	Intronic	0.83	0.84	-	0.3
RGS9	rs11658773	chr17:60615676	A/G	GG	Intronic	0.81	0.84	0.79	0.08
RGS9	Exon15	chr17:60630941	A/G	GG	Intronic	A rare‡	A rare‡	-	-
RGS9	Reg5b	chr17:60641271	C/T	CC	Intronic	T rare‡	T rare‡	-	-
RGS9	Exon18a	chr17:60651675	A/G	GG	Non-Syn. R-H	A rare‡	A rare‡	-	-
RGS9	Exon19b	chr17:60654250	A/G	AA	3'UTR	0.13	0.13	-	1
GNA13	Reg6C	chr17:60486596	C/T	CC	5'	T rare‡	T rare‡	-	-

† UCSC May 2004

$^{\Sigma}$ Minor Allele/Major Allele (estimated from controls or CEU individuals)

* The frequency of the allele carried by the C702 siblings as measured in DNA pools or CEU (HapMap™)

" Allele Association P value

‡ Allele not detected at analysable levels in DNA pools (i.e. MAF <0.03)

Table 4.3: Mutation screening and association analysis result for genes lying in C702 homozygous regions. Table shows in columns (Left to Right): Gene ID from UCSC (UCSC May 2004 freeze), SNP ID (either rs or in-house), UCSC position (UCSC May 2004), SNP alleles (Minor Allele/Major Allele, control or CEU HapMap allele frequencies), sibling 702.06 genotype and the genomic location of variant. Then the case, control and CEPH allele frequencies (if applicable) are given for the C702 allele with an association p value calculated where applicable. Sequence traces (Sequencher) for all novel SNPs are shown in appendices 10.2. When no association p value was given this is either due to an assay failure or because the MAF was too rare to be reliably estimated in DNA pools (Estimate MAF $\leq$0.03).

4.3.3. Homozygosity Mapping using Whole-Genome Association Data

The C702 affected sibling (702-06) passed Affymetrix and WTCCC quality control procedures ((Wellcome Trust Case Control Consortium, 2007) and O'Donovan et al, submitted) and was successfully genotyped for 1028 SNPs across the refined C702 IBD2 region ((Wellcome Trust Case Control Consortium, 2007) and O'Donovan et al, submitted). The resulting marker map shows a smallest inter-marker distance of 14bp and a largest gap of 265,910bp, with 72% of inter-marker distances falling below the mean of 11,100bp. After merging with the microsatellite data across this region (chapter 3), 83 stretches of homozygosity (containing $\geq$ 2 contiguous markers) were identified (table 4.4 and figure 4.2).

To examine whether any of the homozygous C702 SNP diplotypes were rare and so could identify the location of a rare, founder haplotype, each diplotype was analysed in 476 cases and 2938 controls (O'Donovan et al, submitted). Further to this, the diplotypes were analysed for association. The results are shown in table 4.4.

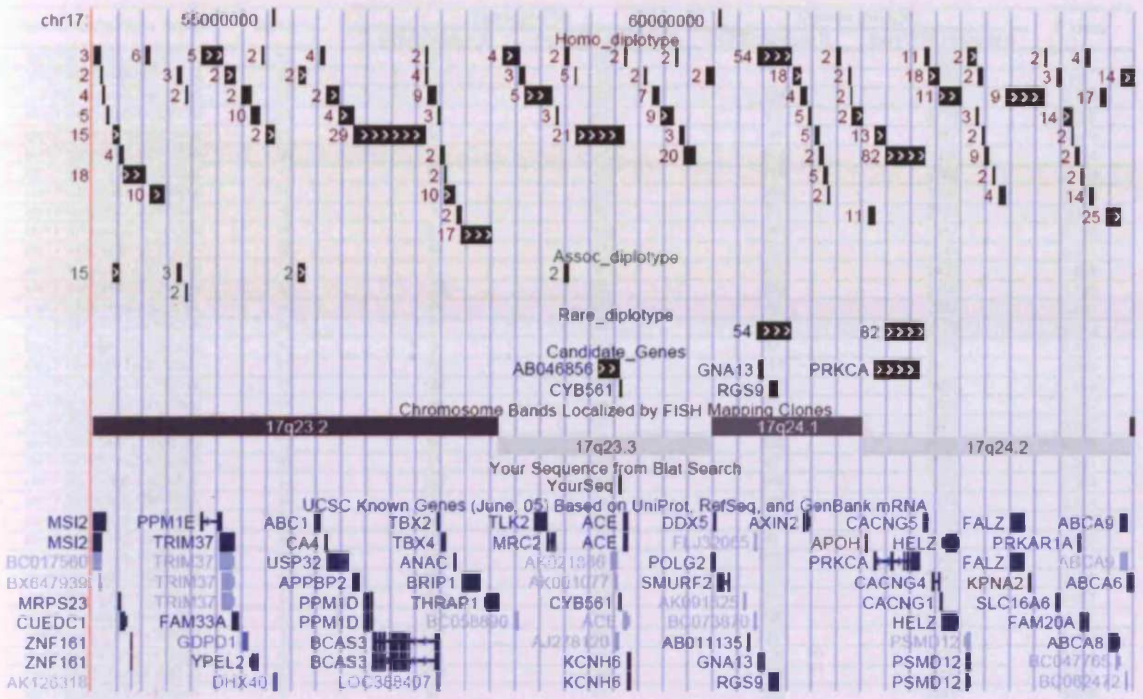


Figure 4.2: All homozygous regions (≥ 2 markers) across the C702 IBD2 region in affected C702 sibling 06, defined by Affymetrix SNP and microsatellite genotyping. The figure shows UCSC genome browser tracks across the C702 IBD2 region including the base position (May 2004 freeze), the chromosomal band and known genes, with added custom tracks. The custom track "Homo_diplotype" shows all homozygous regions with the number of homozygous markers that define that region (table 4.4). The custom track "Assoc_diplotypes" shows the regions showing nominal evidence for association (table 4.4). The track "Rare_diplotype" shows the rare diplotypes (frequency <0.01 in controls. Table 4.4). Also shown are the candidate genes studied in this chapter and chapter 6 ("Candidate_Genes").

Max C702 Diplotype Region	Marker N	Cases (N=476)		Controls (N=2938)		P Value
		C702 diplotype N	Non-C702 diplotype N	C702 diplotype N	Non-C702 diplotype N	
chr17 61851032-62282771	82	0	383	0	2328	NA
chr17 60422644-60799078	54	0	420	3	2553	NA
chr17 55883487-56684750	29	27	422	128	2699	0.168
chr17 64324502-64483170	25	19	432	93	2715	0.33
chr17 58388113-58923464	21	10	442	48	2767	0.448
chr17 59586445-59717248	20	129	313	912	1886	0.154
chr17 53311424-53583002	18	22	438	115	2742	0.449
chr17 62336380-62450618	18	41	424	299	2537	0.256
chr17 60606627-60696841	18	1	451	33	2784	0.08
chr17 57094904-57441290	17	91	359	590	2237	0.753
chr17 64261626-64324001	17	48	414	333	2516	0.417
chr17 53193707-53270939	15	71	365	352	2438	**0.036**
chr17 64483493-64651294	14	120	337	663	2136	0.233
chr17 63831495-63939291	14	15	442	90	2757	0.891
chr17 64140110-64196219	14	7	441	72	2770	0.212
chr17 61748819-61851032	13	9	456	47	2851	0.624
chr17 62465677-62715831	11	42	423	261	2611	0.969
chr17 61663751-61733114	11	198	269	1197	1696	0.669
chr17 62287922-62336380	11	8	454	73	2783	0.287
chr17 53626496-53772286	10	72	371	440	2331	0.842
chr17 56915467-57020199	10	34	424	203	2674	0.776
chr17 54758734-54858683	10	11	450	47	2807	0.261
chr17 63184300-63635027	9	141	299	989	1860	0.363
chr17 59330980-59470994	9	25	442	184	2656	0.354
chr17 58726816-56804418	9	181	288	1068	1828	0.476
chr17 62957456-63020549	9	15	455	81	2809	0.639
chr17 59236755-59306203	7	165	303	1092	1807	0.317
chr17 53583002-53626496	6	105	363	625	2262	0.702
chr17 57806382-58113684	5	53	422	354	2568	0.551
chr17 54195586-54434524	5	127	345	908	2002	0.06
chr17 61164191-61209993	5	160	304	1091	1797	0.173
chr17 61074793-61113688	5	121	349	826	2070	0.214
chr17 58352784-58388113	5	424	40	2670	233	0.663
chr17 60993743-61025491	5	57	411	365	2547	0.829
chr17 53145674-53164517	5	76	392	499	2344	0.487
chr17 57545412-57740993	4	199	266	1310	1601	0.374
chr17 55723603-55883487	4	443	31	2702	226	0.368
chr17 60896841-60988360	4	323	148	1905	999	0.206
chr17 63126561-63184300	4	47	423	234	2682	0.15
chr17 64088034-64140110	4	246	221	1606	1299	0.293
chr17 55522267-55565566	4	10	464	76	2849	0.53
chr17 53091903-53122443	4	96	377	567	2360	0.638
chr17 58698662-58726816	4	374	96	2398	518	0.167
chr17 53260622-53306639	4	157	315	1003	1914	0.634
chr17 57740993-57804047	3	210	265	1288	1635	0.953
chr17 52996269-53057256	3	117	357	755	2174	0.613
chr17 59539496-59586445	3	353	108	2093	793	0.069
chr17 63779245-63823781	3	174	300	1207	1716	0.059
chr17 53926949-53966870	3	159	315	1131	1792	**0.032**
chr17 58142964-58170964	3	151	323	985	1940	0.436
chr17 62935292-62952184	2	52	414	300	2582	0.625
chr17 62870902-62886331	3	192	282	1213	1706	0.667
chr17 56838548-56851642	3	203	272	1343	1582	0.197
chr17 55594527-55723603	2	445	31	2698	229	0.318
chr17 54631492-54758734	2	445	31	2698	229	0.318
chr17 61484260-61594776	2	400	74	2370	544	0.11
chr17 54460005-54564851	2	158	317	1086	1848	0.115
chr17 62776578-62870902	2	177	275	1135	1772	0.963
chr17 54921099-54996857	2	425	46	2568	353	0.147
chr17 59852249-59923805	2	307	166	1897	1027	0.991
chr17 55293281-55351797	2	17	459	185	2751	**0.019**
chr17 57033739-57085167	2	269	206	1587	1343	0.316
chr17 62886331-62930869	2	10	466	80	2844	0.423
chr17 56857101-56899557	2	57	413	384	2532	0.534
chr17 59139636-59181588	2	170	298	1125	1795	0.363
chr17 61299577-61339711	2	230	243	1458	1473	0.662
chr17 63968022-64021487	2	271	198	1666	1254	0.768
chr17 54010871-54043788	2	159	316	1130	1800	**0.034**
chr17 58249995-58281335	2	143	328	1025	1902	**0.048**
chr17 61120331-61149311	2	194	280	1204	1726	0.946
chr17 54870249-54896409	2	9	466	37	2895	0.268
chr17 58923464-58945384	2	107	368	563	2367	0.092
chr17 53070388-53091903	2	176	298	1096	1868	0.738
chr17 61209993-61230568	2	188	288	1275	1658	0.104
chr17 56899557-56915467	2	41	434	274	2661	0.623
chr17 56684750-56698662	2	334	137	2045	886	0.616
chr17 63939597-63952109	2	162	311	931	1999	0.285
chr17 63062132-63073548	2	39	434	234	2686	0.864
chr17 64049634-64057725	2	345	129	2181	744	0.411
chr17 63640160-63647915	2	313	162	2022	907	0.172
chr17 61455793-61463014	2	441	35	2698	233	0.654
chr17 61447793-61454262	2	439	32	2698	216	0.633
chr17 59499460-59500924	2	337	138	2031	901	0.461

Table 4.4: The pedigree C702 sibling 702.06 exhibits 83 diplotypes where every consecutive marker is homozygous. The size of these regions is shown (UCSC May 2004), along with the number of markers in that region (Affymetrix SNPs and microsatellites). The regions are shown in descending order showing the number of markers per region. Also shown are the case and control counts for the C702 diplotype carriers and non-carriers and a chi-squared or fisher's exact test p value. P values were not applicable (NA) to datasets with 0 diplotype counts.

4.4 Discussion

If pedigree C702 represents the convergence of a disease allele originating from a common founder ancestor, the affected individuals would be expected to be homozygous not only for the disease allele but for flanking marker alleles. The search for disease alleles by identifying stretches of marker homozygosity in consanguineous families exhibiting a recessive phenotype is known as autozygosity mapping (Strachan and Read, 2003b). Autozygosity mapping has been successful in the identification of many disease alleles for recessive phenotypes in inbred pedigrees (Moynihan, Bundey *et al.*, 1998, Saba, Montpetit *et al.*, 2005, Strachan and Read, 2003b, Valente, Abou-Sleiman *et al.*, 2004, Vazza, Zortea *et al.*, 2000).

The pedigree C702 siblings show no evidence for increased homozygosity across the genome and so are unlikely to be due to a consanguineous relationship (chapter 3); therefore the siblings should be homozygous by descent only for markers flanking the disease locus. Given that the genomic size of the homozygous region will vary depending

upon many factors including how many generations ago the founder mutation was introduced and also the recombination rate across the region, analysing a dense grid of markers spanning the C702 IBD2 region may reveal areas of homozygosity that may denote the location of the disease alleles. As pedigree C702 is not consanguineous then the common ancestor that founded the disease allele present in the parents coalesced many generations ago, meaning that unless recombination is suppressed around the disease allele the region of homozygosity is unlikely to be large and encompass many microsatellite markers. The analysis has been termed homozygosity mapping for C702 to distinguish it from autozygosity mapping, which entails consanguineous pedigrees (Strachan and Read, 2003b).

Homozygosity mapping using microsatellite markers (chapter 3) identified several single-marker homozygous genotypes in the C702 affecteds and one region containing 5 consecutive markers homozygous, which was also the largest region (table 4.1). However, this region was falsely homozygous (table 4.2) and was demarcated into several regions of homozygosity by further SNP sequencing (table 4.2). The need for a higher density of genotype data for the C702 siblings in the IBD2 region was therefore clear; however candidate gene mutation screening and association analysis was undertaken on five genes within these homozygous stretches.

Genes chosen for mutation screening were *AB046856*, *CYB561*, *GNA13*, *PRKCA* (see chapter 6) and *RGS9*. Bi-directional sequencing of all putative functional regions (exons, promoter region and regulatory motifs) was attempted for all genes in a C702 sibling and unrelated controls. Polymorphisms identified were examined for association by estimating allele frequencies using case and control DNA pools, or were excluded

from analysis in DNA pools if the minor allele was too rare and the C702 sibling exhibited a common allele, as the C702 disease allele is proposed to be homozygous and rare.

Only one allele exhibited evidence for a weak ($p<0.1$) association in case-control pools however this allele has yet to be genotyped individually and is unlikely to explain the C702 phenotype as the siblings are homozygous for the common allele. Furthermore, after resequencing >43kb of known and putative functional DNA sequence no rare homozygous genotypes were observed for the C702 sibling in the candidate genes studied in this chapter. Therefore, it is unlikely that any of the polymorphisms identified in this chapter are responsible for the C702 phenotype.

However, it is possible that the disease alleles have been overlooked by the strategy employed. The microsatellite and SNP marker map may not be dense enough to uncover the true homozygous by descent region, which is particularly true for the microsatellite and sequencing data where the largest inter-marker gap is ~2.3 Mb between D17S1160 and D17S1604, for example. However, the Affymetrix SNP data has much higher-density coverage (see below). Another caveat is the selection of candidate genes for study, as the molecular patholgy of schizophrenia remains unknown and so candidate gene selection is guesswork (Owen, O'Donovan *et al.*, 2005). Furthermore, even if the correct genes have been looked at our mutation screening may have overlooked a true functional sequence and therefore may have missed any disease variant. Finally, analysis for association in DNA pools does not take in to account more complex genetic mechanisms such as haplotypes and genotypic effects.

A secondary genotyping effort involved the assay of 1028 SNPs in a pedigree C702 sibling, giving a mean SNP density of one every 11.1kb. This data identified 83 stretches of homozygous marker genotypes across the 11.7Mb region (table 4.4 and figure 4.2), which included the homozygous microsatellite markers already known. The high-density map still identifies all the five candidate genes screened in this chapter and chapter 6 as being fully or partly in stretches of marker homozygosity (figure 4.2).

The mutation screening of all of these regions would be the ideal methodology to test the hypothesis that a homozygous region harbours a rare, homozygous allele causing the C702 schizophrenia phenotype. A compromise is to examine the regions unlikely to have occurred by chance, which may not be the largest or the most unlikely regions when considering that markers may be in LD or not at a high-density. Using the genotype data for the same SNP marker set genotyped through ~500 cases and ~3000 controls the homozygous C702 identified diplotypes in the IBD2 region were often identified, although no evidence for association that would survive correction for multiple tests was found (best p=0.019). Furthermore, only one of the C702 identified diplotypes showed a nominally significant increase in cases (chr17:53193707-53270939, p=0.035. Case diplotype frequency = 0.16, Control diplotype frequency = 0.13).

The lack of linkage to 17q in the other ASPs that were part of the linkage study that identified the C702 linkage (Williams, Norton *et al.*, 2003), the paucity of widely reported linkage to the region and the IBD2 status of the pedigree points towards a rare, recessive allele causing the C702 linkage. Therefore, extremely rare observations of homozygosity are perhaps more interesting than any population based association

evidence. Two C702 diplotypes are rare in the control sample studied; a diplotype spanning *RGS9* and *GNA13* that occurs in 3 individuals from 2556 controls (table 4.4 and figure 4.2) and a diplotype spanning the 3' of PRKCA that does not occur in 383 cases or 2328 controls. Therefore, the diplotype spanning the *PRKCA* region is an extremely unlikely observation.

The data presented in this chapter shows a multi-step approach to identify the alleles in pedigree C702 causing their linkage. The approaches were all performed under the assumption that the pedigree C702 phenotype is caused largely by a highly penetrant, rare and recessive acting allele, inherited from a common founder ancestor. These assumptions can be justified however, as previous data shows the pedigree are not inbred (chapter 3), the parents are unaffected, the linkage region is due to the IBD2 status of the affecteds and of what little evidence there is of linkage to the same region from other studies (Dick, Foroud *et al.*, 2003, Klei, Bacanu *et al.*, 2005, Lewis, Levinson *et al.*, 2003, McInnis, Dick *et al.*, 2003, Rees, Fenton *et al.*, 1999, Tomas, Canellas *et al.*, 2006) many point to a recessive acting disease locus, particularly in large, multiply affected pedigrees (Klei, Bacanu *et al.*, 2005, Rees, Fenton *et al.*, 1999, Tomas, Canellas *et al.*, 2006). However, potential caveats could be that the disease alleles are disparate, within either the IBD2 or IBD1 regions, or that the family are linked to this region due to a hitherto unidentified phenotype unrelated to their schizophrenia (e.g. a late-onset phenotype).

However, the data presented here supports the rare, recessive allele from a common founder ancestor rationale as a homozygous diplotype has been identified that has been shown to be absent from a sample of ~3500 individuals from the UK. Therefore,

the encompassed region should be prioritised for directed or full resequencing, with the *a priori* hypothesis being that each haplotype carries the C702 disease allele and a homozygous genotype should be identified that is at least as rare an occurrence as the

Chapter 5. Analysis of a Refined Schizophrenia Linkage Region on Chromosome 17 for a Deletion by Comparative Genome Hybridisation

5.1. Introduction

Chapter 3 described the refinement of a genome-wide significant linkage region (Williams, Norton *et al.*, 2003) to an 11.7Mb segment at 17q23.2-q24.3. The affected siblings all share copies of a paternal and maternal haplotype discerned by identical genotypes at a 21 marker segment, whereas the unaffected parents only exhibit one copy of the haplotype which they transmit to all affected offspring. Therefore, a likely explanation of the siblings' phenotype is the presence of a highly penetrant, recessive (or incomplete dominant) acting disease allele on each of the inherited chromosomes. Under the hypothesis that there is a disease allele in the IBD2 genome wide significant linkage region of a single pedigree, it is justifiable to search for Sub-microscopic Structural Variants (SSVs) in this region for affected individuals of pedigree C702. SSVs include sub-microscopic deletions and duplications (Copy Number Variants – CNVs), balanced translocations, insertions, inversions and the inheritance of diploid material from one parent only (Uniparental Isodisomy – UPD (Spence, Perciaccante *et al.*, 1988))

There are several lines of evidence to support the search for a SSV present in the affected C702 siblings' IBD2 region: First, the large inherited paternal and maternal haplotypes in this pedigree hint that recombination has been repressed on chromosome 17

in this pedigree. Although this could occur by chance it is also known that SSVs can inhibit recombination events (Stefansson, Helgason *et al.*, 2005). Secondly, loss of heterozygosity across a region can be indicative of a hemizygous deletion. Therefore, large stretches of homozygous markers, which have been demonstrated to exist across the C702 IBD2 region (chapter 4), may indicate a parental deletion has occurred.

The genomic structure of chromosome 17 makes it highly susceptible to SSVs and there have been a number of reports identifying pathogenic SSVs in this region. Many papers have attempted to study or inadvertently observed the complex structural arrangements of the long arm of chromosome 17 in particular (Chen, Saarela *et al.*, 2004, Dorr, Midro *et al.*, 2001, Monni, Barlund *et al.*, 2001), with reports of gene duplications (*NSF* – (Iafrate, Feuk *et al.*, 2004, Sebat, Lakshmi *et al.*, 2004, Sharp, Locke *et al.*, 2005, Tuzun, Sharp *et al.*, 2005)), common European inversions encompassing a human disease gene (Stefansson, Helgason *et al.*, 2005) and segmental duplications and a human inversion within a multiple sclerosis linkage region, with respect to a syntenic loci in mice (Chen, Saarela *et al.*, 2004). There are numerous papers attempting to catalogue human segmental duplications and deletions that have highlighted 17q as structurally complex region with many rearrangements of syntenic loci in chimpanzee and mouse (Cheng, Ventura *et al.*, 2005, Cheung, Estivill *et al.*, 2003, Newman, Tuzun *et al.*, 2005, Zody, Garber *et al.*, 2006) and segments frequently showing anomalies in cancerous cells (Korshunov, Sycheva *et al.*, 2006, Monni, Barlund *et al.*, 2001).

An attempt to catalogue segmental duplications within the human genome (Bailey, Yavor *et al.*, 2001) resulted in a whole genome directory of segmental duplications (>90% similarity, >1kb) that are accessible via a UCSC genome browser

track (http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=71274077&c=chr7&g=genomicSuperDups). As can be seen in figure 5.1, there is a large potential for chromosomal rearrangement within the C702 IBD2 region on chromosome 17, as facilitated by segmental duplications which can cause non-allelic homologous recombination (NAHR) (Shaw and Lupski, 2004). Indeed, there are a number of CNVs already known across this region (figure 5.1)

The large amount of segmental duplications that have paralog copies to other sites in the IBD2 region shows the promise for non-allelic homologous recombination (NAHR) in the region. The known homozygous regions demonstrated in the previous chapter (chapter 4) could potentially be explained by a deletion. However, the 7 reported and catalogued SSVs in this region do not obviously correspond to any of the homozygous regions or regions flanked by segmental duplications.

In recent years various laboratory techniques have been developed to assay the genome for SSVs. These techniques include array based screens to cover large regions at high density, and qualitative and quantitative PCR techniques (Real Time Quantitative PCR or QPCR). High-density oligonucleotide-array Comparative Genome Hybridisation (oaCGH) allows a high throughput and high resolution screen primarily for the identification of CNVs (Conrad, Andrews *et al.*, 2006, Feuk, Carson *et al.*, 2006, Locke, Sharp *et al.*, 2006, Ylstra, van den Ijssel *et al.*, 2006). The CGH approach involves fluorescently labelling test and reference DNA samples followed by their competitive hybridisation to an array spotted with short, specific oligonucleotides thus allowing high specificity and dense coverage (and therefore resolution) in a region of interest, combined with the technical benefits of a standardised platform. oaCGH services allow the

theoretical coverage of a region of interest down to a single base, which not only has the potential to delineate CNVs but as single base resolution can be attained the breakpoints for balanced translocations and inversions can theoretically be mapped.

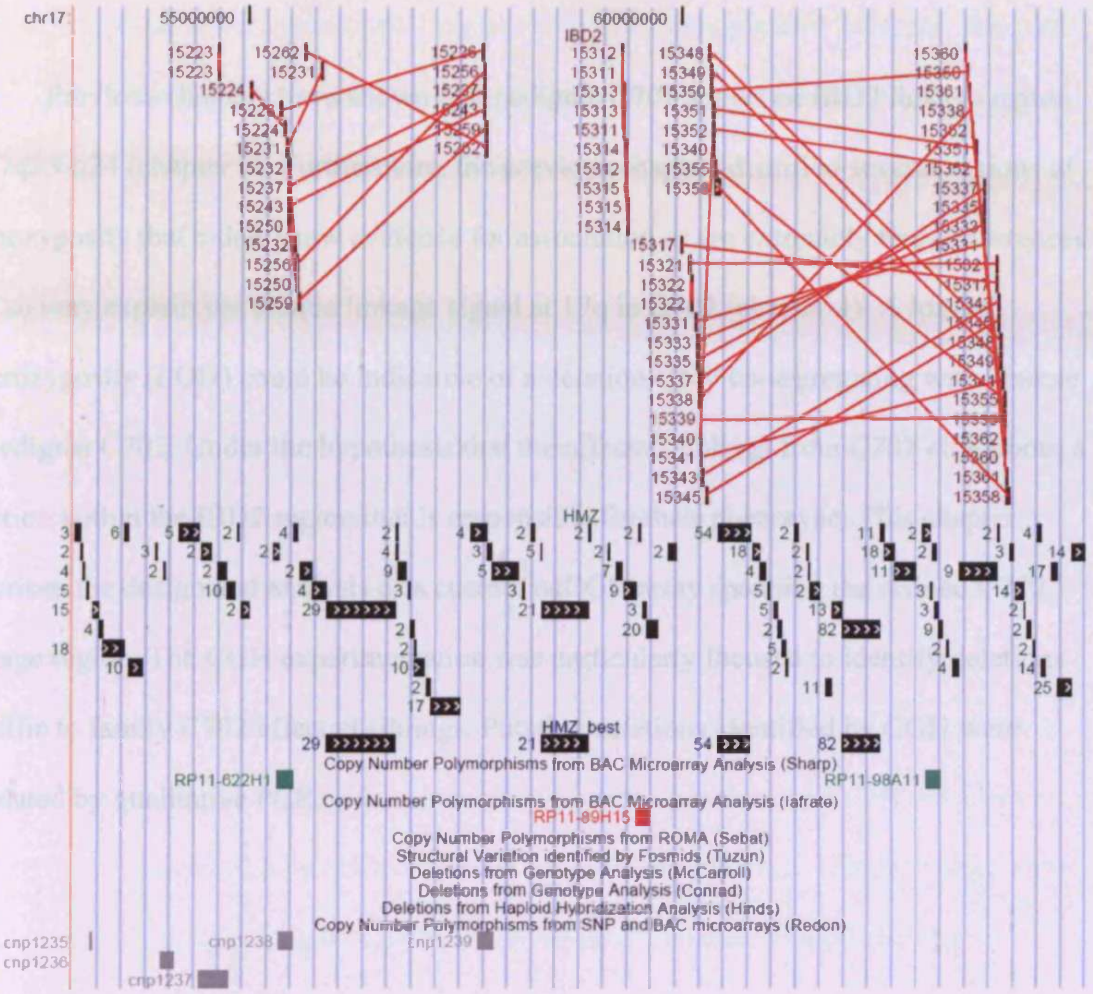


Figure 5.1. Duplications within the C702 IBD2 region. The data from Bailey et al (2001) as accessible via the UCSC genome browser "Segmental Duplications" track (http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=71274077&c=chr7&g=genomicSuperDups). The uppermost track indicates the size of the chromosome 17 region in 5Mb segments shown by black vertical bars. The underlying track titled "IBD2" shows all intra-IBD2 segmental duplications (>1kb, ≥ 90% similarity), corresponding duplications are indicated by joining red lines. The next lower two tracks show the homozygous regions identified in chapter 4 (HMZ) and the most interesting of those stretches (HMZ_best) including the homozygous region at *PRKCA* ("82"). Finally, the next series of tracks list known structural variants and the research article that identified them.

Previous chapters have shown that pedigree C702 shows an IBD2 linkage region at 17q23-q24 (chapter 3). Furthermore, the previous chapter identified several regions of homozygosity that either show evidence for association or are extremely rare occurrences and so may explain the unique linkage signal at 17q in C702 (chapter 4). A loss of heterozygosity (LOH) could be indicative of a deletion CNV co-segregating with disease in pedigree C702. Under the hypothesis that the affected siblings from C702 all harbour a deletion within the IBD2 region that is responsible for their phenotype. This chapter describes the design and analysis of a custom oaCGH array spanning the refined C702 linkage region. The CGH experimentation was particularly focused to identify deletions specific to family C702 affected siblings. Putative deletions identified by CGH were validated by qualitative PCR.

5.2. Materials & Methods

5.2.1. High-Resolution Oligonucleotide-Array Comparative Genome Hybridization (oaCGH)

High resolution oaCGH was performed using NimbleGen Systems Inc Custom CGH Array service. Sample preparation and quality assessment were performed according to their guidelines and details of the CGH experiment, data analysis and resultant output outlined below were obtained from NimbleGen Systems Inc advisers and

supplied information and published literature using the same platform (Conrad, Andrews *et al.*, 2006, Selzer, Richmond *et al.*, 2005, Stallings, Nair *et al.*, 2006, Urban, Korbel *et al.*, 2006).

5.2.2. Microarray Design

Custom tiling-path CGH arrays were designed by NimbleGen Systems Inc at high resolution for fine-tiling analysis of the C702 IBD2 region of interest (UCSC Genome Browser May 2004 freeze. HG17. chr17:52,996,269-64,651,294). Oligonucleotide probes were selected based on several measures of uniqueness and base-pair composition, defined by NimbleGen Systems Inc, starting with DNA sequence masked for repetitive elements in the genome (http://www.repeatmasker.org) (see figure 5.3). Oligos were selected to give maximum density of inter-oligo gaps possible on a ~380,000 oligo array. Oligos were of variable length (45-85bp) and adjusted to match a target melting temperature of 76°C. Probe uniqueness was ascertained by determining the number of 45mer matches (the minimal oligo length) of each oligo in the genome, as well as the average frequency of the 15mers that comprise the longer oligo. Optimal probes were then selected by evaluation using a quality score based upon these facets as defined by NimbleGen Systems Inc. Oligos were incorporated into an array design using ArrayScribe array design software and arrays constructed by maskless array synthesis technology (NimbleGen Systems, Inc), with oligonucleotides being synthesized on to an array by photolithography.

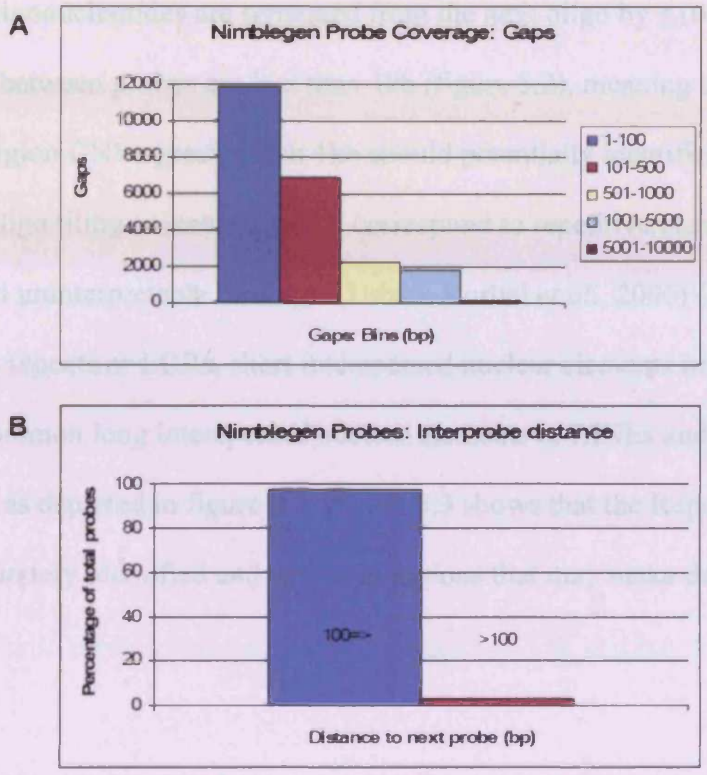


Figure 5.2: Illustration of probe coverage of the IBD2 region. **[A]** The binned size of gaps between tiled oligonucleotides and the number of gaps within each bin. Of the 22,852 gaps in oligo coverage the majority are under 0.5kb. **[B]** The percentage of probes lying within 100 nucleotides (100=>. 97%) and over 100 nucleotides from the next consecutive oligo (>100, 3%).

The array was finalised covering the IBD2 region with 387218 individual oligonucleotide features or spots. The coverage over regions is highly variable; while the mean resolution (mean gap between probes) is 3.67bp, there is a range from 1bp (i.e. complete overlap) to 94686bp (corresponding to a gap in the HG17/NCBI Build 35 sequence spanning chr17:63,492,822-63,587,508, which cannot therefore be tiled).

Therefore, the array provides complete coverage of some regions and no coverage of others. 97% of oligonucleotides are separated from the next oligo by ≤100bp and the vast majority of gaps between probes are less than 1kb (figure 5.2), meaning that for the majority of the region CNVs greater than 1kb should potentially identified.

Gaps in oligo tiling coverage usually correspond to repetitive elements that may give distorted and uninterpretable findings (Urban, Korbel *et al.*, 2006) (Examples include low copy repeats or LCRs, short interspersed nuclear elements or SINEs such as *Alu*s, the more common long interspersed nuclear elements or LINEs and long terminal repeats or LTRs) as depicted in figure 5.3. Figure 5.3 shows that the RepeatMasker software has accurately identified and excluded regions that may make data interpretation problematic.

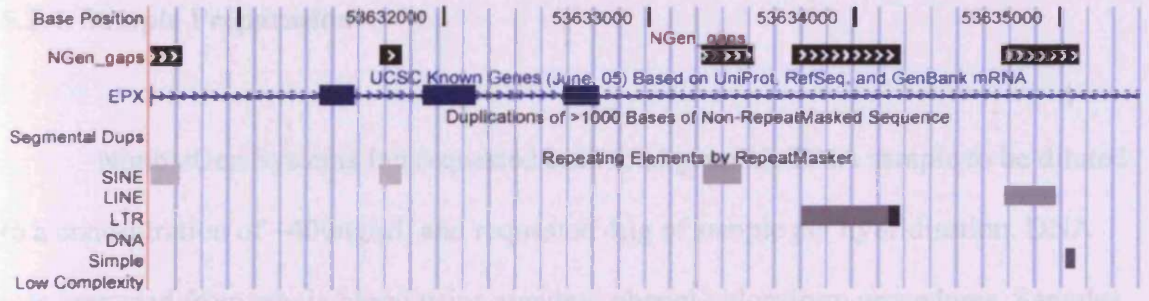


Figure 5.3: A UCSC genome browser track over a small portion of the C702 IBD2 region on chromosome 17. Displayed (in descending order) are; base position, gaps between oligonucleotides (black with white chevron), known genes, segmental duplications (none present) and a RepeatMasker track showing all elements over which CGH tiling would not be attempted and the type of repetitive element. Gaps in the NimbleGen custom design coverage for the C702 IBD2 region are shown in black. All other bases are covered by a complementary oligonucleotide present on the custom array.

5.2.3. Samples

As all the affected C702 siblings share identical paternal and maternal haplotypes in the refined IBD2 region (chapter 3), using any of the siblings as a test sample should be sufficient to detect a CNV in C702, when hybridised to $\geq$ 1 unrelated control. However, to minimise false positives due to experimental variation and also false negatives due to the unlikely occurrence of the same CNV "genotype" (e.g. both case and control are hemizygous) in the controls two siblings were chosen for analysis and three unrelated controls. Male controls were chosen to minimise any potential false positives from differences in chromosome structure as a result of sex. Controls were obtained from National Blood Transfusion Service and are Caucasians of UK origin.

5.2.4. Sample Preparation

NimbleGen Systems Inc requested each total genomic DNA sample to be diluted to a concentration of ~400ng/µl, and requested 4µg of sample per hybridisation. DNA was extracted from whole blood using standard phenol-chloroform procedures. Samples were concentrated to between 200-400ng/µl by drying in a rotary vacuum at 55°C. Samples were then left to enter solution for >24 hours before quantification of a sample dilution using PicoGreen (see materials and methods chapter, section 2.1.3). An aliquot of each concentrated sample was then run on a 0.5% agarose gel to establish that the DNA was not degraded (e.g. high molecular weight bands). Sufficient amounts of the

samples were then shipped using dry ice to NimbleGen Systems Inc where the samples passed their quality controls.

5.2.5. Sample Labelling and Hybridization

Sample labelling and hybridisation was performed by NimbleGen Systems Inc, which is a standardised procedure as described in previous papers (Conrad, Andrews *et al.*, 2006, Locke, Sharp *et al.*, 2006, Selzer, Richmond *et al.*, 2005, Stallings, Nair *et al.*, 2006, Urban, Korbel *et al.*, 2006).

5.2.6. Hybridisation Plan

Each C702 sibling was hybridised to each unrelated control, therefore each control hybridised to a C702 sibling has a sister or duplicate hybridisation experiment which entails the other C702 sibling. The resultant is six hybridisation experiments.

5.2.7. Data Analysis

Raw data generation and analysis was performed by NimbleGen Systems Inc (NimbleGenSystemsInc., 2005) using NimbleScan 2.0 extraction software (NimbleGen Systems Inc). For each spot on the array, log2-ratios of the Cy3-labeled test sample versus the Cy5-labeled reference sample were calculated. Fluorescence intensity normalization was performed using the Qspline algorithm (www.bioconductor.org). The

quality of each array was assessed by measuring the correlation of log2 ratios for each datapoint between duplicate arrays.

The raw data undergoes several types of breakdown analysis including segmentation analysis which uses DNAcopy software that looks for change points in normalised oaCGH data. Contiguous probes are analysed for similarity in fluorescence intensity signal, and depending on the degree of that similarity they can be grouped into a "segment" where all oligonucleotides in that segment show a similar hybridisation score, or can be broken into smaller segments. There are three kinds of output produced by the DNA segmentation analysis: a segmentation table listing all the segments and their corresponding mean log2 ratio (segmentation mean), a PDF plot visualising segments and individual probes across the region (figure 5.5) ,and a GFF file for uploading into the SignalMap software (figure 5.6) (NimbleGenSystemsInc., 2005).

Window analysis data is also supplied, where the mean log2 ratio for every 40bp, 80bp and 200bp window across the region is given, regardless of datapoints within that window. Compared to window base analysis, the segmentation method reduces analysis complexity without sacrificing resolution. Manual inspection of the data (as advocated by NimbleGen Systems Inc) would increase user analysis and may lead to errors, while window based analysis could potentially lower resolution and introduce false results (e.g. if one probe is within a window).

SignalMap software provided by NimbleGen Systems Inc allows a user to visualise probes and their log2 ratios from multiple hybridsations (Figure 5.6). Previous analyses of oaCGH data (Conrad, Andrews *et al.*, 2006, Locke, Sharp *et al.*, 2006, Selzer, Richmond *et al.*, 2005, Stallings, Nair *et al.*, 2006, Urban, Korbel *et al.*, 2006) and

NimbleGen Systems Inc personal communications detail that both hemizygous and homozygous deletions/duplications generally show a log2 ratio of $\geq \pm 0.5$. As the siblings should be identical across the putative IBD2 region any datapoints that achieved a $\geq \pm 0.5$ segmentation mean value are indicative of a CNV at that locus.

The first step in selecting putative CNVs for validation involved the comparison of the segmentation means of duplicate/sister arrays for the region (i.e. those involving the same control but different sibling). Any segment showing a $\leq$ -0.5 log2 ratio was considered as supportive evidence of a deletion in the C702 sibling, however an overlapping segment in the sister hybridisation, also showing a $\leq$ -0.5 log2 ratio, was required for the putative deletion to be considered for further analysis.

To further rule out the presence of false positive findings, for example the presence of a duplicated sequence that is lost in the control sample (therefore appearing as a deletion in the test sample), and common deletion polymorphisms that are unlikely to explain the C702 phenotype, deletions were considered for validation if only they were also identified in another pair of hybridisations involving a separate control (figure 5.7). Finally, to guard against false negative findings only one replicated finding from a separate pair of control hybridisations was required to proceed with PCR validation.

Further to the segmentation analysis, specific regions of interest such as known CNVs and large/interesting regions of homozygosity (see chapter 4) were analysed using 200bp windows across the IBD2 region. Window based analysis is easier to interpret than segmentation analysis when looking at a small region in detail with a specific hypothesis (i.e. the region is a CNV).

5.2.8. Deletion Verification

To examine all putative deletions in C702 for heterozygous or homozygous losses PCR primers were designed exterior to the probe immediately flanking the maximum segment defined by the segmentation mean across arrays showing evidence for a deletion (i.e. external to the maximum putative deleted region). The PCRs were then optimised using one of the unrelated control samples and one of the test C702 siblings over a range of primer annealing temperatures to examine for any evidence of a deletion based upon size analysis on an agarose gel (figure 5.8 and appendices 10.3). After optimisation, PCR followed by bidirectional sequencing of any bands present was performed upon all of the CGH samples. Details of PCR methods and sequencing protocols can be found in the materials and methods chapter (sections 2.2, 2.5.2 and 2.5.3).

5.3. Results

5.3.1. Sample Preparation

Samples were concentrated to 200-400ng/µl, quantified using PicoGreen and >4µg sent in provided tubes packaged in ice to NimbleGen Systems Inc. An aliquot of the concentrated sample was saved and 5µl of each run on two separate 0.5% agarose gels (figure 5.4) to check for sample degradation (<10kb molecular weight bands). The

samples showed no evidence of degradation and NimbleGen Systems Inc confirmed that the samples past their quality control procedures.

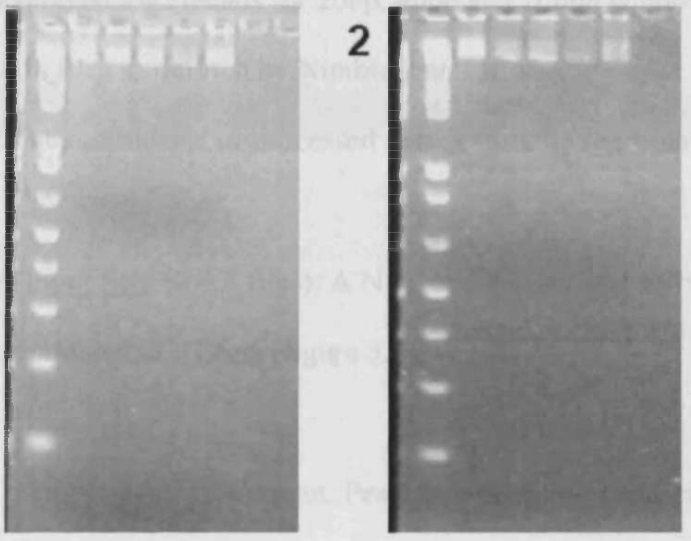


Figure 5.4: Quality control procedure as requested by NimbleGen Systems Inc. Samples were concentrated to 200-400ng/µl by centrifugation and drying in a speed-vacuum at 55°C, then 5µl of each product was loaded with sample loading buffer onto a 0.5% agarose gel which was run at 120V for 1-2 hours in 0.5x TBE buffer. The lanes contain samples (in order, left-to-right) 1kb ladder (Invitrogen), C702.07, C702.10, C0588, C1003, C1492 and loading buffer only. A duplicate gel was also run. No sample showed evidence of low molecular weight bands (i.e. <10kb).

5.3.2. Data Quality and Whole IBD2 Region Analysis

NimbleGen Systems Inc reported that they had successfully performed the six hybridisations requested using the custom designed array. The results supplied included:

1. Raw data files: Un-normalised data as from each array experiment.
2. Processed data files: Including log2 ratios for qspline normalised datapoints: Window averaged log2 means for 20bp, 40bp and 200bp window sizes and segmentation files as defined by NimbleScan software.
3. Pdf plots: The normalised or processed data is visually represented, (see figure 5.5).
4. SignalMap input files (.GFF files): A NimbleGen provided software program for visual inspection of CGH data (figure 5.6).

To assess the quality of the experiment, Pearson correlation coefficients were calculated for each array (SPSS version 14), comparing each datapoint (normalised oligonucleotide log2 ratio) for each array with the same datapoint on the other arrays. Variation between arrays should reflect either technical variation (experimental variation) or biological variation (DNA sequence differences). Hybridisations of C702 siblings involving controls C1003 and C0588 showed consistent correlation coefficients, ranging from 0.37 to 0.46. Initial comparisons indicated that arrays performed upon control C1492 showed either little or negative correlation with the other arrays. NimbleGen Systems Inc were contacted and the hybridisations repeated, although the quality of the arrays based upon this control individual does not show as high inter-hybridisation correlation as the others (table 5.1).

	702.07 vs C1003	702.07 vs C0588	702.10 vs C1003	702.10 vs C0588	702.07 vs C1492	702.10 vs C1492
702.07 vs C1003	1	0.37	0.378	0.371	0.261	-0.027
702.07 vs C0588	0.37	1	0.464	0.42	0.337	-0.018
702.10 vs C1003	0.378	0.464	1	0.458	0.248	-0.073
702.10 vs C0588	0.371	0.42	0.458	1	0.289	-0.024
702.07 vs C1492	0.261	0.337	0.248	0.289	1	0.082
702.10 vs C1492	-0.027	-0.018	-0.073	-0.024	0.082	1

Table 5.1: Pearson Correlation of log2 ratios for the individual probes across all hybridisations (387218 probes). Shown are the correlation coefficients for each hybridisation versus all others. The C702 siblings are 702.07 and 702.10; the controls are C1003, C0588 and C1492. Most experiments show a correlation lying between 0.3-0.5, except for those involving individual C1492 which show less or no correlation even after hybridisations were repeated.

Control individual C1492 passed the quality control procedures as defined by NimbleGen Systems Inc and so the cause of the lack of correlation between sister arrays using this control is currently inexplicable. Despite these results, the arrays using C1492 may be informative for some of the genomic region and so were considered in the analysis, but as putative deletions only had to be present in >1 array to meet the criteria for validation the inclusion of these arrays should only act to increase type I errors.

5.3.3. Identification of Putative Deletions by Segmentation Analysis

All 6 hybridisations were analysed using segmentation means. Comparison of putative deleted segments ($\log^2$ ratio <-0.5) from each duplicate or sister array (same control, different C702 sibling) refined the number of suspected deletions (figure 5.7). This was done for all 3 controls and the resultant compared to the results from other

duplicate arrays (figure 5.7) resulting in deletions that are most likely to represent true deletions in pedigree C702 that are absent in the controls.

23 regions met the criteria for a putative deletion, in that part of a segment with a lower than -0.5 log2 ratio replicated across hybridisation experiments as shown in figure 5.7 and described in section 5.2.7. The details of each putative deletion are shown in table 5.2, which lists all 23 deletions (termed A-W) with their maximum size and the number of probes that characterise the putative deletion. The putative deletions range in size from 77-794bp and contain between 5 and 57 probes each. To exclude the possibility that these regions have a complex sequence make-up (e.g. repetitive) the UCSC tracks "Repeat Masker", "Simple Repeats", "Segmental Duplications", "WSSD Duplications", "Human Chained Self Alignments" and "Structural Variation" (http://genome.ucsc.edu/cgi-bin/hgTracks) were examined for each of the putative deletions. Deletion A (table 5.2) lies within a region showing evidence for copy number gains (Sharp, Locke *et al.*, 2005). Also, potential deletions B, C, G and L lie within potentially duplicated sequence shown by the "Human Chained Self Alignments" track and could again explain the oaCGH findings.

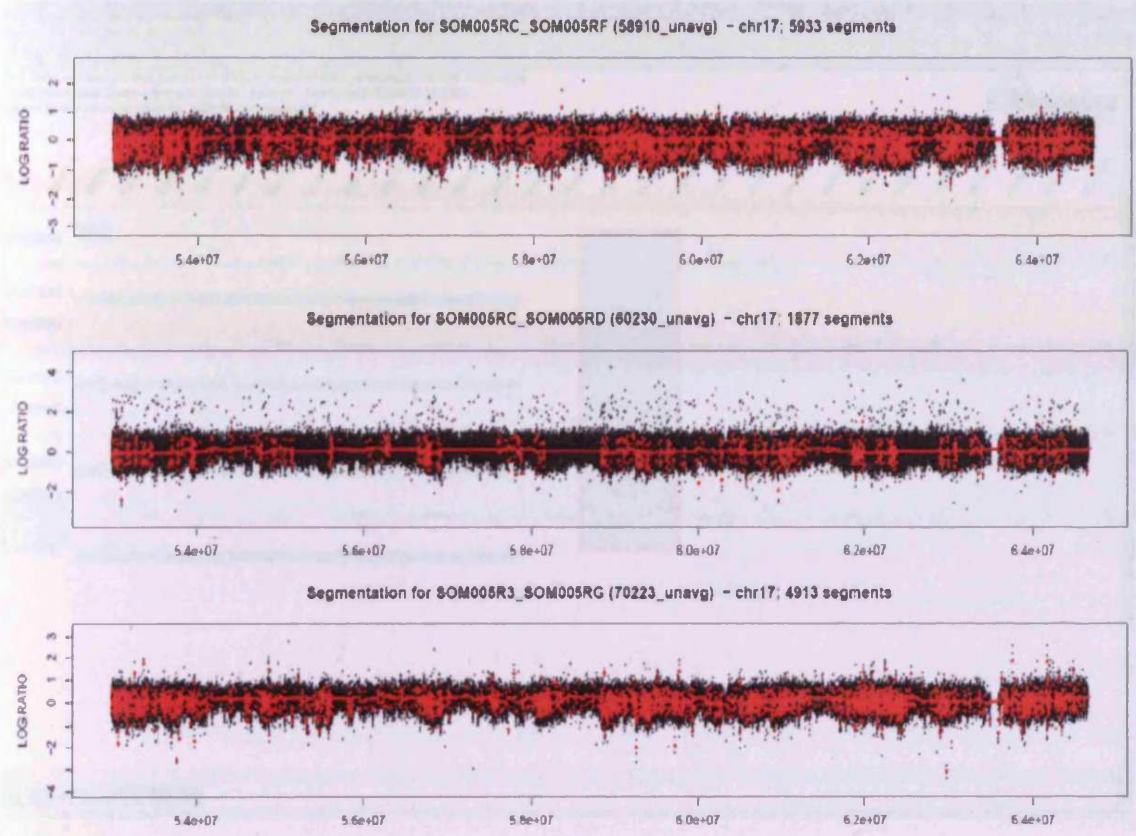


Figure 5.5: Global visualisation of processed CGH data (.pdf plots). Shown are three separate hybridisation experiments where sibling 702-07 was hybridised to controls (in descending order) C0588, C1003 and C1492. Each plot shows the genomic position of the chromosome 17 C702 IBD2 region on the horizontal axis, while the normalised log2 ratio is displayed upon the vertical axis. Black datapoints represent oligonucleotide probes at that position in the IBD2 region and the log2 ratio of that probe. Red features correspond to segments as defined by the NimbleScan software, the genomic position of those segments and the compatible log2 ratio is displayed. A region specific example of this data is shown in figure 5.6 below.

