Genetic Association Studies of Bipolar Disorder

Dr Nicholas James Bass

This thesis is submitted for the degree of Doctor of Medicine (MD) at University College London

2007

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

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

Statement of Conjoint Work

My work was very much part of a team effort. My specific contributions were:

1. UCL bipolar sample collection - recruitment, diagnostic interviewing, blood sampling and data management of 300 bipolar volunteers and over 100 control volunteers.
2. DNA preparation - extraction, quantification and plating out for PCR-based analysis.
3. Marker selection - from literature and using bioinformatic tools.
4. Microsatellite genotyping at BDNF, P2RX7 and COMT loci.
5. Microsatellite genotyping at DISC1 locus in conjunction with Cindy Yang.
6. SNP genotyping at BDNF and COMT loci using Taqman, Amplifluor and RFLP assays.
7. Statistical analysis of data.

Dr Cindy Yang presented some of the microsatellite data relating to the DISC1 locus in her MSc (Molecular Medicine) thesis 2005.

08/04/07.

Abstract

Bipolar disorder is a common and serious mental illness. The occurrence of mania is central to the diagnosis, but affected individuals typically also suffer episodes of depression.

The results of family, twin and adoption studies argue convincingly for genetic susceptibility to bipolar disorder. Linkage studies conducted at the Molecular Psychiatry Laboratory, UCL have previously implicated the regions 12q24, 21q22, 1q42 and 11p14-15 as harbouring susceptibly loci for bipolar disorder.

In this thesis I report fine mapping of the 12q24, 21q22 and 1q42 regions by linkage disequilibrium methods, employing a case-control design. For the 11p14-15 region association with the candidate gene BDNF was tested. I also present attempts to replicate findings of association at the genes DAOA and COMT, located in regions implicated by meta-analysis of the linkage data. I have attempted to put these investigations in context, necessitating consideration of the conceptual developmental of bipolar disorder, the classical techniques for assessing the genetic contribution to aetiology, and mapping strategies.

Fine mapping of the UCL linkage regions implicated two novel susceptibility loci and provided support for two previously identified loci. Association of multiple markers within a 180 kb region of 12q24.3 was found, implicating Slynar and LOC387895. Association was also found with two markers in the more centromeric gene P2RX7, previously implicated in a Canadian sample. Multiple associated markers were found on 21q22.3. Two candidate genes - C21orf29 and TRPM2 - were identified from this region. Initial efforts to fine map the 1q42 region suggested the involvement of the previously implicated DISC1 gene. However association was only found with a single marker. Although haplotypic association was found with BDNF, the complex structure of the microsatellite marker hindered interpretation of the results. Partial replication of the association with DAOA was achieved but the involvement of COMT was not supported.

Contents

List of figures

List of tables

Acknowledgements

I should like to thank the follwing people for their advice and support with this project:
Dr Gulsharan Kalsi
Professor David Curtis
Professor Hugh Gurling
Dr Li Chan
Mr Nick Dand

I should give a special mention to Dr. Andrew McQuillin for his support. I must also thank the Manic Depression Fellowship and all their members who so kindly volunteered for this study. Finally I thank Miss Ellie Bass for all her hard work proofreading.

Chapter 1: What is Bipolar Disorder?

In this opening chapter I will introduce bipolar disorder. I will stray from convention slightly in not taking a strictly chronological approach to the recognition and classification of this disease. Firstly, I will attempt to describe bipolar disorder as it is currently conceptualised. I will then go on to outline the clinical aspects of the disorder, discuss its place in the global burden of ill health and critically consider the efficacy of contemporary treatments. Finally, I will outline the historical development of bipolar disorder as a diagnostic entity.

1.1 Modern conceptualisation

Bipolar disorder, or manic depression as it is still sometimes referred to, is a disorder of mood. Despite the implications of both these terms, modern definitions demand the ocurrence of mania or hypomania but not necessarily depression.

Mania is a syndrome, the core feature of which is a sustained, abnormally elevated or irritable mood. However, changes can be observed across many domains: cognitive, behavioural, physiological and perceptual. Listed below are some of these changes, with accompanying sketches of typical clinical presentations. This is intended to give a flavour of how a patient suffering from a manic episode might appear:

Increased activity

The patient appears restless and their activity becomes unrestrained and urgent. In this state the patient is at risk of physical exhaustion. The patient may, for example, report starting a number of courses or setting up multiple businesses. Initially they may claim increased productivity but as the episode becomes more severe tasks are not finished and activity becomes chaotic.

Flight of ideas and pressure of speech

The patient's speech can be very rapid with increased volume. Accompanying this the patient may rapidly flit from one subject to another. In contrast to the formal thought disorder of schizophrenia one can appreciate the connection between the thoughts even if their rapidity makes them very difficult to follow.

Increased self-esteem and lack of judgment

Generally, patients will experience an increase in confidence and impaired judgment. This often affects social interactions, with the patient becoming over-familiar or indiscreet. The patient may engage in reckless activities, such as driving at high speed. Often there will be a period of profligacy with money.

Somatic symptoms

Somatic symptoms are commonly referred to in the context of depressive disorder. In an analogous but reverse way there are often somatic perturbations in the manic syndrome. One of the earliest clinical signs of mania is decreased need for sleep. It is also common to experience increased libido and appetite.

Clearly many of the symptoms described above are interrelated. However, the relationship between them at an aetiological level remains unknown. Although necessary for a diagnosis of bipolar disorder, it is not established that the mood change is causative of, say, the phenomenon of flight of ideas. Interestingly, patients may be able to exert some control over symptoms for short periods by effort of will, such as in the context of a clinical interview.

Psychotic symptoms, i.e. symptoms in which there is a loss of contact with reality, may also be present during an episode of mania. Essentially these are hallucinations or delusions - fixed beliefs held in contrariness to the evidence. They are usually, although not exclusively, mood congruent, i.e. in keeping with the prevailing mood state of elation. A typical example would be the grandiose delusion of divine birth.

It is now recognised that the psychotic symptoms previously associated with schizophrenia, Schneider's first-rank symptoms, also occur in approximately 20% of manic patients (Gonzalez-Pinto *et al.* 2003). These symptoms include certain types of auditory hallucinations, somatic hallucinations, delusions of control, thought broadcasting and delusional perception. In the case of mania, however, these are transitory phenomenon occurring at the peak of a manic episode.

Insight, often spoken of in psychiatry, refers to the degree to which the patient realises that they are unwell. In patients with mania insight into their problem is invariably impaired, and may be completely absent in the more severe stages. This has obvious implications for the treatment and management of such patients.

Whilst I have described above the typical presentation of a manic episode, some variations can be seen. Mania can, for example, present as manic stupor. This is when patients present in a mute and immobile state but subsequently report that they experienced racing thoughts. However, this condition is now rarely seen, probably due to better surveillance and modern treatments.

I have described above the presentation of mania from the clinician's viewpoint. Below I have inserted a more personal description of mania. Here, one of the bipolar individuals contributing to my study describes the manic behaviour of his girlfriend and fellow sufferer in a book he wrote about his experience of the illness:

> When [her] mania was at its peak, she became so uninhibited and carefree that at such times she put herself at very considerable risk... Anything could happen... She was walking down the middle of the road, drinking a can of Coke. There were cars racing past her in both directions.

I have not referenced this book in order to maintain the participant's anonymity.

Depending on the classification system used a diagnosis of bipolar disorder can be ascribed following a single episode of mania. However, the usual presentation of bipolar disorder is one of recurrent episodes of mania and depression. Indeed, it has been found that 90% of patients experiencing mania will also experience depression (Goodwin and Jamison 1990).

In Western culture, depression is a more familiar concept than mania. I will not, therefore, enter into a detailed explanation here. However, it is important to draw a distinction between transitory feelings of depression, universally experienced in relation to adverse events, and an episode of depression. The core feature of a depressive disorder is persistent low mood. This is typically evident as anhedonia (lack of enjoyment), low energy, slowness and a negative mindset accompanied by the somatic symptoms of disturbed sleep, poor appetite with weight loss, loss of libido and diurnal variation of mood. As is apparent, many of the symptoms are the opposite of those found in mania.

There is considerable variation in the clinical presentation of depression, affected both by cultural and demographic factors. In China, for example, presentation with somatic complaints is more common than in the West (Gelder *et al.* 2001). It would seem that there are no major differences in the symptom profiles of unipolar and bipolar depression (Mansell *et al.* 2005). However, there is a suggestion that psychomotor retardation and the symptoms of atypical depression, namely hyperphagia, hypersomnia and laden paralysis, are particularly associated with bipolar depression.

Here the same volunteer describes his own experience of a depressive episode:

> I know I'm unable to live life in the same apparent matter of fact way that I see most people doing. My whole existence terrifies me. I try to retreat from it even to the point of wishing to end it all.

Diagnosis of bipolar disorder remains purely clinical, as reliable biochemical or genetic markers have yet to be established. The presentation of mania can be very dramatic, with little diagnostic difficulty. However, there can be initial difficulty differentiating bipolar

disorder from schizophrenia when psychotic features are prominent. Diagnostic uncertainty can lead to sub-optimal treatment due to the differing treatments for the two disorders.

The two main psychiatric classification systems currently in use are International Statistical Classification of Diseases, 10th edition, ICD-10, (WHO 1993) and Diagnostic and Statistical Manual of Mental Disorders, 4th edition, DSM-IV(APA 1994). These set out operational criteria by which to diagnose bipolar disorder. The criteria are similar, but in DSM-IV the occurrence of one manic episode is sufficient for a diagnosis of bipolar disorder while another mood disturbance is required for ICD-10. However, if the manic or hypomanic episode is precipitated by treatment it is not counted in DSM-IV. Important exclusion criteria include mood disturbance attributable to psychoactive substance use or any organic mental disorder.

DSM-IV also makes the distinction between bipolar I and II, which will be returned to in discussing the historical development of the diagnosis of bipolar disorder in section 1.5. Personal distress and social dysfunction are important criteria in differentiating bipolar I from bipolar II.

In addition, variants of bipolar disorder are recognised: rapid cycling and mixed affective states. Rapid cycling is defined as at least four distinct episodes of mood disturbance in one year, separated by periods of remission. Rapid cycling has some interesting features. It is associated with hypothyroidism, can be triggered by antidepressant drugs and is relatively unresponsive to lithium treatment. Mixed affective states occur when symptoms of depression and mania co-exist temporally, or follow one another in very quick succession. These states also have some specific correlations with drug response.

1.2 Clinical aspects

I will not describe in detail the clinical aspects of bipolar disorder. Instead, I will give a sketch of the natural history of the disorder. Whilst I present generally accepted figures, it

should be stated that there is considerable variability in estimates of epidemiological facets of the disorder.

A lifetime risk of 0.3-1.5% for bipolar disorder was found in an epidemiological study covering 10 countries from different global regions (Weissman *et al.* 1996). The same study showed that men and women were affected equally. There are a number of controversies over gender differences in bipolar disorder, but it does seem that rapid cycling occurs more commonly in women (Gelder *et al.* 2001).

The mean age of onset of bipolar disorder is approximately 21 years. This is some years earlier than unipolar depression. The average length of a manic episode is about 6 months. Multiple episodes are the norm (Gelder *et al.* 2001). The length of the interval between successive episodes decreases and frequency increases (Angst and Sellaro 2000). There is high comorbidity of bipolar disorder with anxiety disorders and substance use disorder (Krishnan 2005).

A number of prognostic factors have been identified in relation to bipolar disorder, many of which are common to a range of psychiatric conditions. Previous episodes and early onset are the main risk factors. The risk of future episodes is also increased with poor social support, poor physical health, comorbidity, substance misuse, personality disorder and high trait neuroticism. Bipolar II has a relatively good prognosis, while the prognosis for rapid cycling is poor (Gelder *et al.* 2001).

1.3 Is bipolar disorder important?

I will consider the importance of bipolar disorder from two viewpoints: global and personal. To take the former, it is evident from lifetime risk rates of around 1% that bipolar disorder affects a considerable number of people worldwide. The Global Burden of Disease Study used Disability Adjusted Life Years (DALYs) to measure disease burden for many conditions (Murray and Lopez 1997). DALYs are a composite measure of the sum of life years lost to premature death and years lived with disability adjusted for

severity. Out of the 107 disorders considered, bipolar disorder ranked 22nd. According to this study, bipolar disorder accounted for the loss of 1.43×10^7 DALYs in 1990. It has been estimated that the total cost of bipolar disorder in the UK is two billion pounds sterling per year (Das Gupta and Guest 2002).

The personal cost of bipolar disorder can be considered from the perspectives of both mortality and morbidity. Simply put, if you have a diagnosis of bipolar disorder you are more likely to die. Harris and Barraclough (Harris and Barraclough 1997) demonstrated a Standardised Mortality Ratio (SMR) of 15.0 for suicide with bipolar disorder. More recently, Osby (Osby *et al.* 2001) found an all causes SMR of 2.5 for males and 2.7 for females. A much greater risk of suicide was found in females with an SMR of 22.4 compared with 15.0 in males. There has been significant variability in estimates of rates of completed suicide in people with bipolar disorder, ranging between 4% and 19% (Dalton *et al.* 2003).

In terms of morbidity, it is evident from the above clinical characteristics that bipolar disorder generally affects people early in life, involves prolonged periods of illness and is likely to recur and worsen. Although virtually all patients recover from acute episodes, only around 20% have prolonged periods of occupational and social stability (Gelder *et al.* 2001). This is reflected in a number of measures of reduced social functioning. Bipolar sufferers have higher rates of unemployment, increased use of health services and increased marital dysfunction (Sajatovic 2005).

Furthermore, evidence seems to be accumulating of a link between bipolar disorder and cognitive dysfunction (Robinson and Ferrier 2006). However, the link between bipolar disorder and creativity has long been discussed and has recently received some support from empirical evidence (Santosa *et al.* 2006).

Patients suffering manic episodes often spend protracted periods in hospital. Compulsory admission under the mental health act is frequently required. Analysis of patients in Psychiatric Intensive Care Units has revealed that bipolar cases constitute a large

proportion of such admissions, e.g. (Brown and Bass 2004). Use of the Mental Health Act and treatment in more restrictive environments can be particularly distressing to both the patient and carer.

As alluded to above, the impact of bipolar disorder is not necessarily restricted to patients but can also affect friends and family.

1.4 Is bipolar disorder treatable?

In this section, I will briefly comment on the treatments currently available for bipolar disorder. There are important differences in the treatment of the manic and depressive phases of the illness, although some pharmacological agents can be thought of as treatment of bipolar disorder *per se*. I will consider first the treatment of mania, then that of depression and finally maintenance therapy.

Whilst patients experiencing a manic episode may not feel that they need treatment, they often present a risk to themselves, e.g. through risk-taking behaviour. Furthermore, they may pose a risk to others or find themselves at risk from others. Prior to the development of modern pharmacological agents 'manic exhaustion' was a commonly stated cause of death. However, it has been questioned whether this existed as a clinical entity or whether it arose because of physical causes of the mania (Derby 1933). Due to the high level of risk involved, vigorous treatment of mania is required.

Currently, treatment of mania - in terms of trying to alter the aberrant mental state - is essentially pharmacological. The social intervention of hospital admission could also be viewed as treatment. Since its discovery by Cade in 1947, lithium has been considered the first line treatment of bipolar disorder. Indeed, a meta-analysis by Poolsup (Poolsup *et al.* 2000) suggests that lithium is still the agent of choice. In the case of lithium, the number needed to treat (NNT) was calculated as five and would thus generally be considered an effective treatment. By way of comparison with treatments in other branches of medicine, triple therapy for H. Pylori has a roughly equivalent NNT

(McQuay and Moore 1997). Similar response rates were found for the other mood stabilisers sodium valporate and carbamazepine. Carbamazepine may have an advantage in rapid cycling disorder, while valporate may act more swiftly. More recently, lamotrigine and gabapentin have come into clinical use.

Antipsychotics are also widely used in the clinical treatment of mania. The typical antipsychotics haloperidol and chlorpromazine have been demonstrated to be effective in the treatment of mania by randomised control trials. A recent meta-analysis of atypical antipsychotics - including aripiprazole, olanzapine, quetiapine, resperidone and ziprasidone - concluded that they were all superior to a placebo, but only small differences in efficacy were observed between them (Perlis *et al.* 2006). Benzodiazepines are widely used as an adjunct treatment for behavioural disturbance, although they carry the risk of disinhibition. ECT is probably as, if not more, effective a treatment for mania than the pharmacological agents outlined above, yet this procedure is rarely used now in clinical practice.

Treatment of depression as part of bipolar disorder does not differ substantially from that of unipolar depression. Whilst there is an ever-expanding pharmacopeia for the treatment of depression, psychological interventions have also been shown to be effective. Overall, antidepressants are effective treatments for moderate and severe depression with an NNT of 3-4 (Gelder *et al.* 2001). There are now three main classes of anti-depressants: tricyclic antidepressants (TCA), monoamine oxidase inhibitors (MAOI), and serotonin specific reuptake inhibitors (SSRI), plus other miscellaneous antidepressants. Efficacy is similar between antidepressants, although there are notable differences in the side effect profiles. SSRIs are generally used as first line agents because of their higher tolerability and safety in overdose. Whilst it is important to treat depression in bipolar disorder, as it causes much suffering and brings with it the risk of suicide, particular care has to be taken due to the danger of antidepressant induced mania.

For those who do not respond to single agent antidepressants there are a number of effective augmentation strategies available. For instance, a NNT of 3.7 is obtained with

lithium augmentation in non-responders to antidepressants. ECT may be more effective in severely depressed patients than antidepressants. However, as mentioned above, the clinical use of ECT remains limited. There has been considerable evaluation of psychological treatments in depression, and for mild to moderate forms of the disorder CBT is as effective as antidepressants. Interpersonal therapy is also effective (Gelder *et al.* 2001).

Evidence suggests that some pharmacological agents modify the natural history of BP. Many patients are therefore on long-term maintenance medication. It would appear that the mood stabilising drugs - valporate, carbamezipine and lithium - are broadly similar in their prophylactic effect. Most experience has been gained with lithium, with about 50% of patients responding positively. It appears equally effective in the prevention of mania and depression. Importantly, lithium prophylaxis seems to lead to a decrease in mortality from suicide. Currently, the prophylactic effects of lamotrigine, gabapentin, olanzpine and rispedone are being studied. The role for psychological interventions in maintenance appears limited. There is, however, evidence that relapse rates are reduced by education, particularly when this is geared to recognising the warning signs of relapse, thus enabling prompt treatment.

The agents used, particularly in the treatment and maintenance of acute mania, are not well liked by patients. For example lithium, the most commonly used prophylactic agent for bipolar disorder in the UK, is associated with serious side effects. Increased rates of hypothyroidism have been found, particularly in women (Freeman and Freeman 2006). Polyuria secondary to lithium-induced anti-diuretic hormone insensitivity is seen in a significant proportion of people. As lithium has a narrow therapeutic index, there is the danger of lithium toxicity, a condition that is potentially fatal. Risk of lithium toxicity is much increased by a state of dehydration, induced, for example, by diarrhoea and vomiting. Monitoring of these problems has improved, for example through the setting up of lithium clinics, although there remains the need for repeated blood tests. In addition, there are a number of minor side effects, which, nonetheless, can have a significant

impact on quality of life. Patients often describe weight gain, tremor and dulling of mental faculties.

Thus it can be seen that useful, although not brilliant, treatments are available. Whilst some of the response rates are reasonable, care should be taken in their interpretation, as response is rarely defined as a return to the pre-morbid state. Better service provision and standardisation of best treatment practices may improve the outcome in bipolar disorder. This has been the guiding principle behind the development of the National Institute for Clinical Excellence guidelines for bipolar disorder, published in July 2006. However, even with the perfect implementation of current treatment methods there is still a pressing need for the development of better treatments.

1.5 Historical perspective

I wish to consider the historical emergence of depression and mania, both as distinct clinical entities and as a unitary condition. In so doing, I hope to shed light on how the relationship between mania and depression has been perceived historically. Constructing a history of disease is necessarily complex. My aim in this section is therefore modest: to outline a working chronology and highlight salient figures in the history of the diagnosis and classification of bipolar disorder.

Clinical descriptions, recognisable as depression and mania to the contemporary psychiatrist, date back to Antiquity. The first systematic clinical descriptions of melancholia and mania were by Hippocrates (460-337 BC) (Angst and Marneros 2001). However, it was Aretaeus of Cappadocia (150-200 AD) who perceived a relationship between melancholia and mania, recognising that the two states could occur in the same individual. Indeed, Angst maintains that Aretaeus' recognition of a connection between mania and melancholia "can justifiably be considered the first conception of bipolarity", whilst Porter (Berrios and Porter 1995) more reservedly remarks that Aretaeus may have had "an inkling of manic-depressive disorders".

Aetiological theories for both melancholia and mania were developed within the theoretical framework of humoralism, established by physicians in the Classical period and remaining the dominant tradition in Europe throughout the Middle Ages. Simplistically stated, excess blood and yellow bile could lead to mania and excess black bile to melancholy. The treatments of blood letting and diet to balance the humors were thus prescribed in accordance with the model, for instance by Robert Burton in his Anatomy of Melancholy (1621).