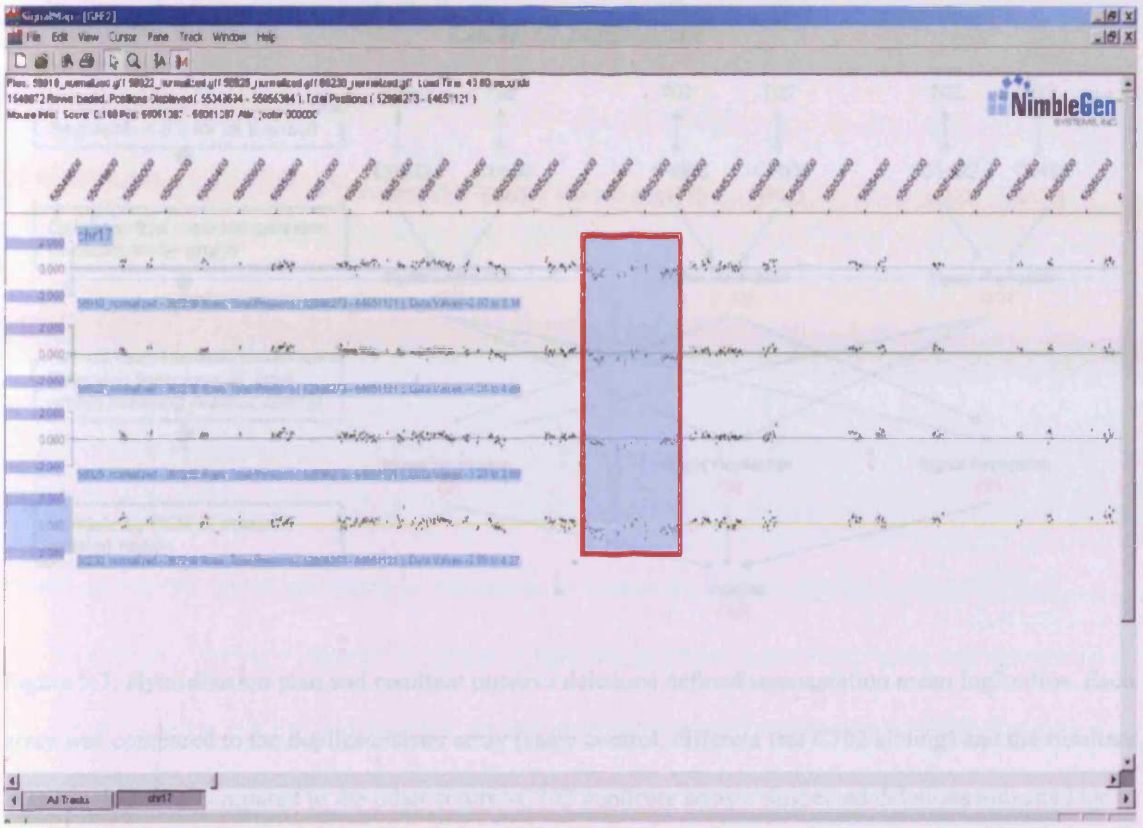


Figure 5.6: SignalMap software window, displaying four arrays. Each array has a y axis of normalised log2 ratio, with genomic position shown on the x axis (shown above all arrays). Each black dot represents an oligonucleotide. The highlighted region over all four arrays (light blue/grey, red box) is the manual delineation of a segment (defined by the segmentation analysis) where the mean log2 ratio is less than -0.5.

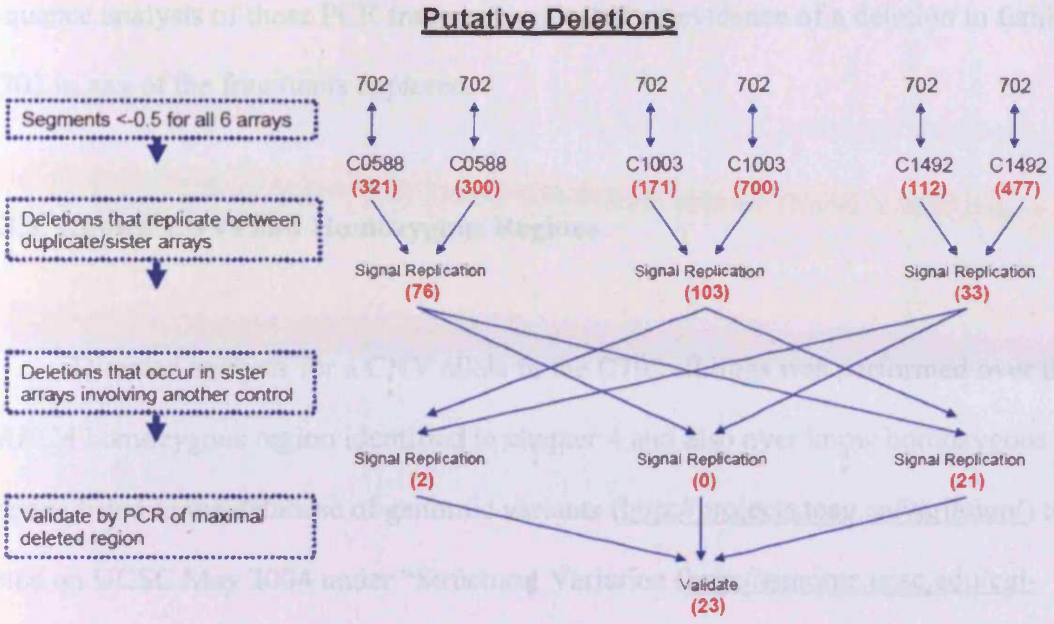


Figure 5.7: Hybridisation plan and resultant putative deletions defined segmentation mean log2 ratios. Each array was compared to the duplicate/sister array (same control, different test C702 sibling) and the resultant putative deletions compared to the other control-C702 duplicate arrays. Suspected deletions indicated by >1 control set of arrays were to be validated by PCR and sequencing. At each stage, the number of potential deletions are shown in red brackets.

5.3.4. Deletion Validation

Successful PCR and sequencing was performed for the 23 fragments encompassing an oaCGH identified potential deletion, covering all bases within the putative deletion up until the next flanking probes that were not contained within the segment (table 5.2). No evidence for a deletion variant present in a C702 sibling was visualised after agarose gel size analysis for any of the fragments over an annealing temperature gradient (data shown in figure 5.8 and appendices figure 10.3). Furthermore,

sequence analysis of these PCR fragments revealed no evidence of a deletion in family C702 in any of the fragments explored.

5.3.5. Known CNVs and Homozygous Regions

Directed analysis for a CNV allele in the C702 siblings was performed over the *PRKCA* homozygous region identified in chapter 4 and also over know homozygous regions listed in the database of genomic variants (http://projects.tcag.ca/variation/) track listed on UCSC May 2004 under "Structural Variation (http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=102894506&c=chr7&g=cnp). The mean log2 ratio for the entire region should be ±0.5 if a true CNV exists at that locus; however none of the regions showed evidence of CNV in any of the hybridisation experiments, the largest deviation from 0 being 0.21 at homozygous region 1 in one hybridisation (table 5.3). Each region should be suitably powered to detect a CNV as the minimum number of datapoints was 359, ten-times that outlined by Locke *et al* (Locke, Sharp *et al.*, 2006).

Putative Deletion ID	Maximum Deleted Region	Maximum Deletion Size	GENEID	PCR primer 1	PCR primer 2
A	chr17:62969798-62969876	78	BC067095	aaaggacatttcatggaaagatg	ccttctccccatticactggt
B	chr17:62201620-62201755	135	PRKCA	ttgcctaacatggcacttttc	cttcctgcctcttttgtgtcc
C	chr17:63407248-63408042	794	FALZ	aagccaaggagtgtgaatcc	gggggcaatttttgaatttagg
D	chr17:64239836-64239929	93	ABCA8	tcttttgctatttcctcttcatgg	gaaggctaaaggcaagaagga
E	chr17:64255550-64255687	137	ABCA9	ccttcgctcatttctgtatttcc	catgtgtgtatgtgtgtgtcattt
F	chr17:56789085-56789162	77	BCAS3	cacaagttggtgcacattcc	ctgaagtgctgggatttacagg
G	chr17:61136715-61136792	77	MGC33887	ccagcagcgttgtaggatct	aaggagagctccatgtttgg
H	chr17:63144255-63144336	81	BC064404	ggagttcccttgcaatttaaaaa	aaccaggaacctgctctcg
I	chr17:56757579-56757664	85	BCAS3	acagaatcgcacttgtgtgg	tggaagccagccagttattc
J	chr17:61083628-61083986	358	MGC33887	ttgcagtttctaaatcagcaca	aactgtccagacttttccaca
K	chr17:53765341-53765442	101	AK090885	tccagaattccagcacaaga	cttctctccgcatagcacct
L	chr17:59392366-59392639	273	SCN4A	gttgtcaaagtgctccgtca	caggggaggtgacagtaagg
M	chr17:56914923-56915090	167	TBX4	gtgtcoggtgatgtggtgt	tgtacggcgacactgaagtc
N	chr17:61072565-61072644	79	MGC33887	ggaaccagctatttgctgaga	cacccaatcttctactcctatgg
O	chr17:63856073-63856232	159	KIAA1001	acctggcttttgcactcact	gacagagtgagacggggaac
P	chr17:59391706-59391983	277	SCNA4	tcctggaaaaatgcaaagg	acatttaggggcagggagat
Q	chr17:64105549-64105702	153	FAM20A	gcttgtctgtgaagagaaaacaa	cagtgaaaccttgcctcaga
R	chr17:56915230-56915370	140	TBX4	ggtgggggaggatcactatt	ctgggacaaggagaaggtt
S	chr17:55052491-55053095	604	CLTC	ccagatttctgccaatttcgtt	aaagggaagccagtcacaga
T	chr17:60883990-60884458	468	AXIN2	tccacctgctctagatgttgc	gcccccaagcagagataga
J	chr17:53130339-53130668	349	MSI2	ccticatgcccacttttgaat	tagctgaggcagggatgg
V	chr17:62988518-62988668	152	BC067095	catcccccatticagaaaaca	gaagaactcaagtgggggaaaaa
W	chr17:61076722-61076896	174	MGC33887	tcacctggcctctggaatac	cttgccaggaaagtgctagg

Table 5.2: Shown is a summary table of the putative deletions in the C702 siblings as defined by oaCGH segmentation analysis. Putative deletions (segments with a <-0.5 log2 ratio) were compared across sister hybridisations and then other hybridisation pairs, giving maximal potential deleted regions. The deletions are listed A-W , with the maximum deletion size shown (UCSC May 2004), the nearest gene to the putative deletion and the PCR and sequencing primers used to assay each deletion (see figure 5.8 and appendices 10.3).

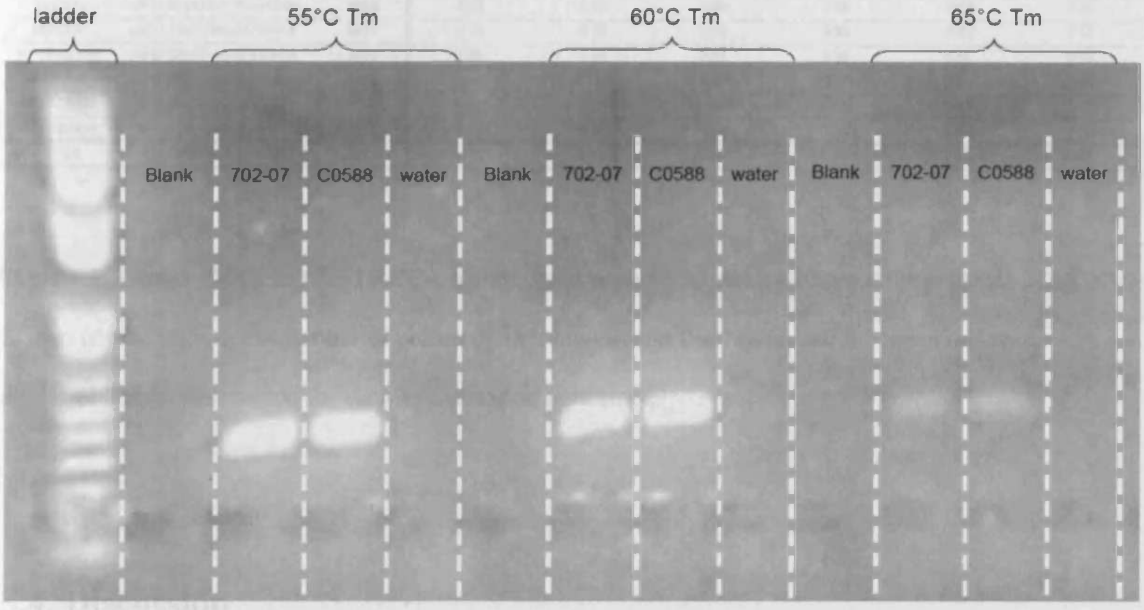


Figure 5.8: All putative deletions were optimised over a range of PCR annealing temperatures (Tm = 55°C, 60°C and 65°C) in a C702 sibling (702-07) and one of the unrelated controls (C0588) that were analysed by oaCGH. No evidence for a SSV was observed for any of the putative deletions during optimisation. Shown is a 1% agarose gel run for 20-45 minutes at 120V, loaded with 1kb plus size standard (Invitrogen) and then the relevant sample PCR (as shown, left-to-right: 702-07, C0588 and a negative water control). The above PCR represents putative deletion "A" and no deletion is visible. Further results are shown in appendices 10.3.

CNV/HMZ	Region (UCSC May 2004)	Datapoints	702.10-C0588	702.07-C0588	702.10-C1003	702.07-C1003	702.10-C1492	702.07-C1492
cnp1235	chr17 53175940-53180824	359	0.06	0.04	0.06	-0.01	-0.03	0.04
cnp1236	chr17 53997376-54149948	4548	0.06	0.06	0.05	0.05	-0.02	0.03
cnp1237	chr17 54437362-54780416	6011	0.12	0.10	0.09	0.09	-0.02	0.11
RP11-622H1	chr17 55328116-55513938	3936	0.09	0.08	0.03	0.06	0.01	0.07
cnp1239	chr17 57650340-57826573	2099	0.15	0.09	0.09	0.07	0.00	0.10
RP11-89H15	chr17 59472574-59627204	9539	0.00	0.03	-0.01	0.01	0.03	0.05
RP11-98A11	chr17 62820089-62981367	6117	0.03	0.05	0.04	0.04	-0.04	0.00
PRKCA	chr17 61851033-62282771	24766	-0.03	0.01	-0.02	0.00	0.03	0.03

Table 5.3: Known CNVs and the PRKCA homozygous region (chapter 4) analysed for evidence of a CNV. Shown is each region, the number of probes (datapoints) within that region and the mean log2 ratio across all datapoints for that region for each hybridisation performed.

5.4. Discussion

A pedigree exhibiting genome-wide significant evidence for linkage due to IBD2 status at 17q23-q24 has been analysed for the presence of a CNV across the region using high-density oaCGH. Preliminary analysis based upon replication of hybridisation signals using multiple affected individuals from the pedigree versus unrelated controls has indicated that it is unlikely that a major chromosomal segment loss explains the linkage results gained in this pedigree and their schizophrenia. Furthermore, the unlikely loss of heterozygosity at *PRKCA* (chapter 4) is also unlikely to represent the loss of genetic material in the pedigree.

There is much fervour in the current literature that SSVs, and particularly CNVs, could explain susceptibility to complex phenotypes (Eichler, 2006, Gonzalez, Kulkarni *et al.*, 2005, MacIntyre, Blackwood *et al.*, 2003, Stefansson, Helgason *et al.*, 2005) and some direct evidence that this is the case in some instances for schizophrenia

(Blackwood, Fordyce *et al.*, 2001, Millar, Pickard *et al.*, 2005, Millar, Wilson-Annan *et al.*, 2000, St Clair, Blackwood *et al.*, 1990). There is no doubt that CNVs exist and account for a proportion of human genetic variation, however the recent estimates of high frequency CNVs common throughout the human genome (Conrad, Andrews *et al.*, 2006, Hinds, Kloek *et al.*, 2006, Locke, Sharp *et al.*, 2006, McCarroll, Hadnott *et al.*, 2006, Moon, Yim *et al.*, 2006) must be tempered with the realisation that only replication of a SSV by independent technologies confirms their existence. Indeed, a review of the current knowledge of CNVs (Eichler, 2006) indicates that many CNVs identified are not re-discovered in the same samples or samples of similar population ancestry. Such an observation implies that the technologies produce many false positives or false negatives and that different technologies or platforms are biased to identify certain SSVs or that SSVs are often uncommon and under negative selection pressure.

The main plus of the NimbleGen Systems Inc custom oaCGH system used in this study is the resolution awarded: with approximately 380,000 oligonucleotide probes and an oligo length of >45 nucleotides a minimum of 17.1Mb of sequence can be covered. However, there are genomic sequences that can alter the hybridisation signals generated by CGH, examples being VNTRs, SINEs and LINEs, which together can occupy up to than 50% of the human genome (http://www.repeatmasker.org/faq.html) and can affect oaCGH analysis (Urban, Korbel *et al.*, 2006). Furthermore, gaps in the human genome sequence are also not included. Therefore, of the 11,655,026Mb of sequence 3,179,713Mb of the region is directly covered by the nucleotides of oligo probes leaving 8,475,313 bases un-tiled.

It is therefore difficult to interpret the resolution of the current study as it differs from region to region. Also, recent oaCGH guidelines indicate that >30 probes are required for confidence in claiming a CNV, further lowering any resolution. However 97% of the nucleotides in the C702 IBD2 region are within 100 nucleotides of the next oligo (figure 5.2). Therefore, if 30 probes are required for minimum confidence of a CNV this would indicate that a >3kb CNV unique to the C702 IBD2 region can be excluded with a high degree of confidence. This confidence however is dependent upon the false negative rate of the technology, which is largely unknown at the time of writing (Eichler, 2006, Locke, Sharp *et al.*, 2006), as the strategy chosen here was that a CNV signal would replicate between sister hybridisations regardless of the control sample used due to the proposed rarity of the C702 disease allele/genotype.

Of the oaCGH studies reported thus far many probes have been used to indicate the presence of a CNV (Conrad, Andrews *et al.*, 2006, Hinds, Kloek *et al.*, 2006, Locke, Sharp *et al.*, 2006, Selzer, Richmond *et al.*, 2005, Urban, Korbel *et al.*, 2006) although the smallest validated CNV by this method being a 70bp deletion as validated by 8 oligonucleotide probes (Hinds, Kloek *et al.*, 2006). However it is worth noting that many of the studies using oaCGH have only reported larger (>1-10kb) CNVs that encompass many more oligos (see review by Eichler (Eichler, 2006)). Some of the studies report that using too few probes increases the false discovery rate drastically, meaning that the smallest of CNVs may not be able to be inferred (Conrad, Andrews *et al.*, 2006), again dependent upon the coverage of that unknown CNV with probes. Unfortunately some studies do not report their rate of false positives and may have biased their analysis by

validating deletions harbouring the most probes showing the largest hybridisation signal disturbances (e.g. (Hinds, Kloek *et al.*, 2006)).

The present study used segmentation analysis, as opposed to window based or manual inspection of the data, to look at the entire C702 IBD2 region. This vastly reduces the number of features that require user attention, and allows the validation of any CNVs by an appropriate method (e.g. qualitative PCR) if the breakpoint is mapped by the segmentation analysis. However there are several potential disadvantages that centre on the performance of the algorithm: Firstly, a segment must contain $\geq$ 2 probes which increases the chance of false positive findings (if adhering to guidelines of >10-30 probes per CNV (Conrad, Andrews *et al.*, 2006, Locke, Sharp *et al.*, 2006)). Additionally, the accuracy of the algorithm in predicting CNV size is questionable; there is no data describing whether the segments defined commonly define the boundaries of CNVs precisely, or whether abutting probes are sometimes miscalled as external to the "segment". However, with these caveats considered a segmentation approach still represents a systematic algorithm for simplifying oaCGH data.

The systematic strategy to identify a rare C702 loss firstly involved the comparison of the two arrays involving both siblings and the same control (sister arrays or sister hybridisations). Segmentation means were compared between sister hybridisations for overlapping segments in both arrays that had a segmentation mean log2 ratio of less than 0.5. This involved the analysis of 300, 171 and 112 segments in hybridisations entailing controls C0588, C1003 and C1492 respectively (Figure 5.7). The resultant was 76, 103 and 33 regions indicative of a copy number loss in C702 (for controls C0588, C1003 and C1492 respectively).

The seemingly high replication of putative deletions between sister hybridisations (25%, 60% and 30% for hybridisations involving C0588, C1003 and C1492) did not however translate into inter-control replication of putative deletion findings. Only 23 deleted regions were identified in one pair of sister hybridisations and then replicated in another pair. These regions could represent several possibilities: True deletions or structural variation that is apparent as a deletion (e.g. a SNP or small insertion), regions of polymorphic duplication where controls exhibit more copies than the C702 siblings, or finally, false positive findings due to technical artefacts.

The 23 putative deletions were analysed by standard PCR, where PCR primers lay exterior to the nearest oligonucleotide probe to the maximal region defined by segmentation analysis (i.e. the maximum region identified as deleted across all hybridisations). However, none of the regions tested showed evidence for variation that could conclusively explain the CGH findings: No deletions (or duplications) were identified during PCR optimisation or in the subsequent sequencing of the amplimer. No other variants were observed that could explain a deleted segment, although some SNPs were identified where controls and the C702 siblings had disparate genotypes that could potentially explain a small proportion of the signal generated.

Considering that 387218 probes were analysed across six arrays to produce hundreds of intra-control potential C702 deletions, 23 false positives may be a pragmatic observation as the 212 total putative deletions identified from the three control replicate hybridisation gives a false discovery rate of 10.9%. Furthermore, recent papers performing detailed CGH studies in many samples have estimated that false positive rates rise unacceptably when <30 oligos define a CNV segment (Conrad, Andrews *et al.*, 2006,

Locke, Sharp *et al.*, 2006). As only 5 of the potential deletions achieve this criterion the false positive rate becomes more acceptable (2.4%).

A further explanation of the false positive findings is that they do not represent deletions but either complex sequences not masked during the repeat masking process (RepeatMasker) or they represent copy number gains in the control individuals. The RepeatMasker algorithm may have overlooked sequences that could potentially produce secondary structure or repetitive sequences that may produce spurious log2 ratios across a small region. The" Human Self Chain Alignment" track (http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=76280281&c=chr17&g=chainSelf) available at the UCSC genome browser indicates deletions B, C, G and L may lie within duplicated sequence. The sequences were not removed by the RepeatMasker software and so may represent an area of improvement for the algorithm.

Finally, the data analysed and the methods of analysis could be implicitly flawed, but it is unlikely however, that the entirety of the oaCGH data itself is erroneous. Many studies are being published entailing the use of the NimbleGen Systems Inc custom array design and oaCGH service and the utility has been used to discover and validate CNVs that are of intermediary size (>10kb) and lower (Conrad, Andrews *et al.*, 2006, Locke, Sharp *et al.*, 2006, Selzer, Richmond *et al.*, 2005, Stallings, Nair *et al.*, 2006, Urban, Korbel *et al.*, 2006). Some of the studies even report that false negative and positive rates are low (<5% and <0.2% respectively (Locke, Sharp *et al.*, 2006)). Furthermore, the present research would concur with a low false positive rate, with only 6 of the 212 identified putative deletions from hybridisation pairs containing >30 probes and only 5 of

the 23 replications where validation was attempted meeting the criteria outlined (Conrad, Andrews *et al.*, 2006, Locke, Sharp *et al.*, 2006).

Therefore, it is more likely that no deletions exist in the C702 IBD2 region that are unique to C702, or the method of data interpretation is flawed. There is no ignominy in the latter, as oaCGH analysis is a recent and therefore inexperienced field of research. The credence of segmentation or window based analysis methods is questionable, with many authors so far choosing their own different methods of statistical analysis or being fortunate enough to encounter large CNVs that can be delineated by manual inspection of the data (as NimbleGen Systems Inc advocate) (Conrad, Andrews *et al.*, 2006, Hinds, Kloek *et al.*, 2006, Locke, Sharp *et al.*, 2006, McCarroll, Hadnott *et al.*, 2006). The preliminary analysis outlined in this study, is still an introduction to oaCGH studies and the data must be re-analysed according to more robust procedures. For example, it may be valuable to analyse each hybridisation individually (either by segmentation or window) to identify putative deletions encompassing >30 probes and attempting to validate these, especially when considering that the false negative rate is unknown and may be higher than the type I error rate (Locke, Sharp *et al.*, 2006).

Although the current chapter has failed to identify a CNV that explains the C702 linkage data the analysis has effectively ruled out several possibilities for the allelic architecture of the C702 schizophrenia phenotype. There are many datapoints across the regions of known CNVs and none of the hybridisations are supportive that a C702 sibling has either gained or lost these sections. Furthermore, the exceptional region of homozygosity at *PRKCA* (chapter 3) could also be explained by a hemizygous deletion in the C702 affecteds, however again the vast number of datapoints and 6 hybridisations are

indicative that, at least for each region as a whole, the data cannot be explained by a deletion. Therefore, the *PRKCA* region of homozygosity must still remain a focus for the directed resequencing of candidate genes to identify the alleles causing the C702 schizophrenia and linkage to chromosome 17.

Chapter 6. Evidence that Rare Variants in *PRKCA* Confer Susceptibility to Schizophrenia in Family C702 and the General Population.

6.1. Introduction

A genome wide significant linkage region on chromosome 17 in a pedigree multiply affected with schizophrenia has been refined to an 11.7Mb region at 17q23.2-q24.3 where affected siblings are IBD2 (chapter 3). The lack of schizophrenia in the parents, the high penetrance of schizophrenia in the offspring and the dearth of linkage to 17q schizophrenia in other UK an Irish samples (Williams, Norton *et al.*, 2003) is consistent with a rare, highly-penetrant disease allele that is acting via a recessive mechanism in this pedigree to produce schizophrenia (chapter 4). The C702 siblings exhibit stretches of homozygous genotypes within the IBD2 region which could indicate the convergence of a disease haplotype from a common ancestor (chapter 4). Rare diplotypes therefore could identify the position of the C702 disease variant. Consistent with this hypothesis, a homozygous diplotype consisting of 82 markers and spanning 801kb that does not occur in ~3500 individuals from the UK was identified spanning the 3' of the gene *PRKCA* (chapter 4).

PRKCA encodes Protein Kinase C Alpha (PKCα), which is a classical isoform of the PKC family of kinases, and is an enzyme with a multitude of interacting partners (Poole, Pula *et al.*, 2004) and cellular roles (Larsson, 2006, Mellor and Parker, 1998,

Michie and Nakagawa, 2005). The primary function of the enzyme when activated by phosphatidylserine, diacylglyerol (or phorbol ester analogues) and Ca^{2+} is the addition of phosphate groups to other proteins at serine and threonine amino acids, thereby activating or inhibiting them (Mellor and Parker, 1998). However, this statement belies the complex role of the ubiquitously expressed PKCα in the cell (Wetsel, Khan *et al.*, 1992); where the catalytic function of PKCα is regulated by multiple signalling molecules, the availability of cofactors and also by sub-cellular localisation (Michie and Nakagawa, 2005). Furthermore, PKCα has some functional redundancy or overlap with the other classical/typical PKC isoforms (α, β, λ) (Kapfhammer, 2004, Mellor and Parker, 1998). Yet despite the extensive literature surrounding the extended PKC family of isozymes it is safe to say that our understanding of the spatial and temporal complexities of PKCα remains limited (Larsson, 2006, Mellor and Parker, 1998, Michie and Nakagawa, 2005).

PKCα has been implicated in a plethora of neural functions including synaptic signalling, long-term potentiation/depression (LTP/LTD) and certain types of memory (Alkon, Epstein *et al.*, 2005, Boehm, Kang *et al.*, 2006, Bonini, Cammarota *et al.*, 2005, Collingridge, Isaac *et al.*, 2004, Craig, Olds *et al.*, 1993, Hussain and Carpenter, 2005, Kennedy, Beale *et al.*, 2005, Leahy, Luo *et al.*, 1993, Leitges, Kovac *et al.*, 2004, Liu, Wei *et al.*, 2005, Moriguchi, Han *et al.*, 2006, Sanchez-Perez and Felipo, 2005). PKCα also has roles in neurite growth and neuronal development (Beaudry, Gendron *et al.*, 2006, Firulli, Howard *et al.*, 2003, Kapfhammer, 2004, Kennedy, Beale *et al.*, 2005, Kim and Yang, 2005, Lopez-Andreo, Torrecillas *et al.*, 2005, Miyazaki, Hashimoto *et al.*, 2006, Roisin and Barbin, 1997, Shimohama and Saitoh, 1998, Xu, Bullock *et al.*, 2005) and also glial activity such as oligodendrocyte function and myelination (Baron, de Jonge

et al., 2000, Barton, Woolmore *et al.*, 2004, Deng, Wang *et al.*, 2004, Kobayashi, Kidd *et al.*, 2001, Saarela, Kallio *et al.*, 2006, Weerth, Holtzclaw *et al.*, 2006). Therefore, regardless of the true molecular neuropathology of schizophrenia, the potential for PKCα to be involved is clear.

Further to the established roles of PKCα, the PKC family and isozyme have directly been implicated in the pathological and therapeutic pharmacology of psychiatric disorders such as schizophrenia (Basta-Kaim, Budziszewska *et al.*, 2002, Birnbaum, Yuan *et al.*, 2004, Dwivedi and Pandey, 1999, Knable, Barcia *et al.*, 2002, Mirnics, Middleton *et al.*, 2001), manic-depressive disorders (Chen, Manji *et al.*, 1994, Chen, Masana *et al.*, 2000, Hahn and Friedman, 1999, Hahn, Umapathy *et al.*, 2005, Hahn, Mazei-Robison *et al.*, 2005, Kirshenboim, Plotkin *et al.*, 2004, Knable, Barcia *et al.*, 2002, Manji and Chen, 2002, Rao, Rapoport *et al.*, 2005, Soares, Chen *et al.*, 2000) and suicide (Pandey, Dwivedi *et al.*, 2004). More specifically, the enzyme is reportedly dysregulated in areas of the frontal lobe in post-mortem studies of schizophrenics (Knable, Barcia *et al.*, 2002, Mirnics, Middleton *et al.*, 2001) and depressive individuals (Pandey, Dwivedi *et al.*, 2004), while studies of peripheral tissues have reported different levels between manic patients and controls (Soares, Chen *et al.*, 2000).

Additionally, reports have shown working memory to be specifically disrupted in primates by over activity of PKCα (Birnbaum, Yuan *et al.*, 2004) and genetic variation in *PRKCA* has been associated with human hippocampal activity and performance on tasks examining episodic memory (de Quervain and Papassotiropoulos, 2006). The PKCα enzyme is therefore a focal point for neuroscientific research and as such there is a huge

volume of circumstantial, but also direct evidence to support a possible role for the enzyme in schizophrenia and related psychiatric illnesses.

The gene encoding *PRKCA* lies at chromosome 17q24.2 spanning ~510kb (chr17:61,729,388-62,237,324. UCSC May 2004). Within the boundaries of the gene there are 25 GenBank mRNAs listed, however only 7 of these transcripts share multiple exons including the first exon, and span the majority of the region. The typical PKCα encoding isoform is given as X52479, although a further 3 isoforms could encode the protein (AY633609, BC109274, BC109274), differing in their UTR length. It is noteworthy then that the AceView transcript database lists a transcript corresponding to X52479 as having an extended 3'UTR (PRKCA.Aug.05. UCSC May 2004). The typical protein encoded by the above 4 transcripts contains two diacylglycerol binding domains, a C2 domain (Ca^{2+} binding) and a protein kinase domain (Pfam. http://www.sanger.ac.uk/Software/Pfam/), however the remaining 3 transcripts do not encode the full length PKCα. The isoform M22119 may represent an incomplete cDNA, exhibiting exons that overlap almost entirely with X52479. However, the remaining transcripts exhibit novel exons and so are unlikely to be incomplete records (UCSC May 2004). Isoform BC053321 lacks the PKC terminal domain, and the AB209475 PKC alpha variant contains only one C1 domain. Therefore, the PRKCA locus contains at least 3 isoforms that may generate alternative proteins (X52479, BC053321 and AB209475), which total 22 exons. However, further variation may exist through extended UTRs.

This chapter describes the mutation screening of exonic and regulatory sequences of *PRKCA* to identify polymorphisms in affected individuals of C702 that could

potentially explain the linkage findings and phenotype. Polymorphisms identified were analysed for association with schizophrenia and related disorders in large case-control samples from the UK, Ireland and also Germany, with further testing in family-based populations.

6.2. Materials and Methods

All nucleotide positions in this chapter are given according to Build 35 (HG17) of the human genome assembly, as recorded by the UCSC genome browser, May 2004 freeze.

6.2.1. Mutation Screening

The gene encoding *PRKCA* lies at chr17:61,729,216- 62,237,328 covering 508,112 bp. All candidate gene exonic DNA sequences were acquired from UCSC Genome Browser (http://genome.ucsc.edu/), and AceView (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/). Exonic sequences selected for screening were from human mRNAs BC053321, AB209475 and X52479 (GenBank human mRNAs) and the extended 3'UTR demonstrated by AceView annotated mRNAs, that has identical sequence to X52479 except for a shorter 5'UTR and larger 3'UTR (PRKCA.Aug.05).

Mutation screening was required for a total of 22 exons, the putative promoter region (1kb prior to initiation of transcription site, common to all identified *PRKCA*

isoforms) and 9 sequences suggested to be harbouring regulatory motifs by the Cluster Buster algorithm (http://zlab.bu.edu/cluster-buster/).

Details of PCR and sequencing methodologies and protocols can be found in the materials and methods chapter (sections 2.2, 2.5.2 and 2.5.3). Sequencing was performed on 14 unrelated schizophrenics from a large UK case and control sample and also on a C702 affected sibling or pool of C702 linked/affected siblings. 14 unrelated individuals from the same population should allow 95% power to detect alleles with a minor allele frequency of >0.1 and 80% power to detect alleles with a minor allele frequency of 0.05.

6.2.2. Bioinformatic resources for analysis of rare alleles identified in pedigree C702

Several datasets were analysed for evidence that the rare C702 alleles disrupted regulatory elements present in the 3'UTR; such as motifs in the RNA that bind proteins and other RNA (i.e. micro-RNAs). UTResource contains several packages that allow a user to screen an untranslated region sequence for such motifs (http://www.ba.itb.cnr.it/UTR/UTRHome.html). A further source of information comes from a comparative analysis of mammalian genomes to identify conserved elements and their corresponding micro-RNAs present within 3'UTR regions (Xie, Lu *et al*, 2005). Also, a database of putative micro-RNA sites is listed on the UCSC March 2006 freeze (TargetScan miRNA regulatory sites), which was devised using a similar approach to Xie and colleagues (Xie, Lu *et* al, 2005). These resources were used to analyse the 3'UTRs of *PRKCA* mRNA isoforms AB209475 and X52479 for regulatory motifs that were disrupted or created by the presence of the rare C702 identified alleles.

6.2.3. Association samples

Details of association samples can be found in the materials and methods chapter, section 2.1.1.

All schizophrenia cases used in this study are unrelated Caucasians born in the UK or Ireland. All 709 cases (483 males and 226 females) met DSM-IV criteria for schizophrenia.

All bipolar disorder cases used in this study are unrelated Caucasians born in the UK or Ireland. All 702 bipolar disorder I cases met DSM-IV criteria (BPI. 257 males and 441 females). Also analysed in this study were individuals with diagnoses of schizoaffective bipolar disorder (SABP. N=78) and individuals with bipolar disorder II (BPII. N=20), psychosis not-otherwise-specified (PNOS. N=24) and a further 28 individuals with bipolar related phenotypes.

All 992 unipolar depression cases are unrelated Caucasians born in the UK or Ireland, (298 males and 656 females) and meet DSM-IV criteria for unipolar depression. Additional to this sample are 98 female cases with DSM-IV post-natal depression.

Blood-donor controls (N=1883. 881 males, 859 females) were all Caucasian, from the United Kingdom and obtained from the UK National Blood Service. A further sample of 1056 UK control subjects was obtained from the Wellcome Trust; a population control cohort (528 males, 528 females) using a sub-set from 17,000 individuals born in England, Scotland and Wales on one day in 1958.

The Bulgarian proband-parent trio sample genotyped in this chapter consisted of 694 trios, where probands with had diagnoses of DSM-IV schizophrenia (N=469),

Bipolar 1 disorder (N=170) and schizoaffective disorder (N=55). Another trios sample analysed was the NIMH autism proband-parent trio sample, consisting of 95 patients and their parents. The probands met a diagnosis of ICD-10 autism and the sample was obtained with permission from the NIMH (United States of America).

A schizophrenia and schizoaffective disorder case-control sample from the Republic of Ireland was made available through collaboration (Derek Morris and Mike Gill. Trinity College, Dublin, Ireland). All schizophrenia cases were unrelated Caucasians of Irish nationality. All 76 schizoaffective, 296 schizophrenia and 1 schizophreniform case (223 males and 127 females) met DSM-IV criteria. The sample also included 812 blood donor controls from the Republic of Ireland health services, consisting of 518 males and 288 females.

A schizophrenia case-control sample from Germany was made available for analysis of data only, through collaboration (Dan Rujescu. Department of Psychiatry, Ludwig Maximilians University, Germany). All schizophrenia cases were unrelated Caucasians of German nationality. All 513 schizophrenia cases (334 males and 179 females) met DSM-IV and ICD-10 criteria for schizophrenia. The sample also included 1332 controls randomly selected from the Munich area of Germany, consisting of 609 males and 721 females.

6.2.4. Individual Genotyping and Association Analysis

Individual genotyping of all samples except the German case-control sample was performed using the Amplifluor (section 2.5.5) or iPlex Sequenom MassARRAY

platforms (section 2.5.6). Individual genotyping of the German case-control sample was performed by Dan Rujescu (Department of Psychiatry, Ludwig Maximilians University, Germany) using a Sequenom MassARRAY platform and iPlex chemistry. Rare E21K and E19A alleles were confirmed by SNaPshot (Chapter section 2.5.4) or sequencing (sections 2.5.2 and 2.5.3) where required to confirm the presence of the C702 identified E21K-E19A haplotype. Assays for SNPs and sequences analysed by myself in this chapter are given in appendices 10.4 with example genotyping.

Association analysis was performed using a standard chi square test (Minitab or Haploview) for cases and controls, or a TDT (Haploview) for familial samples. Odds ratios (ORs) were calculated using the formula AD/BC, where A is allele 1 in cases, B is allele 2 in cases, C is allele 1 in controls and 2 is allele 2 in controls, using a Microsoft Excel formula sheet provided by Dr. Nadine Norton.

6.3.1. Results

6.3.2. Mutation Screening

All amplimers were bidirectionally sequenced in a pedigree C702 affected sibling and also in 14 unrelated schizophrenics, except for 202 bases in the first exon at chr17:61729254-61729455 (UCSC May 2004). 27 novel variants were discovered after sequencing ~20,794 bases and the results are shown in table 6.1. Also, 17 SNPs already catalogued by dbSNP were identified. Association analysis was attempted on all known

and novel variants and details are given in chapter 7. Of the discovered novel variants; 6 are intronic, 3 are 5' to the transcriptional start site of all *PRKCA* isoforms and 18 are exonic (all 3'UTR). A summary of variants discovered is shown in table 6.1. All novel variants are shown with sequence traces in appendices 10.4, except for polymorphisms E19A (figure 6.1) the polymorphisms discovered in exon 21K (figure 6.2).

The a priori hypothesis explaining the C702 linkage results (chapter 3 and (Williams, Norton *et al.*, 2003) and the rare homozygous diplotype seen at *PRKCA* (chapter 4) is that the affected siblings will harbour a rare homozygous disease allele(s) at this locus. The 14 unrelated schizophrenic mutation screening sample has a power of 0.8 to detect alleles with a MAF $\geq$ 0.05 and a power of 0.95 to detect alleles with a MAF $\geq$ 0.1. Therefore, common alleles should be detectable within this sample.

POLY ID	POS	FLANKING	SITE	ISOFORM
P1	61728847	GAGGAAAGAAAAAGGCGGCG[G/A]AGGAGCGCGGACCCGCCACT	5'	NA
P[GCC]	61728913	CTCGGGCCGGCCCTCCCCCG[GCC]TCCCCTTGCCCAGTGTTCCC	5'	NA
P2	61729144	GGTGGGCGGTGCCGCTCCGG[G/A]CCGCCTCTGCCGCCGCGAGT	5'	NA
R2A	61751983	AGTCACGTATTGCTTTCATG[G/A]TCTTTTACAGAGCTCTGTGA	INT	NA
R3	61758980	ACCATGGCCAGCTAATTTTT[G/A]TATTTTTAGTAGAGACAGGG	INT	NA
E20	61981167	TCACTCTCCGAAGGTAAGGT[GGGT]GACCTGTTAATTTTAAAGGA	INT	NA
E10A	62129885	ATTTTTTTTCTGTTAAAAAA[T/A]TTTTTTTTTGACCCTACACA	INT	NA
E10B(1)	62130370	GGGTGACAGACCGAGACTCC[A/G]TCTCAAGAAAAAAAAAAANT	3'UTR	BC053321
E10B(2)	62130389	CNTCTCAAGAAAAAAAAAAA[A]TAGAAAAAATATACTCTTAA	3'UTR	BC053321
E13	62165454	GTGTGAGTATTATTTTTTAA[G/C]TCCTTTAATTTGAACAACCA	INT	NA
E21A_2B(1)	62201079	GGCGTGGTGGTGCACATCTG[T/C]AAACTCAGCTACTTGAGCTA	3'UTR	AB209475
E21A_2B(2)	62201284	GGGGTTGGGTTATATGGTTC[C/T]CTGCAAGTTTGTCTGCAAAG	3'UTR	AB209475
E21C_3(1)	62201766	TGGTCAGTGCCCACCCCCAC[A/G]TAGGACACTTCCGTGGCAAA	INT	NA
E21C_3(2)	62201887	TTAGGGTCTCAGATTTCTCC[G/C]CTAATTCCTCCTCCTTTCAG	3'UTR	AB209475
E21C_3(3)	62201911	ATTCCTCCTCCTTTCAGCC[G/A]AGCACAGATCAGTTACATAG	3'UTR	AB209475
E21D	62202031	AAAGAGGCAGGAAGTATAGC[A/G]TATCTGAGAGCTCAGCAGGG	3'UTR	AB209475
E21G	62203282	TTGGAAGAGGGCCAAGCCGG[C/T]GACTTGAGAGATCTAGTGCA	3'UTR	AB209475
E21K[G]	62204557	CTGGCAGAGAAGTCCCCTTT[G]GGGGGGTGCATTTTGAAGCT	3'UTR	AB209475
E21K(1)	62204820	AAAAACGGTGATGGCTGGGC[C/G]CNGTGGCTCACACCTGTAAT	3'UTR	AB209475
E21K(2)	62204822	AAACGGTGATGGCTGGGCNC[G/A]GTGGCTCACACCTGTAATCC	3'UTR	AB209475
E19A	62230649	GAGAACAAACACCTCCCCAG[C/T]CCCCAGCCCTCCCCGCAGTG	3'UTR	X52479*
E19C(2)	62231088	AGCAGGGAGGGATTGGGGGT[G/A]GGGGAGGCCTCAAAATACCG	3'UTR	X52479*
E19D(1)	62231329	TTTTCATTCTTTGTAAGAGG[C/T]CAAATCGTCTAAGGACGTTG	3'UTR	X52479*
E19D(2)	62231360	AAGGACGTTGCTGAACAAGC[G/A]TGTGAAATCATTTCAGATCA	3'UTR	X52479*
E19F	62232348	TGGTACACACGTGGCATTGC[C/T]GCAGCACCTGGGCTGACCTT	3'UTR	X52479*
E19L	62234830	AACACCGCACAACTAACAAC[T/G]AACACCGCAGTTCCCACCTG	3'UTR	X52479*
E19L_b	62234903	TTTTTTTGTTTTTGTTCCTG[T]GCTGGAATTTGTTTTCTCAG	3'UTR	X52479*

"POLY ID" Polymorphism identifier
"POS" HG17 position
"FLANKING" polymorphism in brackets with flanking sequence. No forward slash denotes a repeat
"N" denotes another polymorphic base
"SITE" is the location within the gene. 5' = 5' to first exon, INT = intron
"ISOFORM" mRNA species where SNP lies (if in exon)

Table 6.1: Novel polymorphisms identified by mutation screening of *PRKCA* in 15 unrelated schizophrenics (including C702). Note (*) X52479 refers to both X52479 and extended 3'UTR transcript (AceView)

6.3.3. Single Nucleotide Polymorphism E19A

Sequencing of 14 unrelated individuals and C702 revealed a C>T transition at position 62230649 (Positive strand. UCSC May 2004), which lies +9174 nucleotides from the transcriptional start site for X52479 and 31 bases into the X52479 3'UTR. The

C702 sibs were homozygous T, while the 14 controls were homozygous C (figure 6.1), therefore the allele is homozygous in the C702 siblings and rare.

CAGCGAGAACAAACACCTCCCCAG[**C/T**]CCCCAGCCCTCCCCGCAGTGGGAA

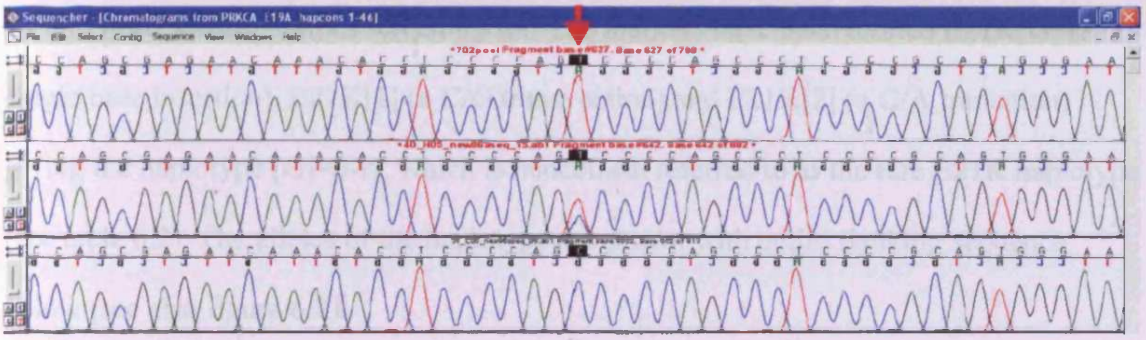


Figure 6.1: Sequence and Sequencher trace of polymorphism E19A. The positive strand sequence of polymorphism E19A is shown. Also shown is a Sequencher output, where the three sequences represent (in descending order) a pool of C702 siblings (homozygous T), a heterozygote and a homozygote wild-type. The red arrow indicates the SNP.

The E19A SNP is unlikely to represent a deletion as a PCR amplimer encompassing both E19A and neighbouring polymorphism (rs2286674) showed that the C702 siblings are heterozygous for rs2286674 (E19A_3'HET PCR and sequencing primer assay. Appendices 10.4).

6.3.4. Variants E21K[G], E21K[1] and E21K[2]

Sequencing of *PRKCA* alternative mRNA isoform AB209475 in 14 unrelated individuals and C702 revealed a haplotype in the 3'UTR consisting of three SNP alleles present in homozygous form in the C702 siblings. None of the alleles was present in the 14 unrelated schizophrenics screening set. The markers have been termed E21K[G] (A single base insertion), E21K[1] (a C/G transversion) and E21K[2] (a G/A transition) giving the haplotype [G]-G-A, which is henceforth referred to as the rare E21K haplotype (see table 6.2). The alleles of the haplotype are shown with entire flanking sequence below (see also figure 6.2):

The identification of individuals in C702 homozygous for the E21K haplotype confirms the phase of the alleles (i.e. the haplotype [G]-G-A for markers E21K[G], E21K[1] and E21K[2] respectively). Sequencing of these alleles in the HapMap CEU individuals showed 7 individuals carry all the E21K alleles; 2 transmissions are consistent with a haplotype and one is unclear (both parents show all the E21K alleles). Therefore, the E21K[G], E21K[1] and E21K[2] alleles are in high LD (r2 and D' of 1. D' confidence interval of 0.67-1). and any one of the SNPs should serve as a proxy for the E21K haplotype [G]-G-A.

TTTCTTCTGGCAGAGAAGTCCCCTTT[**G**/−]GGGGGGTGCATTTTGAAGCTCTGC

CACACTGGCCTCAGACCTCATTTGCAAATTGGCCTGAGGTGATTTACAGGCCT

TATTTAAAGGTTTCCCTACTAAGCTCACGCCCCCTGGAGAGCACCAGGCTCCA

GCTCTCAATAGTTCTGCAAACTTTCTGGCCTGTTTCCAATATTTCCCACACTGC

TTCTCAGAGCCACTTAGCCTGCCTTATATGCATACTTTCAGGTCTCCATCATCT

CTATTAAAAACGGTGATGGCTGGGC[**C**/**G**]C[**G**/**A**]GTGGCTCACACCTGTAATC

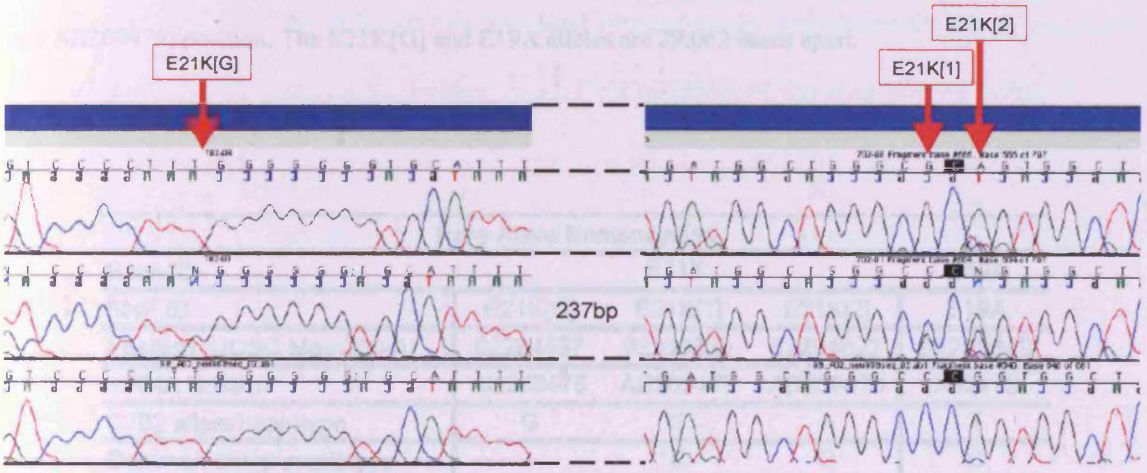


Figure 6.2: : Sequencher traces of polymorphisms composing the E21K haplotype. Polymorphisms are denoted by red arrows and are (left-to-right, indicating the 5'-3' direction of transcription of AB209475) E21K[G], E21K[1] and E21K[2]. 237 bp separate the bases in view from each other. The three sequence traces represent (in descending order) a C702 sibling (homozygous), a heterozygote and a homozygote wild-type.

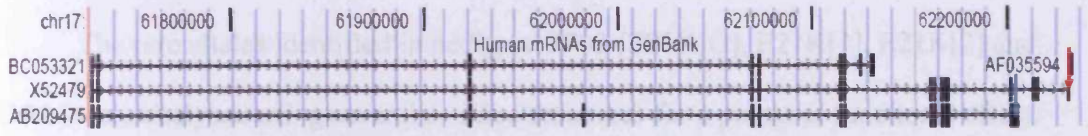


Figure 6.3: UCSC genome browser track (May 2004) showing the genomic position, human mRNAs analysed in this study and the approximate position of the E21K alleles (turquoise arrow) and E19A allele (red arrow) are also shown. The alleles are shown relative to their genomic and mRNA species (X52479 and AB209475) position. The E21K[G] and E19A alleles are 29,062 bases apart.

Rare Allele Nomenclature				
Exon ID		E21K		E19A
SNP ID	E21K[G]	E21K[1]	E21K[2]	E19A
Position (UCSC May 2004)	62204557	62204820	62204822	62230649
mRNA isoform	AB209475	AB209475	AB209475	X52479
C702 allele/haplotype	G	G	A	T
Common allele/ haplotype	–	C	G	C

Table 6.2: Rare allele nomenclature and information. Shown are the exon identifiers where the rare C702 identified alleles were discovered, the SNP identifiers, position in the genome and in mRNA isoforms and the rare and common haplotypes. Therefore, the C702 E21K haplotype is [G]-G-A for the markers E21K[G], E21K[1] and E21K[2] respectively. This forms the C702 identified four SNP haplotype with marker E19A: [G]-G-A-T.

6.3.5. Bioinformatic analysis of E21K and E19A alleles

The rare alleles identified in pedigree C702 (E21K[G], E21K[1], E21K[2] and E19A) were analysed using several *in silico* resources for a potential functional effect. Several resources that list micro-RNA binding sites were examined (Xie *et al,* 2005. TargetScan miRNA regulatory sites (UCSC Genome Browser. March 2006 freeze)) and although micro-RNA binding sites were identified in the 3'UTRs of AB209475 (two motifs) and X52479/aNov07 (19 motifs) none were directly transected by the rare alleles and no micro-RNA binding sites were created by addition of the rare alleles to the reference sequence. Analysis of the wild-type (non carrier of C702 identified alleles) and mutant (C702 allele carrier) 3'UTRs using UTRscan (http://www.ba.itb.cnr.it/UTR/) identified several potentially regulatory RNA-protein and RNA-RNA binding sites in both isoforms, however no sites were disrupted or created by the C702 identified rare alleles.

6.3.6. Linkage Disequilibrium Between the E21K and E19A alleles

The affected members of C702 are homozygous for a rare haplotype, consisting of four SNPs; E21K[G], E21K[1], E21K[2] and E19A, forming the haplotype [G]-G-A-T (table 6.2). The E19A SNP and either the E21K[G] or E21K[2] SNPs were genotyped in 2484 blood donor control individuals from the UK and Ireland. Estimation of LD in this sample gives a D' of 0.53 and an r^2 of 0.12 (Haploview). LD calculation using parent-proband trios from Bulgaria gives a D' of 0.74 and an r^2 of 0.07 between the E21K[2]

and E19A alleles (Haploview). Therefore the E21K and E19A alleles are in appreciable LD.

Individuals heterozygous at E21K and also at E19A may display alleles that are on the same chromosome and form a haplotype (i.e. *cis*), or alternatively the alleles may be on different chromosomes and form a "meta-SNP" (Mira, Alcais *et al.*, 2004) (i.e. *trans*). The Bulgarian trios sample displays 11 transmissions from parent to offspring consistent with the C702 identified E21K-E19A haplotype. Furthermore, there is no conformation consistent with a *trans* arrangement of these alleles in any proband. Therefore, although the phase of the rare alleles at E21K and E19A cannot be confirmed in case and control individuals (without familial genotypes), the most common conformation of the rare alleles at E21K and E19A is as a haplotype. However, there is no reason to believe either a cis or trans arrangement would preferentially show association at this time (as the C702 siblings exhibit both) and so carriers of both the E21K and E19A rare alleles were deemed to carry the C702 identified [G]-G-A-T haplotype and analysed for association.

6.3.7. Association Analysis of E19A, E21K and the E21K-E19A Haplotype

The E19A allele and at least one of the E21K alleles have been genotyped in multiple association samples (see materials and methods 6.2.2 or materials and methods chapter section 2.1.1). All cases studied met DSM-IV or ICD-10 criteria for their psychiatric diagnosis. The UK Caucasian case samples are 709 schizophrenics, 702 bipolar I cases (BPI), 78 schizoaffective disorder (SABP) cases, 20 bipolar II cases

(BPII), 24 individuals with psychosis not-otherwise-specified (PNOS), 28 individuals with a bipolar related phenotype, 992 unipolar depression cases (UPD) and 98 individuals with post-natal depression (PND). The case samples are compared to a blood donor controls (N=1883) and individuals from the 1958 birth cohort (N=1058). Additional case and control samples were from the Republic of Ireland (SABP N=76, schizophrenia N=296, schizophreniform N=1, blood-donor controls N=812) and Germany (schizophrenia N=513, controls N=1332).

Two familial samples were also assayed, a Bulgarian familial sample consisting of 694 trios (schizophrenia N=469, BPI N=170, SABP N=55) and a small sample of autism proband trios (N=95) from the United states.

Analysis of the C702 identified alleles in 2567 controls from the UK results in a minor allele frequency (MAF) for E21K, E19A and the resultant [G]-G-A-T haplotype of 0.031, 0.007 and 0.003 respectively (table 6.3). A rarity that is confirmed in parents from the Bulgarian trios sample (MAFs of 0.067, 0.009 and 0.007 respectively. Table 6.5), and an Irish control sample (MAFs of 0.023, 0.013 and 0.01 respectively. Table 6.4).

In the 8932 individuals of European descent genotyped, there were no occurrences of the C702 identified E21K-E19A diplotype. 5 homozygotes for the E21K alleles were observed (4 cases of mixed phenotype, 1 control), and 2 homozygotes for E19A (post-natal depression case, one parent of unknown phenotype from Bulgaria). Note that the data for the German case-control sample is not presented as the E21K alleles have not been genotyped and also the autism trios sample data is not presented as

the sample was exceedingly small. However, these samples have been genotyped in the search for an individual homozygous for the [G]-G-A-T C702 haplotype.

The results of association analysis of the E21K and E19A SNPs and their haplotype are shown in table 6.3. In the largest sample of affected individuals (including all psychiatric disorder cases from the UK case-control samples) the C702 identified rare E21K-E19A haplotype [G]-G-A-T is significantly associated with disease ($p=0.05$). Although not statistically significant, there is a trend for association in the psychosis sample ($p=0.065$). There is no evidence for association of the [G]-G-A-T haplotype in the smaller schizophrenia sample ($p=0.415$) and there is no evidence of association at the E21K or E19A SNPs in the sample as a whole.

E21K[1]	ALL						MALE						FEMALE					
	CASE N	*CON N*	CASE MAF	CON MAF	P	OR	*CASE N*	*CON N*	CASE MAF	CON MAF	P	OR	*CASE N*	*CON N*	CASE MAF	CON MAF	P	OR
Schizophrenia	*669*	*1715*	0.025	0.031	**0.211**	-	*441*	*823*	0.023	0.032	**0.173**	-	*212*	*760*	0.028	0.036	**0.468**	-
Psychosis[2]	*1319*	*1715*	0.028	0.031	**0.437**	-	*664*	*823*	0.026	0.032	**0.350**	-	*596*	*760*	0.030	0.036	**0.442**	-
Affected[3]	*2420*	*1715*	0.028	0.031	**0.312**	-	*966*	*823*	0.030	0.032	**0.641**	-	*1357*	*760*	0.027	0.036	**0.114**	-

E19A	ALL						MALE						FEMALE					
	CASE N	*CON N*	CASE MAF	CON MAF	P	OR	*CASE N*	*CON N*	CASE MAF	CON MAF	P	OR	*CASE N*	*CON N*	CASE MAF	CON MAF	P	OR
Schizophrenia	*664*	*2937*	0.008	0.007	**0.515**	-	*436*	*1410*	0.011	0.006	**0.074**	**2.0**	*212*	*1392*	0.002	0.008	**0.357**	-
Psychosis[2]	*1470*	*2937*	0.010	0.007	**0.142**	-	*713*	*1410*	0.012	0.006	**0.029**	**2.1**	*695*	*1392*	0.008	0.008	**0.906**	-
Affected[3]	*2567*	*2937*	0.009	0.007	**0.243**	-	*1016*	*1410*	0.012	0.006	**0.020**	**2.1**	*1453*	*1392*	0.007	0.008	**0.548**	-

Haplotype	ALL						MALE						FEMALE					
	CASE N	*CON N*	CASE MAF	CON MAF	P	OR	*CASE N*	*CON N*	CASE MAF	CON MAF	P	OR	*CASE N*	*CON N*	CASE MAF	CON MAF	P	OR
Schizophrenia	*664*	*2937*	0.005	0.003	**0.415**	-	*436*	*1410*	0.006	0.002	**0.063**	**3.2**	*212*	*1392*	0.002	0.004	**1.000**	-
Psychosis[2]	*1470*	*2937*	0.005	0.003	**0.065**	**1.9**	*713*	*1410*	0.006	0.002	**0.021**	**3.6**	*695*	*1392*	0.005	0.004	**0.743**	-
Affected[3]	*2567*	*2937*	0.005	0.003	**0.050**	**1.8**	*1016*	*1410*	0.007	0.002	**0.005**	**3.9**	*1453*	*1392*	0.004	0.004	**0.926**	-

[1] As no association was detected with E21K the entire control sample was not assayed

[2] Includes UK schizophrenia, bipolar I disorder, schizoaffective disorder and psychosis NOS

[3] As [2], including UK unipolar depression, bipolar II disorder and post-natal depression samples

Table 6.3: The rare *PRKCA* alleles homozygous in the C702 siblings (Alleles E21K, E19A and the haplotype [G]-G-A-T) have been assayed through samples of cases and controls from the UK. Shown is the number of cases and controls successfully genotyped (*Case N, Control N*), the Minor Allele Frequency (MAF), the association P value and Odds Ratio (OR) if significant or showing a trend ($p \leq 0.1$) for case-control samples. The case samples are a UK sample of schizophrenics ("Schizophrenia"), this sample is then supplemented by patients exhibiting other psychotic disorders ("Psychosis") and further by depressive disorders ("Affected"). Note that sex information was unavailable for a small number of the UK samples. Association was found in males for the rare E19A SNP allele, which becomes stronger in larger samples. The haplotype is associated in the entire psychiatric disorder sample ($p=.05$) and the effect size is greatest in male affecteds ($p=0.005$, OR=3.9).

	ALL						MALE						FEMALE					
E21K	*CASE N*	*CON N*	CASE MAF	CON MAF	P	OR	*CASE N*	*CON N*	CASE MAF	CON MAF	P	OR	*CASE N*	*CON N*	CASE MAF	CON MAF	P	OR
	353	*769*	0.040	0.023	**0.032**	**1.7**	*209*	*499*	0.038	0.023	**0.110**	**1.7**	*121*	*267*	0.050	0.024	**0.065**	**2.1**
E19A	*CASE N*	*CON N*	CASE MAF	CON MAF	P	OR	*CASE N*	*CON N*	CASE MAF	CON MAF	P	OR	*CASE N*	*CON N*	CASE MAF	CON MAF	P	OR
	357	*782*	0.013	0.013	**0.873**	-	*212*	*506*	0.007	0.012	**0.416**	-	*122*	*272*	0.025	0.017	**0.445**	-
Haplotype	*CASE N*	*CON N*	CASE MAF	CON MAF	P	OR	*CASE N*	*CON N*	CASE MAF	CON MAF	P	OR	*CASE N*	*CON N*	CASE MAF	CON MAF	P	OR
	357	*782*	0.008	0.010	**0.675**	-	*212*	*506*	0.005	0.010	**0.527**	-	*122*	*272*	0.016	0.011	**0.509**	-

Table 6.4: The rare PRKCA alleles homozygous in the C702 siblings were assayed through a case control sample of schizophrenics (N=296) and schizoaffectives (N=76) plus blood donor controls (N=812) from Ireland. Shown is the number of cases and controls successfully genotyped (*Case N, Control N*), the Minor Allele Frequency (MAF), the association P value and Odds Ratio (OR) if significant or showing a trend ($p \leq 0.1$). No association was noted for the E19A allele or haplotype as in the UK sample, however a significant difference in allele frequency was seen in the whole sample for the E21K alleles (E21K[G]).

	N Trios	MAF	T:NT	P Value	T:NT (to son)	P Value
E21K	664	0.067	96:88	**0.442**	-	-
E19A	673	0.009	14:11	**0.578**	10:7	**0.513**
E21K-E19A	673	0.007	11:10	**0.842**	7:6	**0.808**

Table 6.5: The rare PRKCA alleles homozygous in the C702 siblings (alleles E21K and E19A) have been assayed through a large sample of parent-proband trios from Bulgaria (schizophrenia and schizoaffective bipolar disorder). The Number of trios analysed, the Minor Allele Frequency (MAF), the number of transmissions versus non-transmissions (T:NT) of the minor alleles and the resultant TDT p value are shown. The minor alleles are the same as for the UK samples. There was no significant transmission distortion of any of the rare alleles or haplotypes, furthermore there was no evidence for over transmission of the rare allele(s)/haplotype to affected male probands (T:NT (to sons)).

Sex-specific analysis shows a weak trend for association of the minor allele of E19A with schizophrenia when males only are considered (p=.074. OR=2), an association that becomes significant in the larger psychosis sample (p=0.029, OR-2.1) and even more so in the combined affected sample (p=0.02, OR=2.0). When the rare, C702 identified haplotype [G]-G-A-T is considered a larger and more significant effect is seen in all samples, with a trend for association of the haplotype with schizophrenia (p=0.063, OR=3.2) and a significant association in the male psychosis sample (p=0.021, OR=3.6). The largest effect size and greatest significance is seen when all males with a psychiatric diagnosis are considered versus all male controls (p=0.005, OR=3.9)

To replicate these findings a sample of individuals with schizophrenia and schizoaffective disorder (N=372) and controls (N=812) from the Republic of Ireland were assayed for the C702 identified alleles. The expected increase of the C702 identified [G]-G-A-T haplotype in affected individuals and/or males that would replicate the UK association findings was not found (table 6.4), while analysis supported an association with the E21K[G] allele (p=0.032, OR=1.72) that did not appear to be due to a sex-specific effect (Male p=0.11, OR=1.69. Female p=0.065, OR=2.09). Comparison of the UK and Irish control samples revealed a significant difference between the E21K-E19A haplotype frequency in the two samples (p=0.0001) and also for E19A (p=0.0077), with a meagre trend for differences in the E21K allele frequency between the two samples (p=0.1166).

Further replication was attempted in a parent-affected sample of trios from Bulgaria. Analysis of a total sample of 673 trios with a psychiatric disorder

(predominantly schizophrenia and schizoaffective disorder) for all alleles and haplotypes reveals no association (table 6.5). It was not possible to examine for a parent-of-origin effect (e.g. transmissions from mother to son) due to the small number of observations.

6.4. Discussion

Protein Kinase C Alpha (PKCα) is located in the refined 11.7Mb IBD2 region of 17q23-q24 (see chapter 3). Homozygosity mapping of the refined IBD2 region identified a stretch of homozygosity which encompassed the majority of *PRKCA* and was extremely rare (see chapter 4). Mutation screening of the *PRKCA* gene in affected individuals from C702 identified rare exonic polymorphisms that form a haplotype homozygous in affected member of pedigree C702. The C702 haplotype has a frequency of 0.003 in unrelated unaffected individuals from the UK, giving an expected diplotype frequency of 0.000009. The rarity of the C702 diplotype makes this region an excellent candidate for being responsible for the linkage signal in the pedigree. Furthermore, the exonic status of the SNPs makes them functional candidates in this pedigree.

Examination of individuals with a psychiatric disorder diagnosis from the UK versus matched controls, involving the combination of three similarly sized association samples (total cases and controls N=5504. Table 6.3) revealed a significant increase of the rare C702 haplotype in cases as compared to controls (p=.05. OR=1.8) and a trend for association of the haplotype in males with psychosis (p=.06. OR=3.2). The largest effect size and significance is observed in males when combined diagnoses of schizophrenia,

psychosis and unipolar depression are considered (p=.005. OR=3.9), which is striking given that all 6 affected members of C702 are male. Therefore, the data from the UK association samples would suggest that the C702 identified haplotype is a risk allele for a psychiatric diagnosis, possibly in males only where the risk imposed by carrying the allele is almost four-fold that of non-carriers (OR=3.9). Such a finding supports the view that homozygosity for this haplotype would produce even greater risk of the disorder, as seen in pedigree C702.

We attempted to replicate these findings in other European association samples, including a case-control sample of schizophrenics from Ireland and a predominantly schizophrenic proband trios sample from Bulgaria. No replicating association evidence was gained from either sample, however the UK psychiatric disorder sample was much larger than these replication samples. If the rare C702 haplotype [G]-G-A-T exhibits a λ of 2 and has a frequency of 0.005 then ~2400 cases and controls or the same number of trios would be required to have 80% power to detect association (http://pngu.mgh.harvard.edu/~purcell/gpc/). Therefore, it may be the case that the UK sample is the only sample employed in this study with suitable power to detect association with the C702 identified [G]-G-A-T haplotype. Although it may be the case that the UK sample is exhibiting the "winner's curse" (Zollner and Pritchard, 2007), in which case the true effect size of the allele could be much lower than estimated by the UK association.

Independently of any association with the [G]-G-A-T haplotype, ~9000 individuals of European descent have been genotyped in this study and the C702 diplotype for the above haplotype has not been observed. Assuming Hardy-Weinberg

Equilibrium (HWE) for the haplotype and a genotype frequency of 0.000009 ($[MAF]^2 = 0.003^2$), a sample of ~111,000 individuals would be required to observe another such diplotype. However, if the haplotype is functional and detrimental to the reproductive capabilities of carriers then the number of individuals required to observe the C702 genotype at *PRKCA* again may be even higher.

A major caveat of the findings so far is the possibility that all of the individuals identified as carriers of the C702 identified haplotype may not be so. Although the 3 E21K SNPs show high LD, there may be instances where the rare alleles do not co-occur. This could have an impact on the findings presented here, and despite the fact that it is likely that the majority of individuals who harbour one of the E21K alleles harbour them all, only resequencing of this region will allow association testing of all three E21K alleles. This may be important for populations different to the UK where the LD between the alleles may differ (as may be that case given the high frequency of the rare E21K alleles in Bulgarians. Table 6.5). Another drawback to this study is the uncertainty of the phase of the E21K and E19A alleles in the case and control samples. The incomplete LD between the alleles allows for *trans* arrangements and so these must be accounted for to allow association testing of both the E21K-E19A rare haplotype and the even rarer meta-SNP.

It must be noted however, that despite both the admonitions outlined above the mechanism of action of the C702 alleles is unknown. It may be the case that not all the E21K alleles are required for risk, although it would appear that the rare allele at E19A is key to this in the UK (table 6.3). Furthermore, there is no way of knowing at present if a *cis* or *trans* arrangement of alleles has any differential effect on disease risk. However, as

the C702 siblings demonstrate all the rare E21K and E19A alleles in both *cis* and *trans* no risk mechanism can be assumed. However, even if the genotyping outlined in this chapter only represents tagging of a mechanism involving the rare E21K and E19A alleles an association has still been demonstrated with these alleles in a UK case-control sample (table 6.3). Therefore, further work is required to elucidate the mechanism of this association, such as allele specific PCR.

If the association seen in the UK sample is not a type I error, then the rare haplotype in C702 is likely to have a moderate impact upon carriers but only a small population attributable fraction of risk (Carlson, Eberle *et al.*, 2004). However, as seen with *DISC1* (Millar, Wilson-Annan *et al.*, 2000) and early-onset Alzheimer's disease (Lidle, Williams *et al.*, 2002) the identification of a disease relevant allele or locus is extremely valuable in terms of both understanding disease mechanism and identifying other risk alleles in the same or interacting genes (Millar, Pickard *et al.*, 2005, Thomson, Wray *et al.*, 2005). Furthermore, there are a number of linkage studies implicating regions encompassing *PRKCA* as being involved with psychotic disorders (Dick, Foroud *et al.*, 2003, Ewald, Wikman *et al.*, 2005, Klei, Bacanu *et al.*, 2005, Lewis, Levinson *et al.*, 2003, McInnis, Dick *et al.*, 2003, Tomas, Canellas *et al.*, 2006) and furthermore many of these suggest a recessive mode of inheritance for any disease alleles.

The gene encoding *PRKCA* is the conspicuous candidate for involvement in psychiatric illness within the C702 linkage region (Birnbaum, Yuan *et al.*, 2004, Dwivedi and Pandey, 1999, Hahn and Friedman, 1999, Knable, Barcia *et al.*, 2002, Manji and Lenox, 1999, Mirnics, Middleton *et al.*, 2001, Pandey, Dwivedi *et al.*, 2004) and lies

within an exceptionally rare homozygous region (chapter 4). The observation of a very rare homozygous diplotype in a genome-wide significant linkage region due to IBD2 status is consistent with the hypothesis that the linkage signal to 17q in pedigree C702 is due to the presence of highly penetrant, rare and recessively acting alleles inherited from a common ancestor. However, rareness alone is not enough to prove causality as every individual will harbour rare alleles, and so only subsequent observations of the diplotype will confirm or refute the genotype as a risk factor for schizophrenia and related disorders.

A disease susceptibility allele must have a biological consequence. The evidence presented here suggests that the C702 identified exonic polymorphisms and/or haplotype may increase risk of developing psychiatric illnesses. The alleles may be acting as proxies for an un-genotyped functional allele and only re-sequencing of the *PRKCA* region will confirm this. However, the alleles are transcribed (see chapter 8) and so may themselves be functional disease alleles acting via their position in the 3'UTR of different *PRKCA* transcripts.

There is no reason to suggest that either the AB209475 or X52479 transcripts represent the functional disease locus, rather as the associated [G]-G-A-T haplotype extends into both transcripts it is reasonable to assume that both must be anomalous to produce disease risk. For example, the ratio of expression or translation of both transcripts may be key to regulating their function (Andreu, Garcia-Rodriguez *et al.*, 2006, McDade, Hall *et al.*, 2007). The 3'UTR has a key role in gene regulation and variation within the 3'UTR can alter the secondary structure of transcript and thereby alter affinity of binding factors, such as miRNA and RNA binding proteins (He and

Hannon, 2004). Such changes can lead to altered pre-mRNA splicing (Tan, Guo *et al.*, 2007), mRNA editing (Eisenberg, Nemzer *et al.*, 2005), mRNA degradation/half-life (Amrani, Sachs *et al.*, 2006), translational efficiency (de Moor, Meijer *et al.*, 2005) and mRNA sub-cellular localisation (St Johnston, 2005, Van de Bor and Davis, 2004). Also, variation in the 3'UTR that deregulates the above processes may only become functional under certain spatio-temporal conditions (e.g. a cell type or sub-cellular localisation specific function that occurs only at a circumscribed developmental stage).

Therefore, the potential functional consequences of carrying the C702 rare alleles are many and so testing all of the possible consequences may be as difficult as discovering another carrier of the C702 genotype. However, separate support for association at PRKCA with independent alleles (allelic heterogeneity) could also add credence to the locus as a schizophrenia susceptibility gene.

Chapter 7. Population based association studies of PRKCA

7.1. Introduction

Previous evidence has demonstrated that rare variants within the gene encoding Protein Kinase C Alpha (PKCα) may be risk factors for schizophrenia and related psychiatric illnesses (chapters 3-6). An allelic heterogeneity model of schizophrenia would suggest that other variation at this locus may also alter susceptibility to schizophrenia and related disorders (Owen, O'Donovan *et al.*, 2002). Reported here is an attempt to perform association mapping across the putative functional regions of the *PRKCA* gene followed by haplotype analysis for association within a refined region and attempted replication of findings in multiple association samples.

The previous chapter identified exonic variants that may be associated with a psychiatric diagnosis and explain the pedigree C702 linkage signal (chapter 6). To examine the PRKCA region further for variants potentially related to schizophrenia, mutation screening by sequencing was performed across putative functional regions (chapter 6). In this chapter, novel and known variants within and flanking *PRKCA* exons, core-promoter and *in silico* identified functional sequences were tested for association in a UK schizophrenia case-control sample of ~700 cases and ~700 controls by allele frequency estimation using pooled DNA (Norton, Williams *et al.*, 2002). Follow-up individual genotyping was performed to confirm any single-marker association. Nominally associated markers underwent further haplotype analysis in an attempt to

refine the association signal. Replication of any single-marker or haplotype association signal was sought in multiple case-control samples and a trios sample, all of European ancestry.

7.2. Materials and Methods

7.2.1. Association Samples

Details of association samples used in this chapter can be found in the materials and methods chapter, section 2.1.1.

DNA pools were constructed from the UK schizophrenia case and blood donor control sample (see below) and consisted of 709 cases and 710 controls (see section 2.1.1 and 2.1.4). The sample was divided into four matched stages each containing 184 cases and an equal number of controls, except for the fourth stage, which contains 157 cases and 158 controls. Each DNA sample was quantified using Pico Green (Invitrogen) and a Fluoroskan Ascent fluorimeter (Thermo Labsystems) and samples diluted to approximately 4ng/μl (±0.5ng/μl), as detailed in section 2.1.3. The construction of separate case and control pools was performed using a Microlab S (Hamilton) automated sample processor. Samples for DNA pools were quantified by multiple personnel (including myself) and pools were constructed by Anna Preece and Nadine Norton (Norton, Williams *et al.*, 2002).

The UK association samples used in this chapter for individual genotyping were a schizophrenia case (N=709) and blood-donor control sample (N=716), and a bipolar I

disorder case (N=702) and blood-donor control sample (N=708). The UK case samples were also compared to the combined control sample (N=1424).

Further case-control samples were a schizophrenia (N=296) and schizoaffective (N=76) disorder case and control (N=812) sample from the Republic of Ireland and a further schizophrenia case (N=513) and control (N=1332) sample from Germany. The trios sample used in this study was a parent-proband trios sample from Bulgaria (Schizophrenia proband =598 trios. Schizoaffective proband =105 trios).

7.2.2. Mutation Screening

Mutation screening and allele discovery was as reported in chapter 6.

7.2.3. Genotyping and Association Analysis

7.2.3.i. Pooled Genotyping

Methods of association analysis using pooled DNA has been described in section 2.5.4.

All DNA variants identified by sequencing were to be tested for association by estimation of allele frequencies in pooled separate schizophrenia case and control samples by SNaPshot single-base extension analysis (see sections 2.1.1, 2.1.4 and 2.5.4). The 27 novel polymorphisms identified from sequencing and 17 known SNPs identified (chapter 6) were supplemented by UCSC database and HapMap Phase I SNPs (http://www.hapmap.org/) to yield better coverage of the LD around the exons and

putative regulatory motifs of *PRKCA*. A CCG insertion-deletion 258 bp prior to the first exon (P[GCC]), a poly[A] stretch (E10B(2)) and a poly[T] repeat (E19L_b) were also characterised as being polymorphic (chapter 6), however the genotyping of these markers proved problematic and so was not completed.

The criteria for selecting additional SNPs was to gain coverage of $\geq$ 3 SNPs flanking each putative functional sequence (i.e. exon or putative regulatory region) and if frequency data was estimated (either dbSNP or HapMap CEU) then a MAF$\geq$ 0.03 was required. Furthermore, SNPs were genotyped up until flanking genes to examine for any association of distal regulatory elements (e.g. enhancers, silencers) or association that extends into a neighbouring gene. These criteria identified a further 19 SNPs, giving a total of 63 SNPs (novel and known) that were attempted for analysis in case-control DNA pools.

For each variant studied, the case and control pools were simultaneously assayed using PCR followed by SNaPshot chemistry (Applied Biosystems) (see sections 2.2 and 2.5.4) and analysed by capillary electrophoresis using an ABI3100 Genetic Analyser (Applied Biosystems). Fluorescent peak heights for each allele were ascertained through Genescan and Genotyper software (Applied Biosystems).

To account for the possible unequal representation of alleles, the estimated allele frequencies from pools were corrected by using a range of values for k (0.25, 0.32, 0.47, 0.88, 1, 1.14, 2.13, 3.1, and 4) using the formula:

$$\text{Allele Frequency} = A/ (A + (k \times B))$$

Where: A = Peak height of Allele 1 (allele frequency to be estimated)

B = Peak height of Allele 2

k = Average ratio of alleles, e.g. from a heterozygote or simulated

(Peak Height Allele 1/Peak Height Allele 2)

The calculation is applied to each sample pool, then both case and control pool frequencies are combined and a mean taken. The allele frequencies are then converted to allele counts by multiplying by the number of chromosomes present in the case and control groups. The result allele counts are placed into a 2 x 2 contingency for chi square analysis. All pooling allele frequency estimation and case-control analysis was performed using Microsoft Excel macros designed by Dr. Nigel Williams and Dr. George Kirov. Any SNP showing evidence for association at any simulated value of k was assayed again using the same protocol and including two heterozygous samples to estimate k. Therefore, all SNPs showing association from analysis in case and control DNA pools have been repeated and had k estimated from heterozygous individuals.

PCR and extension primers for the SNPs studied are shown in appendices 10.5.

7.2.3.ii. Individual Genotyping

SNPs showing evidence for association with schizophrenia from analysis in case and control DNA pools (section 7.2.3.i) at $p \leq 0.1$ were individually genotyped in the same UK schizophrenia case and control sample. SNPs showing association at $p \leq 0.05$ after

individual genotyping were genotyped in the replication sample outlined in section 2.1.1 and 7.2.1.

Individual genotyping for the UK case-control samples and the Bulgarian trios samples was performed using the Amplifluor (Chapter section 2.5.5), or the hME and iPlex Sequenom MassARRAY platforms (Chapter section 2.5.6). Assays for SNPs analysed in this study are given in appendices 10.5 with example genotyping.

Individual genotyping of the Irish case-control sample was performed using Taqman genotyping assays (Applied Biosystems) by Derek Morris (Trinity College, Dublin, Republic of Ireland). Individual genotyping of the German case-control sample was performed by Dan Rujescu (Department of Psychiatry, Ludwig Maximilians University, Germany) using a Sequenom MassARRAY platform and iPlex chemistry.

7.2.3.iii. Association Analysis of Individual Genotyping

Analysis of SNPs achieving trend significance (allelic $p \leq 0.1$) in the case-control DNA pools was performed by single marker genotyping in the UK schizophrenia case-control sample followed by tests for allelic association by standard chi-squared analysis (Haploview 4.0). Replication was sought in the association samples described in section 7.2.1. Allelic p values were calculated using Haploview 4.0. Haplotype frequencies were estimated and tested for association using Fast EH+ software ((Zhao, Curtis *et al.*, 2000)) for case-control samples. Haplotype analysis in the trios sample was performed using Haploview 4.0.

7.2.3.iv. Further Statistical Analysis

To assess single-marker association result for significance, an experiment wide p value was calculated for all single-markers tested ((Nyholt, 2004) and http://gump.qimr.edu.au/general/daleN/SNPSpD/). Case-control replication samples were assessed for overall evidence of association at single-markers and haplotypes by inverse variance meta-analysis using a Microsoft Excel macro designed by Dr. Valentina Moskvina.

7.3. Results

7.3.1. Association Mapping using DNA Pools

Of the 63 variants attempted for analysis in DNA pools, 58 were successfully analysed in case and control DNA pools (see table 7.1 and figure 7.1). Considering the distance from the most distal 5' to 3' SNPs is 570kb, this gives a mean coverage of 1 polymorphism per 9.8kb, although there is a higher density around exons and putative functional elements (figure 7.1). A post-hoc examination of the genotyped SNPs performance as pairwise tag can be performed as the variants were predominantly found genotyped as part of the Phase II HapMap CEPH data (40 of 58). Of the 40 HapMap markers, 36 were selected as pairwise tag SNPs using Tagger (Haploview) using an $r^2>0.8$ as a threshold.

A list of the variants assayed, their nucleotide position and base change, the minor allele frequency in controls and cases (if a value of *k* was calculated from heterozygotes) the minor allele frequency in the HapMap CEPH individuals (if applicable) and an association p values if significant association was found ($p<0.1$) are shown in table 7.1 and see figure 7.1. Novel variants not assayed could not be optimised for pooling, either due to low MAF or unacceptable assay quality, these are not shown on the given table but their details can be found elsewhere (chapter 6). All polymorphisms assayed and all positively associated variants are shown relative to the physical location of *PRKCA* in the genome (figure 7.1).

The results table (table 7.1) displays that 9 SNPs achieved evidence of nominal significance at $p<0.1$ after analysis in pooled DNA of case and control samples. All the nominally significant SNPs lie within the 3' half of the gene, from rs1010546 to rs733646 (chr17:62,115,027-62,256,511. UCSC May 2004). The association results were assessed for experiment-wide significance using a method designed by Nyholt (Nyholt, 2004). Of the 58 SNPs analysed for association, 40 had genotype data in the CEU individuals of the HapMap. Testing these 40 SNPs resulted in 37.46 independent tests, therefore a single-marker significance of less than $p=0.0013$ would be required to claim significance after correction for multiple tests (and this value is anti-conservative, as only 40 of the 58 SNPs had individual genotype data available). Therefore, the best p value of 0.01 at rs733646 is not experiment-wide significant.

None of the positive SNPs are in $r^2>0.8$ with another positive SNP as estimated in the CEPH individuals assayed by the HapMap and so all were individually genotyped in the same schizophrenia case-control sample.

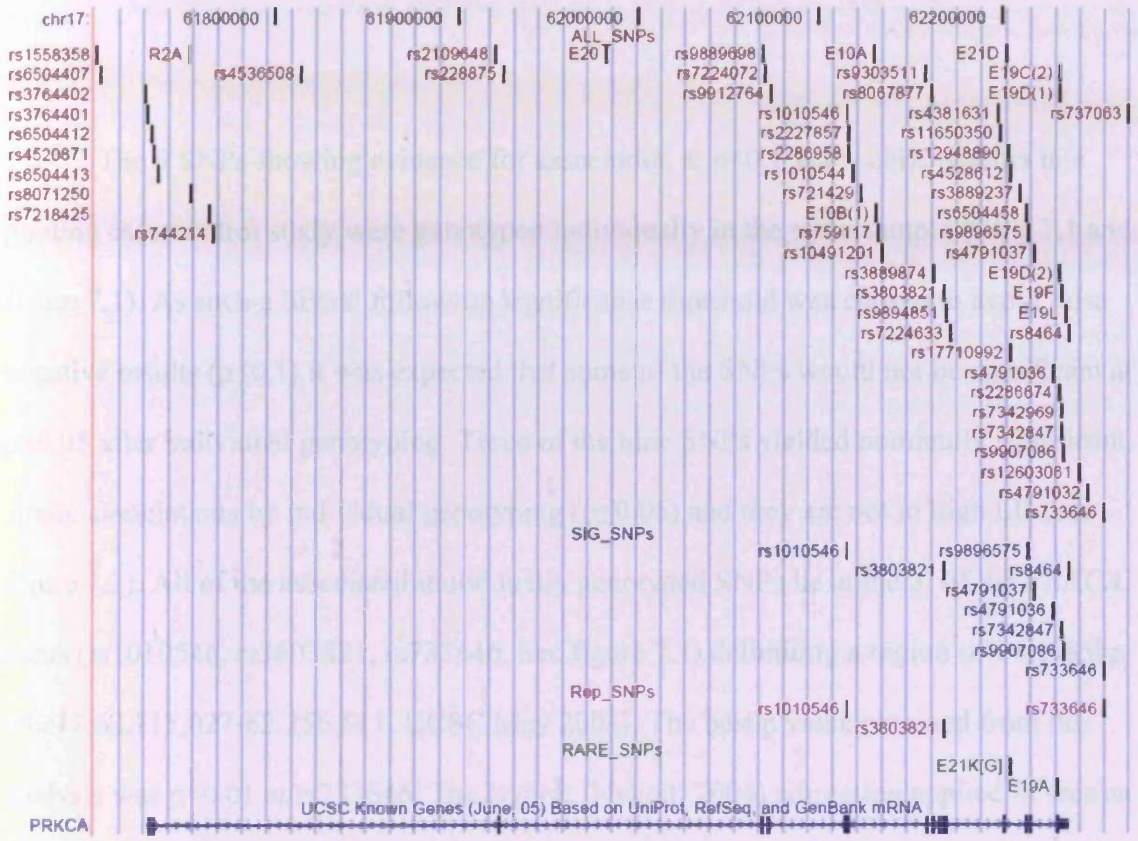


Figure 7.1: UCSC Genome Browser tracks showing (in descending order) the chromosomal base position, all SNPs assayed in the schizophrenia case-control DNA pools ("ALL_SNPs"), SNPs achieving a $p<0.1$ ("SIG_SNPs"), SNPs replicating this association after individual genotyping in the same sample ("Rep_SNPs"), the C702 identified rare SNPs ("RARE_SNPs", see chapter 5) and finally the typical *PRKCA* gene. The figure demonstrates the location of a putative associated region at the 3' of the *PRKCA* locus.

7.3.2. Individual Genotyping Analysis of Putative Associations in Case-Control DNA Pools

The 9 SNPs showing evidence for association at $p<0.1$ with schizophrenia in a pooling case-control study were genotyped individually in the same sample (table 7.1 and figure 7.1). As such a liberal follow-up significance threshold was chosen to avoid false negative results ($p\leq0.1$) it was expected that some of the SNPs would not be significant at $p\leq0.05$ after individual genotyping. Three of the nine SNPs yielded nominally significant allelic associations by individual genotyping ($p\leq0.05$) and they are not in high LD (see figure 7.2). All of the associated individually genotyped SNPs lie at the 3' of the *PRKCA* locus (rs1010546, rs3803821, rs733646. See figure 7.1) delimiting a region of 141,485bp (chr17:62,115,027-62,256,511. UCSC May 2004). The best p value obtained from this analysis was $p=0.01$ at rs733646. The Nyholt (Nyholt, 2004) correction applied in section 7.3.1 is still applicable to this data (a $p<0.0013$ required for significance) and so the individual genotyping data does not support an association at PRKCA that is significant after multiple comparisons are accounted for.

Genotyping in DNA pools						Individual genotyping			
SNP	Position (HG17)	Minor Allele Frequency			Allele P	Minor Allele Frequency		Allele P	Genotypic P
		Case	Control	CEPH		Case	Control		
rs1558358	61700898	0.09	0.10	0.05	*>0.1*	-	-	-	-
rs6504407	61702451	0.23	0.25	0.17	*>0.1*	-	-	-	-
rs3764402	61727915	0.25	0.24	0.27	*>0.1*	-	-	-	-
rs3764401	61728291	-	-	-	*>0.1*	-	-	-	-
rs6504412	61730931	0.04	0.04	0.04	*>0.1*	-	-	-	-
rs4520871	61732117	0.12	0.10	0.04	*>0.1*	-	-	-	-
rs6504413	61734771	0.12	0.11	0.05	*>0.1*	-	-	-	-
R2A	61751983	-	-	-	*>0.1*	-	-	-	-
rs8071250	61752029	-	-	-	*>0.1*	-	-	-	-
rs7218425	61763024	-	-	-	*>0.1*	-	-	-	-
rs744214	61765318	-	-	-	*>0.1*	-	-	-	-
rs4536508	61813793	-	-	-	*>0.1*	-	-	-	-
rs2109648	61920166	-	-	-	*>0.1*	-	-	-	-
rs228875	61924337	-	-	-	*>0.1*	-	-	-	-
E20	61981167	0.37	0.36	-	*>0.1*	-	-	-	-
rs9889698	62068596	-	-	-	*>0.1*	-	-	-	-
rs7224072	62069901	-	-	-	*>0.1*	-	-	-	-
rs9912764	62072848	-	-	-	*>0.1*	-	-	-	-
rs1010546	62115027	0.18	0.15	0.12	**0.03**	0.15	0.12	**0.05**	**0.02**
rs2227857	62115540	-	-	-	*>0.1*	-	-	-	-
rs2286958	62115852	-	-	-	*>0.1*	-	-	-	-
rs1010544	62117908	-	-	-	*>0.1*	-	-	-	-
rs721429	62122367	-	-	-	*>0.1*	-	-	-	-
E10A	62129885	0.15	0.13	-	*>0.1*	-	-	-	-
E10B(1)	62130370	-	-	-	*>0.1*	-	-	-	-
rs759117	62132085	-	-	-	*>0.1*	-	-	-	-
rs10491201	62133625	-	-	-	*>0.1*	-	-	-	-
rs9303511	62158093	-	-	-	*>0.1*	-	-	-	-
rs8067877	62161640	-	-	-	*>0.1*	-	-	-	-
rs3889874	62162275	0.07	0.08	0.11	*>0.1*	-	-	-	-
rs3803821	62168179	0.46	0.50	0.53	**0.07**	0.48	0.52	**0.05**	0.13
rs9894851	62169450	-	-	-	*>0.1*	-	-	-	-
rs7224633	62171849	-	-	-	*>0.1*	-	-	-	-
rs4381631	62196905	-	-	-	*>0.1*	-	-	-	-
rs11650350	62198748	-	-	-	*>0.1*	-	-	-	-
E21D	62202031	0.24	0.23	-	*>0.1*	-	-	-	-
rs12948890	62202951	-	-	-	*>0.1*	-	-	-	-
rs4528612	62203761	0.23	0.23	0.19	*>0.1*	-	-	-	-
rs17710992	62204090	0.27	0.30	0.28	*>0.1*	-	-	-	-
rs3889237	62209892	-	-	-	*>0.1*	-	-	-	-
rs6504458	62213154	-	-	-	*>0.1*	-	-	-	-
rs9896575	62214907	0.08	0.06	0.08	**0.06**	0.09	0.08	0.25	0.46
rs4791037	62218356	0.07	0.05	0.04	**0.06**	0.06	0.05	0.10	0.19
rs4791036	62227597	0.06	0.08	0.12	**0.03**	0.07	0.09	0.26	0.19
rs2286674	62231014	-	-	-	*>0.1*	-	-	-	-
E19C(2)	62231088	-	-	-	*>0.1*	-	-	-	-
E19D(1)	62231329	-	-	-	*>0.1*	-	-	-	-
E19D(2)	62231360	-	-	-	*>0.1*	-	-	-	-
E19F	62232348	-	-	-	*>0.1*	-	-	-	-
rs7342969	62232651	-	-	-	*>0.1*	-	-	-	-
rs7342847	62233006	0.37	0.41	0.38	**0.05**	0.37	0.39	0.39	0.51
rs9907086	62233937	0.07	0.05	-	**0.09**	0.04	0.03	0.17	NA
E19L	62234830	0.06	0.05	-	*>0.1*	-	-	-	-
rs8464	62237178	0.13	0.11	0.15	**0.05**	0.13	0.11	0.12	0.29
rs12603061	62242660	-	-	-	*>0.1*	-	-	-	-
rs4791032	62246927	-	-	-	*>0.1*	-	-	-	-
rs733646	62256511	0.10	0.14	0.08	**0.01**	0.11	0.14	**0.01**	**0.02**
rs737063	62270066	0.16	0.16	0.17	*>0.1*	-	-	-	-

Table 7.1: Results of allele frequency estimation and association analysis in schizophrenia case-control DNA pools and follow-up individual genotyping. Shown is each variant assayed, the genomic position, and the allele frequency as estimated in pools (if a value of *k* was estimated from heterozygotes) and in 60 CEPH individuals (HapMap) and the resultant association p value (allelic and genotypic). Also shown are the results of individual genotyping for SNPs showing evidence for association in pools($p \leq 0.1$).

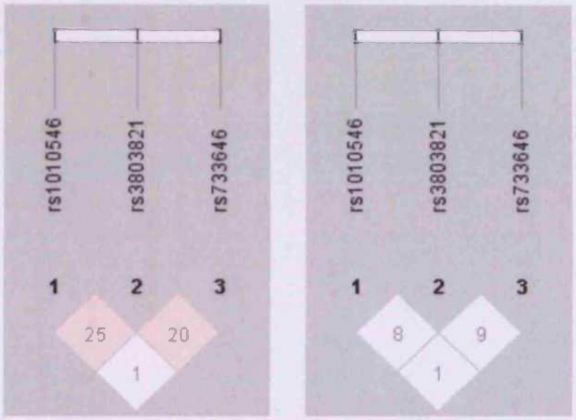
7.3.4. Single Marker Replication Studies

The three SNPs showing evidence for nominal significance in the UK schizophrenia case-control sample after individual genotyping (rs1010546, rs3803821 and rs733646. Table 7.1) were individually genotyped in a UK bipolar I disorder case-control sample, an Irish schizophrenia and schizoaffective disorder case-control sample, a German schizophrenia case-control sample and a sample of schizophrenia and schizoaffective proband trios from Bulgarian (see section 7.2.1 and materials and methods section 2.1.1. The results of test for allelic association are shown in table 7.2.

There was no evidence for replication of the association seen with rs733646 in the UK schizophrenia sample in any of the replication samples (table 7.2). For rs1010546 the German sample shows the same direction of affect (T allele increased in cases) and a trend for association (p=0.095) and a meta-analysis association p value of 0.0390, OR=1.12 (heterogeneity p=0.47, indicating the same direct of effect in across all samples), with a slightly larger effect in males (0.045, OR=1.17), however there was no evidence for replication with the same or alternative allele in the trios sample (table 7.2). The single-marker association at rs3803821 (A allele associated. Minus strand, dbSNP

build 25) was replicated in the German case-control sample with the same associated allele (p=0.051) and gives the most significant meta-analysis p value across all case-control samples (p=0.021, OR=1.096) and also in males only (p=0.0096, OR=1.15) (table 7.2). Note also that each sample used in the meta-analysis of this SNP showed the same direction of effect (heterogeneity p=0.59). The same allele is also significantly overtransmitted to Bulgarian probands (0.031).

None of the SNPs showed evidence of high LD in any of the case or control samples assayed (see figure 7.1).



		Cases (N=2528)		Controls (N=3568)	
SNP1	SNP2	D'	r^2	D'	r^2
rs1010546	rs3803821	0.259	0.012	0.084	0.001
rs1010546	rs733646	0.018	0	0.019	0
rs3803821	rs733646	0.204	0.007	0.095	0.001

Figure 7.2: LD between the three SNPs associated with schizophrenia in the UK, as measured across samples of cases with schizophrenia, schizoaffective disorder or bipolar I disorder (N=2528) and controls (N=3568) from Europe (All UK, German and Irish samples included. Bulgarian trios were not used for this analysis). The data shows the pairwise LD between all three SNPs. The Haploview LD plot shows the *D'* values between each SNP and this is replicated in the table below with r^2 values also given. There is little

LD between any of the SNPs in either sample, however LD is highest between adjacent SNPs in the case sample.

ALL		UK Schizophrenia*			UK Bipolar I*			German Schizophrenia			Irish Schizophrenia/SABP			Combined case-control				Bulgarian Schizophrenia/SABP		
tagSNP	Allele	Case	Control	P	Case	Control	P	Case	Control	P	Case	Control	P	Case	Control	Meta P†	OR	T	NT	P
rs1010546	T	0.15	0.13	**0.038**	0.13	0.13	**0.499**	0.15	0.12	**0.095**	0.13	0.13	**0.851**	0.14	0.13	**0.0390**	**1.12**	130	124	**0.707**
rs3803821	A	0.52	0.49	**0.068**	0.50	0.49	**0.498**	0.55	0.51	**0.051**	0.48	0.48	**0.739**	0.53	0.50	**0.0210**	**1.096**	318	266	**0.031**
rs733646	G	0.11	0.13	**0.027**	0.14	0.13	**0.437**	0.14	0.13	**0.490**	0.12	0.12	**0.680**	0.12	0.13	**0.6770**	**0.98**	121	116	**0.745**

MALE		UK Schizophrenia*			UK Bipolar I*			German Schizophrenia			Irish Schizophrenia/SABP			Combined case-control				Bulgarian Schizophrenia/SABP		
tagSNP	Allele	Case	Control	P	Case	Control	P	Case	Control	P	Case	Control	P	Case	Control	Meta P†	OR	T	NT	P
rs1010546	T	0.16	0.13	**0.062**	0.13	0.13	**0.846**	0.15	0.13	**0.173**	0.13	0.12	**0.595**	0.15	0.13	**0.0450**	**1.168**	66	54	**0.273**
rs3803821	A	0.53	0.49	**0.031**	0.50	0.49	**0.662**	0.54	0.50	**0.098**	0.50	0.48	**0.489**	0.53	0.50	**0.0096**	**1.147**	146	124	**0.181**
rs733646	G	0.10	0.13	**0.032**	0.14	0.13	**0.471**	0.14	0.12	**0.141**	0.11	0.13	**0.274**	0.12	0.13	**0.6460**	**0.963**	65	55	**0.361**

FEMALE		UK Schizophrenia*			UK Bipolar I*			German Schizophrenia			Irish Schizophrenia/SABP			Combined case-control				Bulgarian Schizophrenia/SABP		
tagSNP	Allele	Case	Control	P	Case	Control	P	Case	Control	P	Case	Control	P	Case	Control	P	OR	T	NT	P
rs1010546	T	0.15	0.13	**0.140**	0.13	0.13	**0.709**	0.13	0.12	**0.569**	0.12	0.14	**0.296**	0.13	0.13	**0.635**	–	60	56	**0.710**
rs3803821	A	0.49	0.49	**0.998**	0.50	0.49	**0.540**	0.56	0.52	**0.163**	0.54	0.53	**0.680**	0.53	0.51	**0.290**	–	152	123	**0.080**
rs733646	G	0.12	0.13	**0.311**	0.14	0.13	**0.830**	0.87	0.86	**0.689**	0.14	0.12	**0.429**	0.13	0.13	**0.509**	–	56	46	**0.322**

* Using combined UK control samples

Schizoaffective Bipolar Disorder (SABP)

T (Number of transmissions from heterozygotes parent to proband), NT (number of non-transmission)

† inverse variance meta analysis p value. All psychiatric disorder case and control samples (UK schizophrenia and bipolar I samples treated as one sample)

Table 7.2: SNPs significant in the UK schizophrenia case-control sample were genotyped in a Bipolar I disorder case-control sample from the UK, , a German schizophrenia case-control sample, an Irish schizophrenia and schizoaffective disorder case-control sample and a trios sample of schizophrenia and schizoaffective probands from Bulgaria. Association sample numbers are given in the materials and methods chapter, section 2.1.1. Shown are the allele frequencies in each case and control sample (or transmissions/non-transmission for the familial sample) given with association p values. For marker nominally significant in the UK schizophrenia association sample ($p<0.1$) replication was tested by including the discovery sample and all other samples by inverse variance meta analysis. The same analysis is also split into male and female specific.

7.3.5. Haplotype Analysis of Associated Markers

An exploratory analysis of haplotypes was undertaken to identify if there was any association signal not detected by single-marker analysis. Two and three marker haplotype analysis was performed on the UK schizophrenia sample versus all UK controls. Although it is difficult to correct for multiple comparisons when considering all the single-marker an haplotype tests performed, a Bonferroni correction for 20 independent haplotype tests was considered (all 2 and 3 marker combinations) as a rough guide to replicating haplotype association findings in the UK schizophrenia sample. Therefore, haplotypes with an individual p of ≤ 0.003 were tested for replication and the results are summarised in table 7.3.

The results show seven haplotypes that were significant after corrections for multiple comparisons in the UK sample (table 7.3). In multiple replication samples all haplotypes achieve a meta-analysis p value of <0.02 when all case and control samples are considered (as for the single-marker analysis, the UK schizophrenia and bipolar I disorder samples were considered as one sample). Of these, four show nominal evidence for association in the trios sample, with transmission distortions consistent with a replication of the case-control results.

The most significant two marker meta-analysis result is for markers rs1010546-rs3803821, haplotype C-T ($p=0.00021$, $OR=0.858$), where there is a 7% increase of a haplotype frequency in controls from the UK as compared to UK schizophrenics. This same haplotype is also significantly undertransmitted to Bulgarian trio probands (nominal $p=0.0565$). The best three marker haplotype, a sub-haplotype of the above (C-T-A, table

7.3) gives the most significant meta-analysis results (p=0.000019, OR=0.84) and replicates with the same direction of effect (an increase in controls or transmission to controls) in all samples except the Irish case-control sample.

The haplotype A-A of markers rs3803821-rs733646 is overtransmitted to cases and gives an empirical p value of 0.324. The same haplotype gives a meta-analysis p value of 0.0098 and an OR of 1.11. A three marker sub-haplotype of the above, C-A-A, also shows the same effect with a nominal overtransmission to cases (p=0.0345) and a meta-analysis p value of 0.01 in case-control analysis. The A allele of rs3803821 is the only allele that showed single marker replication across all samples (table 7.2) and so these haplotypes may represent the refined association signal detected by that allele.

Sample	Markers	Hap	Case Freq / T	Con Freq / NT	Global P	Individual P	Meta P	OR
UK (SZ)	1010546-rs3803821	C-T	0.38	0.45	0.0004	<0.000		
UK (BPI)	1010546-rs3803821	C-T	0.41	0.45	0.0120	0.0120	0.00021	0.858
Irish	1010546-rs3803821	C-T	0.40	0.40	0.8146	0.9560		
German	1010546-rs3803821	C-T	0.39	0.42	0.1134	0.0980		
BG	1010546-rs3803821	C-T	291	339	0.1268	0.0565	NA	NA
UK (SZ)	1010546-rs3803821	T-T	0.10	0.07	0.0004	<0.000		
UK (BPI)	1010546-rs3803821	T-T	0.09	0.07	0.0120	0.0020	0.00166	1.275
Irish	1010546-rs3803821	T-T	0.08	0.07	0.8146	0.5210		
German	1010546-rs3803821	T-T	0.06	0.07	0.1134	0.8110		
BG	1010546-rs3803821	T-T	59	69	0.7384	0.3848	NA	NA
UK (SZ)	rs1010546-rs733646	T-A	0.14	0.10	0.0070	0.0020		
UK (BPI)	rs1010546-rs733646	T-A	0.11	0.10	0.8504	0.6640	0.0314	1.147
Irish	rs1010546-rs733646	T-A	0.11	0.12	0.8126	0.6300		
German	rs1010546-rs733646	T-A	0.12	0.11	0.5442	0.1300		
BG	rs1010546-rs733646	T-A	132	134	0.9968	0.9160	NA	NA
UK (SZ)	rs3803821-rs733646	A-A	0.48	0.43	0.0180	0.0030		
UK (BPI)	rs3803821-rs733647	A-A	0.44	0.43	0.8670	0.4610	0.0098	1.109
Irish	rs3803821-rs733648	A-A	0.47	0.46	0.8356	0.9020		
German	rs3803821-rs733649	A-A	0.49	0.45	0.1268	0.0960		
BG	rs3803821-rs733646	A-A	356	293	0.0324	0.0135	NA	NA
UK (SZ)	rs1010546-rs3803821-rs733646	C-T-A	0.32	0.39	0.0004	<0.000		
UK (BPI)	rs1010546-rs3803821-rs733647	C-T-A	0.35	0.39	0.2038	0.0090	0.000019	0.84
Irish	rs1010546-rs3803821-rs733648	C-T-A	0.35	0.34	0.9324	0.8550		
German	rs1010546-rs3803821-rs733649	C-T-A	0.32	0.36	0.3124	0.0390		
BG	rs1010546-rs3803821-rs733650	C-T-A	264	317	0.0868	0.0284	NA	NA
UK (SZ)	rs1010546-rs3803821-rs733646	C-A-A	0.43	0.37	0.0004	0.0010		
UK (BPI)	rs1010546-rs3803821-rs733647	C-A-A	0.40	0.37	0.2038	0.0860	0.01	1.1
Irish	rs1010546-rs3803821-rs733648	C-A-A	0.43	0.42	0.9324	0.6240		
German	rs1010546-rs3803821-rs733649	C-A-A	0.42	0.41	0.3124	0.6870		
BG	rs1010546-rs3803821-rs733650	C-A-A	360	305	0.1078	0.0345	NA	NA
UK (SZ)	rs1010546-rs3803821-rs733646	T-T-A	0.09	0.06	0.0004	<0.000		
UK (BPI)	rs1010546-rs3803821-rs733647	T-T-A	0.07	0.06	0.2038	0.0700	0.017	1.11
Irish	rs1010546-rs3803821-rs733648	T-T-A	0.07	0.07	0.9324	0.9890		
German	rs1010546-rs3803821-rs733649	T-T-A	0.05	0.06	0.3124	0.4760		
BG	rs1010546-rs3803821-rs733650	T-T-A	53	64	0.8136	0.3152	NA	NA

Table 7.3: Haplotype analysis of SNPs achieving nominal significance in the UK schizophrenia case-control sample. All 2 and 3 marker haplotype analysis was performed for each sample (UK (SZ) = UK schizophrenia cases and all UK controls. UK (BPI) = UK bipolar I disorder cases and all UK controls. Irish = Irish schizophrenia-schizoaffective disorder case control sample. German = German schizophrenia case control sample. BG = Bulgarian trios schizophrenia-schizoaffective disorder sample). Shown is each marker and individual haplotype (alleles coded the same as table 7.2). The frequency in cases and controls, or the number of transmissions (T) and non-transmissions (NT) for trios. Then global p values are given (5000 permutations) and individual haplotype p values. An inverse variance meta-analysis p value and odds ratio (OR) was also calculated for the case-control samples (Note, the UK samples were combined into one sample).