Although various writers continued to note the association of melancholia and mania during the 17th and 18th Centuries, it was not until the 19th Century that significant headway was made in the classification of these disorders. Indeed, Berrios argues that the modern concepts of depression and mania arose at this time, developing out of the previous broader concepts of melancholia and mania (Berrios and Porter 1995). Melancholia, previously thought to be a disease of the intellect, was redefined as a primary disorder of emotion. The term depression, taken from cardiovascular medicine, gradually infiltrated medical discourse during the second half of the 19th Century. Mania was similarly refigured as a disorder of emotion and behaviour. It was against this background of conceptual clarification that a new combined concept could emerge.

It was within the tradition of French psychiatry, which had become predominant in Europe through the promotion of careful observation of psychopathology, that the concept of mania and melancholy as a dual disorder took systematic form. Such careful observation over a prolonged period of time led the pre-eminent French psychiatrist Falret to identify a condition he termed "folie circulaire" in 1851. He described a single condition constituting a continuous cycle of melancholy and mania with periods of wellness in between. It was possible that Falret was influenced by the views of German psychiatrist Griesinger, who in 1845 had used the image of a circle to describe the relation of mania and melancholy (Angst and Marneros 2001).

In 1854, the same year that Falret published his theory in full, Jules Baillarger presented his own unitary concept of mania and melancholia, which he referred to as "folie a

double forme". Baillarger's views differed from those of Falret, as he recognised that the symptoms alternated, but accorded no importance to the intervening periods of wellness. In contrast, these intervals were central to Falret's concept of a circular disorder (Angst and Marneros 2001).

These developments found a receptive audience, both within France and throughout Europe. The German psychiatrist Karl Kahlbaum, himself significant in the development of the concept of a single disorder, used both the terms of Falret and Baillarger in his book of 1863 *The Grouping and Classification of Mental disorders* (Angst and Marneros 2001). In France, other descriptions of bipolar disorder were proposed, for example "a double phase" (Billod, 1856) and "folie alterne" (Delaye and Legrand du Salle). It was not until 1883 that some conceptual clarity was afforded by the prize-winning monographs of Ritti and Mordret (Berrios and Porter 1995).

The next significant stage in classification was due to Kraepelin. He coined the term "manic depressive insanity" in 1899. This concept unified the affective disorders, including both the circular and depressive types. A broader category was thus created. More importantly, Kraepelin also differentiated manic depressive insanity, as a form of psychosis, from dementia praecox (Berrios and Porter 1995). This Kraepelinian dichotomy has had a fundamental influence on the search for genes, a topic that will be revisited at various points in the thesis.

Kraepelin's unified view of the affective disorders faced considerable opposition, particularly in Germany. Karl Kleist (1953) challenged Kraepelin's concept of manic-depressive disorders, believing that people who suffered from both manic and depressive episodes were suffering from a different illness to those that experienced only one type of affective state. Kleist's views were in all likelihood influenced by those of his senior colleague Carl Wernicke, who was likewise of the opinion that single or recurrent episodes of either depression or mania should be differentiated from the disorder involving both of these states. It is to Kleist that we owe current nomenclature of "einpolig" (unipolar) and "zweipolig" (bipolar) disorder. Kleist's ideas underwent further refinement by another of his colleagues, Karl Leonhard (Angst and Marneros 2001).

As was mentioned at the start of this chapter, bipolar disorder has become something of a misnomer according to ICD-10 and DSM-IV definitions, as it is mania that is the defining feature of the syndrome. Monomania is classified as bipolar disorder and unipolar disorder now refers to depression. These changes are largely based on the work of Angst, Perris and Winkour in the 1960's, who brought a genetic epidemiological approach to the classification table. The studies of Angst (1959-1963) and Perris (1963-1966) showed that unipolar mania was strongly related in genetic terms to "bipolar disorder" and thus should be more properly included in that category (Angst and Marneros 2001).

In DSM-IV bipolar disorder is split into Bipolar I and II, with Bipolar II being defined as a milder form of Bipolar I. In Bipolar II, people experience hypomania rather than mania. In essence, people with hypomania have fewer and less severe symptoms than those with mania, but the division does seem rather arbitrary. Although Hippocrates used the term, hypomania only really came to be recognised as a different form of mania through the work of Mendel (1881) and has gained in usage over recent years due to the descriptions of Bipolar II. The nosology of Bipolar II is due to Dunner (1976) (Angst and Marneros 2001).

Accompanying this division of bipolar disorder by severity, investigation has taken place into the relationship between bipolar disorder and schizoaffective disorders. It was Kahlbaum who first described schizoaffective disorders as a separate group (1863). Subsequently Kraepelin was exercised by the problem of cases which fell between his categories of manic depressive insanity and dementia praecox (1920). Schizoaffective psychosis was a term coined by Kasanin in 1933. Research since the 1960s has indicated a strong relationship between schizoaffective disorders and affective disorders. Cadoret (1974) separated schizoaffective disorders into schizoaffective bipolar and schizoaffective unipolar. Based on data that relatives of bipolar probands are at elevated risk of schizoaffective bipolar disorder, Marneros (1999) has argued that schizoaffective bipolar disorder should be included in the bipolar group (Angst and Marneros 2001).

Hence in our studies patients with schizoaffective bipolar disorder have been included in the case group.

The development of the concept of bipolar disorder appears to have followed a course of contraction and expansion. This process has involved the gradual refinement of categories - most notably during the 19th Century when the traditional, broader concepts of mania and melancholia were remodelled and the concept of a circular disease took hold. The concept expanded again with the inclusive approach of Kraeplin, and narrowed in opposition to this. In more recent years, there has been increasing interest in subdividing this complex phenotype.

The above description of the development of the working concept of bipolar disorder should not, however, be seen as the sole historical narrative. There has been a parallel movement to consider bipolar disorder as on a continuum from affective disorder to schizophrenia (Berrios and Porter 1995). I will return to this idea of 'unitary psychosis' later.

1.6 Summary

The observation that states of depression and mania could occur in the same individual appears to be long-standing, arguably stretching back to the Classical period. However, almost two thousand years later there is still no unanimous agreement on how these patterns of mood disturbance relate to assumed underlying pathological entities. Leaving aside for a moment the complex classification problems that stem from this uncertainty, bipolar disorder, as currently defined, is a severe psychiatric disorder, which affects approximately 1 in 100 people.

Although there are a number of effective treatments, the global burden of disease remains great. Although the bipolar pharmacopeia has enlarged, no major treatment breakthroughs have been made for 50 years. The development of new treatment and

prevention strategies is crucially impeded by our relative lack of understanding of the basic biology underpinning this disorder.

It is beyond the scope of this thesis to comprehensively review all the biological findings in bipolar disorder – hormonal, neuro-anatomical, neuro-imaging, pharmacological etc. However, I think that it is justified to say that despite much research spanning many disciplines the biological perturbations in bipolar disorder are poorly understood.

In general, the elucidation of the underlying genetic pathology has greatly enhanced our understanding of human disease, if not its treatment. For example, identification of the loci for Duchene Muscular Dystrophy has led to the characterisation of the dystrophin protein and given a molecular explanation for the progressive muscular wasting that is seen clinically. Hence, it would seem a worthy endeavor to try and understand the genetic pathology of bipolar disorder.

Such genetic understanding may allow the classification system to be rationalized, placing it on a more robust aetiological footing. Likewise, it should inform investigations in other disciplines and help clarify the back catalogue of biological findings. In turn, it is hoped that this knowledge will greatly facilitate the rational design of novel treatments and the tailoring of treatments according to predicted response. Furthermore, the potential exists for genetic advances to have a major impact on the developments of strategies to prevent, diagnose and monitor bipolar disorder.

The idea that bipolar disorder has a genetic aetiology is not a recent one. Falret (1854) observed that "circular insanity" seemed to run in families (Maj *et al.* 2002). I hope to demonstrate in the next section that there are compelling epidemiological arguments for searching for genes of aetiological importance.

Chapter 2: Is Bipolar Disorder Genetic?

Before launching into a discussion of the evidence for the genetic aetiology of bipolar disorder, which will be referred to henceforth as BP, I will pause to consider the main methodological problem in examining the role of genes in BP. This is the problem of the definition of 'caseness' – in other words, the difficulties involved in accurately identifying and defining a category of cases.

2.1 The problem of phenotype definition

The definition of caseness is crucial to all forms of genetic investigation. Ideally, the definition would include all cases of a particular genetic aetiology whilst excluding any individuals who had a different genetic aetiology but shared the phenotype. This situation is approached in some simple Mendelian conditions, such as Huntington's Disease. However, even in simple Mendelian disorders genetically homogenous categories often cannot be defined *a priori* based on clinical phenotypes alone.

As was seen in Chapter 1, there has been a great struggle to precisely define BP. Genetic epidemiology has informed modern definitions of BP and supports their general validity for genetic studies. However the precise relationship of the BP diagnostic grouping to putative underlying genetic aetiologies is not yet known. Any deviation from a perfect correlation between genetic aetiology and clinical phenotype will have implications for genetic studies. An over-inclusive category would make it harder to find the aetiology of a particular case whilst an under-inclusive category would mean that not all cases would be identified, which would also hamper gene discovery. This issue is relevant to segregation analysis, twin and adoption studies (discussed below) as well as the mapping strategies discussed in subsequent chapters.

Even if conceptually imperfect, the definition of BP must be consistently applied if meaningful conclusions are to be drawn from genetic studies. From around the 1940's onwards it was increasingly recognised that there was considerable diagnostic variability between clinicians. For example, Beck (Beck *et al.* 1962) found only 54% agreement on specific diagnoses between psychiatrists independently examining a series of 153 patients. This provided the impetus for the development of operational criteria to improve inter-rater reliability of diagnoses.

Initially the St. Louis group developed criteria for 15 psychiatric disorders, called the Feighner Criteria (Feighner *et al.* 1972). With the application to research studies in mind, Spitzer (Spitzer *et al.* 1978) published the Research Diagnostic Criteria (RDC) which extended and modified the Feighner Criteria, and claimed a high reliability. Very shortly after, Endicott (Endicott and Spitzer 1978) published a semi-structured interview based on the RDC: the Schedule of Affective Disorders and Schizophrenia - Lifetime Version (SADS-L). The use of a structured or semi-structured interview further increased reliability (Helzer *et al.* 1977), aiding the objective and comprehensive acquisition of clinical data and facilitating the consistent application of this data to the diagnostic criteria. I should just mention that DSMIII (Diagnostic Statistical Manual of Diseases, 1980), which marked a major change in psychiatric classification systems, can be seen as the child of the RDC, but broadening its perspective to clinical as well as research applications.

The subsequent history of the development of other diagnostic instruments such as the: Diagnostic Interview Schedule (DIS), Mini Neuropsychiatric Interview (MINI), Structured Clinical Interview for DSM-IIIR (SCID), Schedules for Clinical Assessment in Neuropsychiatry (SCAN) and Diagnostic Interview for Genetic Studies (DIGS) is rather involved. However, in essence they have evolved to reflect the changes in operational definitions embodied in subsequent versions of DSM (DSM-IIIR and DSM-IV) and ICD-10. Broadly speaking, diagnostic concordance between the SADS-L and other interviews has been found to be good (Hesselbrock *et al.* 1982; Farmer *et al.* 1993).

While it is not clear to me that there is an appreciable difference in the validity of the various interviews in their application to genetic studies, it has been shown that the definition of caseness can have a major impact on the outcome of genetic studies in psychiatry (McGuffin *et al.* 1991). An attempt has been made to circumvent the problem of 'top down' diagnostic systems by developing an operational criteria checklist for psychotic illness (OPCRIT) in the context of collaborative genetic studies, which could be applied to a number of diagnostic systems (McGuffin *et al.* 1991).

2.2 Evidence for genetic aetiology to bipolar disorder

Despite more than 150 years having elapsed since it was observed that BP runs in families, a genetic contribution to the aetiology has yet to be unequivocally proven.

To understand the seemingly slow progress in unravelling the genetics of BP, general consideration of how it is known that a disease is genetic has to be made. That a disease phenotype follows a Mendelian pattern of inheritance through families, i.e. autosomomal dominant, co-dominant, recessive or X-linked, is strong evidence that the disease has a genetic basis. Mendelian inheritance, such as the vertical transmission affecting both males and females seen in autosomal dominant disorders, may be inferred by inspection of the pedigree alone, e.g. Huntington's Chorea.

However, in less clear-cut cases formal analysis of the segregation ratios within families can be performed to test whether a Mendelian pattern of inheritance is operating. In families where BP recurs, the segregation ratios do not perfectly conform to Mendelian patterns. This has provoked investigations to test the degree of genetic contribution to the aetiology. Classically, three approaches have been employed for this purpose: family, twin, and adoption studies. In this section I will outline the theoretical construct behind each of these approaches and then briefly review the findings in BP. I will also mention in passing some of the major criticisms of these methods.

2.2.1 Family studies

Family studies are studies of the risk of disease(s) in relatives of the affected subject. Of primary interest will be the disease under scrutiny, but it is also usual to look for changes in the risk of other diseases in relatives. Applying this to BP, it is typical that once the index case of BP is identified relatives are screened for psychiatric illness and morbid risk thus determined. The risk estimate can then be compared with the population risk of the disease or the risk derived from relatives of control probands to obtain a relative risk value. This approach can be thought of as formally testing whether a disease is familial. However, one cannot infer with certainty whether this familiarity is due to genetic factors or to the shared environment within the family. However, family studies may give an indication of the genetic relationship between different diseases.

Early family studies made no differentiation between BP and depression and did not routinely assess psychiatric comorbidity (Althoff *et al.* 2005). Therefore little can be inferred about familial risk of BP from these studies. A meta-analysis of eight family studies where such a differentiation was made yielded an odds ratio of 7 for BP in first-degree relatives of BP probands (Craddock and Jones 1999). No heterogeneity was found between studies.

Alzheimer's Disease represents an example, within the realm of psychiatric genetics, of a disorder where the early onset form is more familial than the late onset form. The familial early onset forms have proved more amenable to linkage analysis and the genetic aetiologies are now well characterised. Therefore, one might expect more of an increase in morbidity risk in first-degree relatives of probands with early age of onset BP. Empirical data has supported this assertion, whether the data is analysed by stratifying the sample according to age of onset or specifically recruiting early onset cases: (Rice *et al.* 1987); (Strober *et al.* 1988); (Pauls *et al.* 1992); (Neuman *et al.* 1997); (Grigoroiu-Serbanescu *et al.* 2001); (Somanath *et al.* 2002). However, this analysis is complicated by the rather arbitrary delineation of early and late onset cases and the birth cohort effects found in some studies, e.g. (Rice *et al.* 1987).

Investigators have questioned whether BPI and BPII represent genetically distinct entities. Heun (Heun and Maier 1993) found the risk of BPI to be similarly elevated in relatives of BPI probands and BPII probands. However, the risk of BPII was much greater in relatives of BPII probands than BPI probands. This finding argues for at least some commonality in susceptibility to BPI and BPII.

It has been found that the first-degree relatives of BP probands are at increased risk of developing unipolar disorder, henceforth referred to as UP, (McGuffin and Katz 1989). This has lead to the assertion that BP and a proportion of UP may share a genetic aetiology. The converse has not been convincingly demonstrated; relatives of UP probands do not have increased rates of BP. However it may be harder to detect this effect, as UP depression is much more common and only a small proportion may have a putative UP/BP liability.

The risk of schizophrenia has not been found to be altered by relatedness to a BP proband, but the risk of schizoaffective disorder (BP type) has consistently been found to be elevated (Craddock and Jones 1999). It was found that the risk in relatives was equivalent if the proband had a diagnosis of BP or schizoaffective disorder (BP type) (Rice *et al.* 1987). Hence, as mentioned in chapter 1, schizoaffective disorder (BP type) has been considered to share a genetic aetiology with BP and has been included in many linkage and association studies. Relatives of schizoaffective probands were found to have a higher risk of major affective disorder than relatives of BP probands (Gershon *et al.* 1982). This perhaps suggests that there is a higher genetic loading in schizoaffective disorder.

There has also been interest in the effect of gender (of proband or relative) and nature of relationship (i.e. sibling, parent or offspring) on risk of BP. No clear sex-related effect on risk has been shown. With regard to the effect of nature of the relationship on risk, no consensus has been reached. Either, studies have only presented pooled risks, or the

findings have been contradictory. Furthermore, reliable data is only available for risk in first-degree relatives (Craddock and Jones 1999).

The application of family studies in BP is not without its problems. Some of these relate to the nature of BP, e.g. the already mentioned phenotypic definition and possible birth cohort effects and variable age of onset. Other problems relate to the conduct of the studies, e.g. lack of direct interviews (Faraone *et al.* 1990). Nonetheless, studies from independent groups and different geographical regions have shown a convincing familiarity in BP. Family studies have not, however, really clarified the genetic relationship between BPI, BPII, schizoaffective disorder and UP depression.

Family data can also be used for formal testing of the mode of inheritance by segregation analysis, to the consideration of which we now turn.

2.2.2 Segregation analysis

In a general sense, segregation analysis refers to the statistical testing of the observed distribution of disease within a family identified as containing a BP proband, compared to that predicted by a hypothetical genetic model. As mentioned above, the occurrence of BP in families does not neatly follow Mendelian ratios. The most obvious explanation that the observed ratios do not conform to predicted segregation ratios is that the BP trait does not follow Mendelian inheritance. However, there are other possible reasons for deviations from Mendelian ratios, including: misspecification of the method of ascertainment, differential survival, phenocopies, and incomplete penetrance (Sham 1998).

Allowance must be made for the effect of *ascertainment*. In practical terms this means that a defined systematic method of ascertainment must be employed before a valid segregation analysis can be carried out on family samples. *Differential survival* refers to gametes that contain certain haplotypes that have less chance of surviving. This is a theoretical possibility but there is little evidence of this effect being important in BP.

Phenocopies are environment-caused mimics of the trait under investigation. They probably do have relevance to BP. For instance, it is known that a number of medical conditions can give rise to symptoms that resemble those of BP. It is widely thought that *incomplete penetrance*, i.e. the phenomenon of individuals with the disease genotype not manifesting the disease, is an important explanation for the deviations from the Mendelian ratios. To deal with these issues more complex models of inheritance, incorporating the parameters of phenocopy rate and incomplete penetrance, have been developed.

Polygenic theory is an alternative explanation for non-Mendelian inheritance. For discontinuous traits, it is also necessary to hypothesise a threshold or multiple thresholds; if an individual has enough liability they will cross the threshold and develop the trait. This is the theoretical stance underpinning the Multifactorial Polygenic Model, MFP (Falconer 1965) where the variance is conceptualized as arising from the additive effect of many genes and environmental determinants.

Hence, it can be seen that there has been a great effort to mathematically model inheritance. However, consistent support has not been found for a specific mode of genetic transmission in BP (Faraone *et al.* 1990), with some authors supporting a MFP model, e.g. (Smeraldi *et al.* 1977), and others supporting a single major locus model (SML), e.g. (Rice *et al.* 1987). However, many authors have rejected a SML model, e.g. (Bucher *et al.* 1981; Goldin *et al.* 1983). The SML model has also been criticised on mathematical grounds based on the rapid decline in recurrence risk from monozygotic (MZ) twins to first-degree relatives to the general population (Craddock *et al.* 1995; Craddock *et al.* 1997). Interestingly, an age of onset effect on model fitting has been found. Early onset BP was found to be compatible with a single major gene with a polygenic component, while late onset BP favoured a multifactorial model (Grigoroiu-Serbanescu *et al.* 2001).

There seems to be some evidence that the problem of phenotypic definition could be causing inconsistencies in the fitting of mathematical models to the data. For instance,

transmission of BPI was found to be consistent with autosomal dominant inheritance in a subset of closely related Old Order Amish families (Pauls *et al.* 1995). However, it also transpired that this finding could be rejected when atypical BP, major depressive disorder and hypomania were included in the analysis of these families. A similar conclusion was reached by Spence (Spence *et al.* 1995). However, the inconsistencies may alternatively be due to the operation of more complex modes of inheritance, which are not accurately accounted for by the mathematical models that have been applied to the segregation analysis.

Thus overall the inheritance of BP is considered complex. Jones and Craddock (Craddock and Jones 1999) have reviewed some of the mechanisms invoked for explaining complex inheritance: genetic heterogeneity, interaction, dynamic mutations, imprinting, and mitochondrial inheritance.

Genetic heterogeneity means that the disease has a variety of genetic etiologies. This includes both locus and allelic heterogeneity. Locus heterogeneity refers to a situation where there are a number of loci that act independently to give rise to disease. Thus a particular locus may have a major effect, but only in a proportion of cases (Hodgkinson *et al.* 1987). As will be discussed in the next section, genomewide significant LOD scores have been found at various loci which, it could be argued, is evidence for locus heterogeneity. Allelic heterogeneity refers to the situation where more than one allele at a given locus is relevant to disease manifestation. For example, greater evidence for linkage on 11p in the Old Order Amish data was found with a three-allele model (Sandkuyl and Ott 1989).

Craddock (Craddock *et al.* 1995) proposed that multiplicative models are plausible for the inheritance of BP. However, it is difficult to test these *interaction* models without having identified aetiological mutations.

Dynamic mutations refer to the expansion of certain trinucleotide repeats, which lead to various forms of human disease. Examples of such diseases include Fragile X and Spinal

Muscular Atrophy. Dynamic mutations give rise to unusual patterns of inheritance. For example, they can lead to the phenomenon of anticipation, where age of onset decreases whilst severity increases in successive generations. The molecular explanation for this is that repeat numbers correlate with severity and increase down through the generations. Some evidence has been found for anticipation in unilinear BP families (McInnis *et al.* 1993).

The term 'epigenetic' is used to refer to varying concepts, but here the following definition will be used: heritable information that is encoded by modifications of the genome and chromatin components that affects gene expression. A number of epigenetic mechanisms may have relevance to the inheritance of BP. However, it is genomic imprinting which has been specifically studied.

Imprinting refers to the effect of parental origin on the expression of an allele in the offspring. In relation to human disease the most often quoted example is that of Prader-Willi/Angelman syndrome. The aetiology of Prader-Willi/Angelman syndrome can be traced to genes located in the 15q11-13 region. Prader-Willi syndrome results from lack of, or inactivation of genes that are usually expressed on the paternally inherited chromosome. Conversely, Angelman syndrome occurs from a lack of, or inactivation of genes which are usually expressed on the maternally derived chromosome. A possible parent of origin effect was found in BP, with a higher rate of illness in offspring of affected fathers than mothers (Kornberg *et al.* 2000). This suggests that the father may transmit a more penetrant form of the disorder.

Mitochondrial inheritance denotes the inheritance of the mitochondrial genome. Mitochondrial DNA is only inherited from the mother. Therefore a typical pattern of maternal transmission is seen, with no father to offspring transmission. This is seen in conditions such as Progressive External Ophthalmoplegia. It was found that there was a higher risk of BP in offspring of affected mothers than affected fathers, and lack of paternal transmission of illness in a number of large pedigrees (McMahon *et al.* 1995).

However, if mitochondrial inheritance and/or imprinting are operating one would still have to postulate locus heterogeneity to account for the fact that males and females are equally likely to be affected at a population level.

2.2.3 Twin studies

Both twin studies and adoption studies are used to differentiate genetic and shared environmental contributions to the variance of a trait. In this section I will briefly outline the theory behind twin studies and mention some of the major criticisms. I will then go on to summarise the findings from twin studies in BP.

Theory of twin studies

Within the umbrella of twin studies there are a number of different approaches: classical twin method, co-twin control method, and the method of monozygotic twins reared apart. Consideration of these approaches first requires a brief word on the genetic features of twins that make them so useful in genetic studies.

Monozygotic (MZ) twins share 100% of their DNA while dizygotic (DZ) twins share, on average, 50%. It can readily be appreciated that this has implications for the sharing of a genetic trait between a given twin pair. Twins that share a trait are said to be concordant. Assuming that twins, whether they be MZ or DZ, are exposed to a common environment - the 'equal environment assumption', then one can infer a genetic effect if MZ are more concordant than DZ twins. For example, if one considers a fully penetrant autosomal dominant trait all MZ twins will be concordant while, on average, 50% of DZ twins will be concordant.

In the classical method the ratio of MZ concordance to DZ concordance is estimated. The ratio is reported as either 'proband-wise' or 'pair-wise'. The proband-wise ratio is defined as the proportion of affected twin partners of probands. In other words, it is the probability that an affected proband's twin will have the disease. The pair-wise method

expresses concordance as the proportion of all twin pairs that are concordant. Generally, proband-wise concordance is quoted, as this is directly comparable with the relative risk derived from family studies. Heritability (H^2) estimates can be derived from concordance ratios.

However, care must be taken in interpretation of heritability estimates as they have a circumscribed meaning. Heritability is usually defined as the proportion of the total phenotypic variance that is associated with the genetic variance in a specific sample with a specific genetic composition and environmental context (Vitzthum 2003).

In co-twin control studies, MZ twins discordant for the disease are recruited. They are predicated on the disease not being solely genetic and are used to identify environmental factors relevant to development of the phenotype using the healthy twin as the control.

Criticisms of twin studies

There are various criticisms of the classical twin studies method, both theoretical and practical, to which we now turn.

It has been suggested that MZ twins have a more similar environment than DZ twins, which could also lead to greater concordance. In particular, MZ twins are more likely to share a common intrauterine environment, although approximately 19% will have separate placentas (http://www.twins.org.au/). There are a number of precedents for intrauterine environments modifying disease risk in later life. For instance intrauterine growth retardation (IUGR) predisposes to hypertension in later life (Baum *et al.* 2003).

However, although competition can occur between different placentas, competition effects are more significant between twins that share a blood supply, i.e. share a chorion. Twins that share a chorion are always MZ. This can result in selective IUGR. In monochorionic twins, twin-to-twin transfusion syndrome can occur where there is unequal exchange of blood. In this case both twins will be affected, but asymmetrically.

Therefore it is not necessarily true that the interuterine environment is more alike for MZ twins than DZ twins.

It has been asserted that MZ twins are treated more similarly, e.g. dressed the same way, thus inflating trait concordance. Rarely twins are separated early in life and raised apart. Comparison of MZ twins raised together and MZ twins raised apart can avoid this problem.

A further criticism is the potential for bias to be introduced by the method of sample ascertainment. However, the MZ:DZ ratio of the sample can be used to test for these biases. This ratio is approximately 1:1 in the population, although it may vary slightly from year to year. Therefore the MZ:DZ ratio for a sample of same sex pairs should mirror the population ratio of 1:1. If opposite sex DZ pairs are included the ratio becomes 1:2 according to Weinberg's rule (Hawkes 1997).

By examining these ratios evidence was found for volunteer bias in same sex twin samples obtained through volunteer recruitment (Lykken *et al.* 1987). Samples were found to be ⅔ MZ twins and ⅔ female twins – 'the ⅔ rule'. It is generally thought that there is not only an excess of MZ twins but an excess of concordant monozygotic twins.

Many scenarios are imaginable where ascertainment may be biased towards monozygotic concordant twins. For example, concordant twin pairs for BP may be inherently more likely to respond to adverts for twins with BP, since they are, as a pair, more interested in the disease. Over-recruitment of concordant twins would lead to overestimation of heritability. Strict population based ascertainment can, at least in theory, overcome such biases. This is feasible in a number of countries, such as Denmark, which have national twin registries. Indeed the Danish twin registry has been exploited to perform a twin study in BP e.g. (Bertelsen *et al.* 1977). Denmark also has a register of psychiatric admissions allowing cross-referencing. Not all twins with BP will be thus identified but they should provide an unbiased sample, assuming that concordance does not affect rate of hospitalisation.

Determination of zygosity in same sex twins is, of course, crucial to twin studies. Determination of zygosity is non-trivial. The most common error is that MZ twins are misdiagnosed as DZ due to lack of similarity in appearance. This might inflate the concordance rate in the DZ group. DNA fingerprinting is the most accurate method of determining zygosity, with an error rate of less than 1%. However, analysis of a combination of biomarkers - e.g. blood group, secretor status, haptogobulin types and dermatoglyphics - is also highly discriminative. The inspection of fetal membranes or of chorion is not a reliable method. Neither placenta nor chorion number always correlate with zygosity. MZ twins are often dichorionic and the chorion may be fused in DZ twins giving the appearance of monochorioncity.

Twin studies and bipolar disorder

We turn now to the results of twin studies in relation to BP. Again, early studies did not distinguish between UP and BP and thus concordance rates for BP cannot be calculated. Studies that have applied the modern concept of BP have found consistently higher concordance for MZ twins than DZ twins. Jones and Craddock (Craddock and Jones 1999) estimated MZ concordance for narrowly defined BP as 50% from pooled data of six available studies at the time. However, they commented that this is likely to be an underestimate due to the methodology of three of the studies included.

More recent studies have given very high heritability estimates for BP. Thirty MZ and 37 DZ from the Maudsley twin register were studied and a heritability estimate of 0.85 was obtained using a narrow definition of affectation (McGuffin *et al.* 2003). Meanwhile, a population based twin study of 26 twin pairs obtained a heritability estimate of 0.93 (Kieseppa *et al.* 2004). Non-shared rather than family environment was implicated in explaining the remaining variance by these more recent studies.

Phenotypic definition has caused consternation in twin studies just as it has in segregation analysis. Since family studies have demonstrated an increased risk of UP in relatives of BP probands, some authors have performed analyses where UP co-twins are classified as

affected. As would be expected, broadening the definition of affection in the co-twins results in higher concordance rates for both MZ and DZ pairs, e.g. (Kieseppa *et al.* 2004).

Although discordant MZ twins have been studied in BP they have not quite been used according to the usual co-twin control method to identify environmental risk or protective factors. Investigators have generally measured various biological features in the affected and unaffected twin and normal controls. For example, affected twins were found to be significantly impaired on some measures of visuo-spatial functioning and verbal memory compared with the unaffected twin (Gourovitch *et al.* 1999). Furthermore, the right hippocampus was identified as smaller and less asymmetric in the affected co-twin compared to the unaffected co-twin (Noga *et al.* 2001). However, both these studies found significant differences between the unaffected co-twin and normal controls. This raises the spectre that while the BP phenotype is not expressed in the unaffected co-twin the genotype is being expressed sub-clinically - in these cases in terms of brain morphology and neuropsychological function. These 'hidden' effects are referred to as endophenotypes.

Discordant MZ twins have also been used to identify risk factors at a molecular level by investigating gene expression differences. However, rather counter-intuitively, these differences have been used to try and identify aetiological genes (Kakiuchi *et al.* 2003). I will return to both these issues later.

To my knowledge there are no studies of twins reared apart for BP.

2.2.4 Adoption studies

Adoption studies are a powerful technique to distinguish genetic from environmental influences on a given trait. They are based on the assumption that only genetic factors are transmitted from the biological parents to the child and that the child does not share an environment with its biological parents. Hence, elevated rates of the trait in the biological

parents or offspring compared with the appropriate control provide direct evidence of a genetic aetiology.

In essence, there are two approaches to adoption studies. In the first, BP adoptees are identified and matched with control adoptees without BP. The rates of BP in biological parents and adopted parents are then compared. In the second approach, adopted-away offspring of BP probands are identified and matched with adopted-away offspring of non-BP probands. The rates of BP in the two groups of offspring are thus compared.

However, adoption studies have also been criticised. One criticism is that the environment into which the child is adopted is not infrequently chosen to be similar to that of the biological parents. However, even if this effect were operating a higher rate of BP in the biological versus adoptive parents would still argue for a genetic aetiology. Another criticism, that familial aggregation can arise through the shared family environment prior to adoption, would seem more pertinent. As was mentioned in the twin studies section, intrauterine environment has been shown to have long-term consequences on health. Furthermore, early childhood experiences are implicated in aberrant psychological development by many theorists.

To my knowledge only two adoption studies have been carried out in BP, both of the affected adoptees type. A higher risk of BP was found in the biological parents of BP adoptees than in their adoptive parents (Mendlewicz and Rainer 1977). A subsequent study supported this finding, although only ten BP probands were included (Wender *et al.* 1986).

2.3 Summary

Returning to the question posed in the title of this chapter, is BP genetic? The tests of 'geneticness' are not unequivocal because of intrinsic methodological problems. Furthermore, there are specific difficulties applying these methodologies to BP, which

expose the findings to further criticism. However, taken in their totality I believe there is robust evidence for genetic factors being of prime importance in the aetiology of BP.

However, it is not yet clear what modes of inheritance are operating in BP. This may be related to the problem of phenotypic definition or due to the lack of sufficiently sophisticated models of complex inheritance. It was this sentiment which led Faraone (Faraone *et al.* 1990) to suggest that "genetic modelling studies have reached the limit if their utility" in relation to major affective disorder. Instead, he promoted a shift of focus to genetic linkage studies as a means of bypassing these problems and attempting to map genes influencing BP.

Chapter 3: Mapping Bipolar Disorder I - Chromosomal Aberrations and Linkage Analysis

In this chapter I will first introduce the range of methods which can be used to map genes in humans. I will then discuss the study of chromosomal aberrations and linkage analysis in more detail. I will give a theoretical outline of these techniques, highlight methodological problems encountered when applying them to BP and review the findings for BP.

3.1 Methods of gene mapping

Gene mapping of human traits represents a far greater challenge than gene mapping in organisms amenable to experimental manipulation, such as the fruit fly Drosophila. Nonetheless, mapping efforts for Mendelian traits/diseases have been hugely successful, with approximately two thousand genes mapped (Antonarakis and Beckmann 2006). Some success can also now be claimed for the mapping of susceptibilty loci in complex diseases.

The main techniques of gene mapping in human disease are: the study of cytogenetic abnormalities, linkage analysis, and linkage disequilibrium mapping. The first two of these techniques will be considered in this chapter. It is common for a combination of these techniques to be employed in the mapping of a particular gene, for example in the case of Cystic Fibrosis (Gelehrter and Collins 1990).

Generally, approximate localisation of traits has been achieved through linakge analysis or study of chromosomal abnormalities. Linkage disequilibrium fine mapping and sequencing can then be employed for gene identification. A candidate gene approach, where pathophysiological knowledge regarding the trait of interest is utilised, has often been applied to the fine mapping stage. However in many diseases, such as BP, the

underlying biology is not well elucidated. Increasing a systematic approach to linkage disequilibrium mapping is being adopted, facilitated by recent technological advances.

3.2 Chromosomal aberrations

Chromosomal aberrations have played an important role in the mapping of disease genes. Although there are a variety of chromosomal aberrations, two have a particular place in gene mapping. These are translocations and deletions.

Rarely, a disease co-segregates with a *translocation*. Mapping may then be relatively straightforward if one of the breakpoints directly disrupts a gene or nearby control region. However, it has been proposed that a breakpoint may not necessarily have a direct effect on gene function, but could cause disease by separating genes on the same chromosome which act in unison (Klar 2002). If such mechanisms were in operation, mapping efforts would clearly be more difficult.

The identification of chromosomal *deletions* has, arguably, had an even greater impact on mapping human disease. Even a small chromosomal deletion can affect a number of genes. A consequence of this is that an individual can present with a number of different genetic conditions. Cytogenetic studies of such individuals can prove invaluable, as identification of the deletion directly localises a disease gene to an approximate chromosomal position.

Both of these types of cytogenetic abnormalities were important in the mapping of the gene for Duchenne Muscular Dystrophy (DMD). DMD is an X-linked recessive disorder and therefore is very rare in females. Cytogenetic investigation of rare instances of females with DMD revealed an X-autosome translocation. The autosomes involved varied, but the X chromosome breakpoint always mapped to Xp21. Xp21 was also implicated by a study of a boy presenting with DMD and three other X-linked conditions. Cytogenetic analysis demonstrated an Xp21 deletion. From there, cloning of the DMD gene was possible (Gelehrter and Collins 1990).

From the discussion above it can be seen that there is the potential for a serendipitous chromosomal aberration to unequivocally localise a disease gene to a defined chromosomal region. Thus, the problems of applying linkage and association mapping to a complex disorder could be circumvented. However, there are disadvantages to this reliance upon chance. Informative cases are likely to be rare and even more rarely detected. In some countries, such as Denmark, national psychiatric admission registers and cytogenetic abnormality databases have been set up. Cross-referencing of these can facilitate the identification of potentially informative individuals. However, cytogenetic analysis is not generally considered routine in psychiatric practice. Aetiological genes identified by chromosomal aberrations may only represent atypical forms of the disease and not implicate the genes that are important in the wider population of people with BP.

A number of loci in BP have been implicated by cytogenetic studies. A review of the literature reporting chromosomal aberrations in BP identified four genomic regions of potential interest: 11q21-25; 15q11-13; chromosome 21, and Xq28 (Craddock and Owen 1994). A more general review of chromosomal abnormalities in mental illness found evidence in the literature implicating 1q32, 1q42, 11q23, 15q12-13, 16q22, 18p11, 18q21-22, chromosome 21, 22q11-13 and X specifically in BP (MacIntyre *et al.* 2003). Considerable interest has been aroused by the 1q42 locus. Analysis of a large Scottish kindred showed a reciprocal 1;11 translocation to co-segregate with schizophrenia and severe affective disorder (St Clair *et al.* 1990). Breakpoint analysis identified disruption of two genes, which have been termed DISC1 and DISC2 (Millar *et al.* 2000b). I will return to this region and DISC1 in chapter 5.

Since MacIntyre *et al*'s review, a number of reports of chromosomal abnormalities in BP have been published. Duplication of 15q14 due to a maternal reciprocal translation was identified in a patient with dismorphic features suffering from BP with psychotic features (Reif *et al.* 2004). Positron emission tomography scanning revealed bilateral hypometabolism in many brain regions. This is not typically observed in BP, suggesting that applicability of the finding to BP in general may be limited. Sixty-nine patients with

learning difficulties and co-existing severe psychiatric disorder were screened for subtelomeric deletions (Pickard *et al.* 2004). One patient with comorbid schizoaffective disorder was found to have a 4q35.2 deletion - compatible with the linkage results from a large multiply affected Australian pedigree (Badenhop *et al.* 2003).

Three case reports of BP in patients with sex chromosome abnormalities have been published. One described a patient with a deletion of the long arm of chromosome Y (Olajossy *et al.* 2005), while another cited a patient with 48, XXYY syndrome (Lolak *et al.* 2005). Most recently a patient with a ring Y chromsome was reported (Mouaffak *et al.* 2007). This patient also had gender identity disorder.

In addition abnormalities mapping to 11q and 8q have been described. Bohm *et al* (Bohm *et al.* 2006) reported a young woman with Jacobsen syndrome. The deletion, of approximately 8 Mb, was mapped to 11q24.2 of the paternal chromosome. Magnetic resonance imaging results showed hyperintensities in the frontotemporal regions. Finally, a seven-year-old boy with duplication of 8q22.1-q24.1 was diagnosed with BP and speech delay (Macayran *et al.* 2006).

3.3 Linkage

In this section I will firstly define linkage analysis. I will then highlight some of the methodological developments in linkage analysis that have occurred in response to the challenge of applying the technique to disorders with complex inheritance. Finally, I will address some of the specific issues of using linkage in BP and review the findings in this context.

3.3.1 Basic concept

Since the mapping of the first autosomal gene (Donahue *et al.* 1968), linkage analysis has played, and continues to play, a central role in human gene mapping efforts. While gene

aberrations can be highly informative serendipitously, linkage analysis, in its current form, allows systematic screening of the whole genome to position trait loci.

Mendel's law of assortment states that different inherited traits are inherited independently of one other. A post-Watson and Crick interpretation of this law would be that the segregation of any one allelic pair is independent of other allelic pairs during meiosis. However, linkage analysis is predicated on the fact that the physical proximity of two loci results in non-random assortment of alleles at these loci during gametogenesis. While it is apparent why loci on different chromosomes will randomly assort during meiosis, recombination of homologous chromosomes will also cause random assortment of loci widely separated on the same chromosome. As the distance between loci reduces, the recombination fraction - the proportion of meiotic events resulting in recombination between the two loci - will decrease. Hence, alleles at two closely sited loci will tend to be inherited together, i.e. they will be linked.

This phenomenon can be exploited in the mapping of disease-causing genes, as it means that within families a genetic disease will have a tendency to co-segregate with particular alleles at loci close to the aetiological gene. These patterns of co-segregation can be identified by statistical testing of datasets which include genotypes and affection status from known pedigrees. Either parametric or nonparametric linkage analysis can be performed. The parametric technique of LOD score analysis has been pre-eminent in the history of linkage studies.

3.3.2 Parametric and non-parametric analysis

LOD score analysis was initially developed by Morton (Morton 1955) but now has many derivatives. In essence LOD score analysis tests for linkage between two loci, the disease and the marker loci. This is achieved by postulating differing distances - or more strictly speaking, recombination fractions - between them and testing for compatibility with the observed pattern of segregation of disease and marker within the pedigree(s). An odds ratio is calculated. The numerator is the probability of the pattern of segregation occuring

given that they are linked at a certain recombination fraction (<0.5). The denominator is the probability of the pattern occuring given that the loci are unlinked (recombination fraction =0.5). This calculation is repeated for linkage with different recombination fractions.

The LOD is the log of the odds ratio. The LOD is a measure of the degree of certainty with which the null hypothesis that the loci are unlinked can be accepted or rejected.

Since traits rarely accord to perfect Mendelian inheritance, LOD score analysis has undergone many revisions in an attempt to improve its ability to model traits whose inheritance shows such deviation. Morton (Morton 1956) realised that locus heterogeneity could reduce the power of the LOD score method to detect linkage and developed a test for linkage heterogeneity, the K-test. Other tests of heterogeneity have subsequently been developed, including the Admixture or A test (Smith 1963) and the B test (Risch 1988).

It can be readily appreciated that the phenomenon of incomplete penetrance could have a great effect on likelihood calculations. Hence, a penetrance function was introduced in an attempt to account for this (Ott 1974). However, misspecification of parameters can have important effects on the LOD scores (Clerget-Darpoux *et al.* 1986). Consequently, many authorities have advocated the use of model-free analysis for complex disorders.