7.4. Discussion

An association mapping study using pooled genotyping followed by individual genotyping has identified three SNPs that showed a nominally significant association in the UK schizophrenia case-control sample at rs1010546 (p=0.05), rs3803821 (p=0.05) and rs733646 (p=0.01). While none met criteria for experiment wide correction, all three SNPs were genotyped in a further three schizophrenia association samples to substantiate these findings (table 7.3). Of the three SNPs, two showed evidence of replication as assessed formally by meta-analysis (rs1010546 p=0.039, rs3803821 p=0.021) although there was no overall evidence for association once the discovery sample was removed. One of the associated alleles, at rs3803821, was significantly overtransmitted to schizophrenic and schizoaffective disorder probands (p=0.031). Haplotype analysis showed a more significant replicating association with a haplotype harbouring the same associated allele at rs3803821 (case-control meta analysis p=0.0098. Trios p=0.0135) and separate haplotype associations were also identified. The data therefore suggest that common variation at PRKCA may alter susceptibility to developing schizophrenia and related disorders. However, the effect size is likely to be weak, or the genotyped SNPs are not in high LD with the true disease allele(s).

The initial mutation screening attempted to identify all common variation in exonic and putative regulatory sequence at *PRKCA* (chapter 6). The 27 novel SNPs were merged with 16 known SNPs to create a marker map across the *PRKCA* gene that was aimed at identifying association at putative functional sites. However, analysis of the performance of these SNPs reveals that much of the single-marker information at *PRKCA* remains un-captured by this study design. Phase II HapMap

data indicates that 207 SNPs are required to capture every SNP (MAF>0.001) genotyped in the CEU individuals with an r^2>0.8 between a proxy SNP and ungenotyped SNP (Haploview, Tagger). Although 36 of the SNPs assayed in pools are confirmed as non-redundant according to the same criteria, this still leaves a large proportion of even the common variation across the region unaccounted for.

The incomplete association map across *PRKCA* uncovered three independent association signals in the UK schizophrenia samples (figures 7.1 and 7.2), though none of the nominally significant results at rs1010546 (p=0.05), rs3803821 (p=0.05) or rs733646 (p=0.01) would survive correction for multiple comparisons. However, arguably the most appropriate method to validate association findings is to perform replication studies (Hirschhorn and Altshuler, 2002) which help validate weak but genuine disease associations, as seen for other complex disorders that use many thousands of samples to replicate small effect size associations seen in type II diabetes (Zeggini, Weedon *et al.*, 2007) and forms of cancer (Hunter, Kraft *et al.*, 2007). Replication was sought in 3 independent schizophrenia and schizoaffective disorder association samples from across Europe, as well as UK bipolar I disorder sample obtained from the UK. The bipolar I disorder sample was included as a replication sample in all further studies due to the finding of association of the rare C702 alleles with psychosis (chapter 6) and the literature implicating *PRKCA* in the pathogenesis of bipolar disorders (Birnbaum, Yuan *et al.*, 2004, Einat, Yuan *et al.*, 2007, Hahn and Friedman, 1999, Manji and Lenox, 1999).

Single-marker replication analysis provided support for two of the associations seen in the UK schizophrenia sample. SNP rs1010546 showed a weak overall meta-analysis p value when all case-control samples were considered (p=0.039, OR=1.12) and also in males (p=0.045, OR=1.168), however no independent sample showed

significance when considered alone. Similarly, rs3803821 shows a nominally significant meta-analysis result (p=0.021, OR=1.12), again stronger in males (p=0.0096, OR=1.147), however the same allele shows significant overtransmission to probands in the trios sample (p=0.031) and was significantly associated in the German case-control sample (p=0.051). Therefore, a common SNP allele in the 3' of *PRKCA* shows association with schizophrenia in three independent European association samples and overall association with five samples exhibiting schizophrenia, bipolar I disorder and schizoaffective disorder as assessed by a meta-analysis. Furthermore, the familial sample analysis shows the association in the case-control samples is unlikely to be due to population stratification.

Haplotype analysis may be more informative and powerful than single-marker analysis, and may identify other sources of association not visible from single locus data (Schaid, 2004). As none of the three SNPs are in high LD (figure 7.2) there should be many haplotypes that could serve as tags for un-genotyped variants. Exploratory haplotype analysis in the UK schizophrenia sample identified significant global haplotypes after permutation and also individual haplotype association that would survive conservative corrections for multiple comparisons (table 7.3). However the number of corrections that should be applied to the haplotype data is unclear, given that 58 non-independent SNPs were genotyped to uncover these associations. The nominally significant individual haplotypes were examined in the replication samples, just as for the single-marker results (table 7.3). The results of the three marker haplotype analysis show at least three apparently separate sources of haplotype association, all of which achieve nominal significance after meta-analysis and two of which (C-T-A and C-A-A, table 7.3) show uncorrected significance in the trios sample with transmissions consistent with the case-control data. Intriguingly, the

bipolar I disorder sample that showed no individual evidence of association with single-markers, shows a significant decrease of the C-T-A haplotype compared with controls (p=0.009). For this haplotype, only the Irish cases-control sample does not show significant association, with all other samples showing the same direction of effect giving a meta-analysis p value when the discovery sample is included of 0.000019 and a nominal TDT p value of 0.0284 in the trios.

Therefore, an association mapping strategy at *PRKCA* has identified several common allele associations with schizophrenia in a UK case-control sample that replicate across five association samples of European ancestry. Although the associations are not strong and would not survive rigorous correction for multiple testing, the replication across multiple case-control samples and a familial sample is a noteworthy observation, especially given that the gene may harbour other susceptibility alleles for schizophrenia (chapter 6). As the data indicates common allelic heterogeneity at this locus, the mapping of the variants causing these associations may prove problematic; for example if the wrong set of markers is genotyped the allele associations may be masked (Lin, Vance *et al.*, 2007), also the variants may be in weak LD with an allele of large effect or in strong LD with an allele of small effect size. However, re-sequencing and further association mapping can use the associated alleles identified here as a starting point, e.g. examining SNPs in LD with rs3803821 or occurring on the C-T-A haplotype. As for pedigree C702, the data presented here is unlikely to have any relevance to a rare and highly penetrant disease allele that may be explaining the linkage seen in that pedigree (chapters 4 and 6)

Common allele association at the 3' of PRKCA is remarkable when considering that molecular genetic analysis of pedigree C702 also identified the same

region (chapters 4 and 6). The rare allele association seen with the C702 identified [G]-G-A-T haplotype is a frequency difference between cases and controls of ~0.002-0.005 and so is unlikely to explain the larger frequency differences seen for example at rs3803821 (A allele is increased by 0.03 in all cases compared to all controls. Table 7.2). Independent association at the same region of *PRKCA* indicates allelic heterogeneity but may identify functional homogeneity of disease alleles. Therefore, demonstration of a common neuropathology involving variation at PRKCA may indicate the mechanism by which the C72 identified haplotype or diplotype imbues risk. It is therefore interesting to note that dysregulation of PRKCA transcripts and protein has been reported in small samples of schizophrenic post-mortem brain (Knable, Barcia *et al.*, 2002, Mirnics, Middleton *et al.*, 2001).

There are several lessons to be drawn from this study with regard to the rigorous assessment of candidate genes in complex disorders. First, comprehensive association mapping of the entire gene must be performed to accurately localise any association signal(s) in a suitably powered sample to detect effects, considering the frequency of alleles tested. Subsequently, regions of association must be studied for the source of that association; a potential disease relevant allele. There are few methods that will pin-point the source of a genetic association, although the imputation of un-genotyped markers may prove useful for common variant studies such as this (Marchini, Howie *et al.*, 2007). Therefore, a researcher is left with the difficult task of re-sequencing of an implicated region, or the whole gene, in a random or directed sub-set of individuals, followed by more genotyping to and replication analysis to confirm any disease allele finding.

Chapter 8. Post-genomic study of Protein Kinase C Alpha

8.1. Introduction

The previous chapters and research have generated evidence that implicate *PRKCA* as a schizophrenia susceptibility gene ((Williams, Norton *et al.*, 2003) and chapters 3-7). This chapter describes attempts to establish a functional biological mechanism via which risk variants act at *PRKCA*.

The difficulty in identifying the function of a putative disease allele is using suitable assays to measure the exact biological outcome of the correct genetic functional mechanism (Cirulli and Goldstein, 2007, Strachan and Read, 2003c). A disease allele occurs on a relatively fixed macromolecular sequence upon which other factors act to produce the biological function of the cell. It is through the disruption of this mechanism of gene regulation that disease risk will be altered, but there are many possible mechanisms via which any allele can act (Strachan and Read, 2003c): The allele may occur within one gene sequence but act to alter the transcription of any gene, for example through the *trans* action of a pre-mRNA or acting via a *cis* regulatory element. The allele may alter the transcription rate of an isoform (which may be novel), may alter pre-mRNA splicing, mRNA location or function, translational activity and also protein sequence and function. Furthermore, the actions of the allele may be dependent upon temporal (e.g. developmental time), spatial (e.g. cell type, or regional cell type), epistatic (i.e. require another allele that may act in *cis* or *trans*) and environmental (e.g. epigenetic) conditions. Additionally, and prominently in complex nervous system phenotypes, the effects of an allele may be so slight as to be overlooked at the molecular or cellular level. So, even for a true disease

susceptibility polymorphism, the functional corollary of the variant may be extremely difficult to both find and verify.

Previous evidence indicates that rare alleles within the exonic sequence of *PRKCA* may be highly penetrant schizophrenia susceptibility alleles (chapter 6). These rare alleles identified in pedigree C702 lie within the 3'UTRs of two different mRNA isoforms of *PRKCA* (chapter 6) and the evidence of common variant association at this locus also incriminates the 3' of the gene (chapter 7). Variation within the 3'UTR can influence transcription (Hughes, 2006, Shao and Ismail-Beigi, 2004), post-transcriptional activity such as spatial localisation (Dalgleish, Veyrune *et al.*, 2001, St Johnston, 2005), RNA splicing (Banihashemi, Wilson *et al.*, 2006), RNA editing (Golden and Hajduk, 2006) and mRNA half-life (Hughes, 2006), translation (Hughes, 2006, Subramaniam, Chen *et al.*, 2004) and protein sequence (Kryukov, Kryukov *et al.*, 1999), any of which may be cell or temporally specific (Di Liegro, Bellafiore *et al.*, 2000). Further, any effect of the variation may be dependent upon other genetic or environmental factors (i.e. there may be no direct relationship between allele and measurable effect). Therefore, as the possible functional consequences of variation at this locus are apparently unbounded this chapter focuses upon transcript and protein assays in attempts to screen for any effect of specific genotypes or schizophrenia status upon the biology of *PRKCA*. The assays of steady-state mRNA and protein levels involves the use of diseased and control nervous tissue and lymphoblastoid cell lines.

8.2. Materials and Methods

8.2.1. Cell Culture Samples

For 4 of the affected C702 sibs an Epstein-Barr Virus (EBV) immortalised β-lymphoblast cell line was created from whole blood (The European Collection of Cell Cultures (ECACC)). Cell lines were culture as described in the materials and methods chapter section 2.7.1. Samples were DNA extracted using standard phenol-chloroform methods (chapter section 2.7.2) and genotyped for the rare E19A allele using SNaPshot (Applied Biosystems) chemistry (section 2.5.4 and appendices 10.4); all C702 cell lines were homozygous for the rare E19A allele and control samples used in the study did not carry the E19A allele. Total RNA was extracted as detailed in section 2.7.3. Total RNA was treated with DNAase prior to reverse transcription using random decamers and the RETROscript kit (Ambion).

8.2.2. Cerebral Cortex Samples

Experiments were performed using 126 post-mortem brain tissue samples primarily from the frontal cortices but also including parietal and temporal cortex samples. All samples obtained were from unrelated, anonymized Caucasians. These samples were obtained from 3 sources (The MRC London Neurodegenerative Diseases Brain Bank, UK. The Stanley Medical Research Institute Brain Bank, Bethesda, USA. The Karolinska Institute, Stockholm, Sweden). For all samples, genomic DNA and RNA was extracted as in materials and methods sections 2.1.2, and 2.7.3. RNA and DNA extractions were performed by Dr. Nicholas Bray. All

Reverse Transcription (RT) reactions used in this study were performed by myself or Dr. Nicholas Bray. Genotyping was performed by primer extension using SNaPshot chemistry (Applied Biosystems) or Amplifluor allele-specific fluorescent PCR (see sections 2.5.4, 2.5.5 for protocols and appendices 10.4 and 10.5 for primers used).

8.2.3. Allelic Expression Procedure

PCR, sequencing and extension primers used for genotyping and for the allelic expression assays are given in appendices 10.6. Genotyping and allelic expression assays were performed by primer extension using SNaPshot chemistry (Applied Biosystems) analogous to previous studies (Bray, Buckland *et al.*, 2003a) and as given in the materials and methods chapter, section 2.5.4.

Differences in allelic expression were tested by comparing corrected genomic ratios with cDNA ratios from the same heterozygous samples, using protocols and statistical tests described in the materials and methods chapter, section 2.7.7.

8.2.4. Analysis of Steady-State PRKCA mRNA Levels

Total *PRKCA* mRNA levels were estimated in lymphoblastoid cell lines and cerebral cortex cDNA samples described above (8.2.2) using Real-time Quantitative PCR (QPCR), performed using a TaqMan Gene Expression Assay kit Hs00176973_m1 (Applied Biosystems). PRKCA expression levels were normalised using TaqMan Gene Expression Assay kit probe sets specific to *18S* (hs99999901_s1) or *ACTB* (hs99999903_m1). The means of the two samples were compared using a

two-sample *t*-test. Details of the method are given in the materials and methods chapter, section 2.7.8.

8.2.5. Western Blot

For more detailed description of the western blot procedure, see materials and methods chapter, section 2.7.9.

The EBV transformed lymphoblast cell lines used for RNA analysis were cultured and total protein extracted simultaneously for each sample, using a cell lysis/ protein extraction buffer (Sigma-Aldrich, IP buffer) with Protease Inhibitors (Roche), as detailed in chapter 2, section 2.7.5. The total protein extract for each sample was quantified simultaneously by Bradford Assay using Bradford Reagent (Bio-Rad) measured against a BSA standard curve (Qiagen), as detailed in chapter 2, section 2.7.6. A human Cerebellar cortex sample was obtained (Dr. Angela Hughes), extracted in a Thiourea based buffer, and quantified by the same means, to be used as a positive control.

Primary antibodies to PKCα were acquired: a rabbit polyclonal antibody raised to the C-terminus (AbCam. Ab4118), and a mouse monoclonal antibody raised to the central portion (M4 clone) of PKCα (BD Pharmingen. 610107 or 610108). Both were used at dilutions of 1:500. A Primary antibody to α-tubulin (AbCam. Ab24246) was used to normalise the levels of PKCα, at a dilution of 1:20,000. A further primary antibody to PKCβ was obtained (BD Pharmingen. 610127) to ascertain the specificity of the anti-PKCα antibody. Secondary antibodies were from Pierce Biotechnology Biotechnology: Goat anti-rabbit HRP and goat anti-mouse HRP, used at dilutions of

1:2000 (Anti-rabbit; AbCam anti-PKCα) 1:500 (Anti-Mouse; for BD Pharmingen anti-PKCα/β) and 1:10,000 (Anti-Mouse; for α-tubulin).

Western Blots were performed using the Invitrogen NuPAGE system and standard protocols (section 2.7.9). 100ug of total sample protein was loaded into a gel lane for each of the cases and controls, with 5ul sample buffer, 2ul reducing agent (Invitrogen) and extraction buffer to a total volume of 20ul. Note, only 20ug of cerebellar control sample was loaded due to appreciable expression of PKC isoforms in this tissue. A size standard was loaded into one of the sample lanes (Invitrogen. MagicMark XP). Samples were run on Bis-Tris 4-12% Gels using MOPS running buffer (200V;50mins) and transferred to PVDF membrane (Millipore) using Invitrogen Transfer buffer (30V;40mins). Antibody detection and protein quantitation was performed by using a Pierce Biotechnology Biotechnology auto-luminescence kit, a dynamic integration camera and gel-doc system (UVP) utilising LabWorks software.

For the semi-quantitative analysis of PKCα, levels were taken as a measure of the Optical Density (OD) of the PKCα band, adjusted by the OD of the corresponding samples α-Tubulin band. The experiment was replicated 3 times using protein taken from the same cell lines after re-culturing.

8.3. Results

8.3.1. Allelic Expression Analysis

Figure 8.1 shows a schematic of the locations of the SNPs assayed for differential allelic expression analysis relative to their location in the different mRNA isoforms of PRKCA (AB209475 and X52479).

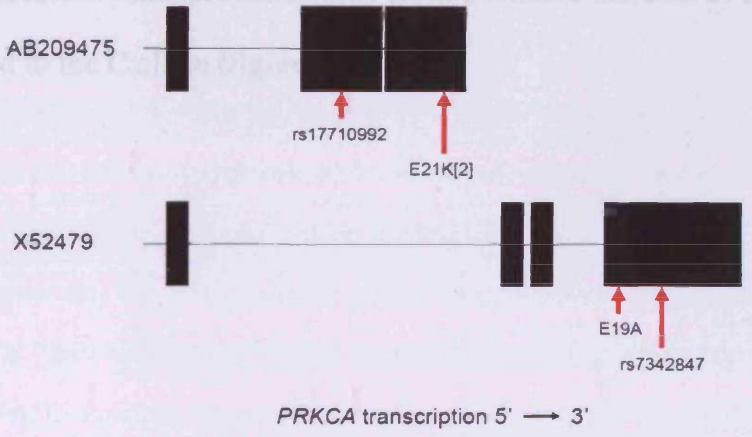


Figure 8.1: Schematic of the 3' exons of PRKCA and the location of SNPs and therefore transcripts assayed for relative allelic expression in this chapter.

SNP rs17710992

Genotyping of SNP rs17710992 (see appendix for genotyping primers) in the gDNA of 66 individuals identified 27 heterozygotes. Allelic expression analysis was performed (see appendix for assay primers) on cDNA generated from the cerebral cortex post-mortem tissue of the same 27 heterozygous samples. 10 samples showed a

standard deviation between duplicate RT reactions of $\geq$ 0.2 and so were removed from further analysis. The mean standard deviation for the remaining 17 samples was 0.07. Figure 8.2 shows the corrected genomic DNA ratios and the corrected averaged cDNA ratios (common allele/rare allele, T/C) for each sample.

A significant difference was observed between the mean of the corrected allelic ratios for the gDNA versus cDNA (2 independent samples t-test p=0.003), which is indicative of a common *cis* acting allele influencing the levels of the *AB209475* mRNA. More detailed analysis of this data showed that the evidence for differential expression came predominantly from a relative increase of the rarer T allele compared to the C allele (figure 8.2).

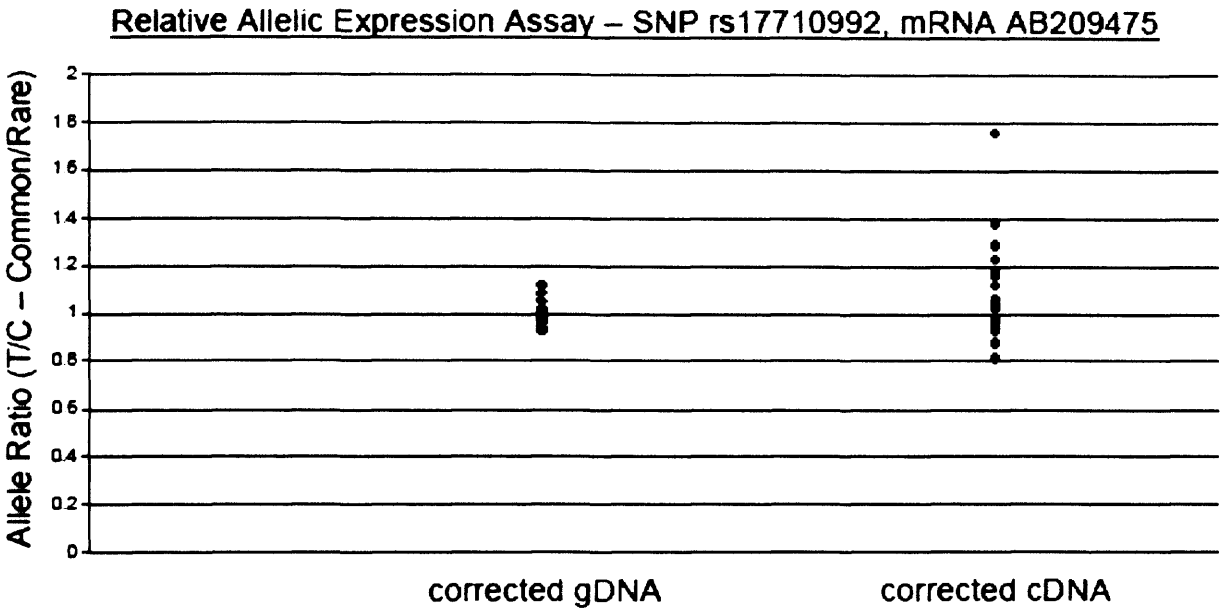


Figure 8.2: Relative Allelic Expression Assay of SNP rs17710992, which exists on mRNA AB209475. Both genomic DNA (gDNA) and complementary DNA (cDNA) were analysed in 17 heterozygous individuals using cDNA gained from post-mortem cerebral cortex. Shown are gDNA and cDNA allelic ratios (common allele/rarer alleles, T/C) corrected by the mean gDNA ratio for all samples, for all assayed individuals. There is evidence to suggest a reasonably common *cis*-acting factor affects the levels of the AB209475 transcript, when the means of the gDNA and cDNA are compared (2 independent samples t-test p=0.003).

SNP E21K[2]

The above evidence of a *cis* acting variation affecting AB209475 expression could incriminate the rare C702 identified alleles as *cis* acting variants. Genotyping of SNP E21K[2] (see appendix for genotyping primers) in the gDNA of 126 individuals identified 8 heterozygotes. Allelic expression analysis for the E21K[2] SNP was performed (see appendix for assay primers) on cDNA generated from the frontal lobe post-mortem tissue of the same 8 samples. No samples showed a standard deviation of

>0.2 and the mean standard deviation for all 8 samples was 0.06. Figure 8.3 shows the corrected genomic DNA ratios and the corrected averaged cDNA ratios (common allele/rare allele, C/T) for each sample.

A significant difference was observed between the mean of the corrected allelic ratios for the gDNA versus cDNA(2 independent samples t-test p=0.04), which is indicative of a *cis* acting allele influencing the levels of the *AB209475* mRNA. More detailed analysis of thesis data showed that the evidence for differential expression came from a relative decrease of the rarer T allele compared to the C allele (figure 8.3).

The genotyping of the E19A and E21K alleles in 126 gDNA samples was followed by sequencing (for PCR/sequencing primers see appendix 10.6) to identify 3 samples heterozygous for the rare haplotype carried by family C702 (see below). The results for the E21K[2] allelic expression analysis were stratified for those individuals homozygous for the common alleles at E19A (haplotype GGAC) and those heterozygous at E19A (haplotype GGAT), see figure 8.3. The results show no significant difference between the means of the GGAC haplotype group and gDNA (2 independent samples t-test p=0.14) and similarly no difference between the mean allelic ratios for the GGAT haplotype carriers when compared to gDNA (2 independent samples t-test p=0.26). Furthermore, there is no difference between the mean allelic ratios for carriers of the more common GGAC haplotype and the C702 identified GGAT haplotype (2 independent samples t-test p=0.83).

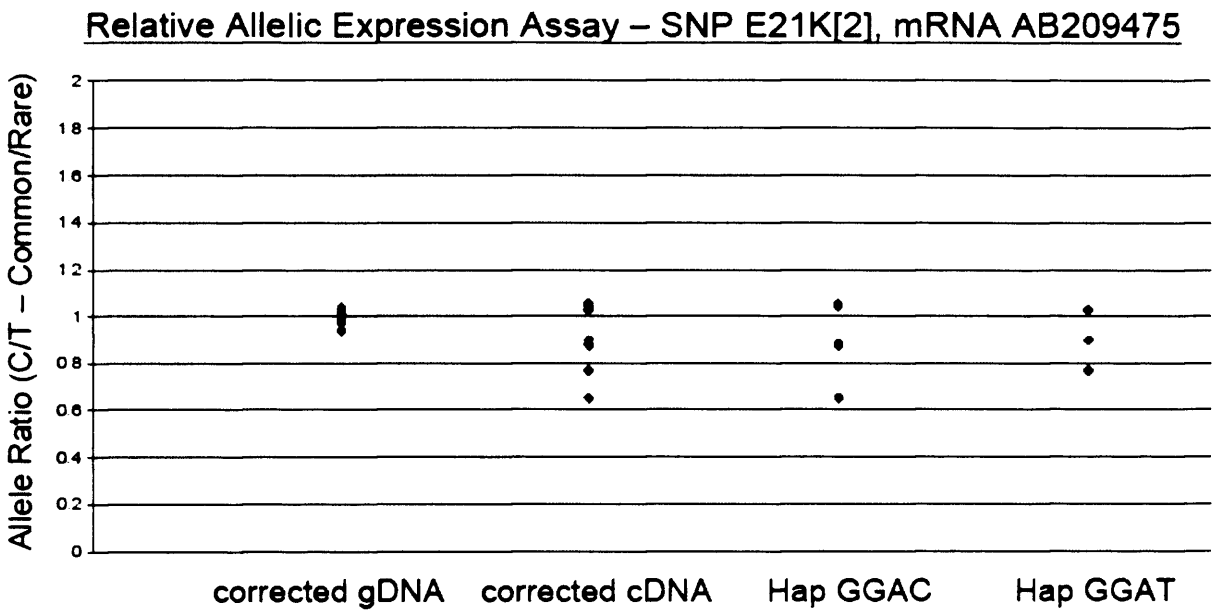


Figure 8.3: Relative Allelic Expression Assay of SNP E21K[2], which exists on mRNA AB209475. Both genomic DNA (gDNA) and complementary DNA (cDNA) were analysed in eight heterozygous individuals, using cDNA gained from post-mortem frontal lobe. Shown are corrected gDNA and cDNA allelic ratios (Common Allele/Rare Allele), for all assayed individuals. Also shown are the same cDNA allelic ratios stratified for being heterozygous for the rare GGAT (i.e. heterozygous at E19A as well as E21K[2]) or carrying the more common haplotype GGAC (i.e. homozygous for the common allele at E19A). The allelic ratios of the cDNA significantly deviate from the expected 1:1 gDNA ratio (2 independent samples t-test p=0.04) implying that the expression of this transcript is under the influence of cis acting variants. No significant difference between the allelic ratios of the haplotype suggests that the risk haplotype is not correlated with these changes.

rs7342847

There is no evidence to suggest the rare C702 identified alleles alter the levels of the AB209475 transcript. To examine the X52479 transcript the genotyping of SNP rs7342847 (see appendix for genotyping primers) in the gDNA of 126

individuals identified 65 heterozygotes. Allelic expression analysis was performed (see appendix 10.6 for assay primers) on cDNA generated from the cerebral cortex post-mortem tissue of the same 65 heterozygous samples. 3 samples showed a standard deviation between duplicate RT reactions of $\geq$ 0.2 and so were removed from further analysis. The mean standard deviation for the remaining 17 samples was 0.03. Figure 8.4 shows the corrected genomic DNA ratios and the corrected averaged cDNA ratios (common allele/rare allele, T/C) for each sample.

A significant difference was observed between the mean of the corrected allelic ratios for the gDNA versus cDNA (2 independent samples t-test p=0.004), which is indicative of a common *cis* acting allele influencing the levels of the *X52479* mRNA. More detailed analysis of this data showed that the evidence for differential expression did not come solely from a relative increase or decrease of either the rare or common allele (figure 8.4).

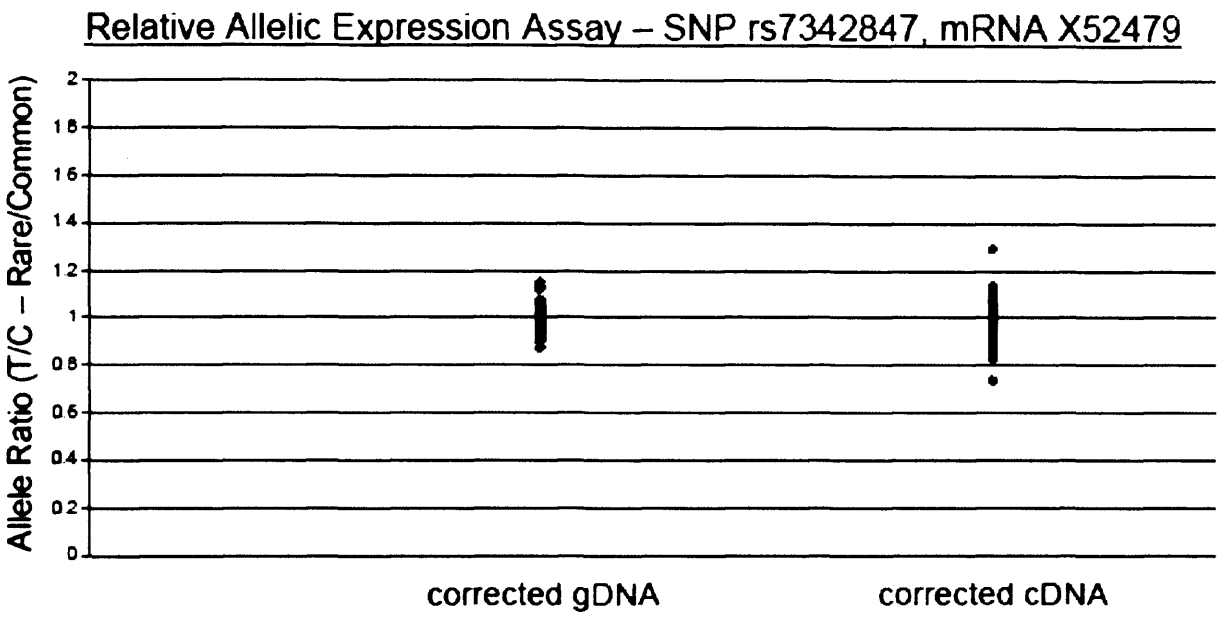


Figure 8.4: Relative Allelic Expression Assay of SNP rs7342847, which exist on mRNA X52479. Both genomic DNA (gDNA) and complementary DNA (cDNA) were analysed in 62 heterozygous individuals using cDNA gained from human cerebral cortex. Shown are gDNA and cDNA allelic ratios (rarer allele/common allele, T/C) corrected by the mean gDNA ratio for all samples, for all assayed individuals. There is evidence to suggest *cis*-acting factors affect the levels of the X52479 transcript (2 independent samples t-test p=0.004).

E19A allele

The differential allelic expression seen at X52479 could be due an effect of the SNP E19A. The genotyping of SNP E19A (see appendix for genotyping primers) was performed in the genomic DNA (gDNA) of 126 individuals identified 3 heterozygotes. Allelic expression analysis was performed using the E19A assay on cDNA generated from the frontal lobe post-mortem tissue of the same 3 heterozygote samples, results are shown in figure 8.5. The standard deviation of the allelic ratios for duplicate RT reactions was calculated for each sample. No samples were removed

from further analysis because of excessive variation (standard deviation for duplicate RT reactions was <0.2). The mean standard deviation for the remaining samples was 0.03 . None of the 3 samples yielded an allelic ratio that significantly deviated from the expected 1:1 ratio of the genomic DNA (figure 8.5).

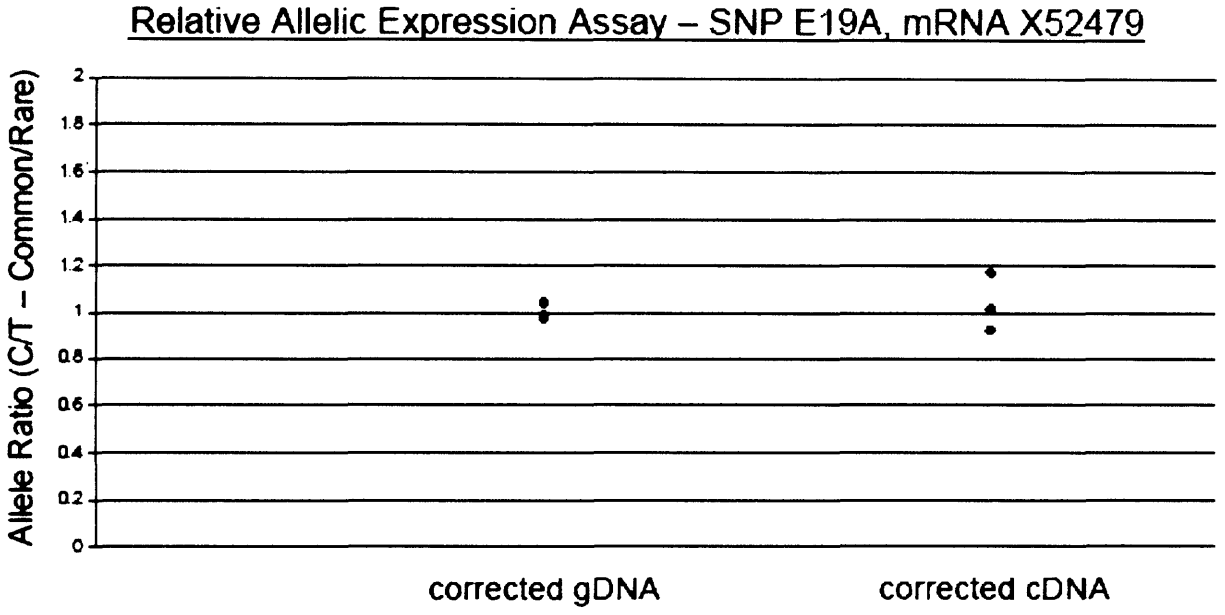


Figure 8.5: Relative Allelic Expression Assay of SNP E19A, which exists on mRNA X52479. Both genomic DNA (gDNA) and complementary DNA (cDNA) were analysed in three heterozygous individuals, using cDNA gained from post-mortem frontal lobe. Shown are gDNA and cDNA allelic ratios (Common Allele/Rare Allele) corrected by the mean gDNA ratio for all samples, for all 3 individuals. There is no evidence that the rare alleles cause differential relative allelic expression in this assay (2 independent samples t-test p=0.31)

8.3.2. Total *PRKCA* mRNA Expression: Neural VS Lymphoblastoid

Total levels of PRKCA were estimated in lymphoblastoid cell lines and neural tissue. Total values for *PRKCA* and *18S* expression together with the normalised PRKCA values are given in table 8.1. The normalised levels of the *PRKCA* mRNA isoform studied (X52479) were low and often immeasurable for the lymphoblastoid cell lines (0.01-0.03) analysed, indicating an extremely low level of this mRNA species in this cell line. However, the normalised level of X52479 in human frontal cortex is shown to be much higher (0.36-0.46) (table 1). Although the standard error of these measurements is likely to be large the data indicate that the level of this PRKCA isoform in the cerebral cortex greatly increased relative to the lymphoblastoid cell line (2 independent sample *t*-test, p=0.02).

	Mean PRKCA[†]	Mean 18S[†]	Normalised PRKCA level
Cortex 1	2700000	7480000	0.36
Cortex 2	2040000	4460000	0.46
Lymphoblast 1	109253	8080000	0.01
Lymphoblast 2	30926	894802	0.03

[†]Arbitrary fluorescence intensity values

Table 8.1: The TaqMan real-time PCR procedure was performed simultaneously on unrelated cortical and lymphoblastoid samples. Arbitrary values obtained for PRKCA levels were compared to a normaliser (18S) to obtain relative measurements of PRKCA (X52479) steady state mRNA in Brodmann area 6 and the lymphoblastoid cell line.

8.3.3. Total *PRKCA* mRNA Expression: A Case-Control Study

Due to the low levels of PRKCA mRNA in the lymphoblastoid cell lines, only PRKCA in the cortex was analysed. Steady-state *PRKCA* (X52479) mRNA levels were estimated in Brodmann Area 6 samples from schizophrenic cases and controls (The Stanley Medical Research Institute Brain Bank, Bethesda, USA). Levels of expression were normalised using a β-actin (*ACTB*) TaqMan Gene Expression Assay (hs99999903_m1). Shown in figure 8.6 are the levels of *PRKCA* mRNA (X52479) in 11 controls and 10 cases, normalised as for table 8.1. There is no significant difference between the two groups levels of PRKCA mRNA (p= 0.35. 2 independent samples *t*-test). As none of the samples analysed carried the rare allele at E19A and therefore do not carry the C702 identified E21K2-E19A haplotype no analysis can be performed on whether the rare alleles affect total expression of *PRKCA* mRNA.

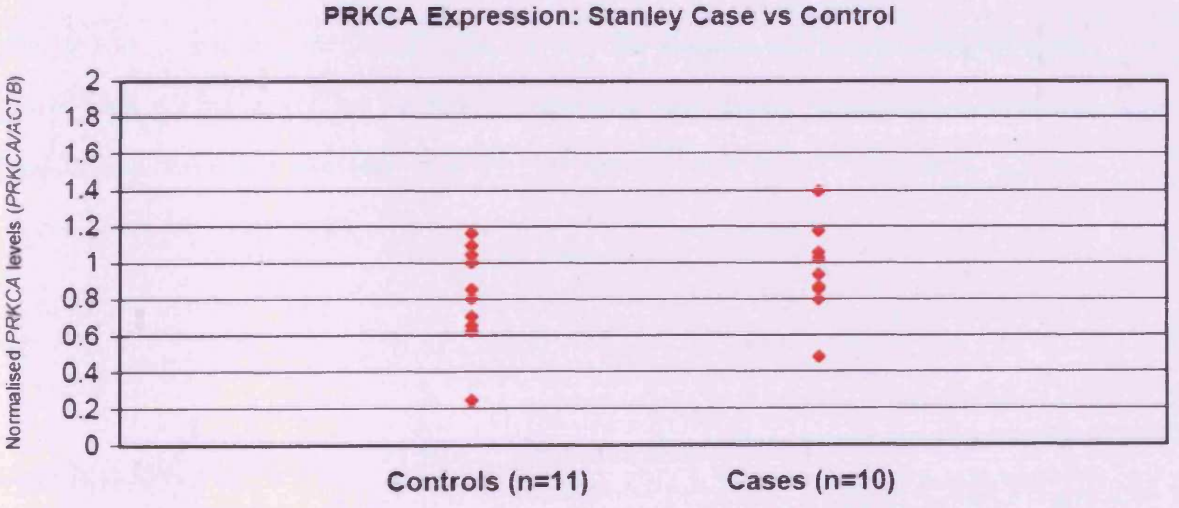


Figure 8.6: Shown are the normalised relative levels of *PRKCA* isoform X52479 (*PRKCA/ACTB*) for 11 unrelated controls and 10 unrelated cases. 2 sample t-test p=0.35. Mean values (±standard deviation) of 0.83 (±0.27) and 0.94 (±0.24) for controls and cases respectively.

8.3.4. Semi-Quantitative Western Blot of C702 Cases and Controls

The western blots and histogram displaying mean levels of PKCα for each sample are presented in figure 8.7. The blot used the mouse primary antibody 610108 (BD Biosciences). The mean normalised levels of PKCα across all C702 siblings was 0.37 (standard deviation 0.13) and across all controls was 0.40 (standard deviation =0.24), which is not significantly different (2 sample t-test p=0.88). There is no difference between the PKCα levels in the lymphoblastoid cell lines of C702 cases and unrelated controls, as detectable by this method. However, the large standard deviation, particularly for the control values, indicates that the assay itself may show variability that may be masking any quantitative difference between cases and controls. Furthermore, the specificity of the antibody may be questionable (see section below).

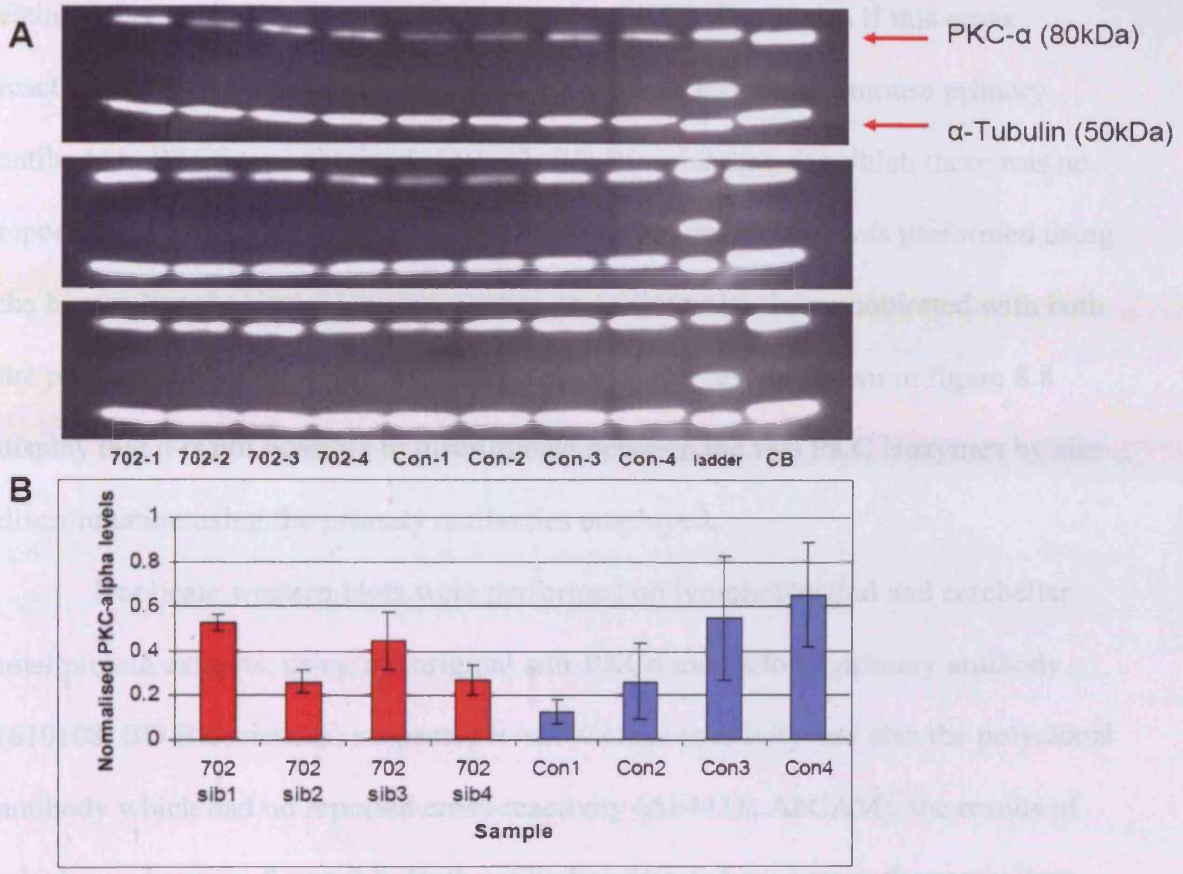


Figure 8.7: Western Blots of using mouse primary anti-PKCα. [**A**] The three technical replicate Western Blots showing anti-PKCα and anti-α-Tubulin. [**B**] Histogram displaying normalised levels protein detected using anti-PKCα (± standard deviation) for each of the C702 siblings (sib) and controls (con). There is no significant difference between the means of the C702 cases and unrelated controls (2 sample t-test p=0.88).

8.3.5. Assessing the Performance of the Anti-PKCα (610108) used for Semi-Quantitative Western Blotting

The semi-quantitative western blots were performed using a mouse primary antibody to PKCα (610108, BD Biosciences); however the product information also

claimed the possibility of cross-reactivity with PKCβ. To discern if this cross reactivity may explain the semi-quantitative western blot data, a mouse primary antibody to PKCβ was obtained (610127, BD Biosciences), for which there was no reported cross reactivity to other PKC isoforms. A western blot was performed using the human lymphoblastoid and cerebellar protein extracts, immunoblotted with both the primary antibodies separately and combined. The results shown in figure 8.8 display that it is not possible to discriminate between the two PKC isozymes by size discrimination using the primary antibodies employed.

Duplicate western blots were performed on lymphoblastoid and cerebellar total protein extracts, using the original anti-PKCα monoclonal primary antibody (610108. BD Biosciences) suspected to show cross-reactivity and also the polyclonal antibody which had no reported cross-reactivity (Ab4118. AbCAM), the results of which are shown in figure 8.8. Both antibodies detected product in the cerebellum, however only the antibody suspected of cross-reactivity (610108. BD Biosciences) detected product in the lymphoblastoid cell extract (figure 8.9). This implies that the initial results were due to cross reactivity and that the level of PKC in the lymphoblastoid cell lines is too low for subsequent analysis.

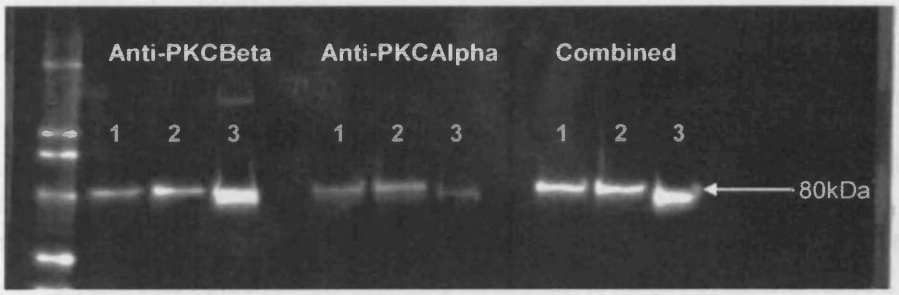


Figure 8.8: Western Blots of PKCα (BD Bioscience mouse monoclonal antibody 610108), PKCβ (BD Bioscience mouse monoclonal antibody 610127) and a combination of the primary antibodies. Shown are the results of blotting for three samples (1=C702 lymphoblastoid. 2=Control lymphoblastoid. 3=Human Cerebellum). The proteins cannot be distinguished by size and are known to exhibit cross-reactivity due to the homology between PKCα and PKCβ.

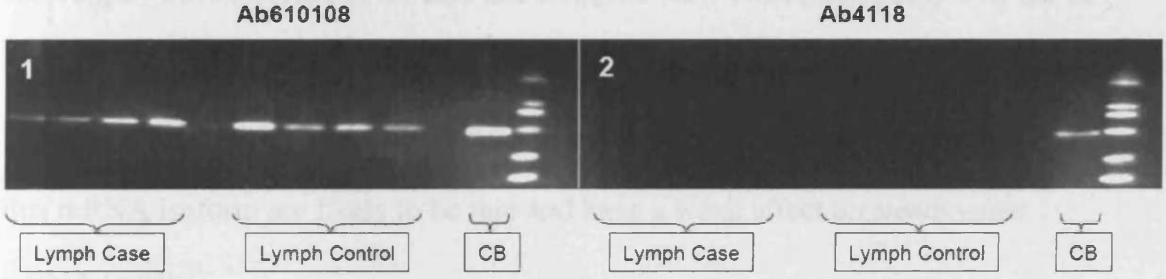


Figure 8.9: Western Blots of PKCα. First showing primary antibody 610108 (BD Bioscience) [1] and secondly showing primary antibody ab4118 (AbCam) [2]. The same samples were prepared and run identically on two gels, the only difference being the primary antibody. The visible samples in image [1] are (left-to-right) C702 sibling lymphoblastoid, unrelated control lymphoblastoid, human cereballer sample (CB) and size standard. This order is replicated in image [2]. The second image using the ab4118 primary antibody is raised to the C-terminus of PKCα and reportedly shows no cross-reactivity, while the primary antibody of image 1 is known to show cross-reactivity.

8.4. Discussion

This chapter aimed to establish the functional consequences of carrying the C702 sibling's rare allele/haplotype. The putative C702 disease alleles (E21K, E19A and the E21K-E19A haplotype) lie within the 3'UTR of two PRKCA isoforms, AB209475 and X52479 respectively.

Allelic expression analysis was performed on common SNPs present in the AB209475 and X52479 isoforms (figure 2 8.and 8.4). These assays demonstrate that both isoforms are under the influence of *cis* acting variation, as mean cDNA allelic ratios differ significantly from the expected 1:1 ratio of gDNA (figures 8.2 and 8.4). The AB209475 transcript is under the control of at least one common *cis* polymorphism that has a reasonably large effect; with 5 of the 17 analysed individuals showing a >20% increase of the rare allele (figure 8.2). Conversely, only 3 of the 62 heterozygotes analysed for rs7342847 in the X52479 transcript show a >20% alteration of the allelic ratio (figure 8.4). Therefore the cis polymorphisms acting on this mRNA isoform are likely to be rare and have a weak effect on steady-state mRNA levels.

To determine whether the differential allelic expression seen at the two PRKCA isoforms examined was due to the rare C702 identified exonic alleles (chapter 6), assays were performed to directly test this hypothesis. SNP E19A, that lies within the transcript X52479, showed no effect on relative allelic expression (figure 8.5). The E21K[2] SNP examined showed evidence for differential allelic expression of the AB209475 isoform (figure 8.3), however this was not found to be caused by the minor E21K[2] allele, or the C702 E21K-E19A haplotype ([G]-G-A-T. Figure 8.3). Therefore, using the assays and tissue samples examined, no evidence

was found to suggest that the C702 identified rare alleles alter steady-state mRNA levels of the PRKCA isoforms examined.

The C702 disease alleles may not act via an allelic mechanism, or in the tissues examined and so a quantitative transcript assay of the PRKCA isoform X52479 was attempted in lymphoblastoid cell-lines from pedigree C702. However, the levels of this PRKCA isoform in the cell-lines tested was too low for subsequent analysis, in both C702 cases and unrelated controls (table 8.1). Therefore, although pre-translational effects such as alternative pre-mRNA splicing and spatial (tissue specific) transcriptional effects cannot be excluded by the allelic expression assays performed, as another possible function of the C702 identified alleles is to alter translation efficiency (Hughes, 2006, St Johnston, 2005). To test this hypothesis a semi-quantitative western blot was attempted, again using the lymphoblastoid cell lines from pedigree C702 and unrelated controls. However, the levels of PKCα protein in the cell lines was too weak to be reliably detectably by one antibody (Ab4118. Figure 8.9) and another antibody showed evidence of cross-reactivity with PKCβ (figure 8.8). Therefore, the effect of the rare C702 PRKCA diplotype on mRNA or protein levels has not been analysed. Furthermore, any effect of the rare C702 alleles on translation of PKCα has not been examined.

Although global quantitative transcriptional assays have identified PRKCA as being dysregulated in schizophrenia ((Mirnics, Middleton *et al.*, 2001) and personal communication from the authors) no support for the dysregulation of PRKCA transcripts in schizophrenia post-mortem cerebral cortex was identified from analysis of a small sample of unrelated cases and controls (figure 8.6).

The research presented here has highlighted some of the potential difficulties in ascribing a biological function to an epidemiologically interesting disease variant.

The associated, C702 identified 3'UTR haplotype may have a function that does not effect mRNA levels and may have a qualitative effect upon mRNA species (i.e. by splicing), or alter protein levels. Also, the C702 alleles may only have a biological function at a specific neurodevelopmental time-point, or only in certain neural or glial cells; or both may be the case. Additionally, the haplotype may have an effect that is so weak as to make it difficult to detect, or any effect may only be seen under certain physiological conditions (e.g. stress, synaptic plasticity). It is even possible that the C702 identified alleles have a functional effect that is unrelated to the typical PRKCA gene analysed in this study and act upon a novel isoform or as *cis* regulatory factors upon another gene. Finally, it may be that only the C702 diplotype has a functional effect, which has not been tested here.

Although no functional effect of the rare C702 alleles has been identified both direct and indirect functional evidence has been published that supports a role for *PRKCA* in psychiatric disorders

A recent study indicated that overactivity of the isoform disrupts prefrontal cortex dependent working memory in primates (Birnbaum, Yuan *et al.*, 2004), and so the C702 haplotype may be acting via a similar route. However, contrary to this finding, microarray analysis may show a down-regulation of *PRKCA* mRNA in schizophrenia as compared to controls (Mirnics, Middleton *et al.*, 2001) and corroborative evidence comes from a study of PKCα in the anterior cingulate cortex (Knable, Barcia *et al.*, 2002), although no evidence to support his was found here (figure 8.6). Further muddying any hypothesis is data indicating that the anti-manic effects of lithium may be via a transient activation of PKCα while chronic use results in down-regulation (Manji and Chen, 2002), although a down-regulation is also reported in suicide victims (Pandey, Dwivedi *et al.*, 2004). There is also data

indicating that the hippocampus may be the site of *PRKCA* deregulation (Manji and Chen, 2002), and there is evidence to suggest that variation within *PRKCA* alters hippocampal structure and episodic memory related performance (de Quervain and Papassotiropoulos, 2006).