Model-free analysis is based on testing for deviations from the expected ratios of allele sharing between affected relatives. Model-free analysis can be performed on affected sib-pairs or more extended pedigrees. I will pause for a moment to describe the theory behind affected sib-pair analysis, as I believe it illustrates the concepts behind the study of allele sharing. Sib-pair analysis requires the genotypes of two affected siblings and their parents. For any given locus the probabilities that the siblings share 0, 1 or 2 alleles identical by descent are 0.25, 0.5 and 0.25 respectively. Evidence that these probabilities do not hold implies that the marker allele is linked to the disease locus.

The Affected Pedigree Member Method (Weeks and Lange 1988; Weeks and Lange 1992) and the Nonparametric Linkage Score (NPL) method, as implemented in the program GENEHUNTER, can be used to look for excess allele sharing in extended pedigrees.

Linkage analysis has traditionally considered loci individually in relation to the disease. However, complex inheritance may arise through gene interaction. Attempts have been made to perform two-locus linkage analysis. To my knowledge, the first report of a two-locus linkage analysis was applied to Multiple Sclerosis families and provided evidence for two unlinked loci (Tienari *et al.* 1994). More recently, a systemic two-dimensional genome scan of essential hypertension was performed and was able to identify novel epistatic loci (Bell *et al.* 2006). In BP, two-locus admixture linkage analysis was employed, with some indication that it increased the power to detect linkage when loci on chromosome 21q and 11p were both taken into account (Smyth *et al.* 1997).

3.3.3 Markers

While the development of statistical techniques has facilitated linkage analysis, the utility of the approach has massively expanded as a wider array of markers has become available. In the first application of linkage analysis it was possible to assign the Duffy blood group locus to chromosome 1 only by the observation of an anomalous uncoiled region on chromosome 1 which co-segregated with the Duffy blood group antigen in three families (Donahue *et al.* 1968). Early linkage studies relied on biological markers to which linkage of the disease could be tested, e.g. the HLA antigens (Dupont *et al.* 1977).

Linkage analysis was revolutionised by the proposal of using anonymous polymorphic DNA markers across the genome, rather than a very limited set of biological markers (Botstein *et al.* 1980). The Restriction Fragment Length Polymorphisms (RFLPs) proposed by Botstein were superceded by the use of microsatellite markers (Olson *et al.* 1989). Microsatellite markers are polymorphic repeat sequences. They offer a number of advantages: they are easier to type than RFLPs, requiring only a one stage PCR; they are

more polymorphic, and they provide more dense coverage of the genome. Whole genome scans employing sets of approximately 300-400 microsatellites have now become routine.

The ability to test markers across the whole genome, while undoubtedly advantageous, brings with it the important issue of interpretation of results. Most notably, the problem of determining the level of statistical significance of linkage in the face of such multiple testing (Lander and Schork 1994). Criteria for significance of genomewide linkage results have been set out (Lander and Kruglyak 1995).

Linkage analysis is able to localise a gene only to a broad region, approximately 20-30 cM (Botstein and Risch 2003). It usually represents the first stage in gene mapping, requiring other techniques for finer mapping - in particular, linkage disequilibrium approaches. However, linkage has proved very important in the mapping of genes causing Mendelian disorders. One of many notable examples of linkage findings leading to cloning of an aetiological gene being the mapping of the Cystic Fibrosis gene (Wainwright *et al.* 1985). The path from linkage to cloning has been much less direct in diseases demonstrating complex inheritance. However, there are successful precedents in complex disorders, such as the implication of HF1 gene in age related Macular Degeneration (Hageman *et al.* 2005).

3.3.4 Linkage findings in bipolar disorder

One of the earliest linkages in genetics was the report of X-linkage to BP using colour blindness as a marker (Reich *et al.* 1969). The first linkage study of BP to employ DNA markers was carried in an Old Order Amish pedigree and showed linkage to chromosome 11p (Egeland *et al.* 1987). However, lateral extension of the family resulted in disappearance of the evidence for linkage (Kelsoe *et al.* 1989). Since Egeland's original paper there has been well over a dozen linkage studies in BP. Generally results from these have shown poor consistency. This inconsistency could be interpreted as evidence of extensive locus heterogeneity in BP. It is beyond the scope of this thesis, and probably of dubious benefit, to exhaustively review each linkage paper. Consequently, I will defer

detailed discussion of the evidence for linkage in particular regions of interest until subsequent chapters.

In the rest of this chapter I will briefly discuss the results of the three meta-analyses of linkage studies and the results which have been published subsequent to this. Before reviewing the results I will consider some of the particular problems of performing linkage analysis in BP. I think this helps to place the discussion of the results in a better critical context. I will emphasise how these difficulties might explain the inconsistent pattern of linkage results which are seen.

Methodological considerations

Challenges are encountered at all stages of linkage analysis in BP. There are problems designing appropriately powered studies, determining the 'caseness' of individuals, and applying appropriate statistical analysis - both in terms of specifying the genetic model to use and the interpretation of results. These issues are interrelated and cannot be dealt with strictly separately.

Many factors will affect the power of linkage analysis to detect disease-causing genes. Power calculations are rendered very difficult, as the mode of inheritance is not known. In general, linkage analysis has greater power to detect genes of major effect than genes of smaller effect (Risch and Merikangas 1996). Segregation analysis has provided some evidence for major effects, but it is not possible to accurately determine the size of these effects in the presence of genetic heterogeneity. While linkage analysis is robust to allelic heterogeneity, locus heterogeneity will reduce power.

As a common disease, it has been postulated that there are relatively high frequency disease alleles. If this is the case, the problem of bilineality in pedigrees is increased, which must be excluded in the pedigree ascertainment methodology for successful linkage analysis. However, retrospectively the apparent genetic heterogeneity of BP shown by linkage makes multiple rare variants more likely to explain the commonness of the disease.

Lack of knowledge about the mode of inheritance has caused much debate as to whether to apply parametric or nonparametric methods of linkage analysis to BP. It has been argued that in the presence of locus heterogeneity, parametric analysis of large multiply affected families is more powerful than nonparametric methods, in particular sib-pair analysis. Assuming that individuals within the multiply affected pedigree share a genetic cause, it can readily be seen that the likelihood of detecting the susceptibility locus will be maximised if the whole family is considered rather than, say, an affected sib-pair subset of it. However, the statistical approach used has power implications and therefore influence on the original design of studies. Underpowered studies increase the probability of type II (false negative) error.

Linkage is an attractive technique for the mapping of BP genes, as it does not require any prior knowledge of biology of the disease. However, linkage does have some pre-requisites, the most significant of these being a secure phenotype. While this problem is not unique to BP, is has already been seen that defining affection status is non-trivial. The response to this problem has been to analyse linkage results under different affection models. These have generally involved a division into broad and narrow phenotypes, e.g. a narrow phenotype classifying only those with BPI as affected and a broad phenotype including those with depressive disorder. Misclassification of the phenotype could result in inconsistent findings by increasing the likelihood of both type I (false positive) and type II error. The multiple testing which results from the use of different affection models would, if left uncorrected, lead to an increase in type I error.

The difficulty with phenotype definition may, at least in part, be due to genetic heterogeneity. As has already been seen with segregation analysis, some authors have attempted to reduce any heterogeneity by using affection criteria likely to select for cases of major gene effect, such as age of onset. Applying this strategy to linkage studies has had striking success in mapping breast cancer genes where the BRCA1 was mapped after age of onset was used to enrich pedigrees (Hall *et al.* 1990). From early in the history of

BP linkage studies, researchers have attempted to incorporate age of onset as a covariable, e.g (Morton and Kidd 1980), but this approach has not been universally adopted.

Other attempts to improve the phenotype definition, and thus potentially reduce heterogeneity, have centred on the identification of endophenotypes. Endophenotypes can be thought of as cryptic phenotypes of BP. Endophenotypes used have included lithium responsiveness, e.g. (Turecki *et al.* 2001), and white matter hyperintensities, e.g. (Ahearn *et al.* 2002). However, there are as yet no clear successes for this strategy in BP, although it would seem to have considerable potential.

Misspecification of the parameters for LOD score analysis is a major problem, as the mode of inheritance is not known in BP. Misspecification of the disease allele frequency, penetrance and dominance can all lead to errors in estimates of the recombination fraction (Clerget-Darpoux *et al.* 1986). Misspecification may also have an effect on the power to detect linkage. Phenotype definition is also intimately related to misspecification, as narrow models may underestimate penetrance while broad models overestimate it. Clerget-Darpoux considered that there may be significant errors, particularly when dominant and recessive models were confused. While this is less likely to happen, the effect of applying any of these models to a complex mode of inheritance is not known.

One response to the problem of misspecification of the model parameters is to analyse under a number of different models with different parameter specification, and this has widely been the *modus operandi* in BP linkage scans. However, this compounds the problem of multiple testing which is inherent in genomewide linkage scans. While methods have been developed to correct for type I error resulting from such LOD maximisation techniques, e.g. (Weeks *et al.* 1990), such effects may not be discussed in some studies.

It has been determined that a LOD score of 3.3 is needed to declare significant linkage in a genomewide linkage scan given the use of multiple markers (Lander and Kruglyak 1995). However, this assumes that only one model of inheritance is tested. Furthermore,

multiple scans have now been performed in BP, perhaps as a consequence of a number of underpowered studies producing negative results. As more scans are performed, the probability that one will, by chance alone, give a significant LOD in the absence of true linkage increases. Rather than attempt to formally correct for this source of multiple testing, researchers have generally considered the best evidence for a formally significant LOD being a true positive result is independent replication of that linkage result. However, this in itself has caused some statistical controversy.

While failure to replicate under formal Lander and Krugylak criteria may cast doubt on the original findings, in the presence of locus heterogeneity much larger sample sizes may be required to replicate an original positive finding. Hence, underpowered studies may falsely report non-replication. It has therefore been argued that the threshold for significance should be lowered in regions where previous evidence for linkage has been obtained (Suarez *et al.* 1994). Overall, while there is clear potential for false negatives if too stringent significance levels are set, it seems likely that at least some of the linkage results reported as significant are false positives.

3.3.5 Review of meta-analyses of bipolar linkage studies

Badner and Gershon (Badner and Gershon 2002a) performed the first meta-analysis of BP genome scans. The technique of Multiple Scan Probability (MSP) was applied to combine reported p-values for individual linkage studies with a correction for the size of region containing the minimum p-value (Badner and Gershon 2002b). Badner and Gershon claimed that this method was a more robust means of studying heterogeneity than pooling of data. They only used whole genome scans, considering linkage reports for particular regions more susceptible to publication bias.

In total 11 genome scans were included. Only chromosome regions with a reported p-value of <0.01 were included in the analysis. They found strongest evidence for a BP locus on chromosomes 13q ($p<6\times10^{-6}$) and 22q ($p<1\times10^{-5}$). These probabilities were considered significant on the basis of simulations. In this report Badner and Gershon also

performed a meta-analysis of schizophrenia linkage scans. Interestingly, significant evidence for linkage to both the 13q and 22q regions was found for schizophrenia as well.

Segurado *et al* (Segurado *et al.* 2003) performed the second meta-analysis of BP genome scans. A rank-based technique referred to as Genome Scan Meta-Analysis (GSMA) was applied (Wise *et al.* 1999). In this analysis the genome was divided into 120 'bins', representing, on average, 29.1 cM. For each study, every bin was given a rank according to the linkage score (or p-value) within that bin. The average rank across the studies was then calculated for each bin. A weighted analysis was also performed where the within study ranks were weighted according to the study size. Two p-values were calculated for each average rank value. Simulations were performed to determine p-value at which significant evidence for a bin containing a BP locus could be declared.

An attempt was made to identify all genome scans in BP, whether published or unpublished. However, scans of less than 20 affected cases were excluded, as they were not thought to increase the power to detect linkage under this methodology. In total, 18 genome scans were included. Data was obtained for every marker in the scan rather than just using published data. The research groups were asked to indicate a primary linkage statistic and this was used for ranking, otherwise the most significant statistic was used. Segurado *et al* stated that there was reasonable homogeneity between studies in terms of diagnostic instruments, genome coverage and statistical analysis used, although homogeneity was not formally tested. Analysis under three affection models was carried out: very narrow, narrow and broad. No bin achieved genomewide significance under any of the models. Nominally significant results were obtained on 14q, 9p, 10q, 18p-q and 8q.

Most recently, McQueen *et al* (McQueen *et al.* 2005) performed a meta-analysis using original genotype data. It was argued that this approach would increase the power to detect linkage and allow control over sources of heterogeneity. Original genotype data from 11 BP scans was included. This represented 5,176 individuals from 1,067 families. There were several important exclusion criteria: studies with large pedigrees; those derived from unique, e.g. founder, populations; those employing non-standard

ascertainment schemes, and those lacking original data. Where possible, all markers were located on a common genetic map. The genotypes were pooled and error detection was performed by checking for Mendelian errors and unlikely genotypes. The affection models were defined as narrow (BPI only) and broad (BPII and BPI). Nonparametric multipoint analysis was performed. The threshold for genomewide significance was calculated and a Bonferroni correction made as two affection models were employed. Significant linkage was found on 6q for the narrow phenotype and 8q for the broad phenotype.

In principle, meta-analysis should provide greater power to detect linkage than individual studies and be more conservative. However, there is little consistency even in meta-analysis, the exception perhaps being the 8q region, which was formally significant in the analysis by McQueen *et al* and nominally significant in the Segurado *et al* analysis.

A number of explanations have been put forward to explain the discrepancies. Firstly, it is worth noting that only four datasets were common to all these meta-analyses. If the discrepancy between the Badner and Gershon and Segurado *et al* findings is considered, either the first reports false positives, or the second false negatives, or both. One mechanism by which false positives can arise is use of data from the same families in more than one linkage study, a possibility that Badner and Gershon could not completely exclude. With Segurado *et al*'s study it should be borne in mind that the method is unable to exclude linkage, thus the interpretation should be that no significant evidence for linkage was found in the dataset used. Segurado *et al* excluded smaller studies, which may have provided some of the evidence for the 13q and 22q loci. However, as commented by Levinson *et al* (Levinson *et al.* 2003), GMSA may be more powerful for detecting genes that have a small effect in most families, and therefore may not be appropriate where there is much locus heterogeneity of genes of major effect.

McQueen *et al* may have succeeded to find linkage where Segurado *et al* failed because of increased power of using original genotypes. However, explanations for the inconsistency of these results with those of Badner and Gershon may be similar to those

proposed for the inconsistency with Segurado *et al.* The exclusion criteria applied by McQueen *et al* may be of particular relevance here. In essence, they were applied on the basis that it was desirable to identify loci which are relevant in the majority of BP cases rather than 'personal loci' present in a small proportion of BP cases. However, as discussed with reference to the Segurado *et al* results, this may not be appropriate in the situation of extreme locus heterogeneity. In other words, it may be that this approach is inherently biased towards the identification of loci with smaller effect. Indeed, the effect sizes estimated for the significant 6q and 8q loci were both small (1.34 and 1.26 respectively).

Further genomewide linkage studies of BP have been published subsequent to McQueen *et al*'s meta-analysis and are summarised in Table 3.3.1 below. Generally these studies have not supported the regions implicated by the meta-analyses. However, this needs to be formally tested by further meta-analysis incorporating the new data. The significant linkage reported in the 6q24-25 region (Schumacher *et al.* 2005b; Cheng *et al.* 2006) is not really compatible with McQueen *et al*'s findings. However, it is feasible that the linkage signal detected in kindred cr201 from Costa Rica (Service *et al.* 2006) originates from the same locus. With regard to chrossome 8q, the linkage found in the large pedigree from Eastern Cuba (Marcheco-Teruel *et al.* 2006) is too centromeric to accord with that reported in the meta-analysis.

Two trends seem to have emerged from recent linkage studies. Increasingly, attempts have been made to condition the linkage analysis on putative subtypes of BP. Other investigators have sought to identify 'common susceptibility' loci by using affection models which span traditional diagnostic boundaries. These approaches have yielded evidence for two 'psychotic' BP loci. 16p12-13 is supported by subanalysis of a large US study (Cheng *et al.* 2006) and linkage analysis of BP affective puerperal psychosis pedigrees (Jones *et al.* 2007). Analyisis of psychotic pedigrees from the National Institute of Mental Health (NIMH) samples yielded a linkage significant peak at the GABRA1 gene on 5q34 (Kerner *et al.* 2007). Significant linakge was found very close to this gene

in a combined analysis of pedigrees from Columbia and Costa Rica (Herzberg *et al.* 2006). Of note, significant linkage to schizophrenia has been found in this region.

Table 3.1: Summary of recent linkage findings in BP

Study	Description of study	Significant or 'best' results	
		Region or marker	Linkage statistic
(Hamshere *et al.* 2005)	24 'schizoaffective' pedigrees identified by at least one member with SA, BP type. SA, BPI and SA, BP type coded as affected. 35 affected sib-pairs analysed.	1q42	LOD=3.54
(Schumacher *et al.* 2005b)	52 BP families of Romany, Spanish and Bulgarian descent used in first stage. 56 German families included in subsequent fine mapping stage.	4q31	NPL=5.49
		6q24	NPL=4.87
		1p35-36	NPL=3.97
(Service *et al.* 2006)	Single kindred (cr201) from the central valley of Costa Rica. 168 persons, 25 BPI. SNP genotyping performed. Two-point parametric analysis only due to computational constraints.	6q14	LOD=2.78
(Cheng *et al.* 2006)	US sample of 1,060 individuals from 154 multiplex families. Analysis by psychosis, suicide attempts, and comorbid panic disorder subtypes also performed.	10q25,10p12, 16q24, 16p13,16p12 (standard models)	Significant according to own simulations
		16p12 (psychosis)	
		6q25 (suicide attempts)	
		7q21 (panic disorder)	
(Kerner *et al.* 2007)	Re-analysed data from the NIMH samples. 'Psychotic' BP pedigrees identified by at least two members with 'psychotic' BP or SA. Family members coded as affected according to the psychosis section of the DIGS, all others were coded as unknown. NPL analysis performed	5q34(GABRA 1)	$p<0.00001$
(Etain *et al.* 2006)	87 sib-pairs ascertained by an early onset BPI proband, (onset below 21 years of age). Siblings diagnosed as affected if reached criteria for: BPI, BPII or SA regardless of age of onset. Siblings experiencing episodes of major depression before the age of 21 also classified as affected. NPL analysis performed.	3p14	$p=0.0002$
(Faraone *et al.* 2006)	Re-analysed NIMH Genetics Initiative BP wave 1 sample, using ordered subset analysis with minimal age of onset of mania in each family as the subset identifying covariate. Subset of 58 families with onset of mania <20 identified.	9q34	LOD=3.21
(Jones *et al.* 2007)	'bipolar affective puerperal psychosis' pedigrees identified by at least one member with manic or psychotic episode within six weeks of delivery.	16p13	LOD=4.07

	BPI, BPII and SA, BP type coded as affected. 54 sib-pairs analysed.		
(Venken *et al.* 2007)	10 multigenerational BP pedigrees from Belgium, including one large Ashenazi Jewish pedigree.	10q21.3-10q22.3	HLOD=3.28
(Zandi *et al.* 2007)	98 pedigrees. 428 subjects genotyped. Covariate analysis performed for age of onset, psychosis and comorbid anxiety disorder.	2q24 (standard models)	p=0.0012
(Cassidy *et al.* 2007)	Irish sample. Most significant result obtained under a narrow (BPI) only affection model.	D14S588	p=0.0006
(Goes *et al.* 2007)	Reanalysis of NIMHGI collaborative cohort (708 pedigrees). Mood-incongruent psychotic features incorporated as covariate.	Increased evidence for linkage 13q21-33 and 2p11-q14	
(Herzberg *et al.* 2006)	Six multiplex BP pedigrees from Antioquia, North West Columbia. Follow up in expanded set of Costa Rican and Columbian pedigrees.	5q31-33	p<0.00004
(Vazza *et al.* 2007)	16 pedigrees from North East Italy identified by either SZ of BP or probands.	D15S1014	p=0.00197
(Marcheco-Teruel *et al.* 2006)	Six generation BPI pedigree from Eastern Cuba. 28 affected members. SNP genotyping using Affymetrix chip. Most significant results NPL analysis using branches of family.	14q11.2-12, 5q21.3-22.3, 5p13.1-q12.3, 6p22.3, 8q13.3-21.13, 10q22.3-23.32	p<0.005

SA, BP type=schizoaffective disorder (BP type)
NIMH=National Institute of Mental Health
NIMHGI=National Institute of Mental Health Genetics Iniative
HLOD=Heterogeneity LOD

3.3.6 Homozygosity mapping

Homozygosity mapping has proved a useful technique for the mapping of some rare recessive disorders. Examples of diseases in which homozygosity mapping has been crucial to mapping efforts include the MRAP2 form of Familial Glucorticoid Deficiency (Metherell *et al.* 2005) and Charcot-Marie tooth disease, type 4C (LeGuern *et al.* 1996).

Homozygosity mapping requires multiply affected pedigrees in which all disease alleles are derived from a common ancestor. The greater the degree of inbreeding within pedigrees, the greater the likelihood of this occurring. The affected individual will therefore be homozygous for the disease allele(s). There will also be varying homozygosity for markers extending from the disease locus. Attempts are therefore made

to identify all regions of homozygosity between the affected individuals on the basis that one region will harbour the disease-causing allele.