In addition to direct findings, the PKCα and the PKC family of isozymes are involved in countless neural functions, including many potentially relevant to schizophrenia: memory consolidation (Bonini, Cammarota *et al.*, 2005), antipsychotic action (Basta-Kaim, Budziszewska *et al.*, 2002, Dwivedi and Pandey, 1999), AMPA receptor trafficking and synaptic plasticity (Collingridge, Isaac *et al.*, 2004), prevention of oligodendrocyte excitotoxicity (Deng, Wang *et al.*, 2004), distribution of neuronal glutamate transporters (Gonzalez, Bannerman *et al.*, 2003), retinoic acid induced cell differentiation (Tonini, Parodi *et al.*, 1991), neurite development (Kapfhammer, 2004, Roisin and Barbin, 1997), hippocampal LTP (Hussain and Carpenter, 2005, Leahy, Luo *et al.*, 1993), oligodendrocyte filopodial growth (Kobayashi, Kidd *et al.*, 2001), PSD/NMDA receptor function (Kennedy, Beale *et al.*, 2005), environmental pollutant neurotoxicity (Kim and Yang, 2005), neuroprotection after anoxic insult (Libien, Sacktor *et al.*, 2005), serotonergic function (Lutz, Mitchell *et al.*, 1993), NR1 NMDA receptor sub-unit phosphorylation (Sanchez-Perez and Felipo, 2005) and many more. However, the direct relevance of the current literature to a disease mechanism for schizophrenia involving *PRKCA* is questionable and no firm conclusions can be drawn.

An abridgement of this chapter would state that no functional evidence supportive of the rare C702 alleles being disease alleles has been found and also there is no functional evidence to support a mechanism via which the levels of a PRKCA isoform alter risk. However, the assays thus far have been on limited sample sizes and

in heterogeneous tissues and none of the assays directly study the C702 diplotype. Therefore, the unlikelihood of being able to assay the C702 diplotype in any given cell at any desired developmental stage or in response to a specific exogenous influence means that functional studies must be approached with more determination. Ideally an *in vitro* or even *in vivo* model system needs to be generated to carry the C702 diplotype. Then, all *PRKCA* isoforms must be studied for the plethora of possible functional consequences discussed, and in as many different spatial, temporal and environmental conditions as possible.

If the genetic epidemiological data continues to support the C702 identified alleles as being risk factors for schizophrenia and/or other psychiatric illness' after further study, then the function of these alleles must be uncovered to aid the understanding of the molecular pathology of schizophrenia and related disorders.

Chapter 9. General Conclusions and Future Directions

9.1. General Conclusions

The aim of this thesis was to identify genetic variants that alter risk of an individual being diagnosed with schizophrenia and related psychiatric disorders. Although there is now unequivocal evidence that the risk of developing schizophrenia is principally due to a person's genetic architecture (Owen, O'Donovan *et al.*, 2002), there have been few, if any, susceptibility alleles identified at the time of writing (Owen, Craddock *et al.*, 2005).

The proposed weak effects of individual alleles on disease risk (Risch, 1990b) has engendered studies of many hundreds of pedigrees in linkage analysis (Badner and Gershon, 2002, DeLisi, Shaw *et al.*, 2002, Lewis, Levinson *et al.*, 2003, Williams, Norton *et al.*, 2003) and several genome-wide significant regions of linkage have been identified (Owen, O'Donovan *et al.*, 2002, Riley, Asherson *et al.*, 2003, Sklar, 2002). In one such study, genome-wide significant evidence for linkage was found at chromosome 17q in a single pedigree of Irish descent (Williams, Norton *et al.*, 2003). Closer inspection of the data revealed that the evidence for linkage was due to 6 male siblings diagnosed with DSM-IV schizophrenia sharing maternal and paternal haplotypes at this locus. The linked region has consistently been reported as linked to schizophrenia and bipolar disorders across populations (Bennett, Segurado *et al.*, 2002, Curtis, Kalsi *et al.*, 2003, Dick, Foroud *et al.*, 2003, Ewald, Wikman *et al.*, 2005, Klei, Bacanu *et al.*, 2005, Lewis, Levinson *et al.*, 2003, McInnis, Dick *et al.*, 2003, Rees, Fenton *et al.*, 1999, Tomas,

Canellas *et al.*, 2006). Therefore, a locus for psychiatric illness susceptibility in the general population is likely to exist on 17q and pedigree C702 represents a singular opportunity to identify a highly penetrant schizophrenia disease mechanism at this locus. The research presented here has attempted to locate the genetic mechanism causing the C702 phenotype and relate this to the general population.

The magnitude of the linkage signal generated by C702 at 17q was a singular occurrence in the pedigree, with little evidence of linkage elsewhere in the genome. Nonetheless, the potential for consanguinity as an explanation for the linkage required analysis and so the pedigree was compared to a sample of affected sibling pairs (ASPs) for excess homozygosity across the genome. Analysis of data for ~350 ASPs and 372 microsatellite markers throughout the genome showed that the siblings were less related than second cousins, and were not therefore inbred (see chapter 3).

The linkage region reported by Williams *et al* (Williams, Norton *et al.*, 2003) encompassed much of chromosome 17, where the affected C702 siblings were apparently IBD2. However, the lack of maternal DNA for this pedigree meant that some of the region could be IBS2. Furthermore, recombination frequencies across the chromosome suggested that in 6 siblings many recombinations should exist across the linkage region (Kong, Gudbjartsson *et al.*, 2002). Therefore, chapter 3 described an attempt to refine the original C702 putative IBD2 region. A further 36 microsatellite markers were genotyped across 17q in pedigree C702 and multiple recombinant marker genotypes yielded a 21 marker region where all 6 siblings showed identical genotypes (figure 3.3). In conclusion, the original linkage region was refined to 11,855,941 bases at 17q23.2-q24.3, where the

siblings are likely to be IBD2. The region is now amenable to molecular genetic analysis and excludes strong candidate genes such as *MAPT* (Pittman, Myers *et al.*, 2005, Stefansson, Helgason *et al.*, 2005) and *PPP1R1B* (Meyer-Lindenberg, Straub *et al.*, 2007).

Linkage due to IBD2 status may indicate a recessive acting disease mechanism, which is supported by the unaffected status of the C702 parents and by the recessive linkage signals reported on 17q (Dick, Foroud *et al.*, 2003, Ewald, Wikman *et al.*, 2005, Klei, Bacanu *et al.*, 2005, Rees, Fenton *et al.*, 1999, Tomas, Canellas *et al.*, 2006). Therefore, regions where the C702 siblings are homozygous for contiguous markers are particularly interesting. A high-density SNP map was generated across the C702 IBD2 region for one sibling and merged with the microsatellite data, giving 1028 SNPs and 21 microsatellites across the 11.7Mb region (chapter 4). The data identified 83 regions where the C702 sibling exhibited a homozygote diplotype; any of which could identify a region consistent with the convergence of a founder mutation. Association analysis of these diplotypes did not identify any empirically associated regions; however one of the diplotypes was extremely rare; not occurring in ~3500 individuals from the UK. The rare diplotype and encompasses the 3' of the functional candidate gene *PRKCA* (chapter 4 and (Birnbaum, Yuan *et al.*, 2004)) and so identified it as the primary candidate to explain the linkage findings in pedigree C702.

Strong linkage in a single, multiply affected pedigree is a rarity in schizophrenia and so an equally rare and unique mechanism may explain the IBD2 linkage, such as the occurrence of a deletion or deletions that may be identified by a loss-of-heterozygosity, as seen in chapter 4. Scanning the IBD2 region using high-density oligonucleotide-array

comparative genome hybridisation (oaCGH) failed to identify any evidence of a CNV unique to pedigree C702 (chapter 5) and furthermore failed to support a deletion variant as an explanation for the homozygous diplotype at *PRKCA* (chapters 4 and 5). Although difficult to estimate, 97% of the IBD2 region is unlikely to contain a deletion CNV >3kb (chapter 5) and so a large loss of genetic material is unlikely to explain the linkage signal in family C702. It must be noted however, that the technology is relatively new and empirical standards of data analysis have yet to become commonplace (Eichler, 2006), and so more sophisticated analysis of the data is required that could potentially identify a small CNV. Nonetheless, the rare region of homozygosity encompassing PRKCA, identified in the affected siblings of pedigree C702, is unlikely to represent a hemi- or homo- zygous loss of DNA (chapter 5).

Several genes in the C702 IBD2 region that lay in stretches of marker homozygosity were screened for mutations and no rare or associated alleles were seen in the C702 affecteds that would be consistent with the convergence of rare, recessive acting disease alleles (chapter 4). However, mutation screening of the rarest homozygous region overlapping with *PRKCA* identified several rare, exonic (3' UTR) alleles in the gene. The exonic alleles formed a haplotype with a minor allele frequency of 0.003 in unrelated UK controls that was homozygous in affected individuals of pedigree C702 (chapter 6). Assuming Hardy-Weinberg Equilibrium for this allele, only 1 person in ~111,000 individuals would exhibit this diplotype and so its occurrence in affected siblings showing genome-wide significant linkage to the region is a conspicuous finding (chapter 6).

Independently of C702, the rare alleles identified were genotyped through ~9000 individuals from association samples collected across Europe and no observation of the diplotype occurred (chapter 6). In the most powerful association sample consisting of individuals affected with schizophrenia and related psychotic and affective disorders, the C702 haplotype was found to be significantly associated with affection status (p=0.05. OR=1.8). In addition, a strong association and moderate effect size was seen with psychosis (p=0.021, OR=3.6) and a psychiatric disorder affection status (p=0.005, OR=3.9) when males only were considered, an observation remarkable given that all the affected siblings in family C702 are male (chapter 6). Thus, the evidence implicating the 3' of *PRKCA* as being responsible for the C702 phenotype is compelling, and as the known exons in the region have been screened and the alleles are exonic their potential as functional disease alleles is promising. However, it must be noted that only when full re-sequencing of the IBD2 region has been accomplished that other variants can be excluded from causing the C702 linkage. Nonetheless, the affected individuals of pedigree C702 are homozygous for a rare, exonic haplotype that is associated with psychiatric illness in the UK (chapter 6)

Multiple studies identify linkage to 17q in schizophrenia and related disorders (Bennett, Segurado *et al.*, 2002, Curtis, Kalsi *et al.*, 2003, Dick, Foroud *et al.*, 2003, Ewald, Wikman *et al.*, 2005, Klei, Bacanu *et al.*, 2005, Lewis, Levinson *et al.*, 2003, McInnis, Dick *et al.*, 2003, Rees, Fenton *et al.*, 1999, Tomas, Canellas *et al.*, 2006) and so other susceptibility alleles may exist that have a greater population attributable factor (Carlson, Eberle *et al.*, 2004) than the C702 alleles. Directed mutation screening of *PRKCA* identified many novel variants that were supplemented with known SNPs to

provide an association marker map of 58 SNPs across the gene that were genotyped in a UK schizophrenia case-control sample (chapter 7). Several SNPs were identified as associated and nominally significant single-marker allelic and haplotypic replication was achieved in multiple independent association samples (chapter 7). Although each association is weak and would not survive corrections for multiple comparisons the replication of the same associated allele of a common SNP in the 3' of *PRKCA,* the same location as the C702 associated haplotype, is a notable finding (chapter 7). In short, a sparse association map analysed in a schizophrenia case-control sample from the UK indicates that there may be common alleles at *PRKCA* that are associated with schizophrenia (chapter 7).

Several functional assays were also attempted to test the ramifications of harbouring the C702 alleles on *PRKCA* mRNA and protein levels (chapter 8). However no effect of the rare alleles was found on mRNA levels in cerebral cortex tissue from post-mortem adult human brain (chapter 8). Furthermore, quantitative western blot results that show no effect of the C702 identified *PRKCA* alleles in lymphoblastoid cell-lines may be assaying a homolog of PKCα (chapter 8). Additionally, despite reports of a dysregulation of PKC mRNA in schizophrenia there was no evidence that this is the case in a small frontal cortex case-control sample (chapter 8). Although no functional corollary of the genetic variation at *PRKCA* has been demonstrated only a tiny proportion of the potential functional consequences have been assayed.

Despite the systematic approach of these studies, several caveats are recognised. The primary focus of this thesis has been on a single pedigree where all affected siblings

share paternal and maternal haplotypes across one region and so generating linkage to the region. However, it is possible that this linkage signal is due to an unrecognised phenotype that the siblings share, such as susceptibility to a late-onset disorder, although the obvious similarity between the siblings is their schizophrenia. Another fair criticism could be the disregard of the IBD1 region as harbouring a phenotypically relevant allele (as the siblings share this as well), although recombination frequencies suggest the IBD1 region is unlikely to be truly the size estimated with the current samples and genotypes available (Kong, Gudbjartsson *et al.*, 2002). Furthermore, it is intuitively more unlikely that the siblings would share two haplotypes and so attending to the IBD2 region primarily is logical, although an interaction between alleles in the IBD1 and IBD2 region is also possible.

It has been argued that the C702 disease allele is likely to lie within a homozygous region and itself be in homozygous form and rare. The reason for looking at the IBD2 region is to search for a convergence of highly-penetrant (in this pedigree) disease alleles, explaining the linkage in this family to 17q (Williams, Norton *et al.*, 2003). As the parents are drawn from the same population and highly penetrant schizophrenia alleles must be rare, it is likely that the alleles are the same and the IBD2 haplotypes represent the convergence of an allele from a founder ancestor. However, if one can accept this is true, this still does not unequivocally implicate the rare *PRKCA* genotype identified, although the rare homozygous region containing an extremely rare diplotype is the best explanation of the current data. Therefore, the entire IBD2 region must be re-sequenced to unambiguously identify the *PRKCA* diplotype as the only probable variant that can explain the C702 linkage. Although the association seen with

the C702 identified allele in the UK would at least suggest that even if other alleles are acting in pedigree C702, the alleles identified are part of the genetic architecture of their phenotype.

Finally, there is no evidence to suggest that the identified C702 haplotype is anything but a tag for the UK case-control association seen, except the exonic location of the alleles. There is no question that the encompassing exons are transcribed, but this does not preclude other functional alleles in introns or as yet unidentified transcripts. Therefore, any re-sequencing efforts need to be directed at the region of homozygosity harbouring the rare *PRKCA* alleles identified in pedigree C702.

9.2 Future Directions

The findings presented here require further analysis before they can be accepted or disproved. Since the start of this thesis there have been many advances in the tools available for molecular genetic research that may facilitate this. Whole-Genome Association (WGA) studies of multiple phenotypes are being published, entailing genotyping platforms that assay thousands of SNPs across the genome in large case and control studies (e.g. (Hunter, Kraft *et al.*, 2007, Wellcome Trust Case Control Consortium, 2007, Zeggini, Weedon *et al.*, 2007)). This data can be coupled with HapMap Phase II data, that allows estimation of the common variation coverage of a WGA study design and also empowers the statistical imputation of missing data (Servin and Stephens, 2007, Wellcome Trust Case Control Consortium, 2007). Association data can then be coupled with projects attempting to uncover the allelic architecture of the

genome and identify functional sequences (ENCODE Project Consortium, 2004). In addition to this, there are several high-throughput mutation screening technologies and also a number of platforms that offer whole-genome resequencing (Topol and Frazer, 2007), allowing the full allelic architecture of an individual to be obtained. These technologies and others may help elucidate the susceptibility to schizophrenia and related disorders conferred by chromosome 17.

Large-scale re-sequencing technologies have very recently become economical alternatives to positional cloning and directed re-sequencing (Topol and Frazer, 2007) and it is now possible to resequence millions of bases in several individuals for thousands (rather than hundreds of thousands) of pounds. The application of this technology to the pedigree C702 IBD2 region will achieve two things: A complete dataset for the region in the pedigree that will allow all interesting variants to be tested and also the rare alleles at *PRKCA* (or other alleles that refine the associated haplotype) may be identified as the only plausible explanation of the families linkage to 17q.

The whole-genome association studies of schizophrenia that are underway will involve unprecedented sample sizes that should allow the detection of alleles of weak effect (O'Donovan et al, submitted). The same strategy must be applied to the C702 identified haplotype at *PRKCA* where the sample must be powerful enough to observe at least one other carrier of the C702 diplotype. If the diplotype is found in an affected individual this would give credence to the genotype being responsible for the pedigree C702 phenotype. Assuming HWE, this may require ~110,000 samples to disprove the alleles as being involved in psychiatric illness, although the figure should be less than this

if the genotype is associated with disease. Alternatively, proving the haplotype imbues disease risk or not would also solve the ambiguity, although this will also require a schizophrenia case sample of many thousands, even more if the risk only occurs in males.

Independently of pedigree C702, whole-genome association platforms and cheaper genotyping technologies, coupled with high-throughput mutation screening or gene re-sequencing should enable a much greater amount of the common genetic variation to be assayed at *PRKCA* than has been attempted in this study (chapter 7). The WGA studies currently being conducted united with the statistical inference of ungenotyped variants (Wellcome Trust Case Control Consortium, 2007) may have greater power than the association mapping design employed here (chapter 7) to detect or disprove association at the *PRKCA* locus. However, the discovery of rare or common associated alleles at *PRKCA* may be easier than confirming an associated allele as a disease allele.

The demonstration that the C702 identified alleles have a biological consequence would give authenticity to the diplotype being a disease variant, at least in pedigree C702. A functional consequence of an allele does not *de facto* mean that allele is a disease relevant allele; the function may be innocuous or related to a disparate phenotype. However, the demonstration that the affected C702 siblings are homozygous for a rare haplotype that has a biological function shows independent and convergent evidence that the alleles are disease relevant. Towards this end several studies are warranted. Molecular *in vitro* studies of the effect of the polymorphisms on reporter gene activity, or analysis for any effect of the alleles on DNA/RNA-protein binding (DNA or RNA electromobility shift assay) in cell-lines are obvious suggestions. Additionally, analysis of different

transcript and protein levels and also localisations in post-mortem neural tissue of carriers of the C702 rare exonic haplotype at *PRKCA* may also be successful. However there are other options such as functional neuropsychological analysis of rare allele carriers versus controls (i.e. fMRI, or working memory tests) that could also identify a functional consequence of the C702 haplotype.

It may be more practical, given the possible disease allele heterogeneity that may exist at *PRKCA*, to search for a common molecular pathology of schizophrenia and related disorders that involves *PRKCA*. For example, it may be that schizophrenics commonly show a dysregulation of PRKCA mRNA (Mirnics, Middleton *et al.*, 2001) or protein (Knable, Barcia *et al.*, 2002). The demonstration of a molecular pathology involving *PRKCA* may help direct functional assays attempting to connect genotype and function to the correct functional and spatiotemporal locus of disease risk. However, such an approach may only be viable for a common allele association at *PRKCA* (chapter 7) as the C702 alleles may produce their effect on disease risk by an separate mechanism.

At the time of writing there are a few examples of psychiatric disorder association samples that have undergone WGA studies (Lencz, Morgan *et al.*, 2007, Wellcome Trust Case Control Consortium, 2007). These studies are being praised as more powerful and informative than linkage and pedigree analysis, however such studies typically involve a fixed marker set and a few thousand cases and controls and so the rare alleles identified in this thesis are unlikely to be assayed by the WGA approach (Bourgain, Genin *et al.*, 2007). However, the WGA study approach may detect the common allele association reported here (chapter 7), although weak associations may be overlooked or dismissed as

chance when so many tests are performed. Furthermore, if some loci engender risk by different, rare alleles acting in many individuals then genetic variation imbuing a large overall population attributable risk may be overlooked (McClellan, Susser *et al.*, 2007), whereas linkage analysis should identify such regions. Therefore, the future of this study and genetic studies of common, complex disorders should be designed to incorporate allowances for rare and common disease risk alleles.

As an example, our research group has recently completed a WGA study consisting of ~500 schizophrenia cases and ~3000 unrelated controls from the UK successfully analysed for ~300,00 SNPs (O'Donovan *et al,* submitted). Such a sample size, even if directly testing the C702 identified alleles, would not be powerful enough to prove or disprove association. Furthermore, no alleles within the original (Williams, Norton *et al.*, 2003) or refined (chapter 3) C702 linkage regions achieved genome-wide significance or were deemed worthy of initial follow-up (O'Donovan *et al*, submitted). However, linkage analysis of the multiply affected pedigree C702 (chapter 3) identifies a small proportion of the genome for study, containing only 1028 analysed SNPs (chapter 4). Therefore, a p value of $\sim 10^{-4}$ obtained within this region would survive conservative allowances for multiple tests. It is noteworthy then, that a *PRKCA* identified SNP not only meets this criterion but is the most significant allelic and genotypic result in the C702 IBD2 region (unpublished observation).

This thesis may serve as an example for future genetic studies of complex disorders, particularly in the era of large projects involving perhaps millions of datapoints for tens of thousands of samples. There are many pathways to disease gene discovery and multiple lines of evidence are required to confirm an allele or gene as responsible for risk

of developing a complex disorder. Therefore, detailed analysis of small but exceptional pedigree may offer insights into the mechanisms of complex disorders, and schizophrenia is no exception.

10. Appendices

10.1. Appendices Chapter 3

Example microsatellite marker traces

Shown are the genotypes for recombinant markers that define the C702 putative IBD2 region.

Each image shows the fluorescent trace for that sample, corresponding on the x axis to size (size of alleles genotyped is shown) and the y axis to fluorescence intensity units. Sample identifiers (in descending order) are: 702-01 (father), and the six affected male siblings; 702-04, 702-06, 702-07, 702-08 and 702-10.

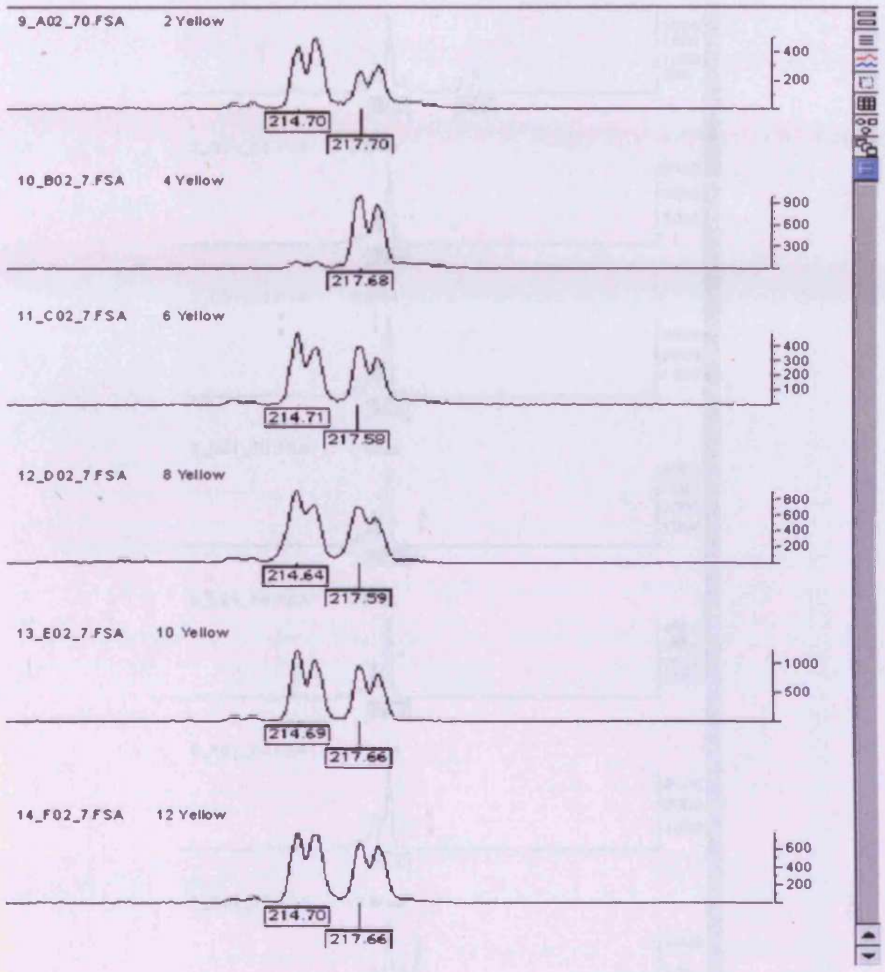


Figure 10.1.A. D17S1295 genotypes in pedigree C702 (Genotyper)

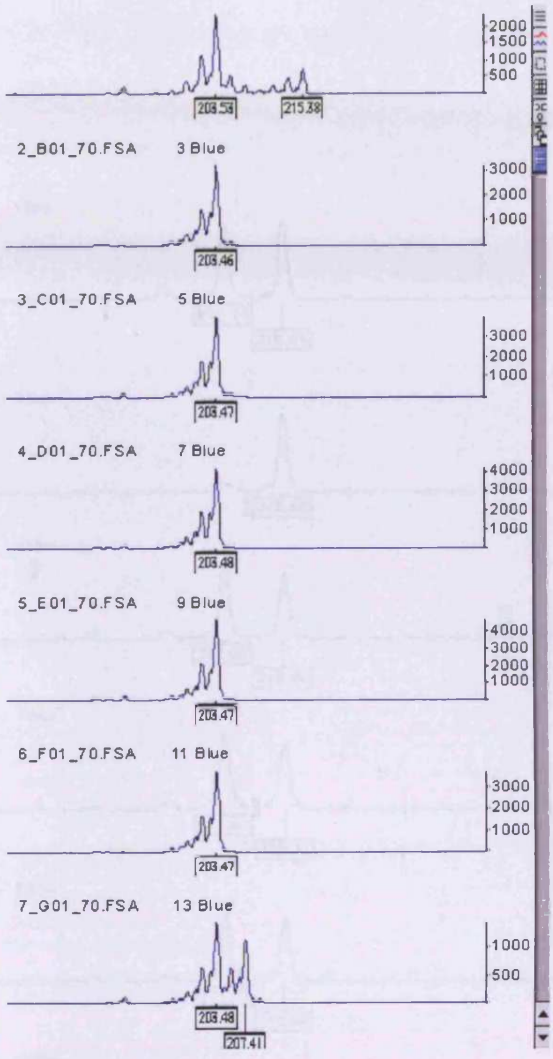


Figure 10.1.B. D17S1160 genotypes in pedigree C702 (Genotyper)

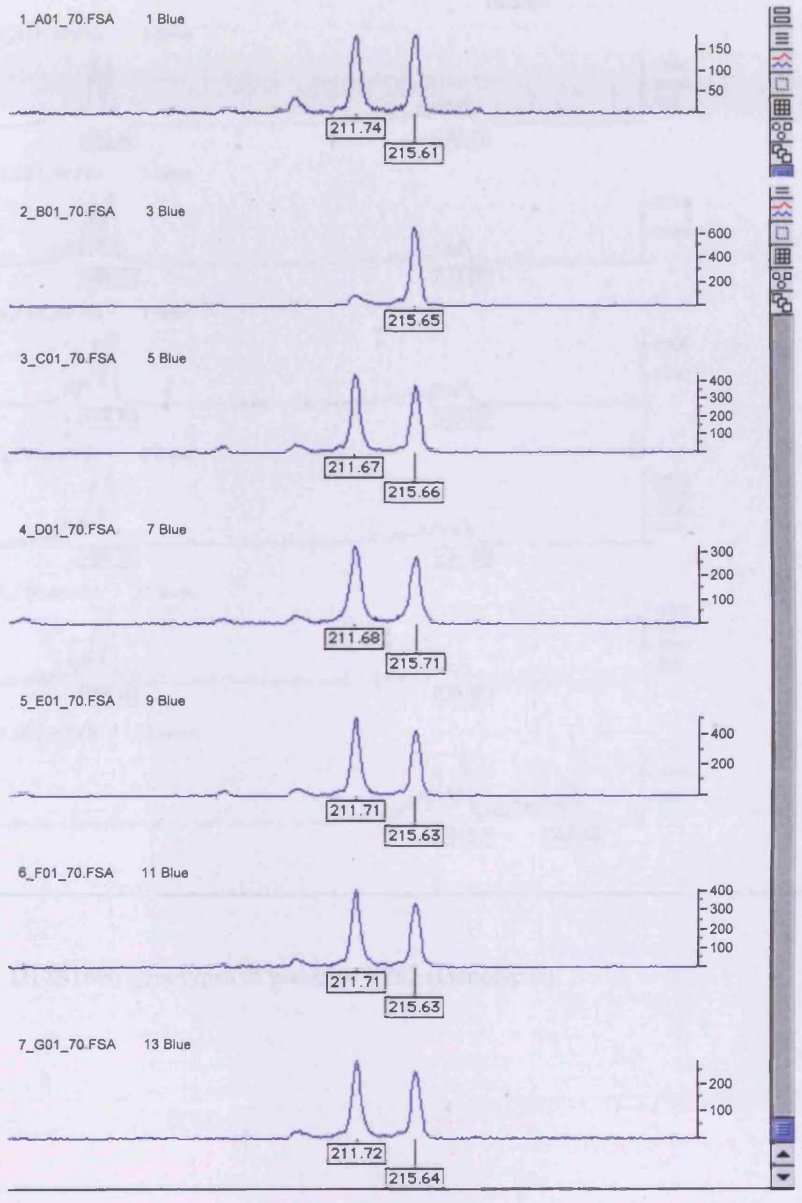


Figure 10.1.C. D17S2182 genotypes in pedigree C702 (Genotyper)

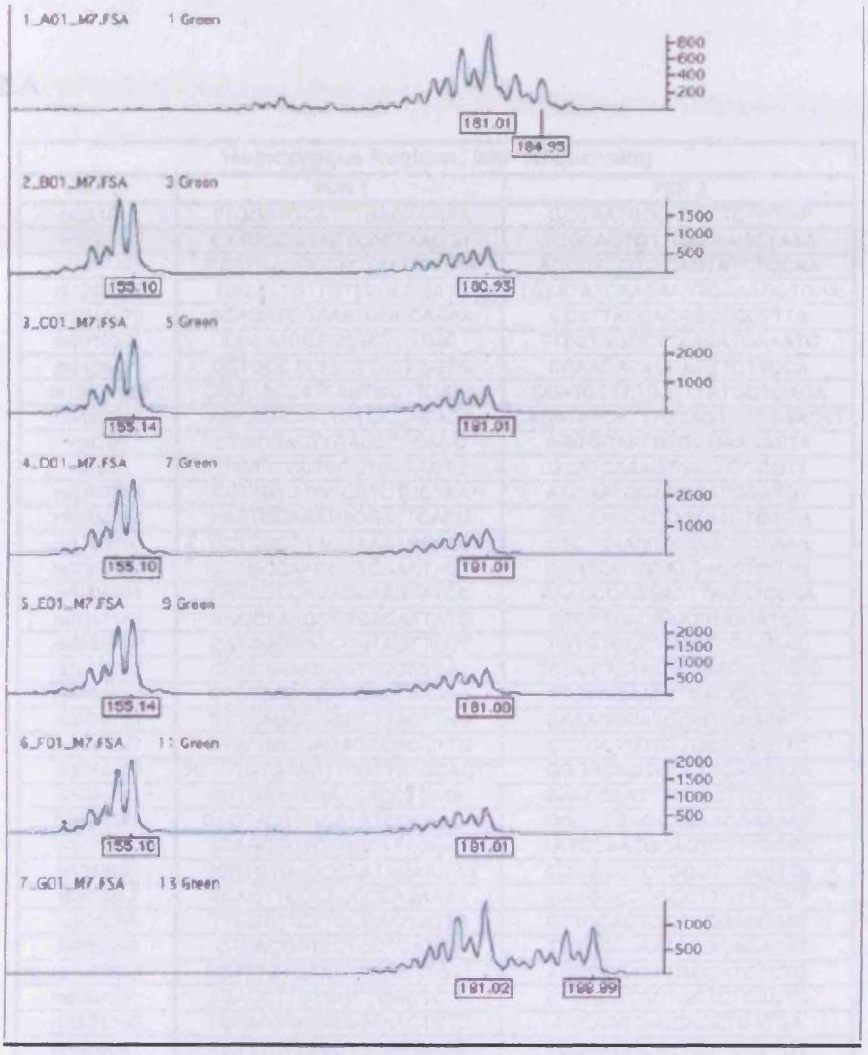


Figure 10.1.D. D17S1606 genotypes in pedigree C702 (Genotyper)

10.2. Appendices Chapter 4

Table 10.2.A

Homozygous Regions: SNP Sequencing		
SNP ID	PCR 1	PCR 2
rs241062	TTGGATGCATTTGAGAGCAA	GCGAATGGGTTTCTCTGTGT
rs958334	CATGCCCGACTCACTAAGGT	TCGCAGTGTTGACAAGCTAAA
rs7214345	CCATGACCCCACTAATCAGAA	AGCCTTTTCTCAGTATCTGCAA
rs1268007	TGGAATGTTCTTCCCCCATA	TGAATATCAAGAAATGAAAGGTGAA
rs3744279	ACAGATGGAAATGGGCAAAA	CCATTACCACAGCCCCTTTA
rs9913301	CCCAACCAGCTCTTCTGC	TTTGTGGCCTTAAAATGAAATC
rs2429426	GCTGCCTCTCCTTACTGGTG	CCAAGACACCAGTTCTTCCA
rs10491168	TGGGACCATTAGTGCTTCATC	CGATCCTATGATTTATGCTGAGA
rs4968648	AGCAGAGGTTGTGCCAGAGT	CAGATATCATTTCCAGTCTGTGACTT
rs4295	CTGGGACTTGACCCTGACC	AGGGCAATGTGGGAACACTA
rs4311	TCCTTTCCTGCCTGAAACTC	GCATCAAAGTGGGTTTCGTT
rs4267385	CCTTGTGTGCCCTCTCCTAA	ATGGATGGATGGATGGATGT
rs6504165	AACTGGAAATGGGATTCACG	ACCAAGCAGCCTAACTGTGG
rs3760252	CCTCGCCTAGGAAAAGGATT	CTCTCAAGTAGGCCGCAAAA
rs2058194	CCAGCCAACGGTGAAGTAGT	GCATGGTGTATGACCTCGTG
rs6504204	CACCGCCACAGAAAGTATCC	AAACCCAGGAATTAGGTCCAA
rs2047153	GGGCAAAGGCTCAGATTATG	GTGTTCACAAATTGCCATGC
rs4968716	CATGGGTGACCGTAGTTCCT	TGTGTGCCTGGTTCTGAGAC
rs1550792	CCACAAAAGGATTCCTGGAT	CCCCCTCTACTCTGGTATCTCC
rs4968723	CTCTCAGGGACCACCAAGAA	TTGTGGAACTTGACCCTGAA
rs2070782	GCTCAGCCGGGCTTACTTAG	GAAAGCCAAGGGTGAGCATA
rs6504227	ATGAGACCAGAGCCAGCTTC	CCTGCTGTGTTGCACAGTTC
rs3744409	TGTCTGTGTAGTTCCTTATGCACTT	GGGACAGTGCCTTCAGGTTA
rs1991402	GGTGGAAGAATCACCTGAGC	GAACCAATTTTGCCTTGTTGA
rs6504248	GACTACAGGCATATGGCAACA	GGCTGGAGCAGAAGAAAAAT
rs1476171	ACAGCCCTCAGGCTTAGGAT	AATCCAATGCAGTGTTTCAGC
rs7218921	GCATGTGAGCCAATAGAACCA	CCAGGGTCTGGGTTCAGTTA
rs299842	ACAGTTCCCACCCCAGAAAT	CATGCCTGGCCTTGTTTACT
rs938299	TTGGTCTCTTGCGAAGACTG	GTTCCACTGCAAGGACCACT
rs908089	CTCACGGTTCTCCCCATCT	TTGCCCAAATAGGGACACAT
rs8077696	GCTTTGTGAATGGCAGTTGTT	AGACCCATTTGCCATCTCTG
rs6504297	AGACCTTGTGGCTGGCTCTA	CTCACGTGCTGCTCTCACTC
rs3926150	TGCAAGGAGGGGTAACTGTC	AGCAACCACCACCTGATGA
rs9908987	TCTTTCTTGGCTGTGTGAGG	TGGAAGCCACTGTCCTCTCT
rs11868547	CACAAACCCCAATCTTCACC	TGCTGGGAGAGTCCTGAGTT
rs1468329	ACAAAAGAGCGGCACTGTCT	GTGCACTCTGACCCAAGGAT
rs876843	TCTGCCTCTAAGGATTCACACA	GCTTTAGGATGGATCGCTGT
rs4377195	ATGAGGCTGCAAGAGATCGT	TCTCATGTGGGCAGTCTGAG
rs2052153	TGGTTTCCCTAAGCCAAAAG	CGATATGGGGCTTCAAACAT
rs972952	CATGCGAGGGATCTAGGTTG	GCCCTGACATCTTTGAGCTG
rs11079641	CTCCTCAACCACAGGCTGAT	CTGAGGCTTGCTTTCAGTCC

Table 10.2.B

AB046856

AB046856				
Amplimer ID	PCR 1	PCR 2	Extension primer	SNP ID
PromoterA	TGTGTAGAAAGAGTTGAAGAAAAGTGA	CTCCTGACCTCAGGTGATCC	NA	NA
PromoterB	ATTATTGGCTGGGTCCAGTG	TGCCAGTTTTTCTGCCTACA	NA	NA
PromoterC/Exon1a	AGCAGGGAGAACTTGTGCTT	GCTTTCAAGGGTAGAGGTTGG	NA	NA
Exon1b	ACTAGCAGCACTGCCTCTCC	AGGTTTCTCTCTCCAGCATCA	NA	NA
Exon2	TCCCCAATGTGTGTTTTTATCA	CCCAAACACAAAAGAACAGC	GGAAACATAAAACCTTGGCA	rs2460111
Reg2a	TCCTGCCTCTCTTTTCCTCA	GAGATGCCTTCCTTCCACCT	NA	NA
Reg2b	GCTGGTATCCAAGTTCCAAGA	GCACCTTCATAAGACCCAAAA	NA	NA
Reg3	CCACTAAGAACAAGGCACTGG	GGATGCTGGTTAGCCAAAAA	NA	NA
Reg4	ACTGCAAAATGGGATTCAGG	TGATAACCCAGTTTCTCCTTGC	NA	NA
Exon3	AGCCCTGATTCCAAAGTTCC	AGCAACTTTTTGGCTTTTGC	NA	NA
Exon4	TCTCAATGTTGACAGTTTGACCT	AACTTCAGCGCCAATGAAAC	NA	NA
Exon5	TTGGGTGGACATGTGTTTTT	TCCTCCACGTTCTCATGGTA	NA	NA
Reg5	CTCCTTGCCCTTGAATGAAA	TGGGGAATGGTGGAAGATAG	NA	NA
Exon7	TGATCAAAGATCCAGTAAGAACTCA	TCGGGAATTATTCTTCATAATGCT	NA	NA
Exon8	GTTTCCCTGTTTGGCCACT	CACATTCAGGAGAGGCATCA	NA	NA
Exon9	CCTGTTTCGCTCTCAGAGGT	TCAATACTCACACTCCATGCAA	CCGAGGTAAGACATATATCTGTGATAA	rs4968765
Reg6	AAGTAGAAAGTTTAGAAGCATGAAACA	TGGTCTCTACTTTTTCATTTCACA	TTACTTCCTTTTGTTTCTCTAATTCC	rs4968768
Exon10	GAGGCTGAGGGAGGAGAATC	GGGAGATGAACCAAGTTTGC	GCATATGAAAGTCTGCTAGAGATGG	rs8065950
Exon11	GGCCACTAATTCACCACTGA	GAAAGACAAGTCAAAGATCAACTAAA	NA	NA
Exon12a	AAAAGATGACTAAAAGGATGTGACTC	TGCAGAACCAGCACTCTCTC	NA	NA
Exon12b	CCCTTTTTCTTTATTTTCGATGC	TGCCCTTGGTCAGTCTGAGT	NA	NA
Exon12c	CCTGTCCTTGAAGACCACCA	CGAGGCCCAGTTATCATTGT	NA	NA
Exon12d	CAGTGATTGTGCACCTGTGA	GCTCTGGGGAAAACAATGAA	TCGGAGGTTTGCTAAAACTC	rs4968644
Exon12e	TTAGCTGGGTCTTTCCTGCT	TAGGACAGGTTTGGCAGAGG	NA	NA
Exon12f	CCTCTCCCTTGCTAACCAGA	GGGCTCAGAGATGTGAGGAA	NA	NA
Exon12g	CCTCTCTCAATCGCACTTTTC	GAGGCAGGGGAATTGCTT	NA	NA
Exon13a	CAGTGGCGGATGTGACTAGA	CTTTCCAACTCGCACCTTTC	TTCCTGGGTCCTTCAATACA	rs2270133
Exon13b	CGTGATCGCTGCTCTGAATA	CAATTCCCTACATTTGCCACT	NA	NA
Exon13c	AAGGCTGTTGTAAGGGATGC	CGAGGTTTGGGACACAAAAG	NA	NA
Exon13d	AGATGCAACTGAAGGCCAGT	CCCGAGTAGCTGGGATTACA	NA	NA
Exon14	GCTCATCATGGAAGCTGTGA	TCTGCATTACAAAAGCCTGAGA	NA	NA
Reg7	GGGTATGCATTTGGGTTAGG	AAGGGTCTGCTGGAGATGTG	NA	NA
Exon15	TGGAAAAGGTGGCAAAAATC	TCCTCAAACCTGATCACTGCT	NA	NA
Exon16	GGTACTTATGGGGGAAGTGGA	GCCTTTCATTTTCCAACCAA	NA	NA
Exon17	CTGGCATACCAGCATCAGTC	TCCACTTTCAACCCCAGGTA	NA	NA
Exon18	GCAGCATAGCCTGGGGTAG	GAGCCTCTCTGTTCACTTGGA	NA	NA
Exon19	GCCTCTCACTGCTGCTTTTT	AGGGAACTCATGCAGGCTAA	NA	NA
Exon20	GCTTCTTGGCAGTGGGAGT	GGCCATCACTGAGGAACAGT	NA	NA
Exon21	TTGGGAGGGTTAGCCCTTAG	TGCTCTGTGAAGGACAACCA	NA	NA
Exon22	TTCCATCCATGAATGTCAAAAA	GATCTTGAGCTGAGGGCTGT	NA	NA
Exon23a/Reg8	TACCCCAGACTCCCACAGAC	GGTGAGCTGGAGATGTAGGC	NA	NA
Exon23b	CCTGGAACAGGATGTTGAAAA	AGATTTGGGTCTGCCAGGA	NA	NA
Exon23c	GCCCTCCAGCTGAAAGTATG	TCGGTATGCTGAGGAAGGTC	NA	NA
Exon23d	GGAATCGTAGGGGATGGAAG	TCCTTGAGTTGGAGACACTGG	NA	NA
Exon23e	ATCTCCTCCACCTCCAACCT	GAGTCCGAGCTACCTTCACG	NA	NA
Exon23f	GTCTCGTGGAGACCTCTTGG	GGTTAAAGCACACAGCACGA	NA	NA
Exon23g	GGGAGCCAGGACTTCAGTTT	TGTTCCTACAATGCCTCTTGG	NA	NA
Exon23h	GCTGGTCTTCCTAGGGGGTA	CAGGAAGAGGGCCACTAGAA	NA	NA
Exon23i	AAAAGGAAACTCTAACCGAAGG	AATAGTGGCCTCGCTCACTC	NA	NA
Exon23j	GTGGGGGCATGAGATAGACA	CTGGCCTGTTTTAGGTTTGG	NA	NA
Exon23k	GACACAGGGAGCATGAACCT	AGCAGGAGCACTTTCCTGAC	NA	NA
Exon23l	GCACTGCTGGCATCTTTACA	AGGGGCAGCTCTGAAATACA	NA	NA
Exon23m	CCTGTGCCCTGTTTGCTACT	TGTGCTATAAATGTTGTGATCCACT	NA	NA
Exon23n	GCAGAACCGCTCTTGTCTTT	TGGAATTTTCCTGAGCCATC	NA	NA
Exon23o	GTATCTTGTCCGGGTTGCAG	GCTACAAAACAAGCGCAGTG	NA	NA
Exon23p	CGTACAGTGGGCTATGTGGA	AGCACTGGTTTTGTCCAAGG	NA	NA
Exon23q	CAGTTCCTCTCAGCCTCCAG	AGCCAGAGCTGAGTTCAAGG	NA	NA
Exon23r	CTGCAAAGACGTGTCATGGA	GCACTTCCGTTCCACAGAGT	NA	NA
Exon23s	ATACCGACGCTGAGCTGACT	TGCTGTGCTAGATGGAAAGG	NA	NA
Exon23t	CTGAGTGAGGCCAGTTCCTC	TGCTCAAGCAAGTACAGCAAA	NA	NA
Exon23u	AGTTGTGAAGCCGTGCAGT	TTTGGCCTTAGCCCTAACCT	NA	NA
Exon24	CATGCCTGTAATCCCAGCTA	AACGTGTAGGAGCCAGAACG	NA	NA

Table 10.2.C

CYB561

CYB561				
Amplimer ID	**PCR 1**	**PCR 2**	**Extension primer**	**SNP ID**
Exon1a	CTCTTGCAGCTCCACCATC	AGCCGGAATGCTACAGAGAG	NA	NA
Exon1b	TCACTTTGGCAAACAGACCA	TTGTAATGGCTGCCCTTTTC	NA	NA
Exon1c	AGGCAGGAGACACAGCAAAA	GCCAGGAGATCAAGACCATC	AGAGTCGTCAGGTTTGGAAA	rs3087776
Exon1d/Reg2a	CTCAGCCTCCCAAGTAGCTG	GGCTCCCTCCTTGTAGCAC	NA	NA
Exon1e/Reg2b	ATAAGGTTTGTGCCCTCTGC	AGGGACAGGCCTGGAAAG	NA	NA
Exon1f/Reg2c	TATTCAGGCTGGAGGGACTC	CATGGACTTCAAGACGCTGA	NA	NA
Exon1g/Reg2d	AGGGGCAGGCAAGAAGAC	AGCAGGATCTGCAGGACAAC	NA	NA
Exon1h/Reg2e	CCAGCTATGGCTCACAGACA	GTGGGCTTCAGCTTCTTCCT	NA	NA
Exon2/Reg2f	ACGGAAAGGAGGAAGATGGT	GAGCTGGGCTGAAATGTACC	NA	NA
Exon3/Reg2g	CAGCTCTCTCACCTGCACAC	CTGGGACTAAAGCCCAGAAG	NA	NA
Exon4	AACCAGCTAGCGGGACTTCT	CAGCTCACATCACCTCTCCA	NA	NA
Exon5a	GTGACAAGACCCCTGCAAAC	GGGGCTGTGCATATCTGG	NA	NA
Exon5b/6a	CGAGATGTACACACCCCTCA	CCTCACCGCATATGGAACTC	GAGCAGCCCCCCTGAGATGC	rs4968773
Exon5b/6a	CGAGATGTACACACCCCTCA	CCTCACCGCATATGGAACTC	CTAGGGGCTGTGCGCATCCGT	Exon5b#2
Exon5b/6a	CGAGATGTACACACCCCTCA	CCTCACCGCATATGGAACTC	CTGAGATGCGGCGCGGCTCC	rs4968646
Exon6b	GGCAGAGGTCTCAAGTCTCG	ATGACGCGTCTGGAGAAAAG	NA	NA
Exon7/8a	TCCCACTCGGGAAACTCC	CAGGTGAGCCCTGACTGC	NA	NA
Exon8b	CGAGGCCGGCTGCTACT	CCGCGTACGCCACTTCC	NA	NA

Table 10.2.D

GNA13

GNA13				
Amplimer ID	**PCR 1**	**PCR 2**	**Extension primer**	**SNP ID**
Ex1b	AGGCATGATTTCACGGATTC	TCAAATTGAGGCTGTTTAGGG	NA	NA
Ex1c/Reg1	CATTTCTGGTATTAAGGCGATCTT	TCACACAAGTGTCTTCATTGACC	NA	NA
Ex1d	AAAACATTTGACAAAACCCTTTC	TTCAACTTAGGCTTGTTTTGACA	NA	NA
Ex1e	AAGGATTCAGTTTTCATGATACAGG	TCGGAATATGTGCACAGTAACA	NA	NA
Ex1f	CTTGAGCCCTCACTGGAGTC	TATGGATCACCCAGCCTTTC	NA	NA
Ex1g	GAAAGGCTGGGTGATCCATA	TCCTGTTCATTCTTGGTGTGTC	NA	NA
Ex1h	AGCTGGACTTTCAACAACAACA	GGAGAGCAAAAGGCAAGCTA	NA	NA
Ex1i	TTCACTCGTTACGTTTGACCA	TGCCAAGTATCAAAGGTGTCC	NA	NA
Ex1j	TGCTATATGTAAGGCGGACAGA	TCTGCATGACAACCTCAAGC	NA	NA
Ex1k	CAAACAGAAAACATCAAAAACACA	CGTTGGTTTGAATGTTTCGAC	NA	NA
Ex1l	TCCATAAGCACCTGGTCAAA	ATAGGCATGACCCACCATGT	NA	NA
Ex2	GCTACCAATCACTCTAAATAAAGCA	GGTTAGCTGTCTGAAATTGGTG	NA	NA
Reg2	CCCAACCAAACTCTCTCCAG	AGCGATTCTTATGCCTCACC	NA	NA
Reg3	GGATTTGAAAAGTGCAAAGCA	AAAGCATAATGGCCAAAGGA	NA	NA
Reg4	ACCCAATCACAGCATTCACA	GTGTGATCTTGGCCCACTG	NA	NA
Ex3	CCTTTCAGTCTGTTCCTCCATC	TCGATGTTGGTCTTCATTTTTG	NA	NA
Reg5	GCTAGAGCTGGCTGTGGTCT	GGGCATTCAAAGAAATCGAA	NA	NA
Ex4b	GCTTCAGGAAGGTGGACTTG	CGAGTCAGTTCGCTGGTTC	NA	NA
Reg6c	CTGTGTGACAGAGTGAGACCCTA	CCAGGCTAGAGTGCAATGG	CCTCTTTGTGTCTTGTTTTTTAGA	Reg6C
Reg7a	GTAGACCCGGCCTTTACCTC	AAAATTAGCAGGGCATGGTG	NA	NA
Reg7b	CACTGCTCACTGCAGCCTTA	CCTTTCCACAACCAGGAGAA	NA	NA

Table 10.2.E

RGS9

RGS9				
Amplimer ID	**PCR 1**	**PCR 2**	**Extension primer**	**SNP ID**
Promoter A	CGCTCAATTGTGCTGTGAAC	TGAAGAAGGAAGAGAGGAAGGTT	NA	NA
Promoter B	CCAACAACCAACTAACCGAAA	CAGCGAGCCTTTTAACTTGG	NA	NA
Exon1a	GGGATGGGAAAGTCCGTATT	CCAGTGGGAGACAAGAGGAG	GGAGGAGGAGGAGGACCTAG	rs402137
Exon1b	GATTCCAGCTGCTTTCCAAG	TCGACACACACAGCAAAGAA	NA	NA
Exon2	GAGTCACTCAGTGCCCACCT	CTCCCAAAGTGCTGGGATTA	NA	NA
Reg1a	CACATGGGGTTATGGAGACC	GTGTGCAGGGAAGGAAAGC	NA	NA
Reg1b	AAAGTGCCGGGATTACAGG	CTGCCGTCCAAGAAGTTAAA	NA	NA
Exon3a	GGGATGTGTCAGGAGGCTAT	AGCATCACTTCTTTCAAGACCA	NA	NA
Exon3b	ACAGGGATAAGTGGGGAGGA	TCAGGAACAGGCCATGTAAA	NA	NA
Exon4	CACCTGTCCTGGAGTTTGG	CCACCATTTCTGAAGTCCTTG	NA	NA
Exon5	GGCTGTTCATATGGGGTTGA	CATGCCCAGTACAGATGTGG	NA	NA
Exon6	CCACATCTGTACTGGGCATGT	TGATGTGGGAATGAGGAAAG	NA	NA
Reg2a	TCTCTGATGCTCTCTTCCTAATCA	TTTTATTGGTTTTCTTTGCTTTTG	NA	NA
Exon8	GAATTACTTGGGCAGGTTGG	GGAGTTGGCAACAGGTATGG	NA	NA
Exon9	AGGCCACAAGCTCTTCTCAG	CGTTGCCCTAACAGACACG	NA	NA
Reg3	GCTATGTGCCAGGGACCTTA	CTTGAGCCCAGGAGTTTGAG	NA	NA
Exon10a	TGCTCTGTTACCCAGGCTAGA	CCCCTGGGGAATAAATACCA	ATCATTCTGTAGGCACTATGCTAAA	rs11658673
Exon10a	TGCTCTGTTACCCAGGCTAGA	CCCCTGGGGAATAAATACCA	GCCAAGATCAATTAAGGACTAAATTTG	rs11654243
Exon10a	TGCTCTGTTACCCAGGCTAGA	CCCCTGGGGAATAAATACCA	TTTAGGTGAGCAAAGTTACAGTGGTAA	Exon10a#3
Exon10a	TGCTCTGTTACCCAGGCTAGA	CCCCTGGGGAATAAATACCA	CTCCTTCGCTTACAAAATCTCTT	rs11658773
Exon10b	CTTGGCCTTTAGGTGAGCAA	GATGCTTCGTCACTCTGTGC	NA	NA
Exon11	TCTGGGTAAATGGCTTCTGG	CCCCAAGAGACTGGTTTTGT	NA	NA
Reg4	TGTCCATTGTTCATCTTTGATTT	TCCCAGGAATAAGCCAAGTTT	NA	NA
Exon12	GCCAATCCCAAGGGTTATTT	AAACCATTGAACTGTATGCTTTGA	NA	NA
Exon13	TTGTCTGTGGCTGCATCTTC	TGCTGAGAAATTCCCCAAAC	NA	NA
Exon14	AATCCATCCCGGTTGACTC	ACAATGGGGTGGTCCTCAG	NA	NA
Exon15	AAAAACCACCACCCTCACC	CACAACCGGCAAAGCTCTAT	GGCTTTCTGTCACTAAGTACCC	Exon15
Exon16	CCTTTCACCCAAGGAGACAA	GCACTCATTCCCTGAGCAAT	NA	NA
Exon17	TGGAAAAAGGAGGGAACAAA	AATGGGGAATGGAACCTTCT	NA	NA
Reg5a	GTGGTGAGTGCTATGGTGGA	TATGCAAGCTGTGTGCTGCT	NA	NA
Reg5b	TGCTCCCTCCAAGTAGGTGT	GAAGGGGCATCTAAATCAGC	CCCTCCAAGTAGGTGTCCTTATAACA	Reg5b
Reg6	TCTGCTTCCAAGCTCATTTG	ACAGGGGCTCATTACCCTTT	NA	NA
Exon18a	CCTAGGAGGCAGCCATATCA	AACCTGCTGAAGGACAGAGC	CCCCTTCTCCTCCTCCTGCC	Exon18a
Exon18b	CTTGGAGCAGAAAGGGGAGT	CTGTTCTCACCCATCCCAGT	NA	NA
Exon19a	AGTCTTTGGCAAATGCACCT	GGTCCACATGCGTTTCTTCT	NA	NA
Exon19b	CAAAGAATGCTCTGGCTGGT	GCGACTCAATGTCCCTTTGT	GACTTTCCCTGCTGCCTTAA	Exon19b

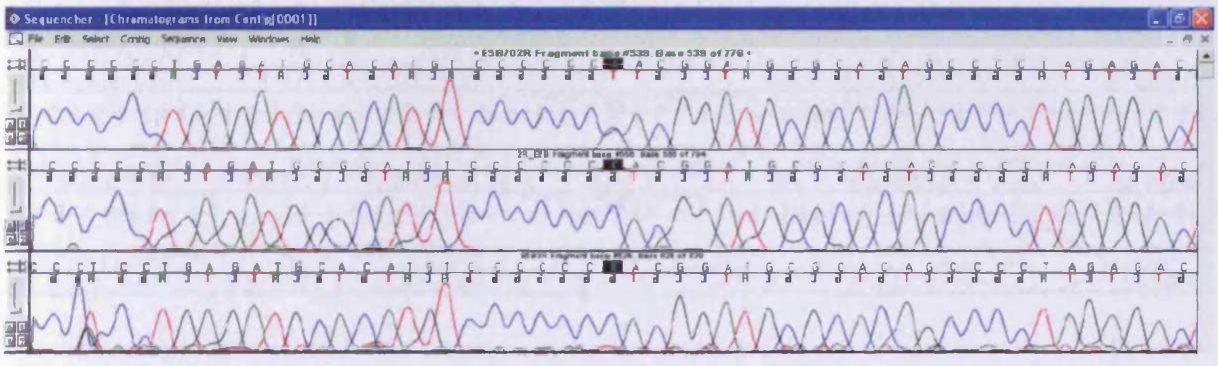
Figures 10.2.A

Sequencing traces for novel SNPs

Note: A C702 sibling is shown in the top sequence trace, unrelated control individuals are shown below. The SNP position is in the centre of the trace (base coloured black).