Only limited evidence for recessive inheritance operating in BP has been drawn from segregation analysis (Faraone *et al.* 1990). However this could be due to preferential of 'dominant' looking families. A few linkage studies have found maximum LOD scores under a recessive model, e.g. (Tomas *et al.* 2006). This may be due the existence of a recessive form of the disorder or be an artefact of testing multiple modes of inheritance.

Ewald *et al* (Ewald *et al.* 2003) performed genomewide linkage of a single pedigree containing two individuals with BP who were offspring of a first cousin marriage. Homozygosity mapping was also performed, employing three affection status models. Evidence for a susceptibility loci at 2q31.3, 10, 12q24 and 21q22.3 was found by linkage analysis and haplotype sharing. Subsequently, homozygosity mapping in 22 consanguineous families was performed using only one affected offspring from each family (Ewald *et al.* 2005). Approximately 1,500 Single Nucleotide Polymorphisms (SNPs) were typed in each of 24 individuals with BP using microarrays. A parametric multipoint LOD of 1.96 was found in the region of 17q24-q25.

3.4 Summary

A convincing aggregation of chromosomal abnormalities in specific chromosomal regions in BP has not, as yet, clearly been seen. However, the chromosomal aberrations and linkage results are mutually supportive in some regions, e.g. 1q42. It could be argued that some of the cytogenetic findings have been somewhat neglected. Identification of chromosomal aberrations remains an attractive approach due to its directness and circumvention of many of the methodological problems of linkage.

Aberrations that cause breakpoints are particularly desirable, as they may exactly identify the aetiological gene. The example of DISC1 in schizophrenia sets a precedent for this in psychiatric genetics. The example of DISC1 also demonstrates that linkage and

chromosomal abnormalities, although considered separately above, can be used together in certain circumstances.

I believe that the study of cytogentic abnormalities may still have a role to play in BP mapping efforts. The resolution of abnormality detection continues to improve and as an increasing number of individuals pass the peak age of onset of BP the repository of chromosomal aberrations is bound to increase.

Despite the methodological problems, and in particular concerns over multiple testing, significant linkage has been found to a number of chromosome regions. This suggests considerable locus heterogeneity, although the sheer number of loci raises concerns about false positives. Meta-analysis does not seem to have resolved this to date but is the logical response to the problem. It is questionable that a sufficiently powerful meta-analytical linkage technique, which can deal with extreme heterogeneity, has yet been developed.

While linkage studies continue to be performed, much of the research effort has now turned to fine mapping. Fine mapping efforts have generally employed techniques based on the phenomenon of linkage disequilibrium, which is the subject of the next chapter.

Chapter 4: Mapping Bipolar Disorder II - Linkage Disequilibrium Mapping

In a similar manner to the last chapter, I will begin with a brief discussion of the theoretical basis of linkage disequilibrium mapping - otherwise known as allelic association - and highlight some of the methodological problems related to this. I will then give a general review of the findings from association studies carried out in BP. Association studies have proliferated in the field and it is beyond the scope of this thesis to deal with each one individually. Hence in this chapter I will consider only two loci as exemplars. These loci were chosen as meta-analysis of association studies support their involvement in the aetiology of BP. I will also refer to the recently published genomewide association studies (GWA). I will defer discussion of association findings pertinent to the loci I have tested until chapter 5.

4.1 Basic concept

Whereas linkage analysis relies on the co-segregation of alleles with a disease within families, linkage disequilibrium mapping is based on co-segregation of alleles with a disease at the population level. This can occur due to the evolutionary phenomenon of linkage disequilibrium (LD). Simply stated, particular alleles at genetic loci that are physically very close to one another will tend to remain together during meiosis. Hence, their assortment will be non-random.

As the physical distance between two loci decreases, the probability that meiotic recombination will disrupt the relationship between the two alleles drops. It was estimated that useful LD is unlikely to extend beyond about 3 kb (Kruglyak 1999b). However, distances over which LD has been found to operate have been revised upwards as more empirical data has been gathered. For instance, it was shown that LD in a population of northern European descent typically extends 60 kb from common alleles

(Reich *et al.* 2001). However, LD has been shown to vary across chromosomal regions, e.g. (Nickerson *et al.* 1998).

Fine mapping of disease genes can be achieved by identification of LD between alleles at marker loci and a disease. LD is identified by testing for association between particular alleles at the marker locus and the disease (allelic association). If LD is present, an unequal distribution of alleles at the marker locus should be seen when a group of individuals with the disease (cases) and a group without (controls) are compared. The associated allele or alleles thus identified could either be directly involved in disease susceptibility or, more likely, be in LD with an aetiological mutation.

To illustrate the processes from an evolutionary perspective let us assume that there is a single (unknown) aetiological mutation increasing susceptibility to BP. The mutation can be termed **B**. This mutation will have occurred on a certain haplotypic background. Consider a polymorphic 'marker' on this mutated chromosome in very close proximity to **B**. Let us assume that an allele referred to as **M** is present on this particular chromosome. Provided that the mutation does not adversely affect reproductive fitness and the population is in expansion, the number of people in the population with the disease will increase. If there are no recombination events between the two loci all individuals with BP in the population will also carry the **M** allele at the marker locus. Hence, if you can measure the number of **M** alleles in a sample of people with the disease and compare it with a sample of people without the disease one will find that **M** should be overrepresented in the case group.

Clearly the above example describes the simplest possible situation, but hopefully it illustrates the underlying concept. In realistic scenarios one could expect multiple susceptibility alleles and different founding mutations, resulting in private associations with neighbouring alleles. This has implications for both the optimal population to be studied and the sample sizes that might be needed to detect susceptibility when there is genetic heterogeneity in susceptibility to BP, as strongly indicated by the linkage results.

It is important to note that the associated allele may be the common allele in the population which is in LD with the aetiological mutation. Hence, there may only be small differences in allele frequencies between the case and control groups.

4.2 LD measures

As a prerequisite to more in-depth consideration of LD, a brief word on the measures of LD is necessary. There are several different measures of LD, each of which has advantages in certain situations.

Lewontin (Lewontin 1964) developed the first measure of LD, denoted *D*. *D* is a simple measure of the difference between the observed frequency of co-occurrence of particular alleles at two loci and that which would be expected on the basis of allele frequencies. Let us again consider the hypothetical loci **B** and **M**. We shall assume they are bi-allelic and denote the alleles $\mathbf{B_1}$, $\mathbf{B_2}$ and $\mathbf{M_1}$, $\mathbf{M_2}$. The expected frequency of the $\mathbf{B_1M_1}$ haplotype is the product of their respective allele frequencies (p_1=frequency $\mathbf{B_1}$ and q_1=frequency $\mathbf{M_1}$). The observed frequency of the $\mathbf{B_1M_1}$ haplotype can be denoted p_{11}. Thus:

$$D=p_{11} - p_1q_1$$

If *D* is greater than 0 then LD is said to exist. However, as the value of *D* is dependent on allele frequencies it is usual to quote the normalised measure *D'*:

$$D'=D / D_{max}$$

A *D'* of 0 means there is no LD while 1 signifies complete LD. *D* can also be normalised by division by the square root of the product of all the allele frequencies. This value is denoted *r*:

$$r=D / \text{square root}(p_1p_2q_1q_2)$$

4.3 Study design

I will expand my explanation of the theory and problems of LD mapping by considering elements of association study design. While I have tried to separate the elements by headings for clarity's sake, it can be appreciated that many of the issues are interrelated.

4.3.1 Power

There are a number of factors which will affect the power of a study to detect association between an allele and BP, including: LD between the marker and disease allele; genetic heterogeneity; sample size and effect size of disease allele.

A priori the extent of LD between the marker allele and a true susceptibility allele cannot be known. Since neither the totality of allelic variation in the human genome nor the extent of LD between them in all human populations is known, attempts at systematic LD mapping can only be approximate. However, the HapMap project is continually expanding knowledge of LD patterns, enabling increasingly comprehensive mapping efforts.

LD between a marker allele and a BP susceptibility allele will be influenced by numerous factors. The degree of LD which exists will primarily be a function of the number of generations which have passed since the mutation occurred and the recombination rate. LD will decline with successive meiosis at any given recombination rate. However neither of these values are known. Estimates of recombination are prone to error due to marked variance in recombination rates across the genome, with some recombination hot spots.

Models regarding the genesis of BP susceptibility alleles are pertinent here. It has been hypothesised that susceptibility alleles for common diseases that are seen across the globe, of which BP is one, may have arisen before the 'Out of Africa' expansion of modern humans approximately 100,000 years ago (Kruglyak 1999a). One would now expect to

observe only short range LD under such conditions, as LD has been eroded by the high number of generations since the mutational event. Thus, more markers would be required for mapping.

In addition to recombination, subsequent mutations of the marker allele, or indeed the susceptibility allele itself, will cause a breakdown in LD. The implication of this for the choice of markers for LD mapping will be further discussed below.

Factors related to population history are the other main determinants of LD patterns and include genetic drift, population growth, admixture and population structure. However, the effect of these processes are not well characterised in human populations and their implication for LD mapping is not entirely clear.

Power to detect LD will be maximised when a single mutation arising on an uncommon haplotypic background accounts for all cases of disease in subsequent generations of that population. Hence, when the cases are sampled from the population all susceptibility alleles will have arisen from a common ancestor. If susceptibility alleles have arisen on more than one haplotypic background the power to detect LD will be reduced. In contrast to linkage analysis, LD mapping is affected by allelic as well as locus heterogeneity. The linkage results suggest that there is considerable locus heterogeneity, while the extent of allelic heterogeneity is unknown.

It logically follows that for any given number of markers tested, the larger the effect size of a susceptibility allele at a population level the greater the power of LD mapping to detect that allele. However, it has been demonstrated that LD mapping has good power to detect genes of small effect (Risch and Merikangas 1996). There is controversy as to the likely effect size of BP susceptibility genes. It has been argued that the genes must be of modest effect size (Craddock *et al.* 1995). However, the linkage results are consistent with there being genes of large effect size which only operate in a small proportion of BP cases and hence have a small effect size on the BP population as a whole.

4.3.2 Sample groups

As alluded to above, the basic requirement for LD mapping is a sample of cases and a sample of controls between which allele frequencies can be compared. Clearly one only wants to identify frequency differences which stem from LD between the marker and susceptibility allele, rather than from spurious causes. Hence, there has been much debate as to the optimal control group. There have been two basic approaches: population based case-control and family based studies.

The main methodological criticism with case-control studies is the potential confounding of results by population stratification. Allele frequencies for non-conserved bases may vary markedly between different ancestral populations. Hence, it can be seen how the unequal distribution of individuals from different populations across the case and control groups could create allele frequency differences unrelated to disease association. There is also the danger that true associations could be masked by population allele frequency effects. For example, large interpopulation differences were found for the DRD2 *Taq*1 allele which had previously been associated with alcoholism (Goldman *et al.* 1993). No association was found between the *Taq*1 allele and alcoholism in a case-control sample derived from a solely Cheyenne Indian population, although it has to be said the numbers were small. The problem of stratification can, however, be addressed at the recruitment stage by using ancestral screening for all volunteers. Furthermore, stratification can be tested for by determining the allele frequencies in the case-control groups at loci unlinked to BP (genomic control).

There has been some debate as to whether it is acceptable to use a control group derived from the general population or a 'supernormal' group which has been screened for BP and related illness. It has been asserted that the loss of power using unscreened controls is minimal (Owen *et al.* 1997). However, supernormal controls were used successfully in the LD mapping of an asthma susceptibility gene, ADAM33 (Van Eerdewegh *et al.* 2002).

These methodological pitfalls have to be set against a number of distinct advantages of the case-controls paradigm. Sample size and cost considerations generally favour case-control designs over family based studies on the basis of ease of recruitment. Furthermore, it could be argued that the results of case-control studies are more generalisable, or at least that different genetic susceptibilities may be identified because of inherent selection bias in the trio design (Brunn and Ewald 1999).

The alternative strategy involves the use of family based samples. It was recognised that the untransmitted haplotypes of the parents of an affected proband could be used as the control (Falk and Rubinstein 1987). It has been noted that if each case haplotype is effectively matched with a control haplotype derived from the same population the risk of false associations due to stratification are negated (Craddock *et al.* 2001). Falk and Rubenstein's original haplotype relative risk method has undergone much modification and development. One important development was that of the transmission disequilibrium test (TDT) to take account of matched pair data (Spielman *et al.* 1993). Subsequent refinements of the TDT have enabled it to deal with both multiallelic markers (Curtis and Sham 1995) and multiple markers (Wilson 1997). In addition, refinements have been made to utilise the various combinations of pedigree data that may be available.

Family based tests are not, however, immune to criticism on theoretical grounds. Firstly, the TDT is a test of both linkage and association. Hence, there is the danger of confounding linkage with association. Secondly, TDT techniques are less efficient as not all alleles typed contribute to the analysis and some individuals will be uninformative. For instance, it was shown that trio TDT is only two thirds as efficient as case-control samples even if all the individuals are fully informative (Morton and Collins 1998).

The discussion above has focused on the control groups, but the characteristics of the case group are of course crucial to successful LD mapping. It can be understood that, for any given case sample size, power to detect a particular susceptibility locus will decrease as locus heterogeneity increases. Approaches to defining more homogenous samples have been mentioned with regard to linkage studies. However, the 'Catch 22' situation of the *a*

priori uncertainty regarding phenotype/genotype correlation persists. Generally, association studies have used ICD-10, DSM-IV, or RDC definitions of BP. However, some investigators have separated BPI and BPII for study on the basis that they may be genetically distinct, e.g. (Del Zompo *et al.* 2007).

4.3.3 Populations

There has been much debate as to the optimal population from which to draw samples for LD mapping. This debate has revolved around the age and extent of inbreeding which is desirable. It has been proposed that 'young' populations that have been genetically isolated, e.g. the Finns, should be used. Founding of a population can be thought of as taking a small sample of the haplotypic variation of a much larger population – the sample being the population founders. Thus, one could consider the founding of a population as the resetting of the LD clock. Generally, higher levels of LD have been demonstrated in younger populations (Peltonen *et al.* 2000). Furthermore, presuming that the disease-causing mutations in the population are not new, it follows that the disease alleles are derived from a small number of founders. Thus, genetic heterogeneity should be reduced with a commensurate increase in power to detect the relevant alleles.

However, it is not universally accepted that genetic isolates are advantageous for LD mapping. Not all investigators have found increased LD in genetic isolates. For example, it was not found that LD was greatly increased in genetic isolates such as the Lapps and Basques (Lonjou *et al.* 1999). It has been argued that heterogeneity would only be reduced if very severe bottlenecks had occurred under a common variant oligenic model. Furthermore, heterozygosity will be reduced in inbred populations, thus reducing the efficiency of TDT (Jorde 2000). In fact, it may be that lower LD is desirable to achieve higher resolution mapping.

Some investigators have advocated the use of admixture populations, e.g (Wright *et al.* 1999). If there are different allele frequencies in the two populations, recombination events following mixture of the populations will set up new LD relationships. LD will

quickly break down between unlinked loci and this phenomenon can be exploited for admixture mapping. However, the need to use a population that is not continuing to undergo admixture limits its practical value.

Ultimately, the choice of population will depend on the aim of the study in terms of the mapping resolution required and resources available. In some situations a two stage mapping approach may be appropriate (Goldstein and Weale 2001). Initial mapping is performed in a younger population requiring a lower marker density. More precise localisation can then be achieved by mapping of associated regions in older populations.

I believe there is also an issue of whether it is most appropriate to follow up linkage findings in an association sample derived from a broadly similar population. Given that the linkage results indicate heterogeneity in BP, and assuming that isolated populations do indeed increase the homogeneity of BP, it is likely that individual susceptibility loci make differing contributions to the totality of disease across different populations. Therefore, attempting to fine map a region implicated in an isolated population might require a sample of higher power from an outbred population where this locus is contributing less to the disease burden.

4.3.4 Markers

There has been some debate as to whether to employ microsatellite or SNP markers for LD mapping. For both historical and theoretical reasons microsatellite markers have generally been preferred in linkage studies. Four of the key issues in marker selection for LD studies are: degree of polymorphism, marker mutation rate, marker density, and ease of typing. Being more polymorphic, microsatellites have a theoretical advantage over SNPs in terms of power. It can be seen that with more potential marker alleles for a new aetiological mutation to associate with, the likelihood of it occurring on a 'rare' haplotypic background is increased. This scenario will increase the power to detect LD (Jorde 2000). However, this advantage may be offset by the higher mutation rate of some microsatellite loci which will cause a breakdown in LD (Weiss 1998).

The Human Genome Mapping project has lead to the identification of thousands of microsatellite markers spanning the genome. However, SNPs account for a far greater proportion of genetic variation, with an estimated ten million SNPs across the genome (http://www.hapmap.org/). Hence, a far higher marker density can be achieved with SNPs. It is the hope of the HapMap project that the information collated will allow selection of SNPs for both systematic and efficient mapping. As the LD relationships between SNPs are elucidated, it becomes possible to identify so called tagging SNPs – SNPs that capture a pre-ordained amount of the SNP diversity in that region. However, the extent of LD needed between the tagging SNPs and those SNPs being tagged in order to capture diversity relevant to the disease is unresolved.

Marker heterozygosity is a particular issue in SNP selection. Theoretically, a SNP with a heterozygosity of 0.5 would maximise the power to detect LD with a susceptibility allele, assuming an equal likelihood of the allele arising on the two backgrounds. This consideration has lead to SNPs with minor allele frequencies greater than 0.05 generally being used for LD mapping. However, it has been argued that SNPs of lower heterozygosity are more likely to be casual or in LD with a casual SNP (Collins *et al.* 1999).

The emphasis on SNPs has fuelled the development of multiple SNP typing methods. Breen (Breen 2002) has provided a useful review of SNP technologies. Very high throughput methods have now been developed, culminating in SNP microarrays which allow testing of up to one million SNPs at a time.

4.3.5 Haplotype analysis

Simply stated, a haplotype is the combination of alleles at a particular set of loci on a given chromosome. An aetiological mutation will have necessarily occurred on a particular haplotype. If LD decay is indeed driven primarily by recombination, this ancestral haplotype will have progressively shortened with recombination events. It is

this truncated ancestral haplotype which should be overrepresented in cases compared to controls. It has been argued that the identification of such mutation-harbouring ancestral haplotypes is the primary goal of LD mapping (Daly *et al.* 2001). It is intuitively clear that the single marker approach, that predominated until recently, will go only as far as to dichotomise the totality of haplotypes.

Daly *et al* (Daly *et al.* 2001) studied a 500 kb region of 5q31 using 103 SNPs. Discrete haplotype blocks ranging from 10-100 kb in length, which contained only a few common haplotypes, were identified. Treating the haplotypes within each block as alleles, a stepwise decay in LD was observed as distance from a given block increased. It was assumed that these steps indicated historical sites of recombination. This pattern was more readily understandable than the pattern of single marker LD where considerable fluctuations in LD were observed. Such 'noise' in single marker analysis reflects the mutational history of individual markers and has important implications for interpreting LD results, as true disease associations may be missed. Thus, Daly *et al* advocated the creation of a comprehensive haplotypic map to facilitate mapping approaches based on haplotypes within LD blocks.

These arguments provided much of the motivation for the HapMap Project (2003). Determination of the haplotypes present in various populations should enable haplotype-tagging SNPs to be identified. These should be able to capture a defined proportion of the haplotype variation.

Despite the theoretical attractions of haplotype analysis, there are both methodological obstacles and the potential for the introduction of errors. Unambiguous haplotype assignment necessitates the use of a trio study design. Case-control samples can be used but haplotype frequencies have to be estimated using the EM algorithm. Haplotype analysis is very sensitive to genotyping error (Moskvina *et al.* 2006). This generally leads to type I error, with highly inflated significance values for haplotypes in the absence of any single marker associations. Various tests can be performed to try to identify genotyping error, including testing for Hardy Weinberg Equilibrium (HWE).

4.3.6 Statistical considerations

I will not enter into a detailed account of the statistical tests which have been applied to association data, but merely point out some of the issues regarding sources of error and interpretation of association results.

The null hypothesis in association studies is that the locus (or loci) being studied does not have an effect on phenotype, and therefore the allele frequencies are expected to be the same in the case and control groups. Hence, one is in essence simply testing the significance of the difference between observed and expected frequencies. This is generally achieved by use of the chi-squared test or a derivation thereof. The chi-squared test generates a p-value (nominal significance). This is the probability of obtaining the observed value of the chi-squared statistic if the null hypothesis is correct, i.e. it is the probability of obtaining a false positive result.

In the majority of association studies more than one chi-squared test is performed. If one follows convention and declares a p-value below 0.05 significant, then if 100 independent tests are carried out one would expect five to render, by chance, significant p-values. Consequently, account must be taken of multiple testing to avoid the probability of type I error increasing. The Bonferroni correction, a simple division of the significance level for each test by the number of tests performed, is often applied. However, it is generally considered conservative because the tests are not fully independent for markers in LD.