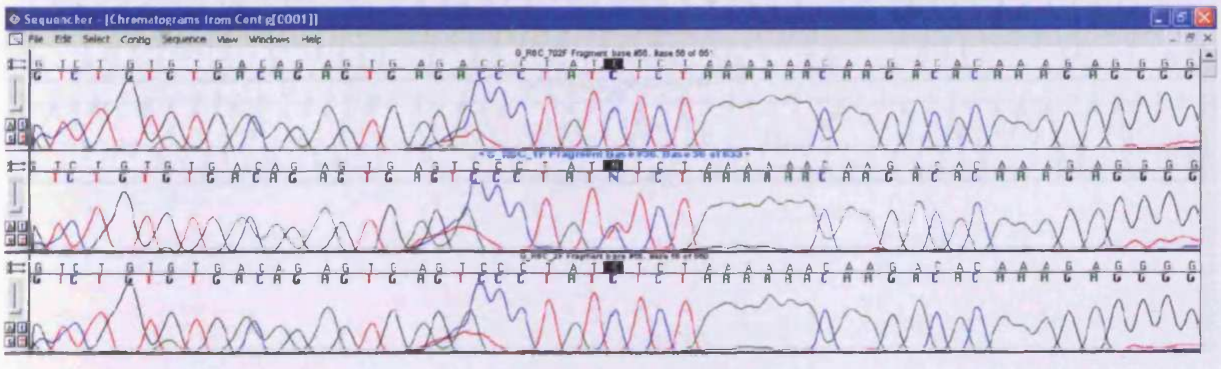
CYB561

SNP: Exon5b#2



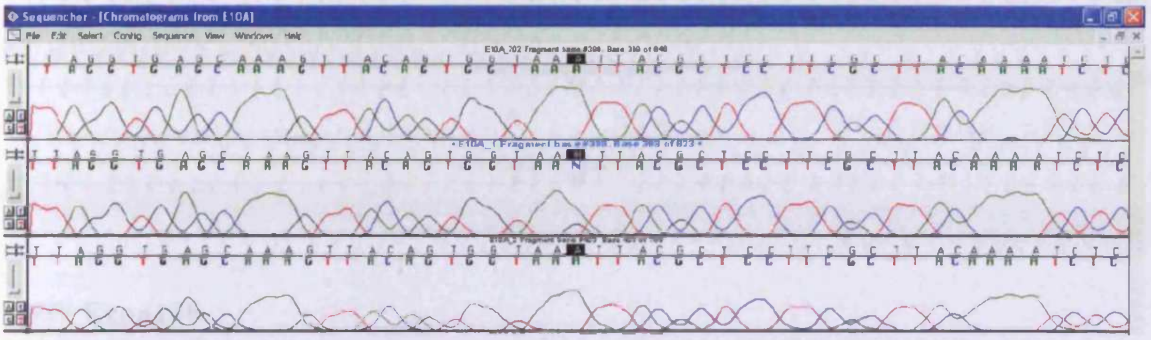
GNA13

SNP:Reg6C

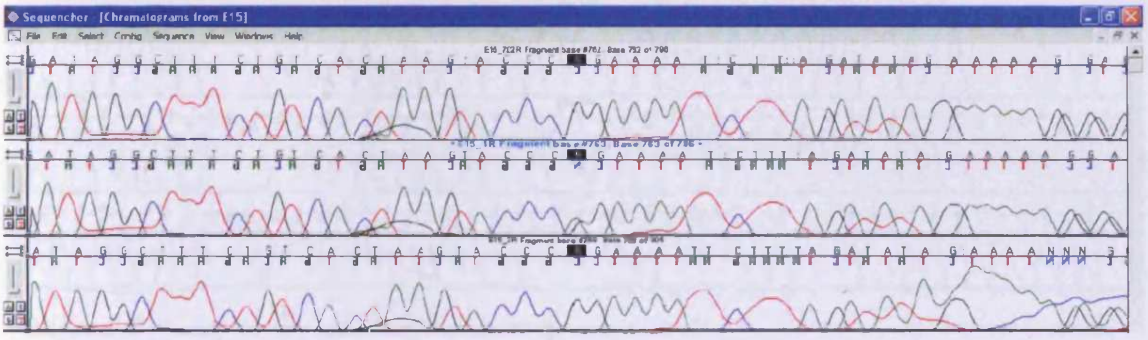


RGS9

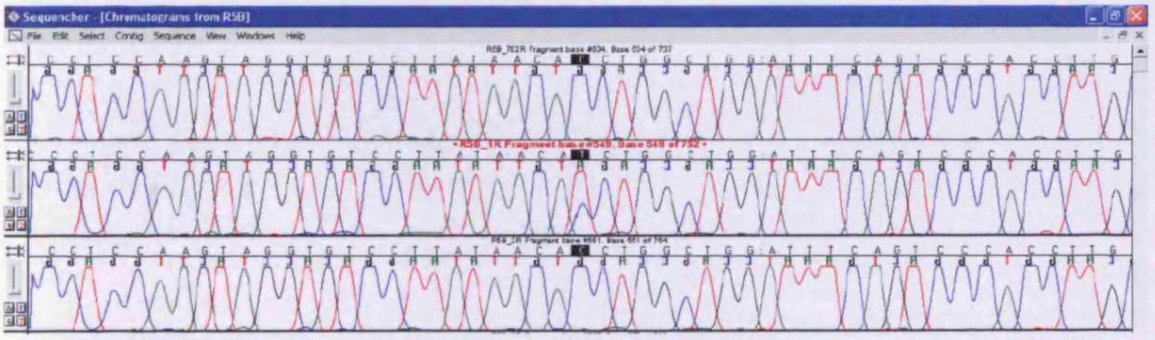
SNP: Exon10a#3



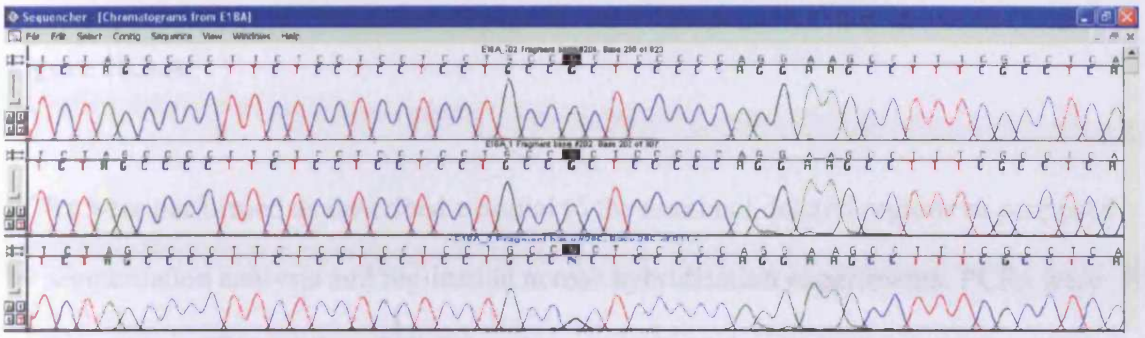
SNP: Exon15



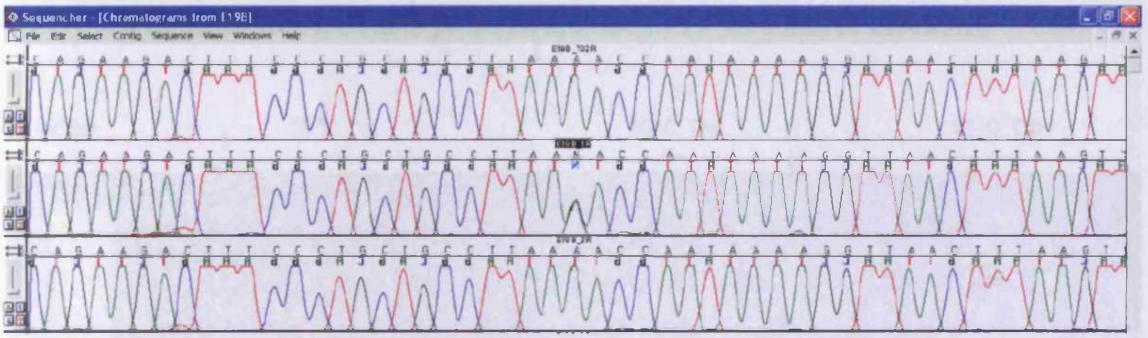
SNP: Reg5b



SNP: Exon18a



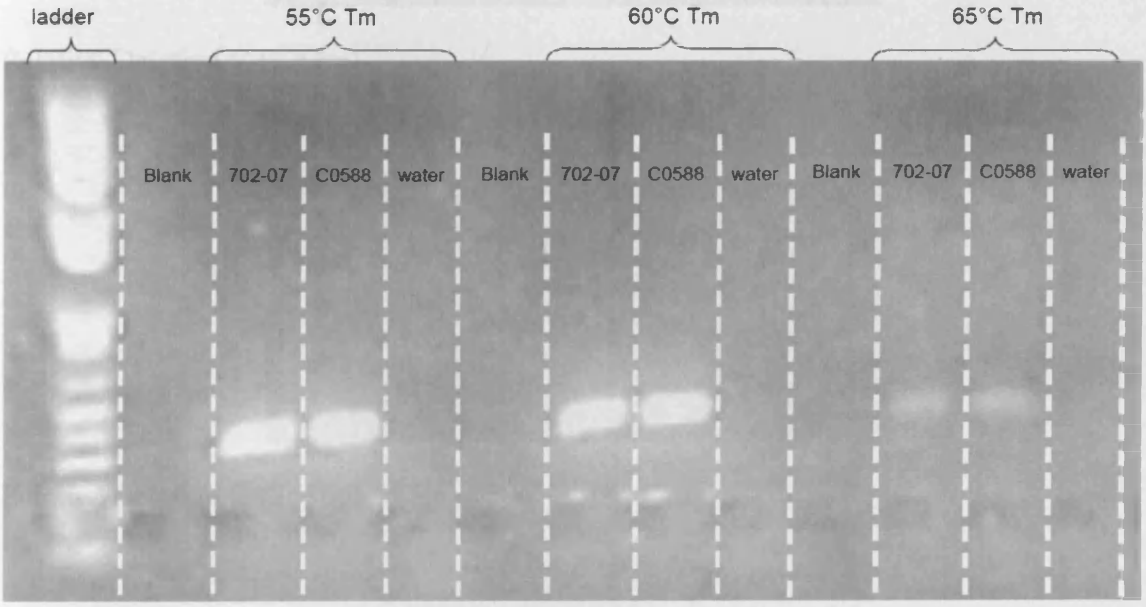
SNP: Exon19b



10.3. Appendices Chapter 5

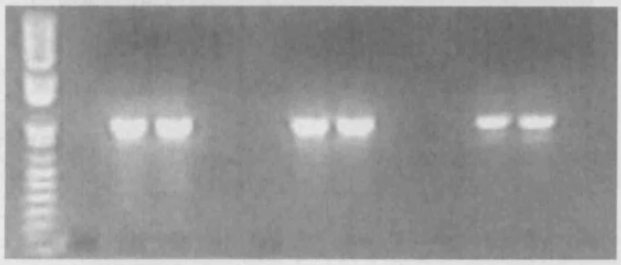
Figure 10.3.A

PCRs were performed as described external to the maximal deleted regions as proposed by segmentation analysis and replication across hybridisation experiments. PCRs were performed in a C702 sibling (702-07) and control (C0588) used in the oaCGH analysis over a range of annealing temperatures using standard PCR conditions (55°C, 60°C and 65°C). All figures show the same protocol applied to all 23 putative deletions.



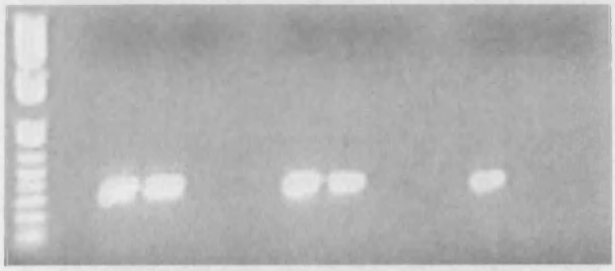
Putative Deletion A (292bp). Note, all figures below were set-up identically.

Putative Deletion B (496bp)

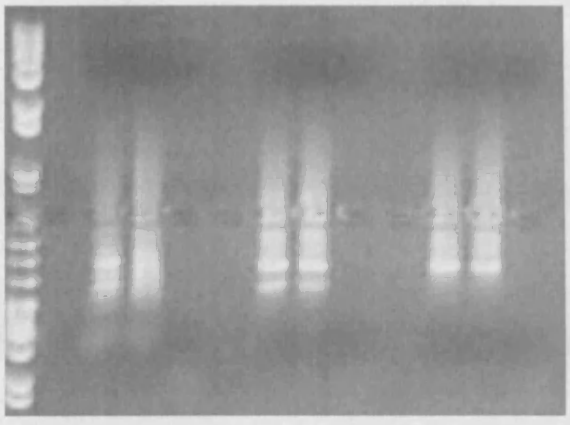


Putative Deletion C (979bp)

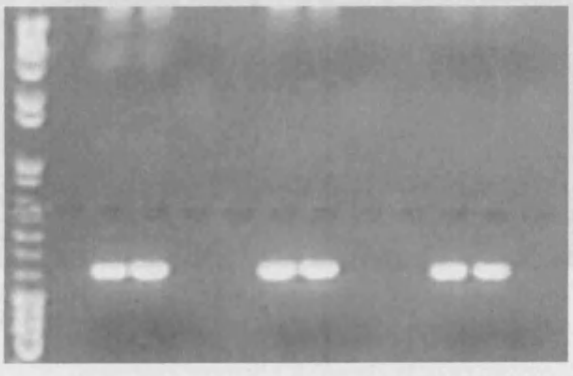
Putative Deletion D (363bp)



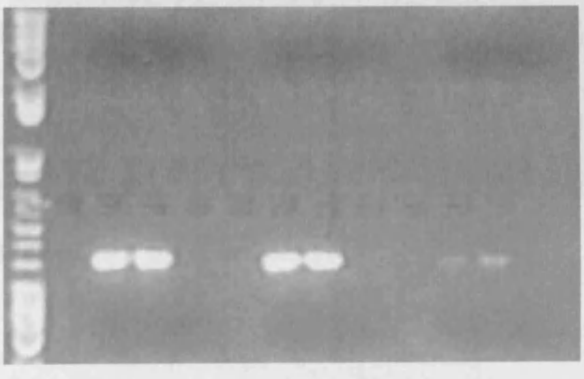
Putative Deletion E (298bp)



Putative Deletion F (375bp)

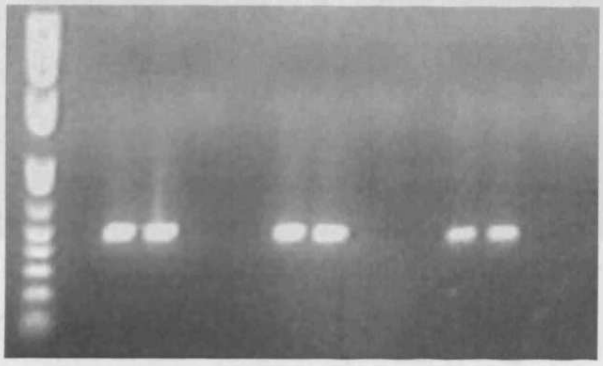


Putative Deletion G (292bp)



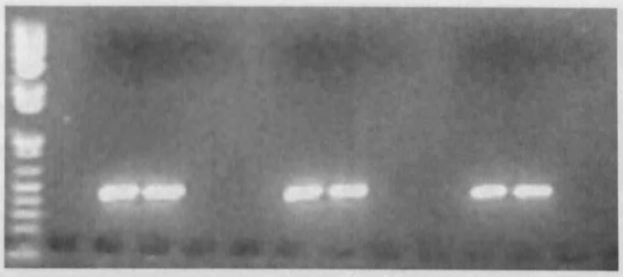
Putative Deletion H (300bp)

Putative Deletion I (472bp)



Putative Deletion J (476bp)

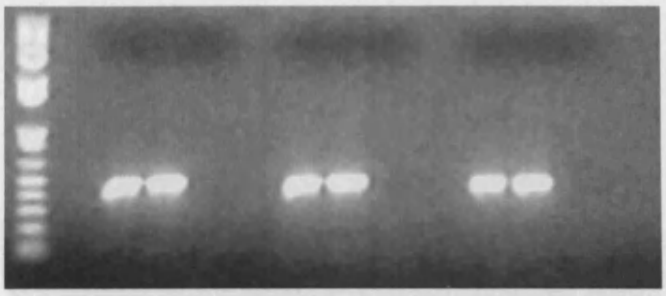
Putative Deletion K (374bp)



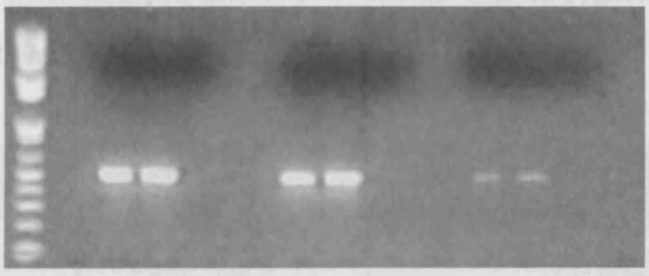
Putative Deletion L (400bp)

Putative Deletion M (498bp)

Putative Deletion N (300bp)

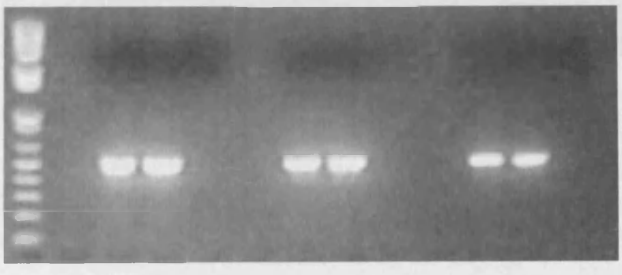


Putative Deletion O (300bp)

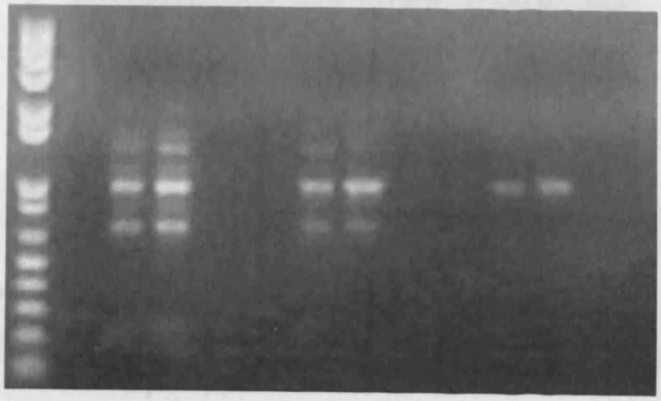


Putative Deletion P (471bp)

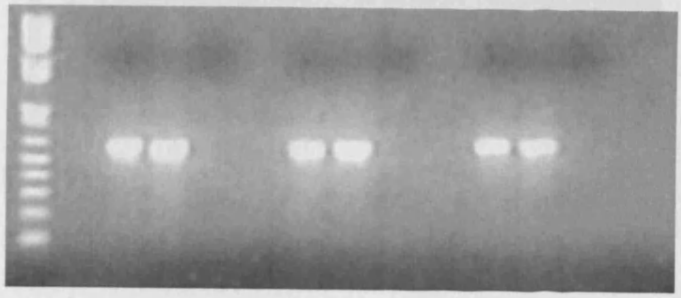
Putative Deletion Q (299bp)



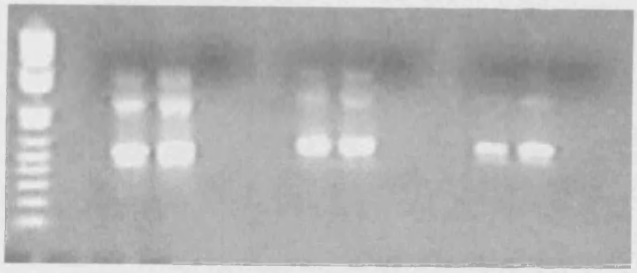
Putative Deletion R (500bp)



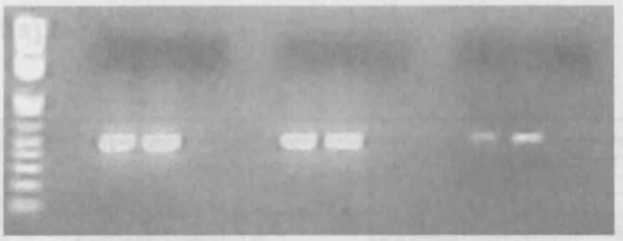
Putative Deletion S (966bp)



Putative Deletion T (595bp)



Putative Deletion U (493bp)



Putative Deletion V (462bp)

Putative Deletion W (391bp)

10.4. Appendices Chapter 6

Amplimer ID	PCR F	PCR R
P1	GTGCTGAGGGGTTAGGAGGT	CGCCTCAGTTCTCCTCCTCT
P2	AGACCGGGAGAGGAGGAG	GGGAACACTGGGCAAGG
P3	GGAAGGGGAGGAAAGAAAAA	CCCAGGCTCACTGACAGC
P4	CTGCCTCCTCCGCACTC	GGAGAGTCGGGCTGGTG
P5	CACCGGGCTGTCAGTGAG	CCCCAACCACCTCTTGC
P6	CACCAGCCCGACTCTCC	GCGATGAATTTGTGGTCCTT
R1/PA	GCTCTGGTGGCCATTGAATA	CACTCCCGCTCTCTCTCTTG
PB	GTGCTGAGGGGTTAGGAGGT	CCCAGGCTCACTGACAGC
PC/E1a	CTCCCCTTGCCCAGTGTT	GCGATGAATTTGTGGTCCTT
E1b	GACTCCACGGCGTCTCAG	ACTCCGGGCGAAGAGTTC
E2	CCCCCAAGGTTGAATTTACA	TGCAGATTAGTAGGCCCCAAT
R2	GCCCAAGCCGATGTTTATTA	GGGGTGAGGATCCCACTATT
R3	GACCATGGGAATTCTGGTGA	TGCAGCTTCTCCACATCTCA
R4	TGCCAAAGCTTTAGACCATAAAC	AGCAAGAACCGCTAGAGGAA
R5	TGAATCTGGGCTTTGAGGAG	CGACCTCAAGATCAGGAAACA
R6	TTTTCCTGCCTGCTTAGGAA	CCCCTCTGGATAATCTGTGG
R7	GTACTTCCACGGGAGGATCA	GAAAATGCTTGGGGATGAAA
R8	CCACCACACCCAGCTAATTT	GCAGTTCTTAAACGACACAACG
R9	ACTTTGTGCAGCCCTACCAT	TCAACTTTTCTGCCCATGCT
E3	TGAAATGCCTTCTGATTGCT	CTGCCCAAAGTGGAACAGTA
E4	CACCCATTGTGTGACTCCAG	TTACTGAGCCTTCCTTGCTCA
E5	ATCGAGTTGGGGGTAACTCA	TGAGACAGTGATGGCTCCAC
E6	TTAGAGGCTTTTGGCTTCCA	GCAGCCTGAGGAGGTCTTAG
E7	CCAGCATAAGGGGTTGGTATT	TGCTGGGATTACAAGCATGA
E8	CTCGACTAGAGGCTGCAGGA	AAAATGTGACCACAACATCTCG
E9	TGCCCCTTTTATATCCCAGTT	TGGTAGCACCTCATTTGCAC
E10a	AGCCACCACAACACTTAGGC	TATAGGCATGAGCCACCACA
E10b	CTCCTGAAGGCTCTGCCTAA	GCCAGTCACACAAAAATACAGC
E11	TCCTTTTCTCACCCCATCC	GCCAGGATGGTCTTGATCTC
E12	GCTGGAGGGACTGAATTTGT	CCAGTCACCTCTGCTTGACA
E13	CACATCTGCCCATGCAGTC	TGGTGGTGTGTGCCTGTAGT
E14	GAGAGCCTCTGGGTCTGACA	TCCTCCTGGGGAGTTATCAA
E15	GGGACTTATAGTTAATATTGCCATTGA	TTTAGTGGCTGTGACCATGC
E16	TCACGGGGTGATAGTTCTCC	GAAGTCACGCCAACACCAG
E17	AACCGTCAAGTTAGCTGAGCA	AAATGCAGTGGTGTGGGTTT
E18	GCTCTCTGAGATGCTGTCCA	CCCGTTTCCTCACCCTGTA
E19a	CCTAACCAGCCCGGAGTT	GCATGACGTACCATCTTCCA
E19b	GCAGTGGGAAGTGAATCCTTA	CAACAGTCTGCTTGCTTTCTAGG
E19c	TCTGGCTCTAGGTTAACCCTTC	TCAGCAACGTCCTTAGACGA
E19d	TCATCTGGTACCCTCCTTGG	CTCCAGGCCTGTCATTCAAC
E19e	TGGATGTCCACTTAGGGATAAAA	TGTTTGCAGGTAGAATCAAATACA
E19f	TCAAAAATGAATGGCTAGTTTACG	TCCGCAGCCTCTGAAAAG
E19g	TGTACTGGACACGCAGGAAG	GTCAGGACACGAACCCAGAT
E19h	CTGCCATCTTTTATGCACCA	CTGGGGGTTTGTTCTGATGT
E19i	CACTGTCTCAGCTGGCTGTT	GATTCGGACGGTGCTAGATT
E19j	GATGCTTTGGGAAGTCCTTG	GCCTAATACGTGGCATCAGA
E19k	TCCCTTCCTTTCATCCACTG	TCCCTGGCAAAACACTCTCT
E19k_b	GGGTGCATCCAGAAATATGC	CAGGAACCTACTGGCCTCTG
E19l	GGTTGGTGCAGATGGAAGTT	ATCACCTTTCCGAGGCTTTT
E19l_b	AGTTCCCACCTGGGTTCC	CAGGAAATGGTGTGATGTGC
E19m	TTTGCTGGAATTTGTTTTCTCA	AGGAGAATCGCTTGAACCTG
E19m_b	ACCAGACAATCGTAATCACAAA	TTGCAGTGAGCCGAGATT
E19o	GGAGTGCGGTGGTACAATCT	TTGTGGTTGACCCCATTTCT
E19p	GTTACCTGACGCACACATGG	ATTCCCACGTGGACTCCATA
E19q	CAATGCCCTGGAGCTTGTAG	CGTTTGCTGCTGCCTCTTAT
E19r	CCACTGGGCAACTCAGAAAT	TGTCCTTCACACACACAAATCA
E19s	TCCCTTCTGTTAGGGTCTTTCA	TGAAAGTGTTCTGCGGAGTG
E20	TGTGAACTTGTCCCAGTGGA	CAACTGGCTGATCCATTCCT
E21a_2a	TTTAGTGCAGCTGGTGTTGG	ACAAACTTGCAGGGAACCAT
E21a_2b	CGGTGGCTCATGGCTATAAT	TGCCAAGGGTGAGATAGGAG
E21c_3	CTGGCCCCTAAGACATCCTT	CCAGTGGAACCTAGGCAGAA
E21d	TGACCCACAGGTTCATTCAA	GGACGTAACATGTGGGAAGG
E21e	TGACTCTCCATAGCCAGAAGC	TTAAACGTGGCTGCACATTC
E21f	TTGGCCAGAGAGGATAAGGTT	GGTGCCTGTAATCCCAGCTA
E21g	GATGGAGTTTTGCTCTGTTGC	CAGACAGCCCAACCCAAG
E21h	TTGGGGTTTTATGCATTGGT	TCTGGAGATGACGTTCATGG
E21i	TGACATTCCTGGTGGGTAGG	GGACTCTTAGGGGGCCTGTA
E21j	GCAGTTCGTACACAGTCACCA	CCAAAGGGGACTTCTCTGC
E21k	CCTTCCTGGGATCTCATTCA	CCACCACACTCGGCTAATTT
E21l	CACACCTGTAATCCCAGCAC	TCTCGAACTCCTGGACTCAA
E21l_2	GCATACTTTCAGGTCTCCATCA	CCACTGCACCTGGTCTAAGG

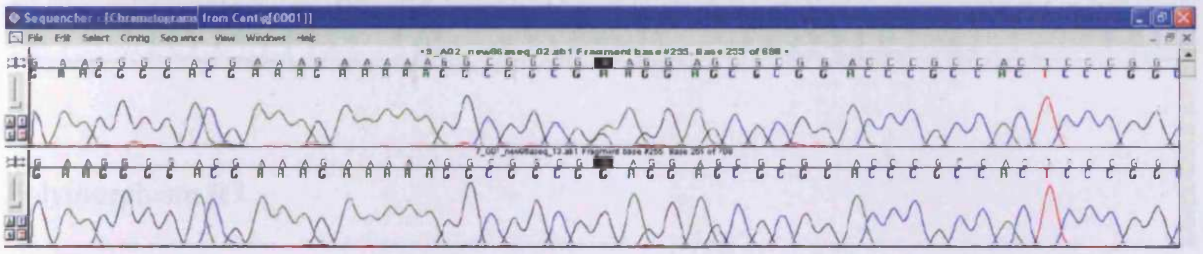
Table 10.4.A: Primer pairs used for PCR and sequencing of *PRKCA* exons and regulatory regions.

Figure 10.4.A

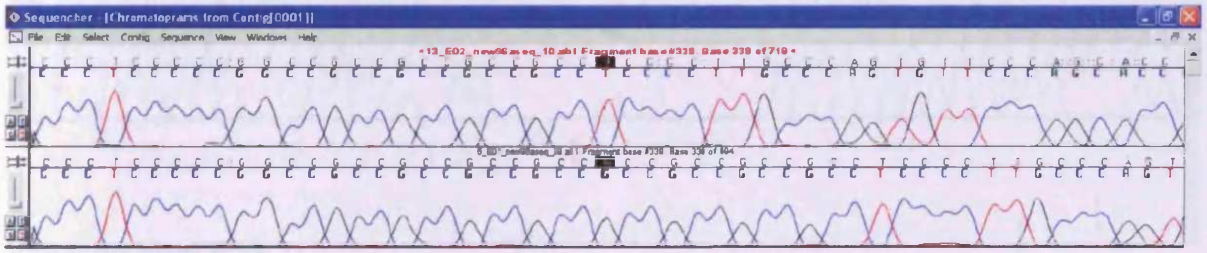
Sequencing traces (Sequencher) for novel SNPs

Note, the SNP or first polymorphic nucleotide is highlighted in black (in the centre of the image). Shown are >1 sample trace and therefore >1 allele.

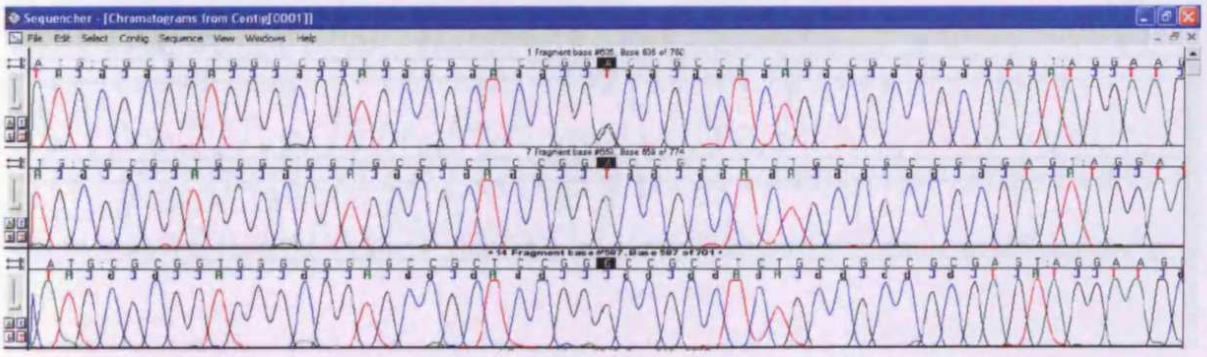
Polymorphism P1



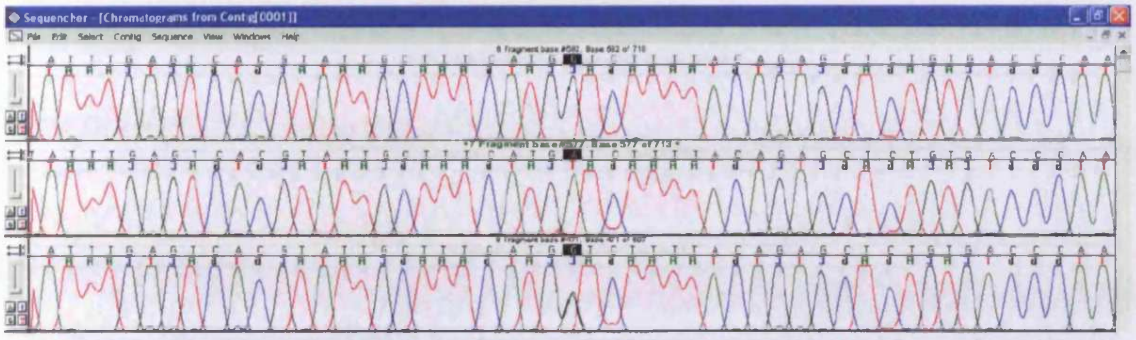
Polymorphism P[GCC]



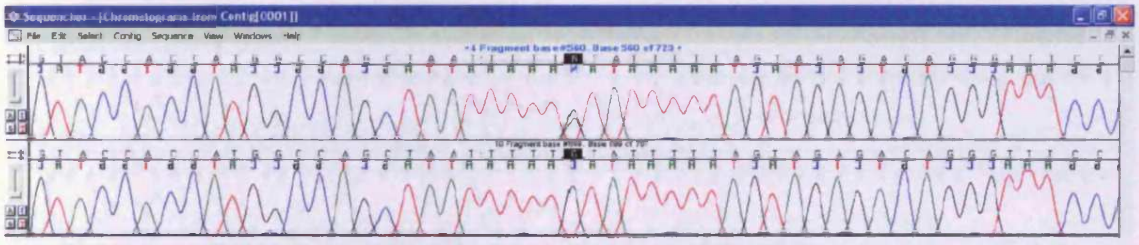
Polymorphism P2



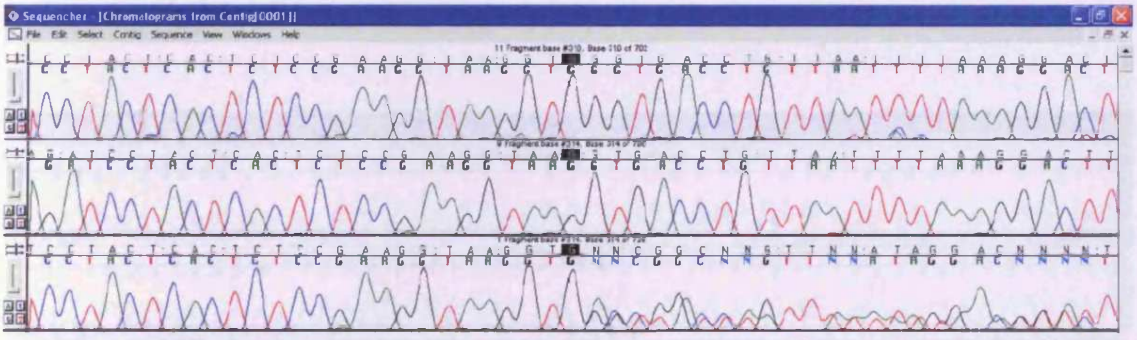
Polymorphism R2A



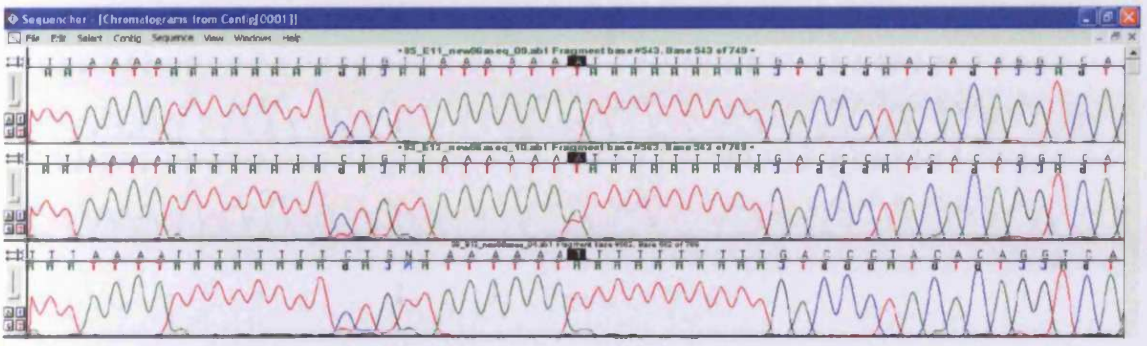
Polymorphism R3



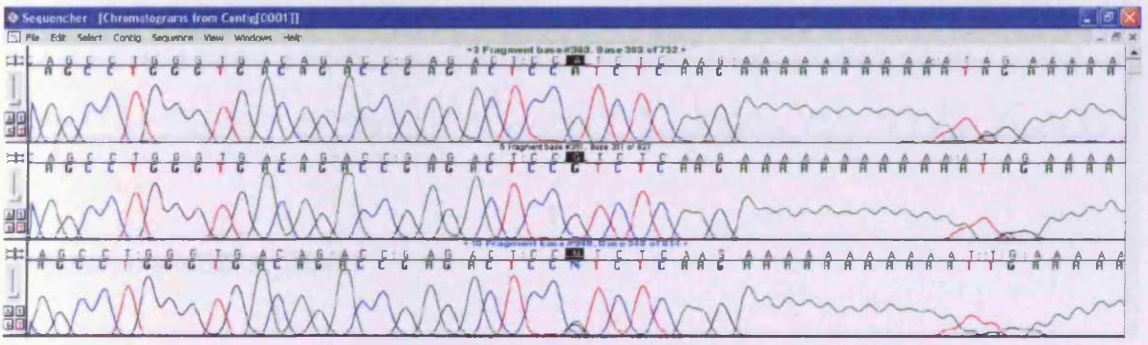
Polymorphism E20



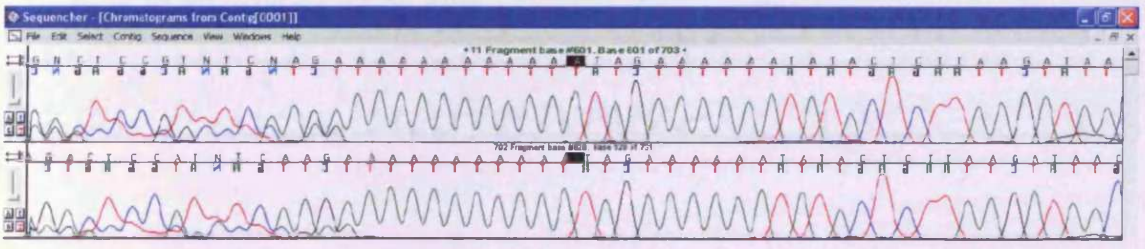
Polymorphism E10A



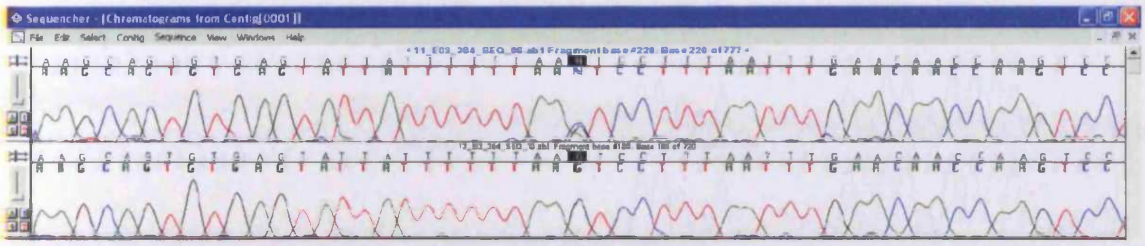
Polymorphism E10B(1)



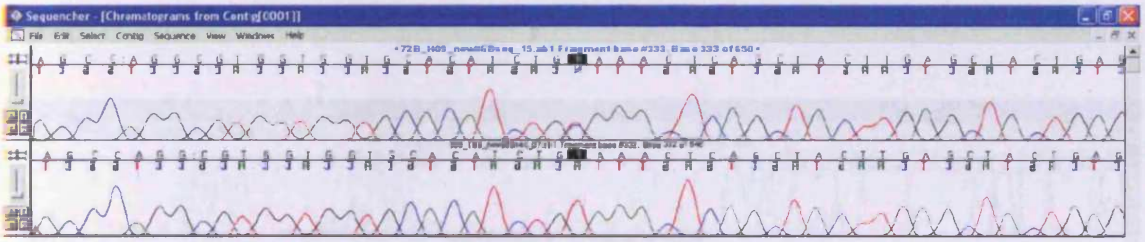
Polymorphism E10B(2)



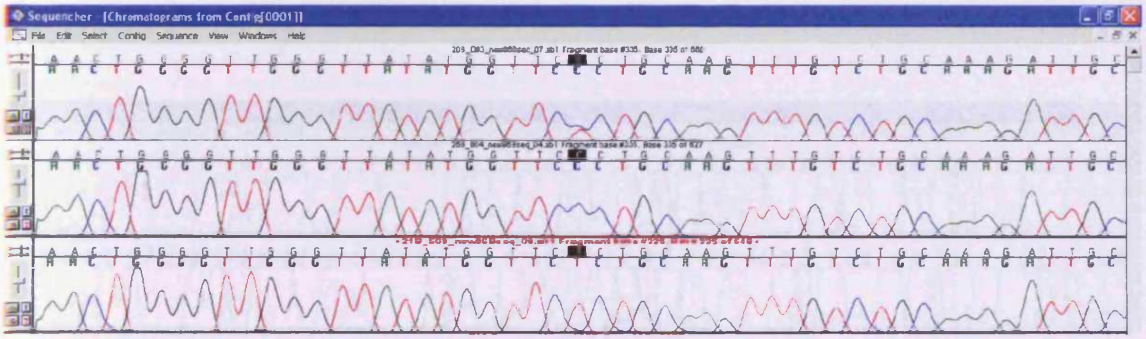
Polymorphism E13



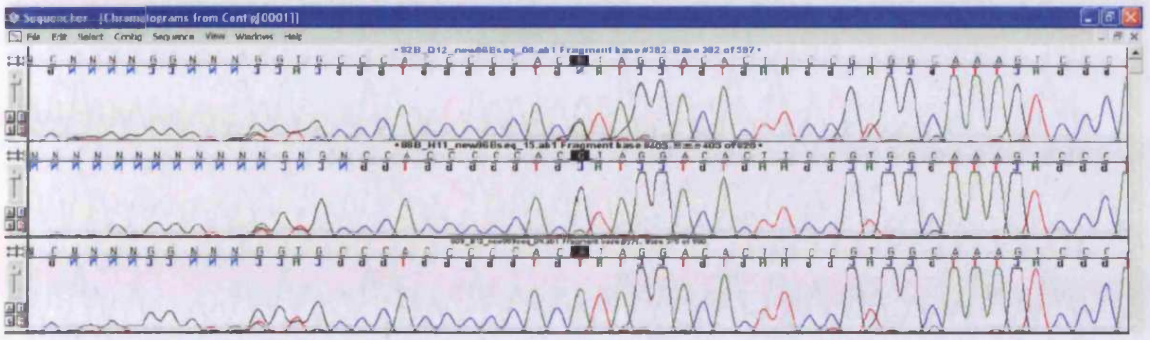
Polymorphism E21A_2B(1)



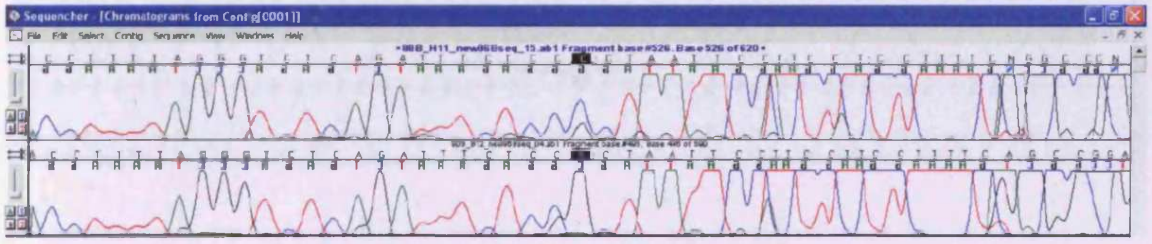
Polymorphism E21A_2B(2)



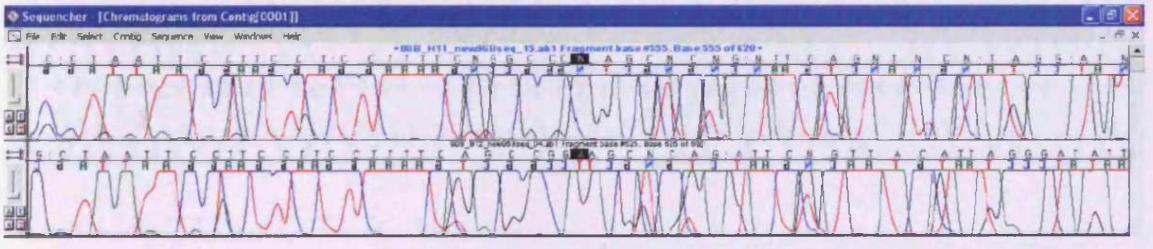
Polymorphism E21C_3(1)



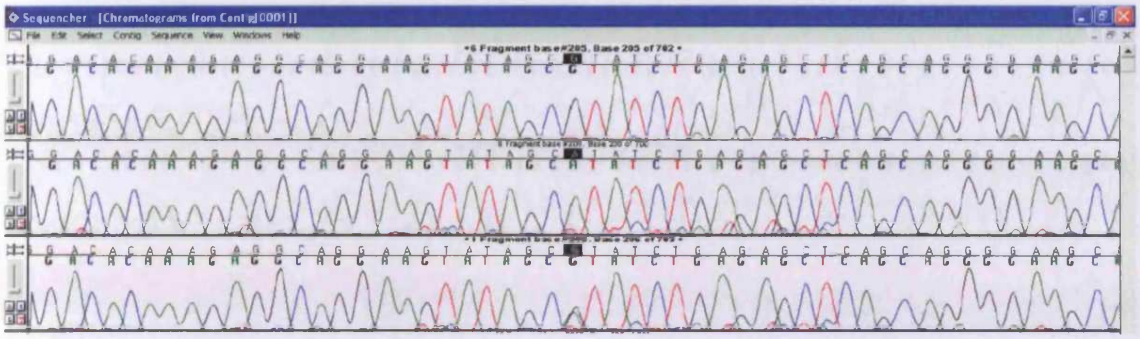
Polymorphism E21C_3(2)



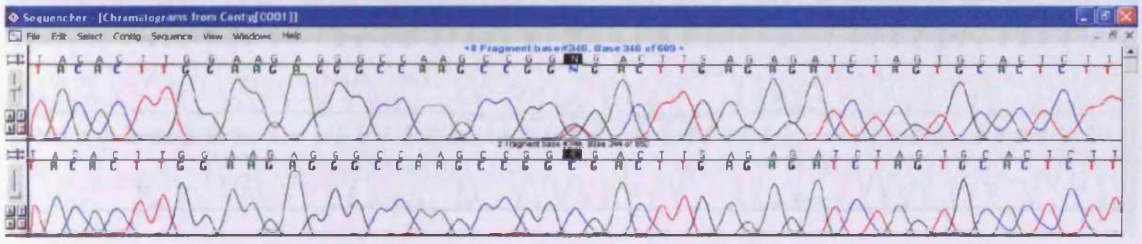
Polymorphism E21C_3(3)



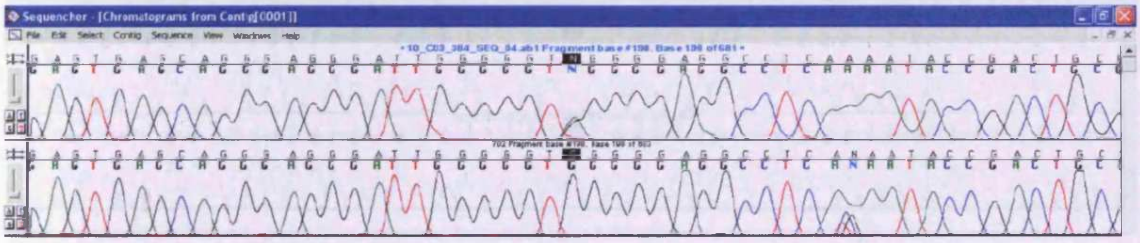
Polymorphism E21D



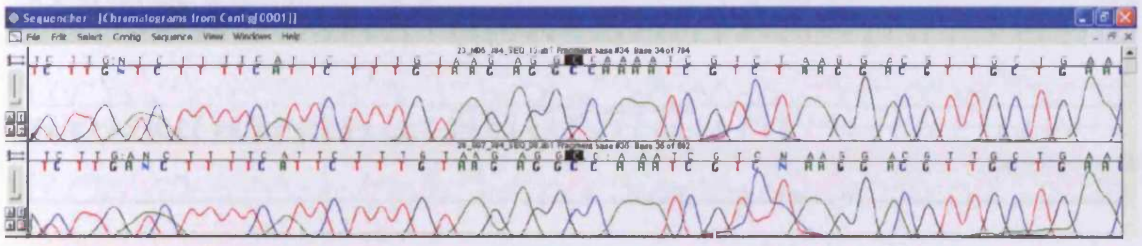
Polymorphism E21G



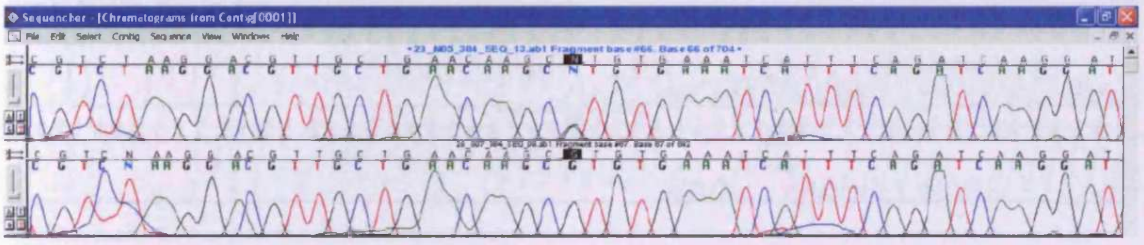
Polymorphism E19C(2)



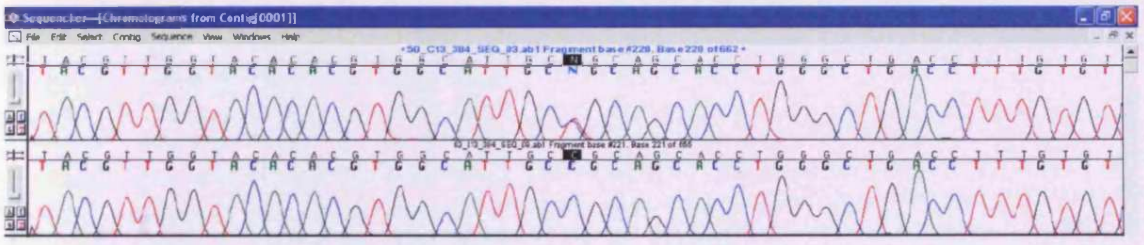
Polymorphism E19D(1)



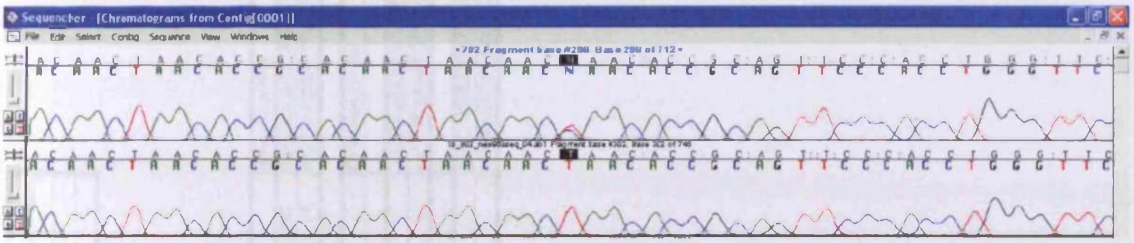
Polymorphism E19D(2)



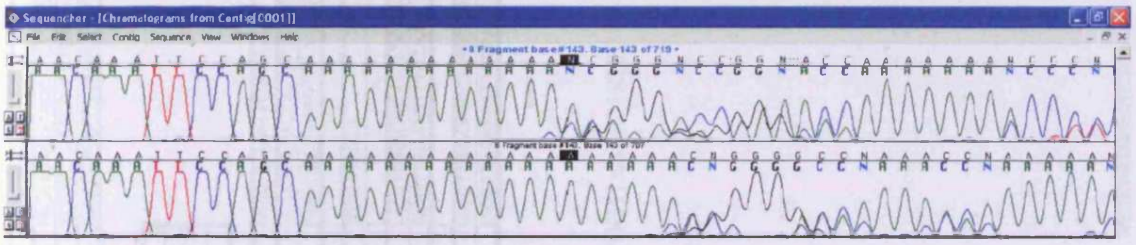
Polymorphism E19F



Polymorphism E19L



Polymorphism E19L_b



PCR/Sequencing assays	PCR primer 1	PCR primer 2
E19A_1	CCTAACCAGCCCGGAGTT	GCATGACGTACCATCTTCCA
E19A_3'HET	TTTGTGCACCCCATCTTACA	CTCCCTGCTCACTCTCTTGG
E21K	CCTTCCTGGGATCTCATTCA	CCACCACACTCGGCTAATTT

SNaPshot® extension primers	
E19A_F	GAGAACAAACACCTCCCCAG
E19A_R	CACTGCGGGGAGGGCTGGGG
E21K[1]_F	TAAAAACGGTGATGGCTGGGC
E21K[2]_R	GGATTACAGGTGTGAGCCAC

Amplifluor® assays	Allele Specific PCR Primer 1	Allele Specific PCR Primer 2	Reverse primer
E21K[2]	GAAGGTCGGAGTCAACGGATTTGGGATTACAGGTGTGAGCCACT	GAAGGTGACCAAGTTCATGCTGGATTACAGGTGTGAGCCACC	AGAGCCACTTAGCCTGCCTTAT
E19A	GAAGGTGACCAAGTTCATGCTAGAACAAACACCTCCCCAGC	GAAGGTCGGAGTCAACGGATTGAGAACAAACACCTCCCCAGT	GGTTAAGGATTCACTTCCCACT

MassARRAY® iPlex® primers	PCR primer 1	PCR Primer 2	Extension primer
Multiplex 1			
E19A	ACGTTGGATGAAGGATTCACTTCCCACTGC	ACGTTGGATGGCAGTATGAAACTCACCAGC	AACAAACACCTCCCCAG
E21K[G]	ACGTTGGATGCAGGGAATGCTTTCTTCTGG	ACGTTGGATGTTTGCAAATGAGGTCTGAGG	ccCTTCAAAATGCACCCCCC
Multiplex 2			
E19A	ACGTTGGATGAAGGATTCACTTCCCACTGC	ACGTTGGATGGCAGTATGAAACTCACCAGC	gcACAAACACCTCCCCAG
E21K[G]	ACGTTGGATGCAGGGAATGCTTTCTTCTGG	ACGTTGGATGTTTGCAAATGAGGTCTGAGG	cTCAAAATGCACCCCCC

Table 10.4.B: Primers and assays used for genotyping SNPs E19A, E21K[G]. E21K[1] and E21K[2].