It is, of course, as important to avoid type II error as type I. Too stringent correction is one potential source of this. Replication of positive findings in independent samples can circumvent this problem - by testing only the associated markers from the original sample the problem of multiple testing is minimised. The debate over the interpretation of LOD scores in regions showing prior evidence for linkage has been recapitulated for association findings. The appropriate significance level to use given *a priori* evidence for association is unresolved. Theoretically, replication, or rather failure of replication,

should detect type I error in the original sample. However, interpretation of a negative result is complex in most realistic situations.

It is also possible to test for gene–gene interaction in association data using multivariate statistical techniques such as logistic regression. However, the study of such epistasis is hampered by lack of confirmed aetiological mutations.

4.3.7 Systematic LD mapping versus candidate gene approach

At the end of chapter 3 I indicated that LD mapping naturally took over where linkage stopped. Linkage can be used for approximate localisation and then fine mapping achieved by LD. Indeed, this paradigm has been successfully applied to gene discovery in Mendelian disorders, e.g. Cystic Fibrosis (Kerem *et al.* 1989). However, the systematic LD mapping of the multiple broad linkage regions implicated in BP has been prohibited by time and cost implications.

The alternative, but often complimentary, candidate gene approach has met with some success in the study of complex diseases, e.g. the identification of a susceptibility locus for Age Related Macular Degeneration (Hageman *et al.* 2005). However, as alluded to previously, the combination of the high proportion of the total number of human genes expressed in the brain and our relative ignorance of the molecular pathology of BP makes it hard to refine candidate genes. The story in BP is essentially one of a hybrid approach where the focus has been on candidate genes, lying within a region implicated by linkage. This has a conceptual attraction from a Bayesian standpoint due to the *a priori* evidence that a region actually harbours a susceptibility gene.

Until very recently systematic association testing for even a single candidate gene has represented a major undertaking. It has been argued that functional polymorphisms should first be targeted when testing any given gene, as prior probability dictates this will be the most efficient strategy for identifying an association (Sobell *et al.* 1992). However, it is far from clear that the aetiology being searched for in BP is a change in protein

sequence and laboratory based methods may be needed to determine the effect of a polymorphism on gene expression.

4.4 Association studies in bipolar disorder

The initial focus of association studies in BP was on candidate genes from the monoamine systems, particularly 5HT and dopamine, noradrenaline. These systems were primarily implicated by the known actions of medications that have been found to be effective in the treatment of BP. Genes coding for receptors of these neurotransmitters as well as critical enzymes in their production and degradation were considered candidates. The emphasis in early studies was on polymorphisms of known functional significance on gene expression or protein structure.

The association results in BP are analogous to the linkage results, in that the findings have not been strongly replicable. It has been proposed that the primary reason for this is that the effects of individual genes in BP are small and that few of the published studies have the power to detect these effects. However, it is equally plausible that the inconsistency may be primarily attributable to genetic heterogeneity. Susceptibility loci may have a major effect at an individual level but a small effect at a population level due to genetic heterogeneity. A similar argument as that used to explain inconsistent linkage can also be invoked here, i.e. that initial positive association requires greater numbers for replication. Of course, a combination of these two factors may be operating. Overall, however, the suggestion is that there are a lot of false negatives in the literature.

The burgeoning number of association studies makes it impossible to talk about each individually. A search of the GAD http://geneticassociationdb.nih.gov/cgi-bin/index.cgi (17.11.2006) for BP gave 272 entries. This will represent an underestimate of the total number of LD studies performed in BP as not all studies are submitted to the database. I will therefore restrict my consideration in this chapter to two functional candidate genes - the serotonin transporter gene (5-HTT) and monoamine oxidase A gene (MAOA) - in which meta-anlysis has supported association. For each gene I will discuss only the meta-

analyses that has been performed. I will deal with these in some detail and hope that this will illustrate the critical appraisal necessary in interpretating association results. I will begin by consideration of some general issues regarding meta-analysis of LD studies in BP.

4.4.1 Meta-analysis of association studies

As far as I am aware, there is no all-encompassing meta-analysis of association results in BP. However, a number of authors have reviewed the association findings, the most recent comprehensive review being that of Craddock *et al* (Craddock *et al.* 2001). Meta-analysis has, though, been conducted for a number of individual loci.

As discussed for the linkage data, meta-analysis is the combining of data derived from different studies. It has primarily been proposed as a solution to the problem of false negative results arising from underpowered individual studies' inability to detect genes of small effect size. However, meta-analysis should also be helpful where there is genetic heterogeneity. The example of CAPN10 and type 2 diabetes has been cited as illustrative of the successful application of meta-analysis to association studies which have yielded conflicting results (Munafo and Flint 2004). In essence, meta-analysis is achieved by determining the effect size that each study attributes to the genetic variant and weighting this according to the study size. The individual weighted effect sizes are then combined.

While the power gains of analysing what is, in effect, a sample with a larger n are obvious, the method employed in the combination also has important implications for power (Munafo and Flint 2004). The fixed effects and random effects model can be employed. The assumption of the fixed effect model is that the variant has a single true effect size and that the variation in effect sizes is due to sampling. The random effects model allows for potential differences in true effect size between studies. It is not known which model is more appropriate in the context of BP, but given the linkage heterogeneity it is highly likely that samples will differ with respect to the proportions of the various susceptibility loci present.

It is important to identify heterogeneity between studies included in the meta-analysis, as the fixed effects model assumes no heterogeneity and heterogeneity can decrease power with the random effects model. Heterogeneity has been found to be common between genetic association studies (Ioannidis *et al.* 2001). Both graphical methods and formal statistical tests of between-sample heterogeneity have been developed. The graphical methods are based on plotting the effect sizes from the individual studies against standard error (SE) or some transformation thereof. Clustering or outliers can indicate heterogeneity. One statistical test, based on a chi-squared distribution, which is often used is the Q test. However, its power to detect heterogeneity is determined by the number of studies analysed. Hence, it is generally recommended that a combination of techniques be used. Sensitivity analysis - analysis with and without the studies identified as possible sources of heterogeneity - can then be performed to assess the effect on the summary effect size.

It has long been recognised that publication bias – the increased likelihood of positive findings coming to publication – may artificially inflate effect size estimates (Sterling 1959). There are a number of approaches to detecting publication bias, including the funnel plot and calculation of the 'fail safe n' (Munafo and Flint 2004).

Meta-analysis is not necessarily restricted to univariate analysis. Meta-regression has been applied to genetic association studies to test the effects of covariates such as ancestry, e.g. (Munafo *et al.* 2005). Although less susceptible to type I error than multiple subanalysis, such *post hoc* meta-regression does bring with it the danger of confounding.

4.4.2 Meta-analysis of 5-HTT association

Perhaps the most intensively studied of BP candidates is the 5-HTT gene. The 5-HTT gene is located on 17q11.2 and the receptor brings about re-uptake of 5HT into the pre-synaptic neuron. It is essentially a functional candidate, as the linkage evidence to this region of 17q is, at best, modest. The 5-HTT receptor has been thought of as a strong

functional candidate for both BP and depression, primarily because it has been found that many antidepressants act to inhibit the re-uptake of serotonin by 5-HTT. A NPL of 1.59 at the marker D17S841 was obtained in a sample of 82 affected sib-pairs, but other linkage studies have not been supportive (Murphy *et al.* 2000).

The most studied polymorphisms at 5-HTT are the intron 2 VNTR and the 5-HTTLPR. Both are thought to be functional polymorphisms. The intron 2 VNTR has 3 alleles: Stin 2.9, Stin 2.10 and Stin 2.12. Stin 2.12 has been found to have strong enhancer properties in the mouse (MacKenzie and Quinn 1999). The 5-HTTLPR is a 44 bp insertion/deletion in the promoter region. Hence there are two allelic forms: the long (L) and short (s) alleles. The allelic form has been shown to have an effect on transcriptional activity, with lower transcription in the presence of the s allele (Collier *et al.* 1996b).

There have been a total of six meta-analyses of the association of 5-HTT with BP. In chapter 3 I dealt with the meta-analyses of the linkage studies chronologically, because there was a certain methodological narrative that I thought was worth tracing. Here, the methodological approach has been broadly similar and successive meta-analyses generally represent a more up-to-date version of the previous analysis. Hence, it seems more logical to concentrate on the most recent, and therefore comprehensive, meta-analyses and refer back to previous meta-analyses where appropriate.

The most recent meta-analysis is that of Cho *et al* (Cho *et al.* 2005). A meta-analysis was performed on both of the commonly studied polymorphisms, 5-HTTLPR and intron 2 VNTR. Separate meta-analyses of family based studies and population based case-control studies were conducted. Unlike previous studies, a pooled meta-analysis including both these datasets was then performed. This analysis gave a significantly elevated odds ratio (OR) for both polymorphisms.

Cho *et al* searched the literature for relevant articles up to September 2004. Studies of the 5HT associations in adults with BPI or BPII were included. ORs and SEs were calculated for the s allele of the 5-HTTLPR and Stin 2.12 for the intron 2 VNTR in each of the studies. No significant heterogeneity was found for either polymorphism, therefore

pooling under a fixed effects model was considered valid. Meta-analysis of the 5-HTTLPR included 17 case-control studies and six family based studies. This gave a total of 1,712 cases, 2,583 controls and 587 trios. A pooled OR of 1.12 (95% CI 1.03-1.21) was obtained for the s allele. Meta-analysis of the intron 2 VNTR included 16 case-control studies and four family studies. This gave a total of 1,764 cases, 2,703 controls and 382 trios. A pooled OR of 1.12 (95% CI 1.02-1.22) was also obtained.

The group also performed a meta-regression. It seems that this was primarily conducted to identify potential heterogeneity due to study type or ethnicity rather than exploring, say, phenotypic sub-groups that may account for the association. Neither ethnicity nor study type were found to contribute significantly to overall heterogeneity between studies. No evidence of publication bias was found using Begg's Funnel Plot, Normal Quantile Plot or Eggers Test. However a 'double bump' in Normal Quantile Plot suggested the presence of two different populations.

The meta-analysis carried out by Lasky-Su *et al* (Lasky-Su *et al.* 2005) was published slightly before that of Cho *et al*, but included more cases and controls for the 5-HTTLPR. Meta-analysis was performed only on case-control studies of the two polymorphisms in both BP and UP, i.e. four separate meta-analyses.

Medline citations as of June 2004 were identified. ORs were calculated and pooled under a random effects model. Meta-analysis of the 5-HTTLPR included 15 studies comprising of 2,774 cases and 3,652 controls. Thus, it was larger than the Cho *et al* meta-analysis, although fewer studies were included. This can be explained by the slightly different composition of studies included in the two meta-analyses.

However, a combined OR of 1.13 (95% CI 1.05 -1.22) was found, which was entirely consistent with Cho *et al*'s findings. In contrast to Cho *et al* the OR did not significantly deviate from one for the intron 2 VNTR. Lasky-Su *et al*'s meta-analysis, however, comprised of less individuals: 1,357 cases and 2,292 controls from 11 studies.

There were some analytical differences from the Cho *et al* study. Allele length was modelled continuously in the logistic regression and a random effects model was applied. In addition, a Bonferroni correction was made for the testing of two polymorphisms. Of note, Lasky-Su *et al* found significant heterogeneity for the intron 2 VNTR, which could not be explained by ethnicity or diagnostic instrument. It is thus pertinent to mention that although Cho *et al* claimed no heterogeneity between the intron 2 VNTR studies the p-value was borderline (p=0.06).

Four other meta-analyses were performed prior to this: (Lotrich and Pollock 2004); (Anguelova *et al.* 2003); (Craddock *et al.* 2001); (Furlong *et al.* 1998b). Unfortunately, no clear pattern has emerged.

Lotrich and Pollock considered case-control and family studies in separate analyses. No significant association with either polymorphism was found. Studies using Asian populations here analysed separately, because of allele frequency differences between populations. For instance, the s allele and Stin2.12 allele are more prevalent in Asian populations. However, significant heterogeneity between studies was still found for both polymorphisms in both population groupings.

On the other hand another meta-analysis of case-control studies found significant associations for both the polymorphisms (Anguelova *et al.* 2003). A combined OR of 1.14 (95% CI 1.03-1.26) was obtained for the 5-HTTLPR polymorphism, where 1,382 cases and 2,085 controls were considered. A combined OR of 1.18 (95% CI 1.05-1.32) was obtained for the Stin2.12 allele. 1,368 cases and 2,252 controls were considered. Significant heterogeneity was not found in either analysis.

Craddock (Craddock *et al.* 2001) performed a meta-analysis of 11 case-control samples which gave a total of 1,139 BP cases and 1,693 controls for the 5-HTTLPR polymorphism only. However, they failed to find evidence of association.

In addition to performing their own association study, Furlong *et al* (Furlong *et al.* 1998b) performed a meta-analysis of the 5-HTTLPR and intron 2 VNTR for both BP and UP. A significant association was found with the 5-HTTLPR, OR 1.21 (95% CI 1.00-1.45), but none with the intron 2 VNTR.

Hence, it can be seen that in the case of these two 5-HTT polymorphisms, at least, meta-analysis does not seem to entirely resolve the issue. Interpretation of meta-analysis is, indeed, not as straightforward as one might have hoped. Unlike meta-analysis of linkage studies, the meta-analytical techniques applied to association studies are well established. Yet there is still methodological variability between meta-analyses, which may explain the lack of consistent results. There are many stages in the meta-analytical process where variability can occur.

There was some variation in the studies included. Although contemporaneous, three studies which where included in Lasky-Su *et al*'s analysis of 5-HTTLPR did not feature in Cho *et al*'s analysis. Furthermore, there is a possibility that some non-independent studies have been included. For example, the study of Mendes de Oliveria *et al* (Mendes de Oliveira *et al.* 1998) was excluded by Anguelova *et al* on the basis that it was not independent from Oliveria *et al* (Oliveira *et al.* 2000). However, both studies were included by Cho *et al.* Generally the random effects model has been employed, but not universally. Cho *et al* used the fixed effects model. It is not clear whether Anguelova *et al,* who also obtained positive findings for both polymorphisms, also used this.

It is of particular interest that the tests of heterogeneity produced different results between meta-analyses. Again, it is difficult to know quite how to interpret this. While it would seem likely that the variation in allele frequencies, at least for the intron 2 VNTR, is culpable there is little empirical evidence for this. There is an indication that the studies of Collier *et al* (Collier *et al.* 1996a); (Collier *et al.* 1996b) may be atypical. Sensitivity analysis by Anguelova *et al* demonstrated that these studies made a large contribution to the positive finding for the 5-HTTLPR and intron 2 VNTR respectively. While

heterogeneity between studies does not necessarily impact on the result of the meta-analysis, it is concerning when the cause of the heterogeneity remains unknown.

It seems reasonable to conclude that the two polymorphisms do have a small effect on susceptibility to BP, with the 5-HTTLPR being the slightly more robust finding. Care has to be taken in interpreting small effect in this context; the effect is small at the population level but may be large at the individual level.

There is some modest evidence that the effect may be attributable to specific subgroups of BP patients where these 5-HTT alleles confer a much higher OR. Coyle *et al* (Coyle *et al.* 2000) genotyped the VNTR in 97 women with BP or schizoaffective disorder who had experienced at least one episode of mania or psychotic affective illness within 14 days of childbirth. The allele frequencies were compared to a control group of 72 women. The Stin2.12 allele was strongly associated, p=0.001. It was found that the s allele was associated with antidepressant induced mania in a small patient only sample (Mundo *et al.* 2001). The alternative explanation is that the polymorphisms do have a small effect in a greater proportion of the BP population. Development of BP would necessitate interaction with other genetic or environmental susceptibilities or a combination of the two.

In the absence of more refined phenotypes, meta-analysis will continue to be important in investigating loci of small effect size. The problem of detecting such effect sizes in individual studies is highlighted by Cho *et al*'s power analysis, which estimated that a sample size of 5,500 would be needed to give 80% power to detect at the 0.05 level.

4.4.3 Meta-analysis of MAOA association

The gene coding for MAOA is located on Xp11.3 (at ~43 Mb). MAOA is a key enzyme in the degradation of the neurotransmitters adrenaline, noradrenaline, dopamine and serotonin. As has already been said, these neurotransmitters are widely thought to be relevant to the aetiology of BP. Circumstantial evidence for the function candidacy of

MAOA comes from different lines of enquiry. Moclobemide, which is a highly specific inhibitor of MAOA, is an effective antidepressant. A behavioural phenotype associated with a nonsense mutation of the MAOA that had some similarities to that of mania was reported (Brunner *et al.* 1993). Subsequently, the behavioural phenotype of MAOA knockout mouse was described (Cases *et al.* 1995). Elevated aggression (in males) was found to be a feature. Whether this is relevant to the BP phenotype is, however, debatable. Positionally, it is not a strong candidate.

MAOA was first implicated by association analysis (Lim *et al.* 1995). Three polymorphisms have been commonly studied: an intron 2 CA repeat (MAO-CA) with 12 alelles; an intron 1 GT repeat, and an exonic *fnu*4HI RFLP site. This RFLP site corresponds to a synonymous G/T substitution at position 941. However, variation at this site has been shown to effect enzyme activity – with the G allele corresponding to high activity (Hotamisligil and Breakefield 1991).

Three meta-analyses have been performed. The most recent is that of Preisig *et al* (Preisig *et al.* 2000), which included their own case-control study of 272 cases (BPI or BPII) and 122 controls from France and Switzerland. They performed the meta-analysis on a total of five studies for the three markers described above. The analysis was restricted to case-control studies employing Caucasian subjects. Consequently, the family based study of Nothen *et al* (Nothen *et al.* 1995) and a Japanese study (Kawada *et al.* 1995a) were excluded. Although not explicitly stated, it would seem that these totalled 535 cases and 386 controls. No significant heterogeneity between studies was found. The MAO-CA analysis, using CLUMP, demonstrated a significant difference in the distribution of alleles between the cases and controls ($p<0.0001$). The a5 and a6 alleles were overrepresented in the cases whereas the a2 was found to be less frequent. Subanalysis by gender demonstrated that the difference was only observed in females. An association with the allele 1 of *fnu*4HI RFLP was also found in the overall sample (OR 1.45, 99.17% CI 1.01–2.08). No association was found with the intron 1 GT repeat.

An association study of the MAO-CA and *fnu*4HI polymorphisms in a sample of 106 BP cases and 250 unscreened controls was performed (Furlong *et al.* 1999a). These results were then included in a meta-analysis of these two markers. In total, seven non-overlapping studies were identified. In contrast to Preisig *et al*, the family based study of Nothen (Nothen *et al.* 1995) was included. The two studies using Asian populations were analysed separately, because of marked differences in allele frequencies for the MAO-CA. ORs were estimated by unconditional logistical regression within Statistical Package for the Social Sciences (SPSS). Rare alleles for the MAO-CA were pooled for the analysis and ORs calculated using allele 122 (a2) as the reference allele.

Significant association with the MAO-CA was found in both populations. The meta-analysis of Caucasian studies comprised of 412 BP cases and 489 controls and gave an overall OR of 1.55 (95% CI 1.06-2.28) comparing allele 116 (a5) against allele 122. Again, this association was only observed in females OR 2.13 (95% CI 1.31-3.48). The meta-analysis of the Japanese studies comprised 118 BP cases and 168 controls. The association was in this case observed when allele 114(a6) was compared against allele 122, OR 2.65 (95% CI 1.29–5.45). Association was also found with the *fnu*4HI RFLP. In Furlong *et al*'s report there is some discrepancy between the table (IIC) and the text below it. However, it seems that allele 2 is associated with decreased risk in females OR 0.7 (95% CI 0.49-0.99). If it were, indeed, in this direction it would accord with Preisig *et al*'s findings.

The MOA-CA and *fnu*4HI polymorphisms in 39 cases and 39 age/sex matched controls were analysed (Rubinsztein *et al.* 1996). A meta-analysis was then carried out which included this data. All data used in the meta-analysis was incorporated into the analyses above. Association with found with both markers which was attributable to females, in keeping with subsequent analyses.

Despite the lack of evidence for Xp from linkage studies, the meta-analyses of association studies consistently demonstrate that variation at the MAO-CA polymorphism has a modest effect on susceptibility to BP in females. Analysis and interpretation of

association at microsatellites is slightly less straightforward but overall the results appear compatible. The MAO-CA polymorphism is not thought to be functional and therefore it is postulated that it is in LD with the aetiological change. The *fnu*4HI RFLP is functional with allele 1 (presence of the RFLP site) being associated with lower MAOA activity. However, theories that invoke a reduction in enzymatic activity as the causative mechanism do not sit happily with the gender effect given that this is an X-linked gene.

4.4.4 Interpretation of meta-analysis

I think that it is reasonable to conclude from the discussion above that specific polymorphisms at the 5-HTT and MAOA genes do have a small effect on susceptibility to BP. However, it has not been adequately demonstrated whether these polymorphisms are aetiological or in LD with an aetiological change.

The associated alleles are common in the controls. Thus, one would have to invoke a theory of gene/gene or gene/environment interaction to account for the development of BP in a given individual. For the above loci an interaction is conceivable, as both associated alleles would be expected to increase 5-HT at the synaptic cleft. However, no evidence for such an interaction was found in a case-control sample from Sweden (Van Den Bogaert *et al.* 2006b). Alternatively, the alleles could be LD with a rare high-risk variant.

Yet, before one proceeds to such over-interpretation of a positive result, caution is advised. Acknowledging the potential sources of error inherent in association studies, the above discussion, particularly with regard to 5-HTT, shows that meta-analyses are not necessarily consistent. Therefore, although it is tempting to consider a meta-analysis as the definitive result, it is important to critically evaluate its methodology. For instance, even a meta-analysis may have low power to detect a small effect and thus the problem of type II error remains.

It is important to stress that 5-HTT and MAOA may not be the only genes in the literature for which there is good evidence of influence on BP susceptibility. It is simply that I have taken the approach of looking at these two genes as exemplars. Furthermore, I would not wish to give the impression that only loci which have been subjected to meta-analysis are worthy of consideration. On the contrary, a rigorous large individual study is preferable to meta-analysis, as the potential danger of combining incompatible studies is avoided. However, for the loci above it is clear that associations found in individual studies have not been consistently reproducible. This phenomenon has been seen in association studies in other common diseases. Lohumeller *et al* (Lohmueller *et al.* 2003) made some important observations on this. Taking a sample of loci for which there were multiple published association studies it was possible to demonstrate that replication occurs much more commonly than expected by chance, suggesting that in general true associations were being found. Association studies for BP were not specifically analysed and so this is not necessarily applicable to the BP data, but nonetheless it is reassuring in terms of the association methodology.

Lohmueller proposed three mechanisms which could account for inconsistency: population specific associations; false associations due to population stratification and underpowered replication samples.

For the loci above there is no evidence for ethnicity contributing to study heterogeneity, although the usual caveats about power to detect heterogeneity have to be made. There are, however, reasons to be concerned about population stratification. The majority of earlier studies neither screened subjects for ancestry nor carried out any form of genomic control. However, lack of difference in allele frequencies between control groups used in meta-analysis mitigates against this. Underpowered replication studies would certainly seem to be a cause of non-replication. The majority of individual studies had poor power to detect an effect size observed in the original positive association even without taking into account the 'winner's curse phenomenon', which predicts that the first positive association will overestimate the effect size.

It was demonstrated that meta-analysis could be successful in overcoming inconsistency in results and this also seems to be true for 5-HTT and MAOA in BP. On one point of detail Lohmuller did not include the original positive association study in the meta-analysis, considering the combined results of all subsequent studies as the replication study. This is at variance to the methodology used in the BP meta-analyses.

Even if an individual association study or meta-analysis has sufficient power to detect a small effect at a particular polymorphism it can be seen that only a very limited inference can be made about the gene's role in BP. Lack of association at that polymorphism does not exclude a role for that gene, as an aetiological change may not be in LD with the marker. Historically, it has only been feasible to look at a limited set of markers for a given gene. Furthermore, the extent of the haplotypic diversity captured is only now becoming apparent through the HapMap data.

However, association studies have very recently undergone a revolution. From being able to test a handful of RFLPs or microsatellite markers - as seen in the studies above - SNP chip technology is now allowing 500,000 or even one million SNPS to be assayed at a time. This has very recently enabled genome wide association studies to be attempted.

4.5 Genomewide association studies

To date, two genome-wide association studies have been published for bipolar disorder (2007a; Baum *et al.* 2007).

In the first a test/replication case-control design was employed to address the issue of multiple testing (Baum *et al.* 2007). The 'test' sample comprised of 461 unrelated BPI cases from NIMH BP families and 563 controls. The independent 'replication' sample comprised of 772 German BPI cases and 876 controls randomly recruited from census data. Controls were screened for a personal history of schizophrenia or affective disorder. The test sample was matched for "exclusively European ancestry". A subset of the cases

and controls was tested for population stratification and cryptic relatedness. 98% of the replication sample was of exclusively German ancestry.

Genomewide association testing was carried out using Human Hap 550 chips on pooled DNA samples. DNA pooling was used to reduce costs. 33,949 SNPs were associated at the $p<0.05$ level in the test sample. Three criteria were set out to refine this result: OR>1.4, OR <0.71, $q>0.05$ in cases and controls, location in or near a gene. Of the 1,877 SNPs which fulfilled these criteria 88 were replicated in the German sample at the p <0.05 level. 37 SNPs - mainly from this subset of replicated SNPs - were taken forward for individual genotyping in both samples.

Only modest effect sizes were observed, the largest being 1.67. The most convincing evidence for association was found at the genes SORCS2 and diacylglycerol kinase eta (DGKH). For each gene 3 SNPs were found to be associated in both samples. However only one SNP, in DGKH, remained positive after Bonferroni correction for multiple testing. SORCS2 and DGKH both lie in BP linkage regions, 4p16.1 and 13q14.11 respectively, and are good biological candidates. DGKH encodes components of the phosphatidyl inistol pathway while the SORCS2 gene product is a developmentally brain expresessed VPS10 domain containing receptor. Evidence for interaction between these loci was also found.

BP disorder was one of the seven common diseases chosen by the Welcome Trust Case Control Consortium (WTCCC) for genomewide association testing in a British population (2007a). Individual genotyping was performed on approximately 2,000 BP cases and 3,000 controls using the Affymetrix GeneChip 500 k Mapping Array set. Cases were derived from five centres. Unscreened controls were derived from the 1958 British birth cohort and UK blood donor service. Individuals with non-European ancestry were excluded. The sample was divided into 12 broad geographical regions and the allele frequencies between regions compared. On the basis of this analysis it was concluded that there was little population stratification within the British population, although 13 chromosoal regions did show geographical variation in allele frequencies.

Only one SNP - rs420259, mapping to 16p12 - exceeded the $p<5x10\text{-}7$ threshold set to declare genomewide significance. Interestingly, linkage of puerperal psychosis to this region has recently been reported by the Cardiff group, who contributed the majority of the samples to the GWA (Jones *et al.* 2007). For most of the other common diseases studied - Crohn's disease, rheumatoid arthritisis, type 1 diabetes and type 2 diabetes - the GWA results were remarkably consistent with previously replicated loci in these diseases. Unfortunately this was not evident for BP.

Despite the comparatively large sample sizes used in the GWA studies, power calculations demonstrate ability to detect association under certain scenarios may still be low. Therefore, meta-analysis of GWA data is indicated. However the GWA results do argue that susceptibly genes operating in BP are individually of modest effect size.

4.6 Summary

I have presented the association findings for two major candidate genes rather than assessed the utility of LD for fine mapping in BP. There are many examples where 'fine mapping' by LD has been used to follow up linkage studies and resulted in findings of association. However, it is difficult to synthesise the results of this approach as no gene so identified has been unequivocally implicated in BP aetiology.

Overall, one could perceive that, so far, association studies in BP have had some success in identifying polymorphisms of small effect. It is possible that this is an example of the predicted power of association to detect genes of small effect over linkage, as there was little evidence for linkage in either region. However, it is also apparent that the basis for studying functional candidates is often tenuous, and this may have contributed to the negative findings for many candidate loci. Thus, committing resources to study purely 'biological' candidates may be a potentially inefficient strategy.

Research at UCL Molecular Psychiatry laboratory, the subject of the next chapter, has attempted to increase the prior probability of finding association by focusing only on regions where there is good evidence for linkage.

Chapter 5: Background to Loci Studied

In this chapter I will put the experimental work I undertook in context. This work can be divided into that which stemmed from the previous linkage findings of the UCL laboratory and that which arose from meta-analysis of the linkage data.

Firstly, I will describe the linkage results from the UCL laboratory, which formed the basis for the subsequent fine mapping efforts on chromosomes 12, 21, 1 and 11. I will refer briefly to the preliminary association results on chromosomes 12, 21 and 11 as of the time I joined the effort. I will then go on to critique the evidence for the location of BP loci in these regions. Finally, I will explain why COMT and DAOA/G30 were chosen for attempted replication.

5.1 Regions implicated by the UCL research

The UCL laboratory has been engaged in linkage analysis and, more recently, association analysis in BP.

Since there appears to be considerable locus heterogeneity in BP, the UCL laboratory has chosen to focus on the linkage analysis of large multiply affected kindreds. Such kindreds should favour the finding of linkage in this context. Linkage analysis has been performed on 23 pedigrees from the UK (17) and Iceland (6) or subsets of these. The use of pedigrees from Iceland may further reduce locus heterogeneity in the sample as it is a relatively genetically isolated population. Families were interviewed using the SADS-L. Attempts were made to ensure that transmission was unilineal and that schizophrenia did not occur in the families selected.

The seven largest/most densely affected pedigrees (five Icelandic and two British pedigrees) were selected for a whole genome scan (Curtis *et al.* 2003). 365 microsatellites spaced at approximately 10 cM across the genome were typed. Initial two-point and

three-point parametric and nonparametric analysis was performed. Regions yielding linkage at $p<0.01$ level were subjected to four-point analysis under an assumption of heterogeneity. Overall the strongest evidence for linkage was found on 12q, 1q, 1p and 3q. The highest LOD was obtained at D12S342 – four-point heterogeneity LOD (HLOD) 2.8 using broad affection status and dominant model, nominal significance 0.0007. D12S342 maps to ~124 Mb, in the 12q24.31 region. The telomeric 1p marker D1S243 gave a HLOD of 2.1, nominal significance 0.004. Both of the markers - D1S251 (1q42.2 at ~30.0 Mb) and D3S1265 (3q29 at ~197 Mb) - gave HLODs of 2.0, $p=0.005$. The 1q42 marker D1S251 yielded a LOD of 2.5 in pedigree six.

Prior to the linkage scan, analysis had been performed in the full set of pedigrees with markers on 11p and 21q (Smyth *et al.* 1996; Smyth *et al.* 1997). A maximum LOD of 1.43 was obtained with a tetranucleotide repeat at the tyrosine hydroxylase locus, which maps to near the teleomere of 11p (Smyth *et al.* 1996). Subsequently three further markers across the 21q22 region were genotyped: pfkl, D21S171 and D21S49 (Smyth *et al.* 1997). An overall two-point LOD of 1.28 was obtained at D21S171, with one of the Icelandic pedigrees individually giving a LOD of 1.74. Affected sib-pair (ASP) analysis demonstrated increased allele sharing at this marker ($p=0.001$) and analysis using MFLINK gave a MLOD/MALOD of 2.25. A two-locus admixture linkage analysis was also performed. LOD scores from the 21q22 markers and the tyrosine hydroxylase markers were considered simultaneously. This analysis yielded a maximum LOD score of 3.87.

The general approach of the laboratory has been to follow up its own linkage findings by LD mapping using a case-control design. These mapping efforts have, however, not only been guided by the strength of the UCL linkage evidence, but due weight has also been accorded to the prior evidence for linkage in the implicated regions. This strategy has been adopted to maximise the chances of identifying susceptibility loci.

Fine mapping has focused on the 12q24 and 21q22 regions. At the time of my joining the project, preliminary results indicated multiple associated markers in these regions

(unpublished). Evidence for allelic association in the 12q24 region was first found by our collaboraters in Aarhus in a small Danish association sample. Further evidence was found in a subset of the current UCL association sample. With regard to 11p, four markers across the tyrosine hydroxylase gene and one within the tryptophan hydroxylase gene had been tested but no association found (McQuillin *et al.* 1999a; McQuillin *et al.* 1999b).

5.2 Other evidence for bipolar loci on chromosome 12q

In this chapter I will generally discuss the evidence for linkage to a chromosomal region of interest and then proceed to discuss association studies performed in this linked region. However, in this particular section it makes more narrative sense to consider the linkage and association results together.

A pedigree in which co-segregation of major affective disorder and Darier's disease, mapped to the 12q23-q24.1 region, has been described (Craddock *et al.* 1994). Dawson (Dawson *et al.* 1995) went on to investigate whether this region harbours a BP susceptibility locus by typing seven markers across the Darier's locus in 45 BP pedigrees. Suggestive linkage (maximum LOD 2.11) was obtained under a heterogeneity model with a marker at the PLA2A gene. This marker was also found to be associated in a small case-control sample (Dawson *et al.* 1995). However, this finding was not replicated in a larger British case-control sample (Jacobsen *et al.* 1996). Furthermore, mutational screening and association analysis provided no support for the involvement of the Darier's disease gene itself - ATP2A2 - in the susceptibility to BP.

A further pedigree exhibiting co-segregation of Darier's disease and major affective disorder was reported (Jones *et al.* 2002). Recently, linkage and haplotype analysis using both of these pedigrees has been performed (Green *et al.* 2005). Under a broad affection status model which included all mood disorders, multipoint parametric analysis gave a maximum LOD of 4.77 across the D12S1127–D12S1646 interval. This 6.5 Mb interval was found to delineate a haplotype which was shared by all affected individuals but did

not occur in unaffected individuals. However, denaturing high performance liquid chromatography mutation screening and DNA sequencing of 25 genes within this 'critical' region did not reveal any likely aetiological variants.

The Cardiff group went on to investigate the critical D12S1127-D12S1646 interval by LD mapping of the central 1.6 Mb portion (Glaser *et al.* 2005b). 17 microsatellites were genotyped using a DNA pooling technique in a UK population based sample consisting of 347 cases and 374 controls. Individual genotyping was subsequently performed for seven of these microsatellites, of which one - M19 - showed alleleic association (p=0.0016).

The region around M19 was screened for polymorphisms and the LD structure interrogated using 22 SNPs. Of these, five SNPs were genotyped in the full sample. Four of these - rs3847953, rs933399, rs933305 and the 9 bp insertion/deletion polymorphism rs3840795 - yielded nominally significant p-values (0.002, 0.004, 0.02 and 0.005 respectively). The associations of rs3847953 and rs933399 survived correction for multiple testing. rs3847953 maps to intron 18 of the gene CUX2, while rs933399 maps to the promotor region of the adjacent hypothetical protein FLJ32356. The insertion allele of rs3840795 results in an insertion of three glycine residues in FLJ32356. This finding was not formally supported by an attempted replication in a UK trio sample (n=110).

The Cardiff group has also performed detailed association analysis of the transcription factor gene RFX4 (Glaser *et al.* 2005a). Although it lies just outside their critical region, it is within the broad region of interest as defined by Green *et al* (Green *et al.* 2005). Furthermore, it was perceived to be a strong functional candidate based on the role of RFX4 proteins in the regulation of circadian rhythms.

Taking a similar approach to that for the M19 region, RFX4 was screened for polymorphisms and the LD structure determined using 22 SNPs. Ten hapoltype-tagging markers were identified and genotyped in 676 cases and 690 controls. No individual allelic association was found. However, haplotypic analysis revealed association with the

two-marker haplotype, rs10778502-ss24735177. A significant difference was found between the global haplotype frequencies (p=0.017) and the specific haplotype C-ins was overrepresented in cases (p=0.002). Both these associations were restricted to male cases. Although only modest evidence of association was obtained from the SNP analysis, testing of a microsatellite - D12S2072, located in intron 17 of the gene - did yield association at the p=0.001 level.

Two other independent linkage studies, which have employed more traditional affection status models, have reported significant linkage with markers in the 12q24.2-3 region: (Ewald *et al.* 1998; Ewald *et al.* 2002); (Morissette *et al.* 1999); (Shink *et al.* 2005b).

Ewald *et al* (Ewald *et al.* 1998) genotyped 16 microsatellite markers across the 12q22-24 region in two multiply affected Danish pedigrees. A two-point LOD of 3.37 at D12S1639 was obtained under a dominant model with narrow affection status definition. These pedigrees were subsequently used in a genomewide linkage scan (Ewald *et al.* 2002). The results for the 12q24.3 region were generally consistent with Ewald *et al*'s original report, yielding a LOD of 3.63 at D12S1639 (~124.9 Mb). This slight increase may be attributable to the diagnostic changes made in light of updated clinical information.

Morisette *et al* (Morissette *et al.* 1999) performed a genome scan on two large kindreds, termed VL and BM, from the Saguenay-Lac-St-Jean (SLSJ) population of Quebec. It was proposed that study of this population would reduce the problem of genetic heterogeneity as the SLSJ region was settled by French immigrants in the 1830s and has remained genetically 'rather' homogenous. Multipoint nonparametric analysis yielded a Z-score of 3.92 (p<0.01) at D12S82. This marker maps to approximately 89.3 Mb which is probably not compatible with the UCL linkage signal. However, a further 18 families were recruited from this population and parametric analysis of the pooled sample gave a maximum LOD of 3.35 at the telomeric marker D12S378 (Shink *et al.* 2005b). This was obtained under a recessive model with broad affection status, including individuals with recurrent major depression. Nonparametric analysis supported this finding with increased IBD allele sharing by affected sibs at D12S378.

Shink *et al* (Shink *et al.* 2005a) genotyped a further 20 microsatellite markers in this region in the previously reported SLSJ families and an additional 21 small pedigrees. A MLOD of 3.85 was obtained at D12S1634 under a narrow affection status model. All but one of the markers which spanned the 7.7 Mb region yielded an MLOD with a $p<0.05$.

Shink *et al* then undertook LD fine mapping of the 12q24.31 region implicated by the linkage results from the SLSJ families (Shink *et al.* 2005a). A population based sample consisting of 213 cases and 188 controls was recruited from the same geographic region. The same 20 microsatellites as used in the linkage part of the study were typed. Allelic association was found with three of these: D12S2075 ($p=0.023$), NBG6 ($p=0.008$) and NBG12 ($p<0.0001$).

In the NBG12 area, 32 genes were screened for the presence of polymorphisms in coding sequences and intron/exon junctions. 22 non-synonymous SNPs were found and genotyped. No significant allelic association was found. One uncommon polymorphism in the KIAA1595 gene and another in the FLJ22471 gene showed genotypic association with p-values <0.05 (T1 CLUMP statistic). However, the FLJ22471 polymorphism did not satisfy HWE in the control group – no heterozygotes were observed. Haplotypic association was found ($p=0.030$) with the gene HM74.

NBG6 maps to intron nine of the gene P2RX7. This gene and the adjacent genes, P2RX4 and CAMKK2, were screened for polymorphisms (Barden *et al.* 2006). 24 SNPs were selected on the basis of predicted functionality for association analysis. Allelic association ($p=0.018$) was found for one SNP - rs11065501, located approximately 75 bp downstream of P2RX4 in the slightly enlarged population based sample described above.

Haplotype analysis was performed both according to 'blocks' and using a sliding window approach. Analysis according to blocks revealed significant differences in haplotype frequencies of an eight-SNP haplotype spanning the last six exons of P2RX7 and a four-SNP haplotype, including rs11065501 (global p-values of 0.0410 and 0.0390

respectively). The sliding window analysis also gave significant Likelihood Ratio Statistics for two seven-SNP haplotypes, which were broadly compatible with the block findings.

The study proceeded to type 12 SNPs in a family sample derived from the SLSJ pedigrees and perform association analysis using FBAT. Allelic association was found with the following non-synonymous SNPs: rs2230912 in exon 13 of P2RX7 (p=0.000708) and rs3817190 in CAMKK2 (p=0.01548). The rs2230912 association withstood correction for multiple testing. rs2230912 is a Gln460Arg substitution at position 460. This amino acid is localised in a SH-3 like domain. The glycine residue (A allele) is conserved between humans and rodents. The mutated G allele (arginine) was over-transmitted to affected offspring. Recently, association has been found between the G allele and major depressive disorder in a large German case-control sample (Lucae *et al.* 2006).

P2RX7 is a member of the ATP receptor family. Depending on the level of stimulation, it acts as a cation selective channel or a non-selective large conductance pore (Li *et al.* 2005). P2RX7 agonism causes an increase in intracellular calcium that in some circumstances can lead to cell death (Zhang *et al.* 2005). Sites of brain expression include the hippocampus and pituitary gland. Study of P2RX7 knockout mice has demonstrated that the receptor mediates ATP-induced GABA and glutamate transmission in the hippocampus (Papp *et al.* 2004). The functional effect of the arginine mutation is not well elucidated.