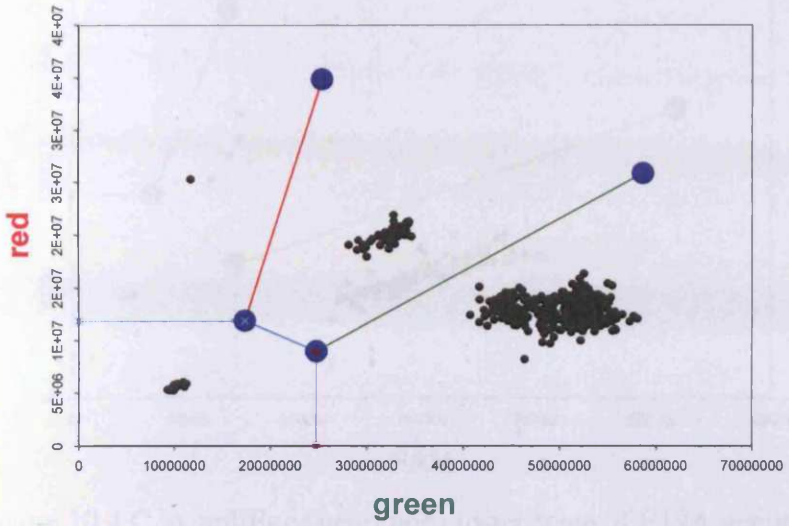


Figure 10.4.B. Amplifluor genotype cluster trace of E21K[2] genotyped in 384 samples. The y axis represents fluorescence of the red fluorophore corresponding to the rare A allele, while the x axis denotes the fluorescence intensity of the green fluorophore, corresponding to the common G allele. The red and green lines demarcate the 3 genotype clusters while the dropouts are shown clustering near the origin.

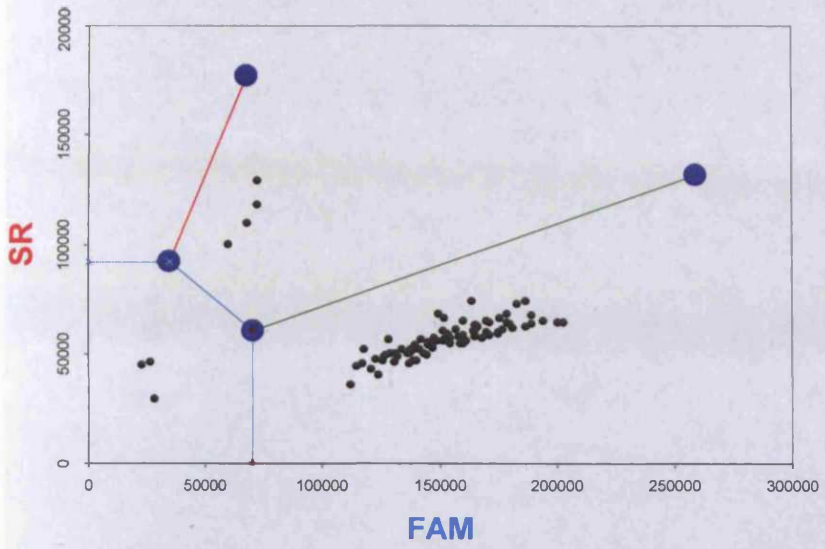


Figure 10.4.C. Amplifluor genotype cluster trace of E19A genotyped in 96 samples. The y axis represents fluorescence of the red fluorophore corresponding to the rare T allele, while the x axis denotes the fluorescence intensity of the green fluorophore, corresponding to the common C allele. The red and green lines demarcate the 3 genotype clusters while the dropouts are shown clustering near the origin. A positive control heterozygote was always included when genotyping the E19A allele.

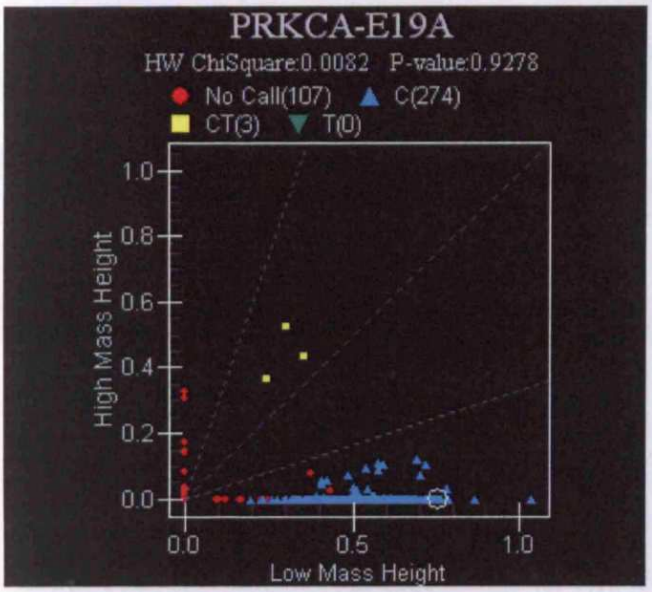


Figure 10.4.D. Sequenom MassARRAY iPlex genotyping of E19A (multiplex 1) in 384 samples. The y axis represents log peak allele height of the rare T allele, while the x axis denotes the log peak height of the common C allele. Yellow genotypes indicate heterozygous individuals, blue genotypes represent homozygotes for the common allele and red genotypes are no-calls. A positive control heterozygote was always included when genotyping the E19A allele.

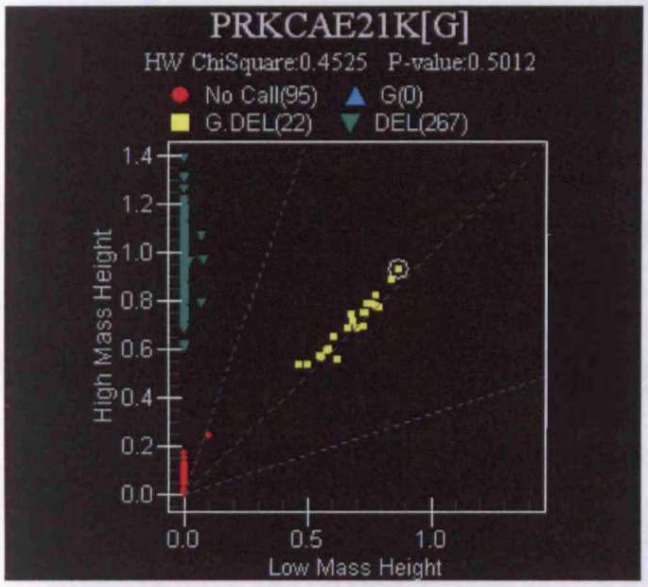


Figure 10.4.E. Sequenom MassARRAY iPlex genotyping of E21K[G] (multiplex 1) in 384 samples. The x axis represents log peak allele height of the rare inserted [G] allele, while the y axis denotes the log peak height of the common deleted allele. Yellow genotypes indicate heterozygous individuals, green genotypes represent homozygotes for the common allele and red genotypes are no-calls.

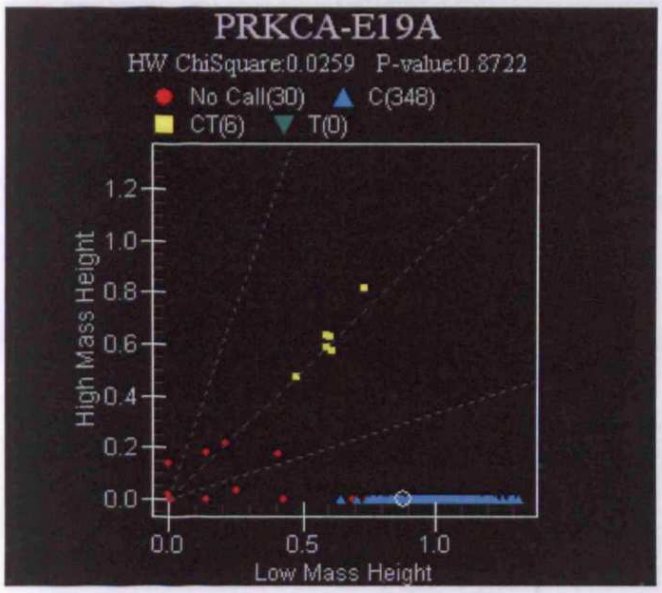


Figure 10.4.F. Sequenom MassARRAY iPlex genotyping of E19A (multiplex 2) in 384 samples. The y axis represents log peak allele height of the rare T allele, while the x axis denotes the log peak height of the common C allele. Yellow genotypes indicate heterozygous individuals, blue genotypes represent homozygotes for the common allele and red genotypes are no-calls. A positive control heterozygote was always included when genotyping the E19A allele.

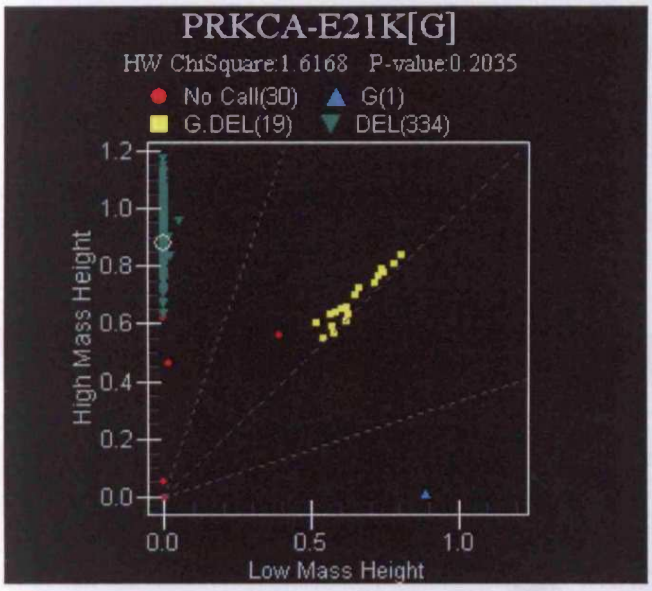


Figure 10.4.G. Sequenom MassARRAY iPlex genotyping of E21K[G] (multiplex 1) in 384 samples. The x axis represents log peak allele height of the rare inserted [G] allele, while the y axis denotes the log peak height of the common deleted allele. Yellow genotypes indicate heterozygous individuals, green genotypes represent homozygotes for the common allele and red genotypes are no-calls.

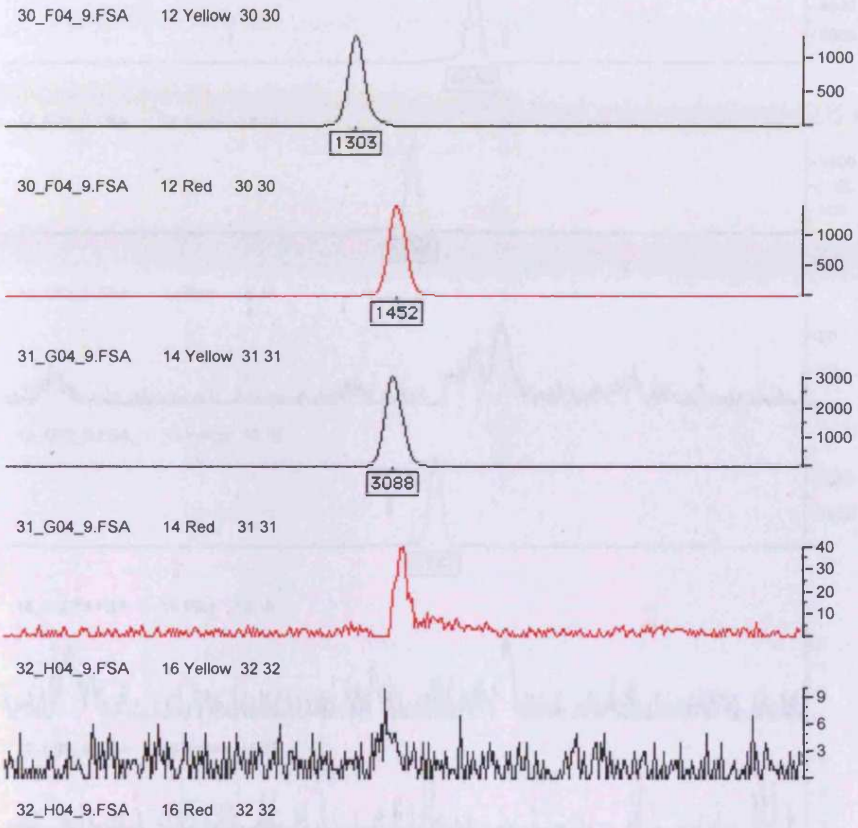


Figure 10.4.H. Genotyper (Applied Biosystems) analysis of a SNaPshot (Applied Biosystems) assay of polymorphism E19A. PCR and extension primers are given in table 10.4.B. Shown is a heterozygote (sample 30), a homozygote C (sample 31) and a blank (sample 32).

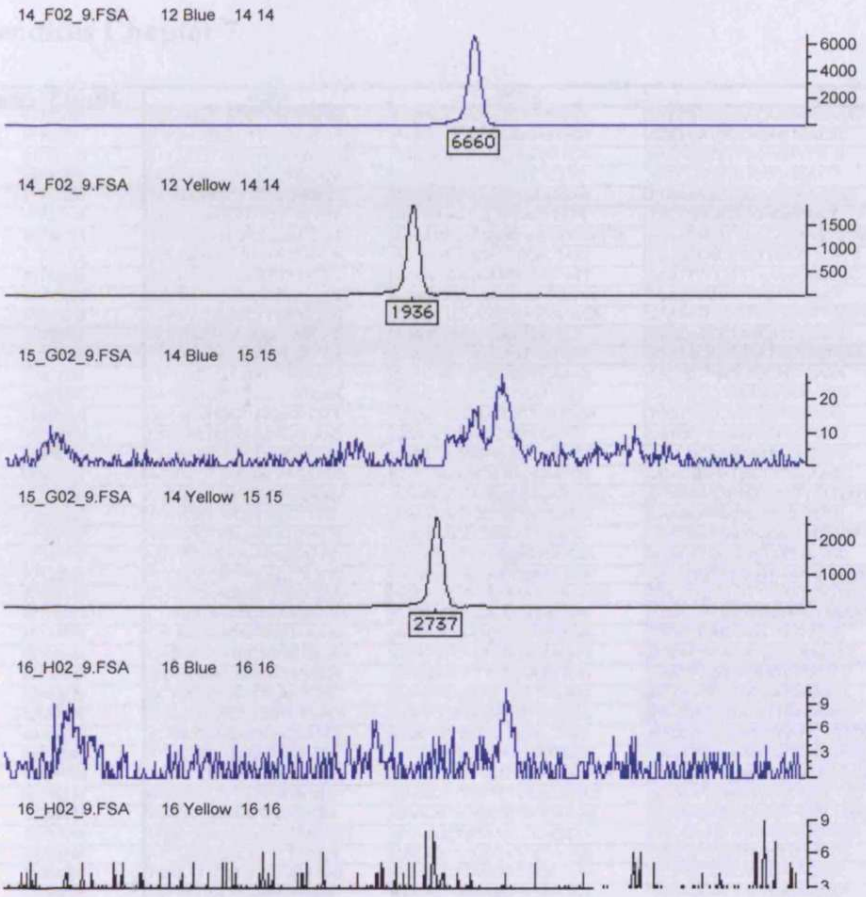


Figure 10.4.I. Genotyper (Applied Biosystems) analysis of a SNaPshot (Applied Biosystems) assay of polymorphism E21K[1]. PCR and extension primers are given in table 10.4.B. Shown is a heterozygote (sample 14), a homozygote C (sample 15) and a blank (sample 16).

10.5. Appendices Chapter 7

SNP	Position (UCSC HG17)	PCR F	PCR R	EXT
rs1558358	61700898	TCCAGCCTAGGTGACAGTGA	CAGAGCAAGACTCCGTCTCA	CATTGCACAAACAAGAAGAAAG
rs6504407	61702451	CTGGCACCACTTGCAGACTA	AAAGGCCACTGGGGAATACT	GTGTGATGCCACAGTGACTC
rs3764402	61727915	CTTGGCCTATGTTACATGCTCT	CTGGAGGCCGCTAAAATACA	AAGTCGTTTTGGGGATTTCG
rs3764401	61728291	GCTCTGGTGGCCATTGAATA	CACTCCCGCTCTCTCTCTTG	AATTTGTACGTCGGGCATTT
rs6504412	61730931	TAGAAGGGCTTCAGGGGTCT	CGTCGGGCAACTTAAAAGAA	TTGAAAGCTGGAACTAAAGGA
rs4520871	61732117	CACGTCAGGGTTCTCTGCTT	GCCAAACGGCATCACTTTTA	TGTTTGCATTTGAGATGACC
rs6504413	61734771	GGCTGCATGTGTGGATGTAT	AAAAATGAAGAGGAAGTCCCATTA	ACAATAGATAGCCAGAAGTGTGTATG
R2A	61751983	GCCCAAGCCGATGTTTATTA	GGGGTGAGGATCCCACTATT	TGAGTCACGTATTGCTTTCATG
rs8071250	61752029	GCCCAAGCCGATGTTTATTA	GGGGTGAGGATCCCACTATT	AGATTTTTTTTTAAAGTGGGGC
rs7218425	61763024	GGTGATCCCTAACCCAACCT	AGAGGGGGAAAAAGAGCAAA	TGAGCTCCATTTCATCCCCT
rs744214	61765318	TGAATCTGGGCTTTGAGGAG	CGACCTCAAGATCAGGAAACA	CAAAAACATCTTATTGGTTGGTTATAGTA
rs4536508	61813793	ACTTTGTGCAGCCCTACCAT	TCAACTTTTCTGCCCATGCT	CTGCAGCCTAGCATTTCACT
rs2109648	61920166	TTGGGAAAGACCATCCTTTG	CAGGCTGTTGCTGTTGTTGT	CTGTACAGTGAATGCTTGTGAA
rs228875	61924337	TGTCAGGCAGGCTTAGAAGG	CCTCAGAGCACACTGCAAAA	TAGTCTTGCCTGCACTGAGA
E20	61981167	TGTGAACTTGTCCCAGTGGA	CAACTGGCTGATCCATTCCT	CACTCTCCGAAGGTAAGGTG
rs9889698	62068596	GGTGGGTGCTGTCACTTTTT	TTCCTCCTGGTAGTTGATGGA	TGCTTGTCAGTGGTCCTCAG
rs7224072	62069901	CTTCACTCCTTGGGCAGAAC	CCAGCCTCCACATCTCATTT	CATCCAAAAATCCTCATTCAT
rs9912764	62072848	TACAAACAGCCTGGCTATGG	AACATGGAGTGGGTCATGGT	AACCACAGGAAAAAGCCATG
rs1010546	62115027	ACTGGGATCGAACAACAAGG	CCTCAGAGCTCAAGCAATCC	CCAGTGGATGGTATGGAAGC
rs2227857	62115540	CTCGACTAGAGGCTGCAGGA	AAAATGTGACCACAACATCTCG	GTAGTACTCACCTTCTTCTTGGTTAAG
rs2286958	62115852	TGTTCGTTCCCTTTCCTTTG	AGAACAATCCTGCCCTCAGA	CAACCATGCTCTCAGTTCTG
rs1010544	62117908	ACCAGCTTGCAGTCTCTGCT	CCCAAACCCTGACTTTCATC	TGATAGATAGACAAATAGTAGTTACACAGATACA
rs721429	62122367	ACTTTTGCTGCACACCATTG	TCTTGTTTTCTTGGGTCCTGA	ACACCTGGGGCTTTGTCTGC
E10A	62129885	AGCCACCACAACACTTAGGC	TATAGGCATGAGCCACCACA	CCTGTGTAGGGTCAAAAAAAAA
E10B(1)	62130370	CTCCTGAAGGCTCTGCCTAA	GCCAGTCACACAAAAATACAGC	TTTTTTCTATTTTTTTTTTTCTTGAGA
rs759117	62132085	CTGAGCAGACCCAAAGGAAG	CACAAGGCCATCCCTGTAAC	TTCATTAGACAGACAACTAGAATGCA
rs10491201	62133625	TGGGGAAACAAAAGCAAAAG	AGAAGGGAGGGTCTTTCCAA	TGTGTGAGTTGATGCTTCTG
rs9303511	62158093	GAGGTCTAGCCAGGGACCAG	CCATGTCAAAGGCAAAAGGT	GTCAGGCAGATGTTCAAAATT
rs8067877	62161640	GAAGGGATGCTCAGAAATGC	GTGCACCTTTGTCAATGTGG	CAATTTGCCTGTATCTCCATT
rs3889874	62162275	GCTGGAGGGACTGAATTTGT	CCAGTCACCTCTGCTTGACA	GGAATCCCTGCGATGCAGTA
rs3803821	62168179	GAGAGCCTCTGGGTCTGACA	TCCTCCTGGGGAGTTATCAA	TGATGCTTGAATTTGAGAATG
rs9894851	62169450	TCAACCAGGACTCCCAAACT	ACATTTTCCCAAGCAACTGG	GTCCAGTTTAGTCCATTTAGTGG
rs7224633	62171849	TTGCCTCTGCAAAATCCTTT	CCCAAAGCACTGGGATTGTA	CAGATGTGAAAGTATCTAGAGCAAGA
rs4381631	62196905	TTCATTCATGCATCCAGTCA	TTCCCAATTGCCTAGATTAAAAA	TAAAAGTAGAATGCTGGAGTAGGAAA
rs11650350	62198748	CAGTCCTTTTGGGAGAGCAA	CAACCTTTGTTCTGCACATTCT	AAGAAATGAGTCTCCAGTCTATTACC
E21D	62202031	TGACCCACAGGTTCATTCAA	GGACGTAACATGTGGGAAGG	ACAAAGAGGCAGGAAGTATAGC
rs12948890	62202951	TTGGCCAGAGAGGATAAGGTT	GGTGCCTGTAATCCCAGCTA	GCCCAGGCTGGAGTCGGTGG
rs4528612	62203761	ATTGGGAGACTGCCTTTCCT	CACAGCCTCCCATCCTAGAG	CTCCTGCCCTGCCCATGTCC
rs17710992	62204090	TGACATTCCTGGTGGGTAGG	GGACTCTTAGGGGGCCTGTA	GCAGTTCGTACACAGTCACC
rs3889237	62209892	TCTTATTCCACCAGGCAAGC	CCCACTGAAGGCTGGATACT	TGCAGGACTTCTCAGGGCCT
rs6504458	62213154	TCTGACCACAAAGCCATCTG	TTGACGGTTGTGTGTTTTTCA	TAATGTTTTTAATGATTTGCAAATAAA
rs9896575	62214907	GTTCCTGCTTTGGTTTGAGG	GCATCACAGAGGATGCCATA	GGAAGGCGACTAGTTAATTCC
rs4791037	62218356	TTCAGCGGTCTCTGAGGACT	TTCACCAAGGAAGCTGGAAC	CGTATTGTGCAGCTTCGACTT
rs4791036	62227597	TCTTGCTCATCTTTGCATGG	TTTGGAGCAAGTGTGACAGC	CCTTTTGGAACCTGATTCATC
rs2266674	62231014	TCTGGCTCTAGGTTAACCCTTC	TCAGCAACGTCCTTAGACGA	GACGTCTTGGGTGGAGATGT
E19C(2)	62231088	TCTGGCTCTAGGTTAACCCTTC	TCAGCAACGTCCTTAGACGA	CGGTATTTTGAGGCCTCCCC
E19D(1)	62231329	TCATCTGGTACCCTCCTTGG	CTCCAGGCCTGTCATTCAAC	TGATACTTTTCATTCTTTGTAAGAGG
E19D(2)	62231360	TCATCTGGTACCCTCCTTGG	CTCCAGGCCTGTCATTCAAC	TCTAAGGACGTTGCTGAACAAGC
E19F	62232348	TCAAAAATGAATGGCTAGTTTACG	TCCGCAGCCTCTGAAAAG	TGGTACACACGTGGCATTGC
rs7342969	62232651	TGTACTGGACACGCAGGAAG	GTCAGGACACGAACCCAGAT	TTTCCAGTCTTGGGAGGAGC
rs7342847	62233006	CTGCCATCTTTTATGCACCA	CTGGGGGTTTGTTCTGATGT	GGGTCCTGGCTCCCCTGTCC
rs9907086	62233937	GATGCTTTGGGAAGTCCTTG	GCCTAATACGTGGCATCAGA	CCTTTCCCTGAGGCTGCAGG
E19L	62234830	GGTTGGTGCAGATGGAAGTT	ATCACCTTTCCGAGGCTTTT	AACACCGCACAACTAACAAC
rs8464	62237178	TCCCTTCTGTTAGGGTCTTTCA	TGAAAGTGTTCTGCGGAGTG	TACACATCCGTCTCCTCTGC
rs12603061	62242660	CTCCCAAAGTGCTGGGATTA	TGTCAATTTCATTCCCAAAACA	CCATAATACTTGTCTTACAGTTTCCA
rs4791032	62246927	CCTGGAGGAAAAGAACAATGA	AAGATCCCCCAAAGCAGAAT	CTGTGGGTGGGAATATAAATTGGTAC
rs733646	62256511	TTTCCTCCCTGTTCCTGATG	CTTCTGGCTCCCCTAAGACC	CCTTAGTAACTGGATGATTGCTTA
rs737063	62270066	TTGCTCATGCTTCAGTGACC	ACAGGGGAAGATGGGAAACT	CATTTTCTACAAATTCAGATTTACTCTC

Table 10.5.A. SNPs analysed in UK schizophrenia case-control DNA pools. Given is SNP name and position (UCSC May 2004 freeze), and the PCR and extension primers used for each assay.

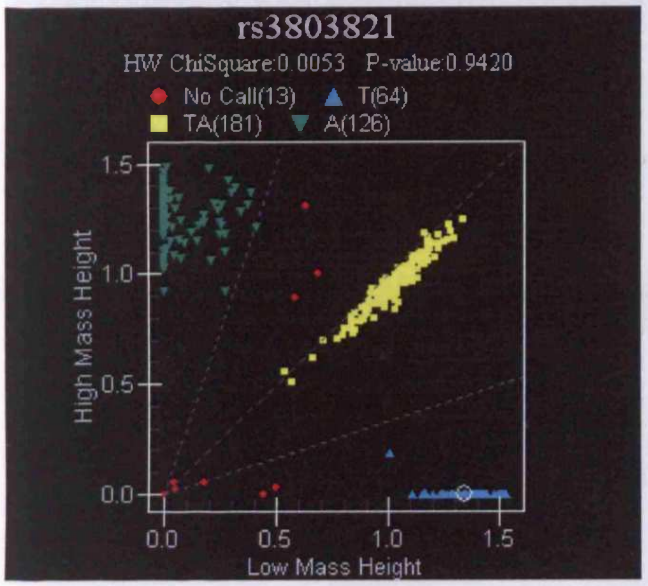


Figure 10.5.A. An example of the Sequenom MassARRAY iPlex genotyping of rs3803821 in 384 samples. The x axis represents log peak allele height of the A allele, while the y axis denotes the log peak height of the T allele. Yellow genotypes indicate heterozygous individuals, while green genotypes are homozygous AA, blue genotypes are homozygous T for the and red genotypes are no-calls.

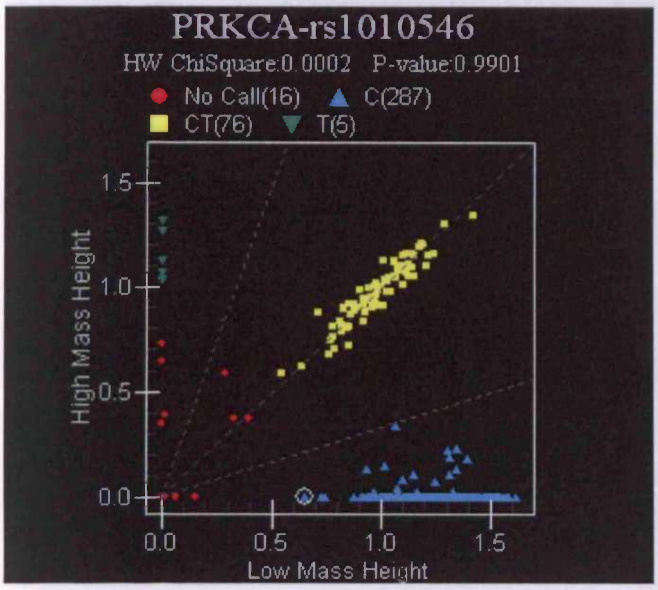


Figure 10.5.B. An example of the Sequenom MassARRAY iPlex genotyping of rs1010546 in 384 samples. The x axis represents log peak allele height of the C allele, while the y axis denotes the log peak height of the T allele. Yellow genotypes indicate heterozygous individuals, while green genotypes are homozygous T, blue genotypes are homozygous C for the and red genotypes are no-calls.

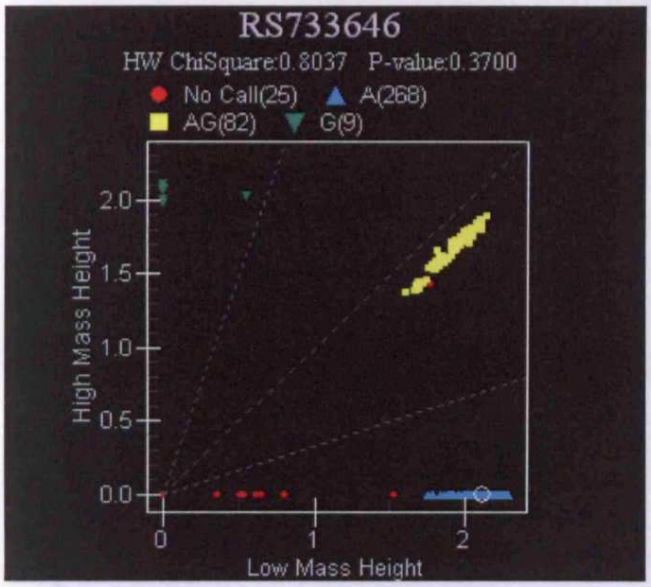


Figure 10.5.C. An example of the Sequenom MassARRAY hME genotyping of rs733646 in 384 samples. The x axis represents log peak allele height of the A allele, while the y axis denotes the log peak height of the G allele. Yellow genotypes indicate heterozygous individuals, while green genotypes are homozygous G, blue genotypes are homozygous A for the and red genotypes are no-calls.

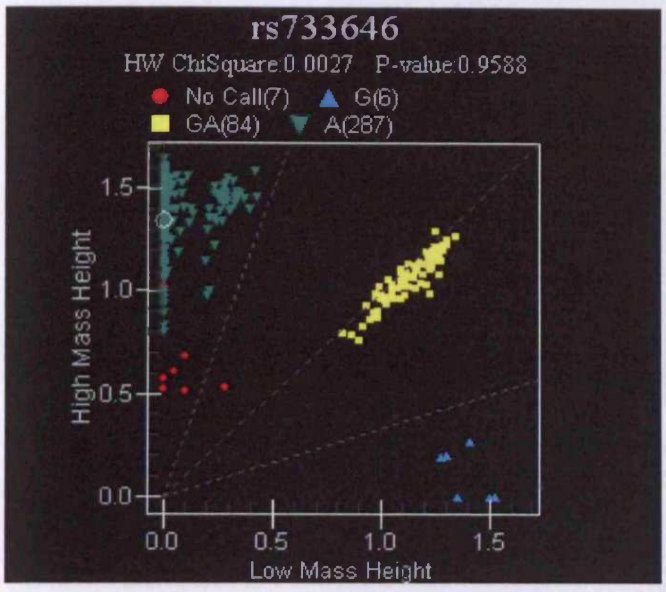


Figure 10.5.D. An example of the Sequenom MassARRAY iPlex genotyping of rs733646 in 384 samples. The x axis represents log peak allele height of the G allele, while the y axis denotes the log peak height of the A allele. Yellow genotypes indicate heterozygous individuals, while green genotypes are homozygous A, blue genotypes are homozygous G for the and red genotypes are no-calls.

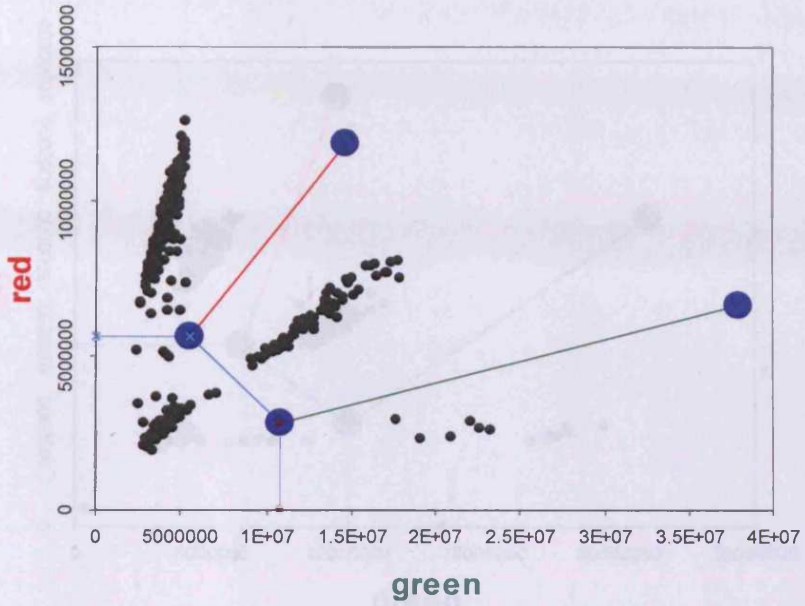


Figure 10.5.E. Amplifluor genotype cluster trace of rs1010546 genotyped in 384 samples. The y axis represents fluorescence of the red fluorophore corresponding to the common C allele, while the x axis denotes the fluorescence intensity of the green fluorophore, corresponding to the rare T allele. The red and green lines demarcate the 3 genotype clusters while the dropouts are shown clustering near the origin.

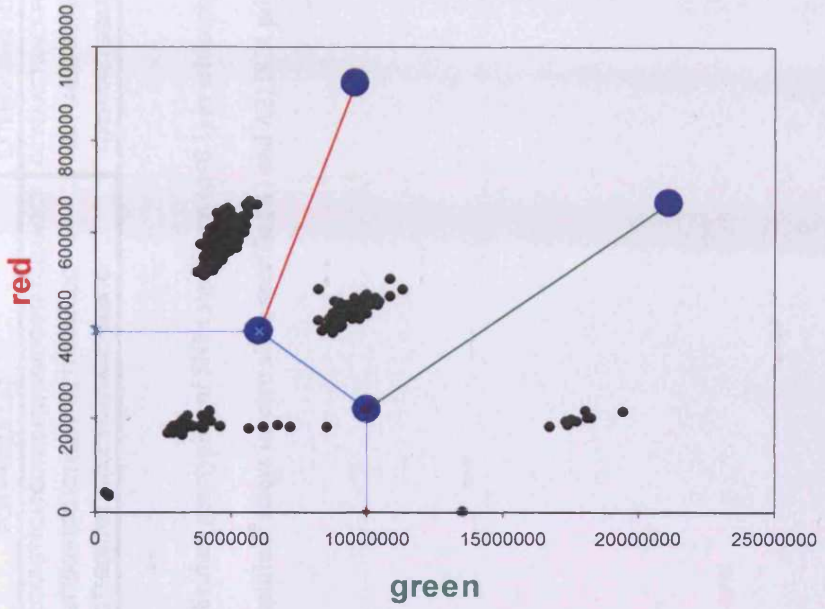


Figure 10.5.F. Amplifluor genotype cluster trace of rs733646 genotyped in 384 samples. The y axis represents fluorescence of the red fluorophore corresponding to the common A allele, while the x axis denotes the fluorescence intensity of the green fluorophore, corresponding to the rare G allele. The red and green lines demarcate the 3 genotype clusters while the dropouts are shown clustering near the origin.

SNP	Method	PCR PRIMER 1/A1	PCR PRIMER 2/A2	EXTENSION/REVERSE PRIMER
rs1010546	Amp	GAAGGTCGGAGTCAACGGATTCAGTGGATGGTATGGAAGCC	GAAGGTGACCAAGTTCATGCTCCAGTGGATGGTATGGAAGCT	CCTCAACTGACATTTGAGAGACT
rs3803821	iPlex	ACGTTGGATGAGCCTCTGGGTCTGACATTC	ACGTTGGATGCACTTGCTTCCTTAGATGCC	CCTTAGATGCCTGTATACTTT
rs733646	iPlex	ACGTTGGATGATTTTGCTGATGGGAGCCCT	ACGTTGGATGTTCCCCTTAGTAACTGGATG	TTAGTAACTGGATGATTGCTTA

Table 10.5.B. SNPs, genotyping technology and corresponding assay primers used for individual genotyping of SNPs significant ($p<0.1$) in schizophrenia-control DNA pools. Note that sequenom assays use PCR and extension primers, while amplifluor assays involve allele specific (A1 and A2) PCR primers in one orientation and a common alternate PCR primer (reverse).

UK Schizophrenia

SNP	Method	PCR PRIMER 1/A1	PCR PRIMER 2/A2	EXTENSION/REVERSE PRIMER
rs1010548	Amp	GAAGGTCGGAGTCAACGGATTCAGTGGATGGTATGGAAGCC	GAAGGTGACCAAGTTCATGCTCCAGTGGATGGTATGGAAGCT	CCTCAACTGACATTTGAGAGACT
rs3803821	iPlex	ACGTTGGATGAGCCTCTGGGTCTGACATTC	ACGTTGGATGCACTTGCTTCCTTAGATGCC	CCTTAGATGCCTGTATACTTT
rs733646	iPlex	ACGTTGGATGATTTTGCTGATGGGAGCCCT	ACGTTGGATGTTCCCCTTAGTAACTGGATG	TTAGTAACTGGATGATTGCTTA

UK Bipolar I

SNP	Method	PCR PRIMER 1/A1	PCR PRIMER 2/A2	EXTENSION/REVERSE PRIMER
rs1010548	Amp	GAAGGTCGGAGTCAACGGATTCAGTGGATGGTATGGAAGCC	GAAGGTGACCAAGTTCATGCTCCAGTGGATGGTATGGAAGCT	CCTCAACTGACATTTGAGAGACT
rs3803821	iPlex	ACGTTGGATGAGCCTCTGGGTCTGACATTC	ACGTTGGATGCACTTGCTTCCTTAGATGCC	CCTTAGATGCCTGTATACTTT
rs733646	hME	GAAGGTCGGAGTCAACGGATTCCCCTTAGTAACTGGATGATTGCTTAT	GAAGGTGACCAAGTTCATGCTTCCCTTAGTAACTGGATGATTGCTTACA	ATTTTGCTGATGGGAGCCCT

Bulgarian trios

SNP	Method	PCR PRIMER 1/A1	PCR PRIMER 2/A2	EXTENSION/REVERSE PRIMER
rs1010548	iPlex	ACGTTGGATGCTGTCTAAATGGTCTGATGC	ACGTTGGATGTTCGGAGCTGATGAAGATGC	GGGTGGATGGTATGGAAGC
rs3803821	iPlex	ACGTTGGATGAGCCTCTGGGTCTGACATTC	ACGTTGGATGCACTTGCTTCCTTAGATGCC	CCTTAGATGCCTGTATACTTT
rs733646	Amp	GAAGGTCGGAGTCAACGGATTCCCCTTAGTAACTGGATGATTGCTTAT	GAAGGTGACCAAGTTCATGCTTCCCTTAGTAACTGGATGATTGCTTACA	ATTTTGCTGATGGGAGCCCT

Table 10.5.C. SNPs, genotyping technology and corresponding assay primers used for individual genotyping of Entropy defined tagSNPs through the three association samples studied. Note that sequenom assays use PCR and extension primers, while amplifluor assays involve allele specific (A1 and A2) PCR primers in one orientation and a common alternate PCR primer (reverse).

10.6. Appendices Chapter 8

mRNA	SNP	Forward PCR primer	Reverse PCR primer	Amplimer size	Extension primer	Numerator allele	Denominator allele
X52479	E19A	ACAGCCCGTCTTAACACCAC	CCATATGGAATCAGACACAAGG	233	GAGAACAAACACCTCCCCAG	T	C
AB209475	E21K[2]	CCTTCCTGGGATCTCATTCA	GTGCTGGGATTACAGGTGTG	378	GGATTACAGGTGTGAGCCAC	C	T
X52479	rs7342847	TGGCAACCCTGAGTAGAAGG	GTCAGGACACGAACCCAGAT	170	GGGTCCTGGCTCCCCTGTCC	T	C
AB209475	rs17710992	CACCCCAAATCAAAGCATTC	GGGACTTGGCTTTTTGTGAC	228	TCCAGCCATTGACTCACTGA	T	C

Table 10.6.A. Assay details for allelic expression assays used for studying of SNPs in *PRKCA* isoform AB209475 and X52479. These primers were used for the genomic DNA and complementary DNA genotyping.

Bibliography

Abdolmaleky, H. M., Cheng, K. H., et al. 2006. Hypomethylation of MB-COMT promoter is a major risk factor for schizophrenia and bipolar disorder. *Hum Mol Genet,* 15(21), pp3132-45.

Addington, A. M., Gornick, M., et al. 2004. Polymorphisms in the 13q33.2 gene G72/G30 are associated with childhood-onset schizophrenia and psychosis not otherwise specified. *Biol Psychiatry,* 55(10), pp976-80.

Aggleton, J. P. and Brown, M. W. 1999. Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behav Brain Sci,* 22(3), pp425-44; discussion 444-89.

Alkon, D. L., Epstein, H., et al. 2005. Protein synthesis required for long-term memory is induced by PKC activation on days before associative learning. *Proc Natl Acad Sci U S A,* 102(45), pp16432-7.

Amrani, N., Sachs, M. S., et al. 2006. Early nonsense: mRNA decay solves a translational problem. *Nat Rev Mol Cell Biol,* 7(6), pp415-25.

Andreu, N., Garcia-Rodriguez, M., et al. 2006. A novel Wiskott-Aldrich syndrome protein (WASP) complex mutation identified in a WAS patient results in an aberrant product at the C-terminus from two transcripts with unusual polyA signals. *J Hum Genet,* 51(2), pp92-7.

Badner, J. A. and Gershon, E. S. 2002. Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Mol Psychiatry,* 7(4), pp405-11.

Bailey, J. A., Yavor, A. M., et al. 2001. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res,* 11(6), pp1005-17.

Ban, T. A. 2004. Neuropsychopharmacology and the genetics of schizophrenia: a history of the diagnosis of schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry,* 28(5), pp753-62.

Banihashemi, L., Wilson, G. M., et al. 2006. Upf1/Upf2 Regulation of 3' Untranslated Region Splice Variants of AUF1 Links Nonsense-mediated and A+U-rich Element-mediated mRNA Decay. *Mol Cell Biol.*

Baron, W., de Jonge, J. C., et al. 2000. Perturbation of myelination by activation of distinct signaling pathways: an in vitro study in a myelinating culture derived from fetal rat brain. *J Neurosci Res,* 59(1), pp74-85.

Barrett, J. C., Fry, B., et al. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics,* 21(2), pp263-5.

Barton, A., Woolmore, J. A., et al. 2004. Association of protein kinase C alpha (PRKCA) gene with multiple sclerosis in a UK population. *Brain,* 127(Pt 8), pp1717-22.

Basta-Kaim, A., Budziszewska, B., et al. 2002. Chlorpromazine inhibits the glucocorticoid receptor-mediated gene transcription in a calcium-dependent manner. *Neuropharmacology,* 43(6), pp1035-43.

Beaudry, H., Gendron, L., et al. 2006. Involvement of PKC{alpha} in the early action of angiotensin II AT2 effects on neurite outgrowth in NG108-15 cells: AT2-receptor inhibits PKC{alpha} and p21ras activity. *Endocrinology.*

Bennett, P., Segurado, R., et al. 2002. The Wellcome trust UK-Irish bipolar affective disorder sibling-pair genome screen: first stage report. *Mol Psychiatry,* 7(2), pp189-200.

Birnbaum, S. G., Yuan, P. X., et al. 2004. Protein kinase C overactivity impairs prefrontal cortical regulation of working memory. *Science,* 306(5697), pp882-4.

Black, J. E., Kodish, I. M., et al. 2004. Pathology of layer V pyramidal neurons in the prefrontal cortex of patients with schizophrenia. *Am J Psychiatry,* 161(4), pp742-4.

Blackwood, D. H., Fordyce, A., 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), pp428-33.

Blouin, J. L., Dombroski, B. A., et al. 1998. Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. *Nat Genet,* 20(1), pp70-3.

Boehm, J., Kang, M. G., et al. 2006. Synaptic incorporation of AMPA receptors during LTP is controlled by a PKC phosphorylation site on GluR1. *Neuron,* 51(2), pp213-25.

Bonini, J. S., Cammarota, M., et al. 2005. Inhibition of PKC in basolateral amygdala and posterior parietal cortex impairs consolidation of inhibitory avoidance memory. *Pharmacol Biochem Behav,* 80(1), pp63-7.

Bourgain, C., Genin, E., et al. 2007. Are genome-wide association studies all that we need to dissect the genetic component of complex human diseases? *Eur J Hum Genet,* 15(3), pp260-3.

Braff, D. L., Geyer, M. A., et al. 2001. Human studies of prepulse inhibition of startle: normal subjects, patient groups, and pharmacological studies. *Psychopharmacology (Berl),* 156(2-3), pp234-58.

Bray, N. J., Buckland, P. R., et al. 2003a. Cis-acting variation in the expression of a high proportion of genes in human brain. *Hum Genet,* 113(2), pp149-53.

Bray, N. J., Buckland, P. R., et al. 2003b. A haplotype implicated in schizophrenia susceptibility is associated with reduced COMT expression in human brain. *Am J Hum Genet,* 73(1), pp152-61.

Cannon, T. D., Hennah, W., et al. 2005. Association of DISC1/TRAX haplotypes with schizophrenia, reduced prefrontal gray matter, and impaired short- and long-term memory. *Arch Gen Psychiatry,* 62(11), pp1205-13.

Cardno, A. and McGuffin, P. 2002. Quantitative genetics In: Gottesman, I. I. ed. *Psychiatric Genetics & Genomics.* First ed. Oxford Medical Publications. pp. 31-53.

Cardno, A. G., Holmans, P. A., et al. 2001. A genomewide linkage study of age at onset in schizophrenia. *Am J Med Genet,* 105(5), pp439-45.

Cardon, L. R. and Palmer, L. J. 2003. Population stratification and spurious allelic association. *Lancet,* 361(9357), pp598-604.

Carlson, C. S., Eberle, M. A., et al. 2004. Mapping complex disease loci in whole-genome association studies. *Nature,* 429(6990), pp446-52.

Chen, C. H., Lee, Y. R., et al. 1999. Systematic mutation analysis of the catechol O-methyltransferase gene as a candidate gene for schizophrenia. *Am J Psychiatry,* 156(8), pp1273-5.

Chen, D. C., Saarela, J., et al. 2004. Segmental duplications flank the multiple sclerosis locus on chromosome 17q. *Genome Res,* 14(8), pp1483-92.

Chen, G., Manji, H. K., et al. 1994. Chronic sodium valproate selectively decreases protein kinase C alpha and epsilon in vitro. *J Neurochem,* 63(6), pp2361-4.

Chen, G., Masana, M. I., et al. 2000. Lithium regulates PKC-mediated intracellular cross-talk and gene expression in the CNS in vivo. *Bipolar Disord,* 2(3 Pt 2), pp217-36.

Chen, Q. Y., Chen, Q., et al. 2006. Case-control association study of Disrupted-in-Schizophrenia-1 (DISC1) gene and schizophrenia in the Chinese population. *J Psychiatr Res.*

Chen, X., Dunham, C., et al. 2004. Regulator of G-protein signaling 4 (RGS4) gene is associated with schizophrenia in Irish high density families. *Am J Med Genet B Neuropsychiatr Genet,* 129(1), pp23-6.

Chen, X., Wang, X., et al. 2004. Variants in the catechol-o-methyltransferase (COMT) gene are associated with schizophrenia in Irish high-density families. *Mol Psychiatry,* 9(10), pp962-7.

Cheng, Z., Ventura, M., et al. 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature,* 437(7055), pp88-93.

Cheung, J., Estivill, X., et al. 2003. Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. *Genome Biol,* 4(4), ppR25.

Chowdari, K. V., Mirnics, K., et al. 2002. Association and linkage analyses of RGS4 polymorphisms in schizophrenia. *Hum Mol Genet,* 11(12), pp1373-80.

Chumakov, I., Blumenfeld, M., 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), pp13675-80.

Cirulli, E. T. and Goldstein, D. B. 2007. in vitro assays fail to predict in vivo effects of regulatory polymorphisms. *Hum Mol Genet.*

Clark, A. G. 2004. The role of haplotypes in candidate gene studies. *Genet Epidemiol,* 27(4), pp321-33.

Clark, A. G., Boerwinkle, E., et al. 2005. Determinants of the success of whole-genome association testing. *Genome Res,* 15(11), pp1463-7.

Cohen, J. C., Kiss, R. S., et al. 2004. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science,* 305(5685), pp869-72.

Collier, D. A. and Li, T. 2003. The genetics of schizophrenia: glutamate not dopamine? *Eur J Pharmacol,* 480(1-3), pp177-84.

Collingridge, G. L., Isaac, J. T., et al. 2004. Receptor trafficking and synaptic plasticity. *Nat Rev Neurosci,* 5(12), pp952-62.

Conrad, D. F., Andrews, T. D., et al. 2006. A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet,* 38(1), pp75-81.

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), pp2463-8.

Corvin, A. P., Morris, D. W., et al. 2004. Confirmation and refinement of an 'at-risk' haplotype for schizophrenia suggests the EST cluster, Hs.97362, as a potential susceptibility gene at the Neuregulin-1 locus. *Mol Psychiatry,* 9(2), pp208-13.

Craddock, N., O'Donovan, M. C., et al. 2005. The genetics of schizophrenia and bipolar disorder: dissecting psychosis. *J Med Genet,* 42(3), pp193-204.

Craddock, N. and Owen, M. J. 2005. The beginning of the end for the Kraepelinian dichotomy. *Br J Psychiatry,* 186364-6.

Craig, A. M., Olds, J. L., et al. 1993. Quantitative distribution of protein kinase C alpha, beta, gamma, and epsilon mRNAs in the hippocampus of control and nictitating membrane conditioned rabbits. *Brain Res Mol Brain Res,* 19(4), pp269-76.

Curtis, D., Kalsi, G., et al. 2003. Genome scan of pedigrees multiply affected with bipolar disorder provides further support for the presence of a susceptibility locus on chromosome 12q23-q24, and suggests the presence of additional loci on 1p and 1q. *Psychiatr Genet,* 13(2), pp77-84.

Cutting, J. 2003. Descriptive Psychopathology In: Weinberger, D. R. ed. *Schizophrenia.* Second ed. Blackwell Science Ltd. pp. 15-24.

Dalgleish, G., Veyrune, J. L., et al. 2001. mRNA localization by a 145-nucleotide region of the c-fos 3'--untranslated region. Links to translation but not stability. *J Biol Chem,* 276(17), pp13593-9.

Daly, M. J., Rioux, J. D., et al. 2001. High-resolution haplotype structure in the human genome. *Nat Genet,* 29(2), pp229-32.

Davis, K. L., Stewart, D. G., et al. 2003. White matter changes in schizophrenia: evidence for myelin-related dysfunction. *Arch Gen Psychiatry,* 60(5), pp443-56.

de Bakker, P. I., Yelensky, R., et al. 2005. Efficiency and power in genetic association studies. *Nat Genet,* 37(11), pp1217-23.

de Chaldee, M., Corbex, M., et al. 2001. No evidence for linkage between COMT and schizophrenia in a French population. *Psychiatry Res,* 102(1), pp87-90.

de Moor, C. H., Meijer, H., et al. 2005. Mechanisms of translational control by the 3' UTR in development and differentiation. *Semin Cell Dev Biol,* 16(1), pp49-58.

de Quervain, D. J. and Papassotiropoulos, A. 2006. Identification of a genetic cluster influencing memory performance and hippocampal activity in humans. *Proc Natl Acad Sci U S A,* 103(11), pp4270-4.

DeLisi, L. E., Shaw, S. H., et al. 2002. A genome-wide scan for linkage to chromosomal regions in 382 sibling pairs with schizophrenia or schizoaffective disorder. *Am J Psychiatry,* 159(5), pp803-12.

Deng, W., Wang, H., et al. 2004. Role of metabotropic glutamate receptors in oligodendrocyte excitotoxicity and oxidative stress. *Proc Natl Acad Sci U S A,* 101(20), pp7751-6.

Devlin, B. and Risch, N. 1995. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics,* 29(2), pp311-22.

Devon, R. S., Anderson, S., et al. 2001. Identification of polymorphisms within Disrupted in Schizophrenia 1 and Disrupted in Schizophrenia 2, and an investigation of their association with schizophrenia and bipolar affective disorder. *Psychiatr Genet,* 11(2), pp71-8.

Di Liegro, C. M., Bellafiore, M., et al. 2000. 3'-untranslated regions of oxidative phosphorylation mRNAs function in vivo as enhancers of translation. *Biochem J,* 352 Pt 1109-15.

Dick, D. M., Foroud, T., et al. 2003. Genomewide linkage analyses of bipolar disorder: a new sample of 250 pedigrees from the National Institute of Mental Health Genetics Initiative. *Am J Hum Genet,* 73(1), pp107-14.

Ding, K., Zhang, J., et al. 2005. htSNPer1.0: software for haplotype block partition and htSNPs selection. *BMC Bioinformatics,* 638.

Dorr, S., Midro, A. T., et al. 2001. Construction of a detailed physical and transcript map of the candidate region for Russell-Silver syndrome on chromosome 17q23-q24. *Genomics,* 71(2), pp174-81.

Duan, J., Martinez, M., et al. 2005. Neuregulin 1 (NRG1) and schizophrenia: analysis of a US family sample and the evidence in the balance. *Psychol Med,* 35(11), pp1599-610.

Dwivedi, Y. and Pandey, G. N. 1999. Effects of treatment with haloperidol, chlorpromazine, and clozapine on protein kinase C (PKC) and phosphoinositide-specific phospholipase C (PI-PLC) activity and on mRNA and protein expression

of PKC and PLC isozymes in rat brain. *J Pharmacol Exp Ther,* 291(2), pp688-704.

Eichler, E. E. 2006. Widening the spectrum of human genetic variation. *Nat Genet,* 38(1), pp9-11.

Einat, H., Yuan, P., et al. 2007. Protein Kinase C Inhibition by Tamoxifen Antagonizes Manic-Like Behavior in Rats: Implications for the Development of Novel Therapeutics for Bipolar Disorder. *Neuropsychobiology,* 55(3-4), pp123-131.

Eisenberg, E., Nemzer, S., et al. 2005. Is abundant A-to-I RNA editing primate-specific? *Trends Genet,* 21(2), pp77-81.

Ekelund, J., Hennah, W., et al. 2004. Replication of 1q42 linkage in Finnish schizophrenia pedigrees. *Mol Psychiatry,* 9(11), pp1037-41.

ENCODE Project Consortium 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science,* 306(5696), pp636-40.

ENCODE Project Consortium 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature,* 447(7146), pp799-816.

Ewald, H., Wikman, F. P., et al. 2005. A genome-wide search for risk genes using homozygosity mapping and microarrays with 1,494 single-nucleotide

polymorphisms in 22 eastern Cuban families with bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet,* 133(1), pp25-30.

Feuk, L., Carson, A. R., et al. 2006. Structural variation in the human genome. *Nat Rev Genet,* 7(2), pp85-97.

Firulli, B. A., Howard, M. J., et al. 2003. PKA, PKC, and the protein phosphatase 2A influence HAND factor function: a mechanism for tissue-specific transcriptional regulation. *Mol Cell,* 12(5), pp1225-37.

Forton, J., Kwiatkowski, D., et al. 2005. Accuracy of haplotype reconstruction from haplotype-tagging single-nucleotide polymorphisms. *Am J Hum Genet,* 76(3), pp438-48.

Fukui, N., Muratake, T., et al. 2006. Supportive evidence for neuregulin 1 as a susceptibility gene for schizophrenia in a Japanese population. *Neurosci Lett,* 396(2), pp117-20.

Fuller, R. L. M., Schultz, S. K., et al. 2003. The symptoms of schizophrenia In: Weinberger, D. R. ed. *Schizophrenia.* Second ed. Blackwell Science Ltd. pp. 25-33.

Funke, B., Finn, C. T., et al. 2004. Association of the DTNBP1 locus with schizophrenia in a U.S. population. *Am J Hum Genet,* 75(5), pp891-8.

Gardner, M., Gonzalez-Neira, A., et al. 2006. Extreme population differences across Neuregulin 1 gene, with implications for association studies. *Mol Psychiatry,* 11(1), pp66-75.

Georgieva, L., Moskvina, V., et al. 2006. Convergent evidence that oligodendrocyte lineage transcription factor 2 (OLIG2) and interacting genes influence susceptibility to schizophrenia. *Proc Natl Acad Sci U S A,* 103(33), pp12469-74.

Glatt, S. J., Faraone, S. V., et al. 2003. Association between a functional catechol O-methyltransferase gene polymorphism and schizophrenia: meta-analysis of case-control and family-based studies. *Am J Psychiatry,* 160(3), pp469-76.

Golden, D. E. and Hajduk, S. L. 2006. The importance of RNA structure in RNA editing and a potential proofreading mechanism for correct guide RNA:pre-mRNA binary complex formation. *J Mol Biol,* 359(3), pp585-96.

Gonzalez, E., Kulkarni, H., et al. 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science,* 307(5714), pp1434-40.

Gonzalez, M. I., Bannerman, P. G., et al. 2003. Phorbol myristate acetate-dependent interaction of protein kinase Calpha and the neuronal glutamate transporter EAAC1. *J Neurosci,* 23(13), pp5589-93.

Green, E. K., Norton, N., et al. 2006. Evidence that a DISC1 frame-shift deletion associated with psychosis in a single family may not be a pathogenic mutation. *Mol Psychiatry,* 11(9), pp798-9.

Grossman, M. H., Szumlanski, C., et al. 1992. Electrophoretic analysis of low and high activity forms of catechol-O-methyltransferase in human erythrocytes. *Life Sci,* 50(7), pp473-80.

Hahn, C. G. and Friedman, E. 1999. Abnormalities in protein kinase C signaling and the pathophysiology of bipolar disorder. *Bipolar Disord,* 1(2), pp81-6.

Hahn, C. G., Umapathy, et al. 2005. Lithium and valproic acid treatments reduce PKC activation and receptor-G protein coupling in platelets of bipolar manic patients. *J Psychiatr Res,* 39(4), pp355-63.

Hahn, M. K., Mazei-Robison, M. S., et al. 2005. Single nucleotide polymorphisms in the human norepinephrine transporter gene affect expression, trafficking, antidepressant interaction, and protein kinase C regulation. *Mol Pharmacol,* 68(2), pp457-66.

Hakak, Y., Walker, J. R., et al. 2001. Genome-wide expression analysis reveals dysregulation of myelination-related genes in chronic schizophrenia. *Proc Natl Acad Sci U S A,* 98(8), pp4746-51.

Hall, D., Gogos, J. A., et al. 2004. The contribution of three strong candidate schizophrenia susceptibility genes in demographically distinct populations. *Genes Brain Behav,* 3(4), pp240-8.

Hampe, J., Schreiber, S., et al. 2003. Entropy-based SNP selection for genetic association studies. *Hum Genet,* 114(1), pp36-43.

Harrison, P. J. and Lewis, D. A. 2003. Neuropathology of schizophrenia In: Weinberger, D. R. ed. *Schizophrenia.* Second ed. Blackwell Science Ltd. pp. 310-325.

Harrison, P. J. and Weinberger, D. R. 2005. Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Mol Psychiatry,* 10(1), pp40-68; image 5.

Hattori, E., Liu, C., 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), pp1131-40.

He, L. and Hannon, G. J. 2004. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet,* 5(7), pp522-31.

Hennah, W., Varilo, T., et al. 2003. Haplotype transmission analysis provides evidence of association for DISC1 to schizophrenia and suggests sex-dependent effects. *Hum Mol Genet,* 12(23), pp3151-9.

Hill, S. J. 2006. G-protein-coupled receptors: past, present and future. *Br J Pharmacol,* 147 Suppl 1S27-37.

Hinds, D. A., Kloek, A. P., et al. 2006. Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nat Genet,* 38(1), pp82-5.

Hirschhorn, J. N. and Altshuler, D. 2002. Once and again-issues surrounding replication in genetic association studies. *J Clin Endocrinol Metab,* 87(10), pp4438-41.

Hirschhorn, J. N. and Daly, M. J. 2005. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet,* 6(2), pp95-108.

Hodgkinson, C. A., Goldman, D., et al. 2004. Disrupted in schizophrenia 1 (DISC1): association with schizophrenia, schizoaffective disorder, and bipolar disorder. *Am J Hum Genet,* 75(5), pp862-72.

Hughes, T. A. 2006. Regulation of gene expression by alternative untranslated regions. *Trends Genet,* 22(3), pp119-22.

Hunter, D. J., Kraft, P., et al. 2007. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet,* 39(7), pp870-4.

Hussain, R. J. and Carpenter, D. O. 2005. A comparison of the roles of protein kinase C in long-term potentiation in rat hippocampal areas CA1 and CA3. *Cell Mol Neurobiol,* 25(3-4), pp649-61.

Hyde, T. M. and Lewis, S. W. 2003. The secondary schizophrenias In: Weinberger, D. R. ed. *Schizophrenia.* Second ed. Blackwell Science Ltd. pp. 187-202.

Iafrate, A. J., Feuk, L., et al. 2004. Detection of large-scale variation in the human genome. *Nat Genet,* 36(9), pp949-51.

Ingason, A., Soeby, K., et al. 2006. No significant association of the 5' end of neuregulin 1 and schizophrenia in a large Danish sample. *Schizophr Res,* 83(1), pp1-5.

Ishii, M. and Kurachi, Y. 2003. Physiological actions of regulators of G-protein signaling (RGS) proteins. *Life Sci,* 74(2-3), pp163-71.

Iwata, N., Suzuki, T., et al. 2004. No association with the neuregulin 1 haplotype to Japanese schizophrenia. *Mol Psychiatry,* 9(2), pp126-7.

Jablensky, A. 2003. The epidemiological horizon In: Weinberger, D. R. ed. *Schizophrenia.* Second ed. Blackwell Science Ltd. pp. 203-231.

Kamiya, A., Kubo, K., et al. 2005. A schizophrenia-associated mutation of DISC1 perturbs cerebral cortex development. *Nat Cell Biol,* 7(12), pp1167-78.

Kapfhammer, J. P. 2004. Cellular and molecular control of dendritic growth and development of cerebellar Purkinje cells. *Prog Histochem Cytochem,* 39(3), pp131-82.

Keller, M. C. and Miller, G. 2006. Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best? *Behav Brain Sci,* 29(4), pp385-404; discussion 405-52.

Kennedy, M. B., Beale, H. C., et al. 2005. Integration of biochemical signalling in spines. *Nat Rev Neurosci,* 6(6), pp423-34.

Kim, S. Y. and Yang, J. H. 2005. Neurotoxic effects of 2,3,7,8-tetrachlorodibenzo-p-dioxin in cerebellar granule cells. *Exp Mol Med,* 37(1), pp58-64.

Kirov, G., Ivanov, D., 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), pp971-5.

Kirov, G., O'Donovan, M. C., et al. 2005. Finding schizophrenia genes. *J Clin Invest,* 115(6), pp1440-8.

Kirshenboim, N., Plotkin, B., et al. 2004. Lithium-mediated phosphorylation of glycogen synthase kinase-3beta involves PI3 kinase-dependent activation of protein kinase C-alpha. *J Mol Neurosci,* 24(2), pp237-45.

Klei, L., Bacanu, S. A., et al. 2005. Linkage analysis of a completely ascertained sample of familial schizophrenics and bipolars from Palau, Micronesia. *Hum Genet,* 117(4), pp349-56.

Knable, M. N., Barcia, B. M., et al. 2002. Abnormalities of the cingulate gyrus in bipolar disorder and other severe psychiatric illnesss: postmortem findings from the Stanley Foundation Neuropathology Consortium and literature review. *Clinical Neuroscience Research,* 2171-181.

Kobayashi, M., Kidd, D., et al. 2001. Protein kinase C activation by 12-0-tetradecanoylphorbol 13-acetate in CG-4 line oligodendrocytes stimulates turnover of choline and ethanolamine phospholipids by phospholipase D and induces rapid process contraction. *J Neurochem,* 76(2), pp361-71.

Kockelkorn, T. T., Arai, M., et al. 2004. Association study of polymorphisms in the 5' upstream region of human DISC1 gene with schizophrenia. *Neurosci Lett,* 368(1), pp41-5.

Kong, A., Gudbjartsson, D. F., et al. 2002. A high-resolution recombination map of the human genome. *Nat Genet,* 31(3), pp241-7.

Korshunov, A., Sycheva, R., et al. 2006. Genetically distinct and clinically relevant subtypes of glioblastoma defined by array-based comparative genomic hybridization (array-CGH). *Acta Neuropathol (Berl),* 111(5), pp465-74.

Kovoor, A., Seyffarth, P., et al. 2005. D2 dopamine receptors colocalize regulator of G-protein signaling 9-2 (RGS9-2) via the RGS9 DEP domain, and RGS9 knock-out mice develop dyskinesias associated with dopamine pathways. *J Neurosci,* 25(8), pp2157-65.

Kryukov, G. V., Kryukov, V. M., et al. 1999. New mammalian selenocysteine-containing proteins identified with an algorithm that searches for selenocysteine insertion sequence elements. *J Biol Chem,* 274(48), pp33888-97.

Kuner, R., Swiercz, J. M., et al. 2002. Characterization of the expression of PDZ-RhoGEF, LARG and G(alpha)12/G(alpha)13 proteins in the murine nervous system. *Eur J Neurosci,* 16(12), pp2333-41.

Lachman, H. M., Pedrosa, E., et al. 2006. Analysis of polymorphisms in AT-rich domains of neuregulin 1 gene in schizophrenia. *Am J Med Genet B Neuropsychiatr Genet,* 141(1), pp102-9.

Laird, N. M. and Lange, C. 2006. Family-based designs in the age of large-scale gene-association studies. *Nat Rev Genet,* 7(5), pp385-94.

Lander, E. and Kruglyak, L. 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet,* 11(3), pp241-7.

Larsson, C. 2006. Protein kinase C and the regulation of the actin cytoskeleton. *Cell Signal,* 18(3), pp276-84.

Leahy, J. C., Luo, Y., et al. 1993. Demonstration of presynaptic protein kinase C activation following long-term potentiation in rat hippocampal slices. *Neuroscience,* 52(3), pp563-74.

Leitges, M., Kovac, J., et al. 2004. A unique PDZ ligand in PKCalpha confers induction of cerebellar long-term synaptic depression. *Neuron,* 44(4), pp585-94.

Lencz, T., Morgan, T. V., et al. 2007. Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol Psychiatry,* 12(6), pp572-80.

Levinson, D. F., Mahtani, M. M., et al. 1998. Genome scan of schizophrenia. *Am J Psychiatry,* 155(6), pp741-50.

Lewis, C. M., Levinson, D. F., et al. 2003. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet,* 73(1), pp34-48.

Lewis, D. A. and Levitt, P. 2002. Schizophrenia as a disorder of neurodevelopment. *Annu Rev Neurosci,* 25409-32.

Li, T., Stefansson, H., et al. 2004. Identification of a novel neuregulin 1 at-risk haplotype in Han schizophrenia Chinese patients, but no association with the Icelandic/Scottish risk haplotype. *Mol Psychiatry,* 9(7), pp698-704.

Libien, J., Sacktor, T. C., et al. 2005. Magnesium blocks the loss of protein kinase C, leads to a transient translocation of PKC(alpha) and PKC(epsilon), and improves recovery after anoxia in rat hippocampal slices. *Brain Res Mol Brain Res,* 136(1-2), pp104-11.

Liddle, P. and Pantelis, C. 2003. Brain imaging in schizophrenia In: Weinberger, D. R. ed. *Schizophrenia.* Second ed. Blackwell Science Ltd. pp. 403-417.

Lidle, B. M., Williams, J., et al. 2002. The dementias In: McGuffin, P., Owen, M. J., et al. ed. *Psychiatric Genetics & Genomics.* First ed. Oxford Medical Publications. pp. 341-393.

Lin, P. I., Vance, J. M., et al. 2007. No gene is an island: the flip-flop phenomenon. *Am J Hum Genet,* 80(3), pp531-8.

Liu, L. Y., Wei, E. Q., et al. 2005. Protective effects of baicalin on oxygen/glucose deprivation- and NMDA-induced injuries in rat hippocampal slices. *J Pharm Pharmacol,* 57(8), pp1019-26.

Liu, P. Y., Zhang, Y. Y., et al. 2005. A survey of haplotype variants at several disease candidate genes: the importance of rare variants for complex diseases. *J Med Genet,* 42(3), pp221-7.

Liu, X., He, G., et al. 2004. Association of DAAO with schizophrenia in the Chinese population. *Neurosci Lett,* 369(3), pp228-33.

Liu, Y. L., Fann, C. S., et al. 2006. A single nucleotide polymorphism fine mapping study of chromosome 1q42.1 reveals the vulnerability genes for schizophrenia, GNPAT and DISC1: Association with impairment of sustained attention. *Biol Psychiatry,* 60(6), pp554-62.

Locke, D. P., Sharp, A. J., et al. 2006. Linkage Disequilibrium and Heritability of Copy-Number Polymorphisms within Duplicated Regions of the Human Genome. *Am J Hum Genet,* 79(2), pp275-90.

Lohmueller, K. E., Pearce, C. L., et al. 2003. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet,* 33(2), pp177-82.

Lopez-Andreo, M. J., Torrecillas, A., et al. 2005. Retinoic acid as a modulator of the activity of protein kinase Calpha. *Biochemistry,* 44(34), pp11353-60.

Lutz, E. M., Mitchell, R., et al. 1993. Functional expression of 5-HT1c receptor cDNA in COS 7 cells and its influence on protein kinase C. *FEBS Lett,* 316(3), pp228-32.

Macgregor, S., Visscher, P. M., et al. 2004. A genome scan and follow-up study identify a bipolar disorder susceptibility locus on chromosome 1q42. *Mol Psychiatry,* 9(12), pp1083-90.

MacIntyre, D. J., Blackwood, D. H., et al. 2003. Chromosomal abnormalities and mental illness. *Mol Psychiatry,* 8(3), pp275-87.

Maki, P., Veijola, J., et al. 2005. Predictors of schizophrenia--a review. *Br Med Bull,* 73-741-15.

Manji, H. K. and Lenox, R. H. 1999. Protein kinase C signaling in the brain: molecular transduction of mood stabilization in the treatment of manic-depressive illness. *Biol Psychiatry,* 46(10), pp1328-51.

Manji, H. K. and Chen, G. 2002. PKC, MAP kinases and the bcl-2 family of proteins as long-term targets for mood stabilizers. *Mol Psychiatry,* 7 Suppl 1S46-56.

Marchini, J., Howie, B., et al. 2007. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet,* 39(7), pp906-13.

McCarroll, S. A., Hadnott, T. N., et al. 2006. Common deletion polymorphisms in the human genome. *Nat Genet,* 38(1), pp86-92.

McClellan, J. M., Susser, E., et al. 2007. Schizophrenia: a common disease caused by multiple rare alleles. *Br J Psychiatry,* 190194-9.

McDade, S. S., Hall, P. A., et al. 2007. Translational control of SEPT9 isoforms is perturbed in disease. *Hum Mol Genet,* 16(7), pp742-52.

McGrath, J. J. and Murray, R. M. 2003. Risk factors for schizophrenia: from conception to birth In: Weinberger, D. R. ed. *Schizophrenia.* Second ed. Blackwell Science Ltd. pp. 232-250.

McInnis, M. G., Dick, D. M., et al. 2003. Genome-wide scan and conditional analysis in bipolar disorder: evidence for genomic interaction in the National Institute of Mental Health genetics initiative bipolar pedigrees. *Biol Psychiatry,* 54(11), pp1265-73.

Mellor, H. and Parker, P. J. 1998. The extended protein kinase C superfamily. *Biochem J,* 332 (Pt 2)281-92.

Meyer-Lindenberg, A., Mervis, C. B., et al. 2006. Neural mechanisms in Williams syndrome: a unique window to genetic influences on cognition and behaviour. *Nat Rev Neurosci,* 7(5), pp380-93.

Meyer-Lindenberg, A., Straub, R. E., et al. 2007. Genetic evidence implicating DARPP-32 in human frontostriatal structure, function, and cognition. *J Clin Invest,* 117(3), pp672-82.

Michie, A. M. and Nakagawa, R. 2005. The link between PKCalpha regulation and cellular transformation. *Immunol Lett,* 96(2), pp155-62.

Millar, J. K., Wilson-Annan, J. C., et al. 2000. Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Hum Mol Genet,* 9(9), pp1415-23.

Millar, J. K., Pickard, B. S., et al. 2005. DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling. *Science,* 310(5751), pp1187-91.

Mira, M. T., Alcais, A., et al. 2004. Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature,* 427(6975), pp636-40.

Mirnics, K., Middleton, F. A., et al. 2001. Analysis of complex brain disorders with gene expression microarrays: schizophrenia as a disease of the synapse. *Trends Neurosci,* 24(8), pp479-86.

Miyazaki, T., Hashimoto, K., et al. 2006. Disturbance of cerebellar synaptic maturation in mutant mice lacking BSRPs, a novel brain-specific receptor-like protein family. *FEBS Lett,* 580(17), pp4057-64.

Moghaddam, B. and Krystal, J. H. 2003. The neurochemistry of schizophrenia In: Weinberger, D. R. ed. *Schizophrenia.* Second ed. Blackwell Science Ltd. pp. 349-364.

Moises, H. W., Zoega, T., et al. 2002. The glial growth factors deficiency and synaptic destabilization hypothesis of schizophrenia. *BMC Psychiatry,* 28.

Molina, V., Sanz, J., et al. 2005. Dorsolateral prefrontal cortex contribution to abnormalities of the P300 component of the event-related potential in schizophrenia. *Psychiatry Res,* 140(1), pp17-26.

Monni, O., Barlund, M., et al. 2001. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc Natl Acad Sci U S A,* 98(10), pp5711-6.

Moon, H. J., Yim, S. V., et al. 2006. Identification of DNA copy-number aberrations by array-comparative genomic hybridization in patients with schizophrenia. *Biochem Biophys Res Commun,* 344(2), pp531-9.

Moriguchi, S., Han, F., et al. 2006. Decreased calcium/calmodulin-dependent protein kinase II and protein kinase C activities mediate impairment of hippocampal long-term potentiation in the olfactory bulbectomized mice. *J Neurochem,* 97(1), pp22-9.

Morris, D. W., Rodgers, A., et al. 2004. Confirming RGS4 as a susceptibility gene for schizophrenia. *Am J Med Genet B Neuropsychiatr Genet,* 125(1), pp50-3.

Moskvina, V., Holmans, P., et al. 2005. Design of case-controls studies with unscreened controls. *Ann Hum Genet,* 69(Pt 5), pp566-76.

Moynihan, L. M., Bundey, S. E., et al. 1998. Autozygosity mapping, to chromosome 11q25, of a rare autosomal recessive syndrome causing histiocytosis, joint contractures, and sensorineural deafness. *Am J Hum Genet,* 62(5), pp1123-8.

Mueller, J. C. 2004. Linkage disequilibrium for different scales and applications. *Brief Bioinform,* 5(4), pp355-64.

Munafo, M. R., Thiselton, D. L., et al. 2006. Association of the NRG1 gene and schizophrenia: a meta-analysis. *Mol Psychiatry,* 11(6), pp539-46.

Murphy, K. C. and Owen, M. J. 2001. Velo-cardio-facial syndrome: a model for understanding the genetics and pathogenesis of schizophrenia. *Br J Psychiatry,* 179397-402.

Mutsuddi, M., Morris, D. W., Waggoner, S. G., Daly, M. J., Scolnick, E. M., Sklar, P. Analysis of high-resolution HapMap of DTNBP1 (Dysbindin) suggests no consistency between reported common variant associations to schizophrenia. *Am J Hum Genet.* Advance Online Publication.

Myers, S. R. and McCarroll, S. A. 2006. New insights into the biological basis of genomic disorders. *Nat Genet,* 38(12), pp1363-4.

Nagase, T., Kikuno, R., et al. 2000. Prediction of the coding sequences of unidentified human genes. XVIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res,* 7(4), pp273-81.

Neale, B. M. and Sham, P. C. 2004. The future of association studies: gene-based analysis and replication. *Am J Hum Genet,* 75(3), pp353-62.

Newman, T. L., Tuzun, E., et al. 2005. A genome-wide survey of structural variation between human and chimpanzee. *Genome Res,* 15(10), pp1344-56.

Newton-Cheh, C. and Hirschhorn, J. N. 2005. Genetic association studies of complex traits: design and analysis issues. *Mutat Res,* 573(1-2), pp54-69.

NimbleGenSystemsInc. 2005.

North, B. V., Curtis, D., et al. 2005. Application of logistic regression to case-control association studies involving two causative loci. *Hum Hered,* 59(2), pp79-87.

Norton, N., Kirov, G., et al. 2002. Schizophrenia and functional polymorphisms in the MAOA and COMT genes: no evidence for association or epistasis. *Am J Med Genet,* 114(5), pp491-6.

Norton, N., Williams, N. M., et al. 2002. Universal, robust, highly quantitative SNP allele frequency measurement in DNA pools. *Hum Genet,* 110(5), pp471-8.

Numakawa, T., Yagasaki, Y., et al. 2004. Evidence of novel neuronal functions of dysbindin, a susceptibility gene for schizophrenia. *Hum Mol Genet,* 13(21), pp2699-708.

Nyholt, D. R. 2004. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet,* 74(4), pp765-9.

O'Donovan, M. C. and Owen, M. J. 2002. Basic molecular genetics In: Gottesman, I. I. ed. *Psychiatric Genetics & Genomics.* First ed. Oxford Medical Publications. pp. 3-29.

Owen, M. J., O'Donovan, M. C., et al. 2002. Schizophrenia In: McGuffin, P., Owen, M. J., et al. ed. *Psychiatric Genetics & Genomics.* First ed. Oxford Medical Publications. pp. 247-266.

Owen, M. J., Craddock, N., et al. 2005. Schizophrenia: genes at last? *Trends Genet,* 21(9), pp518-25.

Owen, M. J., O'Donovan, M. C., et al. 2005. Schizophrenia: a genetic disorder of the synapse? *Bmj,* 330(7484), pp158-9.

Pandey, G. N., Dwivedi, Y., et al. 2004. Decreased catalytic activity and expression of protein kinase C isozymes in teenage suicide victims: a postmortem brain study. *Arch Gen Psychiatry,* 61(7), pp685-93.

Pe'er, I., de Bakker, P. I., et al. 2006. Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nat Genet,* 38(6), pp663-667.

Peirce, T. R., Bray, N. J., et al. 2006. Convergent evidence for 2',3'-cyclic nucleotide 3'-phosphodiesterase as a possible susceptibility gene for schizophrenia. *Arch Gen Psychiatry,* 63(1), pp18-24.

Petrie, A. and Sabin, C. 2005. *Medical Statistics at a Glance.* Second ed. Blacwell Publishing Ltd.

Petryshen, T. L., Middleton, F. A., et al. 2005. Support for involvement of neuregulin 1 in schizophrenia pathophysiology. *Mol Psychiatry,* 10(4), pp366-74, 328.

Pittman, A. M., Myers, A. J., et al. 2005. Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration. *J Med Genet,* 42(11), pp837-46.

Poole, A. W., Pula, G., et al. 2004. PKC-interacting proteins: from function to pharmacology. *Trends Pharmacol Sci,* 25(10), pp528-35.

Pritchard, J. K., Stephens, M., et al. 2000. Inference of population structure using multilocus genotype data. *Genetics,* 155(2), pp945-59.

Pritchard, J. K. 2001. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet,* 69(1), pp124-37.

Puri, V., McQuillin, A., et al. 2006. 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.*

Rang, H. P., Dale, M. M., et al. 1999. Antipsychotic drugs In: Ritter, J. M. ed. *Pharmacology.* Churchill Livingstone. pp. 539-549.

Rao, J. S., Rapoport, S. I., et al. 2005. Decrease in the AP-2 DNA-binding activity and in the protein expression of AP-2 alpha and AP-2 beta in frontal cortex of rats treated with lithium for 6 weeks. *Neuropsychopharmacology,* 30(11), pp2006-13.

Redon, R., Ishikawa, S., et al. 2006. Global variation in copy number in the human genome. *Nature,* 444(7118), pp444-54.

Rees, M. I., Fenton, I., et al. 1999. Autosome search for schizophrenia susceptibility genes in multiply affected families. *Mol Psychiatry,* 4(4), pp353-9.

Reich, D. E. and Lander, E. S. 2001. On the allelic spectrum of human disease. *Trends Genet,* 17(9), pp502-10.

Reif, A., Fritzen, S., et al. 2006. Neural stem cell proliferation is decreased in schizophrenia, but not in depression. *Mol Psychiatry,* 11(5), pp514-22.

Riley, B., Asherson, P., et al. 2003. Genetics and schizophrenia In: Hirsch, S. R. and Weinberger, D. R. ed. *Schizophrenia.* Second ed. Blackwell Science Ltd. pp. 251-276.

Risch, N. 1990a. Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet,* 46(2), pp222-8.

Risch, N. 1990b. Genetic linkage and complex diseases, with special reference to psychiatric disorders. *Genet Epidemiol,* 7(1), pp3-16; discussion 17-45.

Roisin, M. P. and Barbin, G. 1997. Differential expression of PKC isoforms in hippocampal neuronal cultures: modifications after basic FGF treatment. *Neurochem Int,* 30(3), pp261-70.

Rosenberg, N. A. and Nordborg, M. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat Rev Genet,* 3(5), pp380-90.

Saarela, J., Kallio, S. P., et al. 2006. PRKCA and multiple sclerosis: association in two independent populations. *PLoS Genet,* 2(3), ppe42.

Saba, T. G., Montpetit, A., et al. 2005. An atypical form of erythrokeratodermia variabilis maps to chromosome 7q22. *Hum Genet,* 116(3), pp167-71.

Sachs, N. A., Sawa, A., et al. 2005. A frameshift mutation in Disrupted in Schizophrenia 1 in an American family with schizophrenia and schizoaffective disorder. *Mol Psychiatry,* 10(8), pp758-64.

Salyakina, D., Seaman, S. R., et al. 2005. Evaluation of Nyholt's procedure for multiple testing correction. *Hum Hered,* 60(1), pp19-25; discussion 61-2.

Sanchez-Perez, A. M. and Felipo, V. 2005. Serines 890 and 896 of the NMDA receptor subunit NR1 are differentially phosphorylated by protein kinase C isoforms. *Neurochem Int,* 47(1-2), pp84-91.

Sanders, A. R., Rusu, I., et al. 2005. Haplotypic association spanning the 22q11.21 genes COMT and ARVCF with schizophrenia. *Mol Psychiatry,* 10(4), pp353-65.

Sawyer, S. L., Mukherjee, N., et al. 2005. Linkage disequilibrium patterns vary substantially among populations. *Eur J Hum Genet,* 13(5), pp677-86.

Schaid, D. J. 2004. Evaluating associations of haplotypes with traits. *Genet Epidemiol,* 27(4), pp348-64.

Schumacher, J., Jamra, R. A., 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), pp203-7.

Schwab, S. G., Knapp, M., et al. 2003. Support for association of schizophrenia with genetic variation in the 6p22.3 gene, dysbindin, in sib-pair families with linkage and in an additional sample of triad families. *Am J Hum Genet,* 72(1), pp185-90.

Sebat, J., Lakshmi, B., et al. 2004. Large-scale copy number polymorphism in the human genome. *Science,* 305(5683), pp525-8.

Selzer, R. R., Richmond, T. A., et al. 2005. Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes Chromosomes Cancer,* 44(3), pp305-19.

Servin, B. and Stephens, M. 2007. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet,* 3(7), ppe114.

Sham, P. and McGuffin, P. 2002. Linkage and association In: Gottesman, I. I. ed. *Psychiatric Genetics & Genomics.* First ed. Oxford Medical Publications. pp. 55-73.

Shao, Y. and Ismail-Beigi, F. 2004. Control of Na+-K+-ATPase beta 1-subunit expression: role of 3'-untranslated region. *Am J Physiol Cell Physiol,* 286(3), ppC580-5.

Sharp, A. J., Locke, D. P., et al. 2005. Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet,* 77(1), pp78-88.

Shaw, C. J. and Lupski, J. R. 2004. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum Mol Genet,* 13 Spec No 1R57-64.

Shifman, S., Bronstein, M., et al. 2002. A highly significant association between a COMT haplotype and schizophrenia. *Am J Hum Genet,* 71(6), pp1296-302.

Shih, R. A., Belmonte, P. L., et al. 2004. A review of the evidence from family, twin and adoption studies for a genetic contribution to adult psychiatric disorders. *Int Rev Psychiatry,* 16(4), pp260-83.

Shimohama, S. and Saitoh, T. 1998. Differential expression of protein kinase C - alpha and -betaII in rat septum and changes following fimbria-fornix lesion. *Brain Res,* 781(1-2), pp343-7.

Sklar, P. 2002. Linkage analysis in psychiatric disorders: the emerging picture. *Annu Rev Genomics Hum Genet,* 3371-413.

Soares, J. C., Chen, G., et al. 2000. Concurrent measures of protein kinase C and phosphoinositides in lithium-treated bipolar patients and healthy individuals: a preliminary study. *Psychiatry Res,* 95(2), pp109-18.

Song, J. H., Waataja, J. J., et al. 2006. Subcellular targeting of RGS9-2 is controlled by multiple molecular determinants on its membrane anchor, R7BP. *J Biol Chem,* 281(22), pp15361-9.

Spence, J. E., Perciaccante, R. G., et al. 1988. Uniparental disomy as a mechanism for human genetic disease. *Am J Hum Genet,* 42(2), pp217-26.

Spielman, R. S., McGinnis, R. E., et al. 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet,* 52(3), pp506-16.

Srivastava, M., Pollard, H. B., et al. 1998. Mouse cytochrome b561: cDNA cloning and expression in rat brain, mouse embryos, and human glioma cell lines. *DNA Cell Biol,* 17(9), pp771-7.

St Clair, D., Blackwood, D., et al. 1990. Association within a family of a balanced autosomal translocation with major mental illness. *Lancet,* 336(8706), pp13-6.

St Johnston, D. 2005. Moving messages: the intracellular localization of mRNAs. *Nat Rev Mol Cell Biol,* 6(5), pp363-75.

Stallings, R. L., Nair, P., et al. 2006. High-resolution analysis of chromosomal breakpoints and genomic instability identifies PTPRD as a candidate tumor suppressor gene in neuroblastoma. *Cancer Res,* 66(7), pp3673-80.

Stefansson, H., Sigurdsson, E., et al. 2002. Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet,* 71(4), pp877-92.

Stefansson, H., Sarginson, J., et al. 2003. Association of neuregulin 1 with schizophrenia confirmed in a Scottish population. *Am J Hum Genet,* 72(1), pp83-7.

Stefansson, H., Helgason, A., et al. 2005. A common inversion under selection in Europeans. *Nat Genet,* 37(2), pp129-37.

Strachan, T. and Read, A. P. 2003a. *Human Molecular Genetics.* Third ed. BIOS Scientific Publishers Ltd.

Strachan, T. and Read, A. P. 2003b. Genetic mapping of mendelian characters *Human Molecular Genetics.* Third ed. BIOS Scientific Publishers Ltd. pp. 397-414.

Strachan, T. and Read, A. P. 2003c. Molecular pathology *Human Molecular Genetics.* Third ed. BIOS Scientific Publishers Ltd. pp. 461-485.

Stram, D. O. 2004. Tag SNP selection for association studies. *Genet Epidemiol,* 27(4), pp365-74.

Straub, R. E., MacLean, C. J., et al. 1995. A potential vulnerability locus for schizophrenia on chromosome 6p24-22: evidence for genetic heterogeneity. *Nat Genet,* 11(3), pp287-93.

Straub, R. E., Jiang, Y., 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), pp337-48.

Straub, R. E., MacLean, C. J., et al. 2002. Genome-wide scans of three independent sets of 90 Irish multiplex schizophrenia families and follow-up of selected regions in all families provides evidence for multiple susceptibility genes. *Mol Psychiatry,* 7(6), pp542-59.

Subramaniam, K., Chen, K., et al. 2004. The 3'-untranslated region of the beta2-adrenergic receptor mRNA regulates receptor synthesis. *J Biol Chem,* 279(26), pp27108-15.

Sullivan, P. F. 2005. The genetics of schizophrenia. *PLoS Med,* 2(7), ppe212.

Talkowski, M. E., Seltman, H., et al. 2006. Evaluation of a Susceptibility Gene for Schizophrenia: Genotype Based Meta-Analysis of RGS4 Polymorphisms from Thirteen Independent Samples. *Biol Psychiatry.*

Tan, S., Guo, J., et al. 2007. Retained introns increase putative microRNA targets within 3' UTRs of human mRNA. *FEBS Lett,* 581(6), pp1081-6.

Tang, J. X., Zhou, J., et al. 2003. Family-based association study of DTNBP1 in 6p22.3 and schizophrenia. *Mol Psychiatry,* 8(8), pp717-8.

Tang, J. X., Chen, W. Y., et al. 2004. Polymorphisms within 5' end of the Neuregulin 1 gene are genetically associated with schizophrenia in the Chinese population. *Mol Psychiatry,* 9(1), pp11-2.

Terwilliger, J. D. and Hiekkalinna, T. 2006. An utter refutation of the "Fundamental Theorem of the HapMap". *Eur J Hum Genet,* 14(4), pp426-37.

The International HapMap Consortium 2003. The International HapMap Project. *Nature,* 426(6968), pp789-96.

The International HapMap Consortium 2005. A haplotype map of the human genome. *Nature,* 437(7063), pp1299-320.

Thiselton, D. L., Webb, B. T., et al. 2004. No evidence for linkage or association of neuregulin-1 (NRG1) with disease in the Irish study of high-density schizophrenia families (ISHDSF). *Mol Psychiatry,* 9(8), pp777-83; image 729.

Thomson, P. A., Wray, N. R., et al. 2005. Association between the TRAX/DISC locus and both bipolar disorder and schizophrenia in the Scottish population. *Mol Psychiatry,* 10(7), pp657-68, 616.

Thomson, P. A., Christoforou, A., et al. 2007. Association of Neuregulin 1 with schizophrenia and bipolar disorder in a second cohort from the Scottish population. *Mol Psychiatry,* 12(1), pp94-104.

Tishkoff, S. A. and Verrelli, B. C. 2003. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu Rev Genomics Hum Genet,* 4293-340.

Tomas, C., Canellas, F., et al. 2006. Genetic linkage study for bipolar disorders on chromosomes 17 and 18 in families with a high expression of mental illness from the Balearic Islands. *Psychiatr Genet,* 16(4), pp145-51.

Tonini, G. P., Parodi, M. T., et al. 1991. Expression of protein kinase C-alpha (PKC-alpha) and MYCN mRNAs in human neuroblastoma cells and modulation during morphological differentiation induced by retinoic acid. *FEBS Lett,* 280(2), pp221-4.

Topol, E. J. and Frazer, K. A. 2007. The resequencing imperative. *Nat Genet,* 39(4), pp439-40.

Tosato, S., Dazzan, P., et al. 2005. Association between the neuregulin 1 gene and schizophrenia: a systematic review. *Schizophr Bull,* 31(3), pp613-7.

Tuzun, E., Sharp, A. J., et al. 2005. Fine-scale structural variation of the human genome. *Nat Genet,* 37(7), pp727-32.

Urban, A. E., Korbel, J. O., et al. 2006. High-resolution mapping of DNA copy alterations in human chromosome 22 using high-density tiling oligonucleotide arrays. *Proc Natl Acad Sci U S A,* 103(12), pp4534-9.

Valente, E. M., Abou-Sleiman, P. M., et al. 2004. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. *Science,* 304(5674), pp1158-60.

Van de Bor, V. and Davis, I. 2004. mRNA localisation gets more complex. *Curr Opin Cell Biol,* 16(3), pp300-7.

Van Den Bogaert, A., Schumacher, J., et al. 2003. The DTNBP1 (dysbindin) gene contributes to schizophrenia, depending on family history of the disease. *Am J Hum Genet,* 73(6), pp1438-43.

van den Oord, E. J., Sullivan, P. F., et al. 2003. Identification of a high-risk haplotype for the dystrobrevin binding protein 1 (DTNBP1) gene in the Irish study of high-density schizophrenia families. *Mol Psychiatry,* 8(5), pp499-510.

Vazza, G., Zortea, M., et al. 2000. A new locus for autosomal recessive spastic paraplegia associated with mental retardation and distal motor neuropathy, SPG14, maps to chromosome 3q27-q28. *Am J Hum Genet,* 67(2), pp504-9.

Waddington, J. L., Kapur, S., et al. 2003. The neuroscience and clinical psychopharmacology of first- and second-generation antipsychotic drugs In: Weinberger, D. R. ed. *Schizophrenia.* Second ed. Blackwell Science Ltd. pp. 421-441.

Walss-Bass, C., Liu, W., et al. 2006. A novel missense mutation in the transmembrane domain of neuregulin 1 is associated with schizophrenia. *Biol Psychiatry,* 60(6), pp548-53.

Walss-Bass, C., Raventos, H., et al. 2006. Association analyses of the neuregulin 1 gene with schizophrenia and manic psychosis in a Hispanic population. *Acta Psychiatr Scand,* 113(4), pp314-21.

Wang, X., He, G., et al. 2004. Association of G72/G30 with schizophrenia in the Chinese population. *Biochem Biophys Res Commun,* 319(4), pp1281-6.

Weerth, S. H., Holtzclaw, L. A., et al. 2006. Signaling proteins in raft-like microdomains are essential for Ca(2+) wave propagation in glial cells. *Cell Calcium.*

Weinberger, D. R. and Marenco, S. 2003. Schizophrenia as a neurodevelopmental disorder In: Weinberger, D. R. ed. *Schizophrenia.* Second ed. Blackwell Science Ltd. pp. 326-348.

Wellcome Trust Case Control Consortium 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature,* 447(7145), pp661-78.

Wetsel, W. C., Khan, W. A., et al. 1992. Tissue and cellular distribution of the extended family of protein kinase C isoenzymes. *J Cell Biol,* 117(1), pp121-33.

Williams, H. J., Glaser, B., et al. 2005. No association between schizophrenia and polymorphisms in COMT in two large samples. *Am J Psychiatry,* 162(9), pp1736-8.

Williams, N. M., Rees, M. I., et al. 1999. A two-stage genome scan for schizophrenia susceptibility genes in 196 affected sibling pairs. *Hum Mol Genet,* 8(9), pp1729-39.

Williams, N. M., Norton, N., et al. 2003. A systematic genomewide linkage study in 353 sib pairs with schizophrenia. *Am J Hum Genet,* 73(6), pp1355-67.

Williams, N. M., Preece, A., et al. 2003. Support for genetic variation in neuregulin 1 and susceptibility to schizophrenia. *Mol Psychiatry,* 8(5), pp485-7.

Williams, N. M., Preece, A., et al. 2004a. Identification in 2 independent samples of a novel schizophrenia risk haplotype of the dystrobrevin binding protein gene (DTNBP1). *Arch Gen Psychiatry,* 61(4), pp336-44.

Williams, N. M., Preece, A., et al. 2004b. Support for RGS4 as a susceptibility gene for schizophrenia. *Biol Psychiatry,* 55(2), pp192-5.

Williams, N. M., O'Donovan, M. C., et al. 2005. Is the dysbindin gene (DTNBP1) a susceptibility gene for schizophrenia? *Schizophr Bull,* 31(4), pp800-5.

Williams, N. M., Green, E. K., et al. 2006. Variation at the DAOA/G30 locus influences susceptibility to major mood episodes but not psychosis in schizophrenia and bipolar disorder. *Arch Gen Psychiatry,* 63(4), pp366-73.

Williams, N. M., O'Donovan M, C., et al. 2006. Chromosome 22 deletion syndrome and schizophrenia. *Int Rev Neurobiol,* 731-27.

Wing, J. K., Babor, T., et al. 1990. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Arch Gen Psychiatry,* 47(6), pp589-93.

Wing, J. K. and Agrawal, N. 2003. Concepts and classification of schizophrenia In: Weinberger, D. R. ed. *Schizophrenia.* Second ed. Blackwell Science Ltd. pp. 3-14.

Winterer, G. and Weinberger, D. R. 2004. Genes, dopamine and cortical signal-to-noise ratio in schizophrenia. *Trends Neurosci,* 27(11), pp683-90.

Wood, L. S., Pickering, E. H., et al. 2006. Significant Support for DAO as a Schizophrenia Susceptibility Locus: Examination of Five Genes Putatively Associated with Schizophrenia. *Biol Psychiatry.*

Xu, S. Z., Bullock, L., et al. 2005. PKC isoforms were reduced by lead in the developing rat brain. *Int J Dev Neurosci,* 23(1), pp53-64.

Xie, X. Lu., J.Kulbokas, E. J., Golub, T. R., Mootha, V., Lindblad-Toh, K., Lander, E., S.Kellis, M. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature*, 434(7034), pp338-45.

Yang, J. Z., Si, T. M., et al. 2003. Association study of neuregulin 1 gene with schizophrenia. *Mol Psychiatry,* 8(7), pp706-9.

Yau, D. M., Yokoyama, N., et al. 2003. Identification and molecular characterization of the G alpha12-Rho guanine nucleotide exchange factor pathway in Caenorhabditis elegans. *Proc Natl Acad Sci U S A,* 100(25), pp14748-53.

Ylstra, B., van den Ijssel, P., et al. 2006. BAC to the future! or oligonucleotides: a perspective for micro array comparative genomic hybridization (array CGH). *Nucleic Acids Res,* 34(2), pp445-50.

Zeggini, E., Rayner, W., et al. 2005. An evaluation of HapMap sample size and tagging SNP performance in large-scale empirical and simulated data sets. *Nat Genet,* 37(12), pp1320-2.

Zeggini, E., Weedon, M. N., et al. 2007. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science,* 316(5829), pp1336-41.

Zhang, F., Sarginson, J., et al. 2006. Genetic association between schizophrenia and the DISC1 gene in the Scottish population. *Am J Med Genet B Neuropsychiatr Genet,* 141(2), pp155-9.

Zhang, X., Tochigi, M., et al. 2005. Association study of the DISC1/TRAX locus with schizophrenia in a Japanese population. *Schizophr Res,* 79(2-3), pp175-80.

Zhao, J. H., Curtis, D., et al. 2000. Model-free analysis and permutation tests for allelic associations. *Hum Hered,* 50(2), pp133-9.

Zhao, X., Shi, Y., et al. 2004. A case control and family based association study of the neuregulin1 gene and schizophrenia. *J Med Genet,* 41(1), pp31-4.

Zhu, X., Fejerman, L., et al. 2005. Haplotypes produced from rare variants in the promoter and coding regions of angiotensinogen contribute to variation in angiotensinogen levels. *Hum Mol Genet,* 14(5), pp639-43.

Zody, M. C., Garber, M., et al. 2006. DNA sequence of human chromosome 17 and analysis of rearrangement in the human lineage. *Nature,* 440(7087), pp1045-9.

Zollner, S. and Pritchard, J. K. 2007. Overcoming the winner's curse: estimating penetrance parameters from case-control data. *Am J Hum Genet,* 80(4), pp605-15.

Zondervan, K. T. and Cardon, L. R. 2004. The complex interplay among factors that influence allelic association. *Nat Rev Genet,* 5(2), pp89-100.