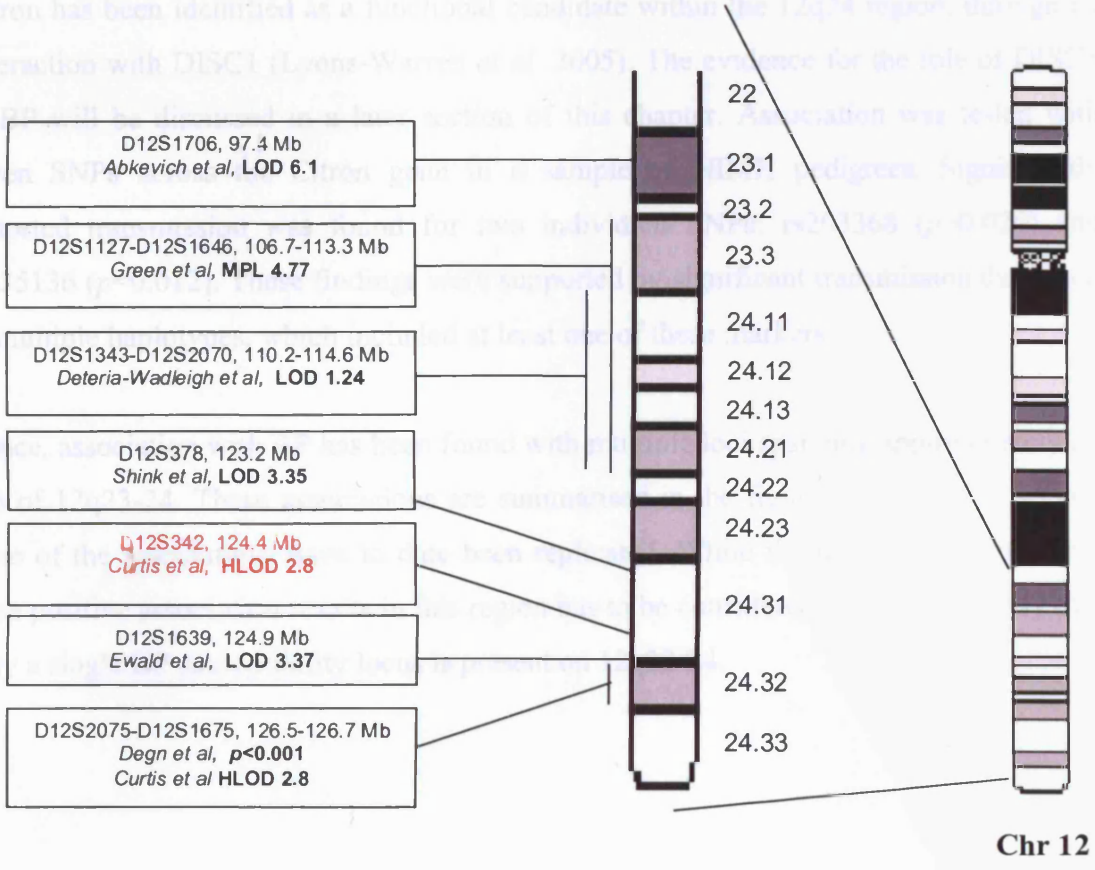
Degn *et al* (Degn *et al.* 2001) adopted a slightly different approach to mapping the 12q23-24 region. The genetically isolated population of the Faroe Islands was studied in an attempt to identify chromosomal segments shared by BP subjects. 17 microsatellite markers across the region were typed in 14 distantly related BP patients and 43 controls. Significant allelic and haplotypic association were found using CLUMP. The most significant single marker association was found at D12S342 (0.007). Analysis of two-

marker haplotypes gave two significantly associated regions: D12S1614-D12S342 (0.007 and 0.003) and D12S2075-D12S1675 (p<0.001).

In light of the increased rates of UP depression in the relatives of BP probands, the finding of highly significant linkage of UP to markers in the 12q22-q23.2 may be relevant. Linkage analysis of a large cohort of families multiply affected with UP from Utah yielded a multipoint HLOD score of 6.1 at D12S1706, at ~97.4 Mb (Abkevich *et al.* 2003). However, it should be noted that this impressive LOD was obtained under a phenotypic definition which only included affected males.

Overall, I conclude that there is strong evidence for linkage of BP to the 12q23-24 region, as summarised in figure 5.1 below. However, it has to be said that with three significant linkage studies the region is not supported by the meta-analyses of the linkage data.

Figure 5.1: Summary of linkage results compatible with UCL signal on 12q24.



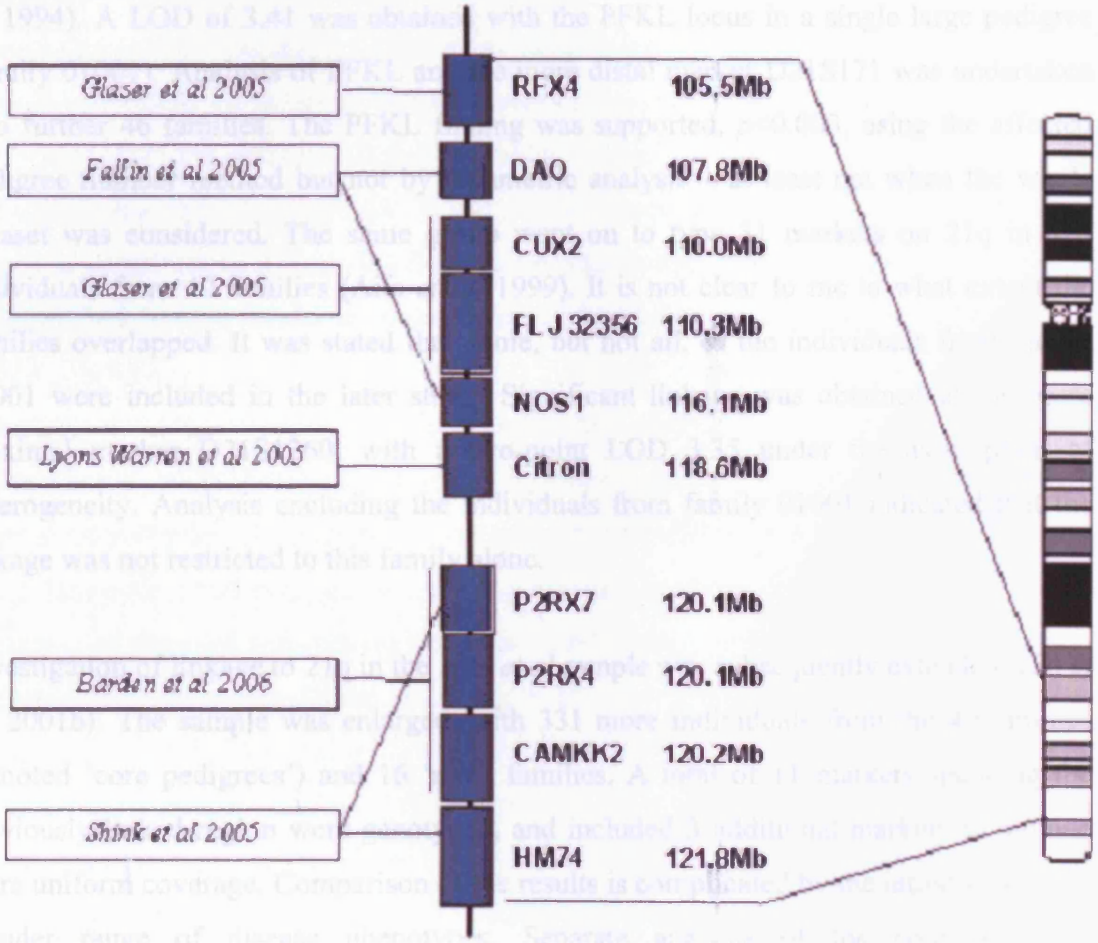
A number of other genes which are compatible with the linkage signal found in the UCL sample have been investigated by association analysis. These include DAO (107.8 Mb), NOS1 (116.1 Mb) and Citron (118.6 Mb).

DAO was identified as an interacting protein with the G72 gene implicated in schizophrenia (Chumakov *et al.* 2002). Five SNPs across the DAO region were tested as part of a screen of 64 candidate genes in BPI and schizophrenia (Fallin *et al.* 2005). In an attempt to reduce genetic heterogeneity, family samples of Ashkenazi Jewish descent were used. The BP sample consisted of 323 trios. Single SNP and haplotype based TDT analysis was performed. Association of single SNPs or haplotypes at the <0.01 level was reported as highly suggestive, while association at the <0.05 level was reported as suggestive. On this basis, DAO was reported as showing highly suggestive association in BP. However, I was unable to glean from the report any details of the association. In this study nine SNPs were also tested across the NOS1 gene and suggestive association was found.

Citron has been identified as a functional candidate within the 12q24 region, through its interaction with DISC1 (Lyons-Warren *et al.* 2005). The evidence for the role of DISC1 in BP will be discussed in a later section of this chapter. Association was tested with seven SNPs across the Citron gene in a sample of NIMH pedigrees. Significantly distorted transmission was found for two individual SNPs: rs203368 (p=0.020) and rs435136 (p=0.012). These findings were supported by significant transmission distortion of multiple haplotypes, which included at least one of these markers.

Hence, association with BP has been found with multiple loci spanning approximately 20 Mb of 12q23-24. These associations are summarised in the figure 5.2 below. However, none of the associations have to date been replicated. While the possibility of multiple false positive association results in this region has to be considered, it seems unlikely that only a single BP susceptibility locus is present on 12q23-24.

Figure 5.2: Summary of association results for the 12q23-24 region



5.3 Other evidence for a bipolar locus on chromosome 21q

5.3.1 Linkage findings for 21q

Significant linkage on chromosome 21q to BP was first reported by Straub *et al* (Straub *et al.* 1994). A LOD of 3.41 was obtained with the PFKL locus in a single large pedigree (family 01001). Analysis of PFKL and the more distal marker D21S171 was undertaken in a further 46 families. The PFKL finding was supported, $p<0.003$, using the affected pedigree number method but not by parametric analysis – at least not when the whole dataset was considered. The same group went on to type 31 markers on 21q in 373 individuals from 40 families (Aita *et al.* 1999). It is not clear to me to what extent the families overlapped. It was stated that some, but not all, of the individuals from family 01001 were included in the later study. Significant linkage was obtained at the more proximal marker D21S1260, with a two-point LOD 3.35 under the assumption of heterogeneity. Analysis excluding the individuals from family 01001 indicated that the linkage was not restricted to this family alone.

Investigation of linkage to 21q in the Aita *et al* sample was subsequently extended (Liu *et al.* 2001b). The sample was enlarged, with 331 more individuals from the 40 families (denoted 'core pedigrees') and 16 'new' families. A total of 11 markers spanning the previously linked region were genotyped, and included 3 additional markers to provide more uniform coverage. Comparison of the results is complicated by the introduction of a broader range of disease phenotypes. Separate analysis of the core pedigrees demonstrated slightly increased evidence for linkage at D21S1260 with a maximum two-point LOD of 3.56. The threshold for suggestive linkage was almost achieved in the new pedigrees at the slightly proximal marker D21S266, LOD 1.89. Combined analysis of the pedigrees did not result in an appreciable change of the LOD score at D21S1260.

Linkage to 21q has been supported by multiple subsequent linkage studies. Detera-Wadleigh *et al* (Detera-Wadleigh *et al.* 1996) initially genotyped 18 markers on chromosome 21q in 22 multiplex families – the Clinical Neurogenetics Initiative (CNG)

pedigrees. ASP analysis demonstrated increased allele sharing at nine of these marker loci (p=0.049-0.0008). Multilocus ASP analysis revealed two intervals showing excess allele sharing: D21S270–D21S171 and D21S1436–D21S65. Subsequently, a high density genome scan was performed using these 22 families (Detera-Wadleigh *et al.* 1999). In this analysis the maximum evidence for linkage on 21q came from the proximal region bounded by the markers D21S1254-D21S65 (LOD 1.85, p=0.0018). Detera-Wadleigh *et al* also carried out linkage analysis in an independent sample of 97 families from the National Institute of Mental Health Genetics Iniative (NIMHGI). Four markers - D21S1254, D21S65, D21S1440 and D21S1255 - exhibited excess allele sharing with p-values ranging from 0.041-0.008. These markers overlap both the proximal and distal intervals identified in Detera-Wadleigh's original study. Multilocus analysis provided stronger evidence for increased allele sharing with a minimum p-value of 0.0009.

Independent evidence was found for linkage to the original marker, PFKL (Kwok *et al.* 1999; Kelsoe *et al.* 2001). Kwok *et al* performed analysis of six 21q markers in 12 Australian pedigrees. Nonparametric analysis using SIMIBD gave linkage at the $p<0.001$ level. However, other nonparametric and parametric analyses did not support this result. Kelsoe *et al* obtained a maximum two-point LOD score of 2.04 under a broad phenotype definition and dominant model at PFKL in a genome survey of 20 North American families.

Linkage analysis of families from Eastern Quebec has also provided support for a 21q locus. As referred to in section 5.2, a genomewide scan was performed on a single very large pedigree from the SLSJ region (Morissette *et al.* 1999). Multipoint nonparametric analysis using GENEHUNTER yielded a Z score of 4.81 ($p<0.01$) at D21S263. This peak was found to be mainly due to one branch of the pedigree. Furthermore, linkage analysis to BP and schizophrenia was carried out in 21 multi-generational pedigrees using 220 markers on 13 'target' chromosomes (Maziade *et al.* 2001). Seven of the pedigrees were predominately schizophrenic, six predominately BP and eight mixed. Linkage to BP was tested in the BP and mixed pedigrees. A maximum LOD score of 2.03 was obtained at

D21S1893 using a narrow phenotypic definition. However, this did not reach the threshold for suggestive linkage according to their correction for multiple testing.

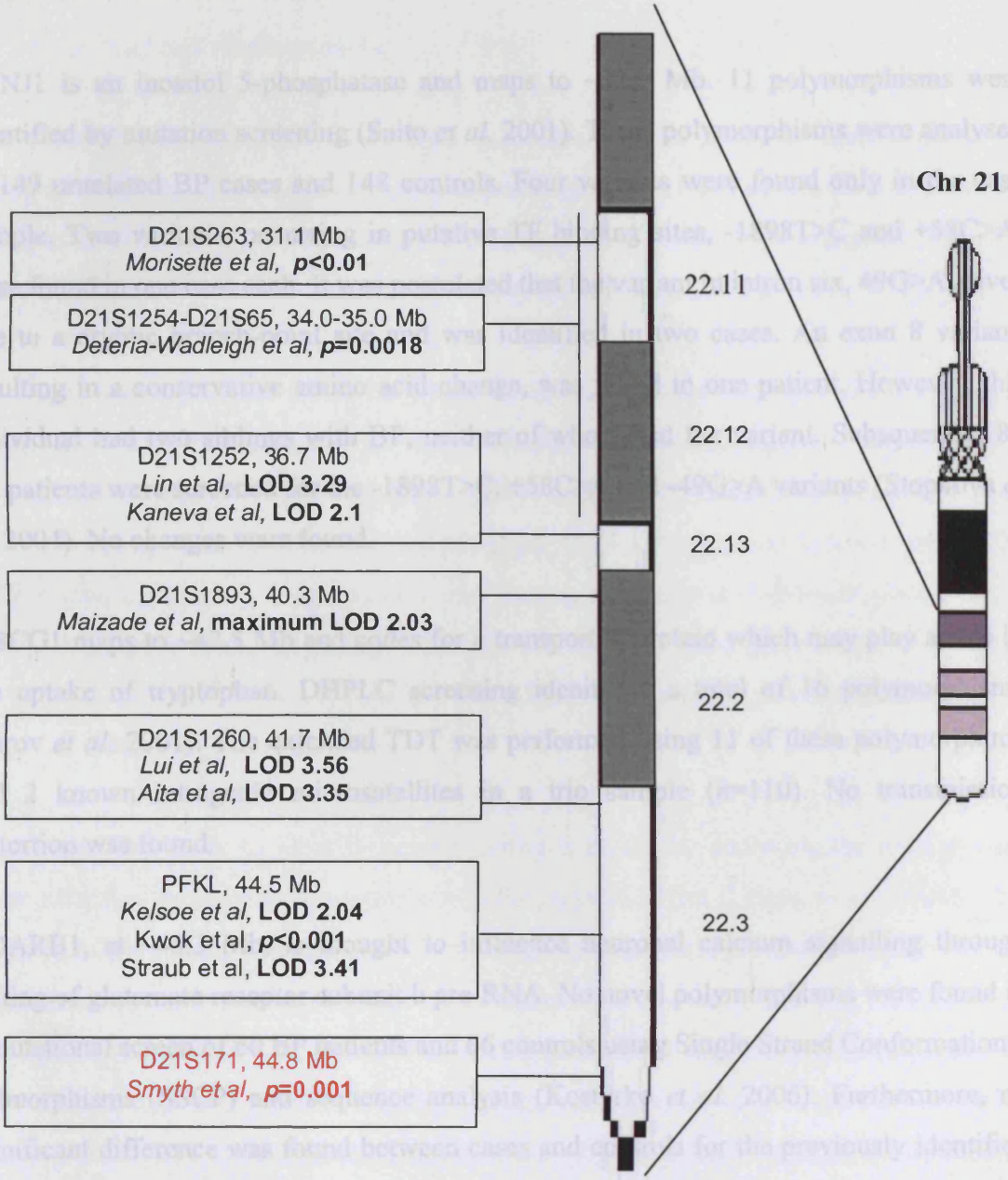
Linkage analysis was performed using 11 markers across the 21q22-23 region in 18 small to medium sized Bulgarian pedigrees (Kaneva *et al.* 2004). Suggestive linkage was observed with the marker D21S1252, with a multipoint LOD of 2.1, p=0.0003 under a narrow phenotype definition.

Most recently, the combined linkage data from the first and second waves of the NIMHGI for BP was re-analysed incorporating age of onset as a covariate (Lin *et al.* 2005). Analysis using the affected relative pair approach produced a LOD of 3.29 near D21S1252, with increased linkage among subjects with early age of onset (defined as below 21). This result was corroborated by the reanalysis of an independent replication sample of 65 pedigrees with a maximum LOD of 2.88 near D21S156. However, this finding was not supported by a genome scan of 70 European families ascertained through early onset probands (Etain *et al.* 2006).

Meta-analysis yielded a single analysis MSP of 0.006 and a best analysis MSP of 0.01 at 31 cM on 21q (Badner and Gershon 2002a). However, neither replication analysis gave significant results, suggesting the linkage may be due to one study. Furthermore, subsequent meta-anlyses have not implicated this region.

The linkage results are summarised in figure 5.3 below. Care needs to be exercised in interpreting the linkage results for 21q, as there is some overlap of samples between studies. Furthermore, while there is often large variation in the LODs obtained with different analyses this seems particularly pronounced in relation to the 21q results. Having said this, the overall evidence for a susceptibility locus in the 21q22 region is compelling.

Figure 5.3. Summary of linkage results compatible with UCL signal on 21q22



5.3.2 Association findings for 21q

LD mapping of the 21q22 linkage region has not, to date, been extensive. Association and mutation analysis has been undertaken for three genes which would be compatible with the UCL linkage results: synaptojanin 1 gene (SYNJ1), ATP-binding cassette transporter G1 gene (ABCG1) and ADARB1.

SYNJ1 is an inositol 5-phosphatase and maps to ~32.9 Mb. 11 polymorphisms were identified by mutation screening (Saito *et al.* 2001). These polymorphisms were analysed in 149 unrelated BP cases and 148 controls. Four variants were found only in the case sample. Two variants, occurring in putative TF binding sites, -1898T>C and +58C>A, were found in one case each. It was postulated that the variant in intron six, 49G>A, gives rise to a cryptic branch-point site and was identified in two cases. An exon 8 variant, resulting in a conservative amino acid change, was found in one patient. However, this individual had two siblings with BP, neither of whom had the variant. Subsquently, 84 BP patients were screened for the -1898T>C, +58C>A and -49G>A variants (Stopkova *et al.* 2004). No changes were found.

ABCG1 maps to ~42.5 Mb and codes for a transported protein which may play a role in the uptake of tryptophan. DHPLC screening idenitified a total of 16 polymorphisms (Kirov *et al.* 2001). The extended TDT was performed using 11 of these polymorphims and 2 known intragenic microsatellites in a trio sample (n=110). No transmission distortion was found.

ADARB1, at ~45.3 Mb, is thought to influence neuronal calcium signalling through editing of glutamate receptor subunit b pre-RNA. No novel polymorphisms were found in a mutational screen of 60 BP patients and 66 controls using Single Strand Conformational Polmorphisms (SSCP) and sequence analysis (Kostyrko *et al.* 2006). Furthermore, no significant difference was found between cases and controls for the previously identified synonymous exon ten variant, +2234A>G.

While testing of rare SYNJ1 variants in larger samples would seem to be warranted, none of the genes above have been associated with BP.

5.4 Other evidence for a bipolar locus on chromosome 1q

5.4.1 Linkage findings on 1q

Under traditional diagnostic models no individual linkage study or meta-analysis of linkage studies has, to date, yielded genomewide significant LOD scores on 1q. However, multiple studies have reported suggestive linkage that would be compatible with the telomeric signal obtained by the UCL laboratory (Curtis *et al.* 2003).

A LOD of 1.98 at D1S103 (~229 Mb) was obtained in a series of moderately sized North American pedigrees (Gejman *et al.* 1993). Increased allele sharing was found between affected individuals at the same marker, $p<0.0001$, in the Old Order Amish (LaBuda *et al.* 1996). A linkage scan of 22 multiplex pedigrees yielded suggestive LODs in the 1q32-1q41 region under both nonparametric and parametric analysis (Detera-Wadleigh *et al.* 1999). Multipoint analysis using GENEHUNTERPLUS(fam) gave a LOD of 2.67 ($p=0.00022$) at the D1S1660-D1S1678 (196.9-201.8 Mb) interval and a parametric LOD of 2.37 at GATA124F08 (~207 Mb).

An attempt was made to identify genetic subtypes of BP by analysing the linkage data under affection models with greater phenotypic specification (Cheng *et al.* 2006). The presence of psychosis, suicidal behaviour and panic disorder were used to subdivide the traditional diagnostic models. Multipoint nonparametric analysis yielded an NPL score of 2.01 near D1S1609. NPL scores of >1.63 were considered suggestive of linkage on the basis of their own simulations. This NPL score was obtained under a traditional model which included BPI and BPII individuals as affected. In this region phenotypic dissection did not reveal a possible genetic subtype.

The debate continues as to whether schizophrenia and BP often occur in the same family. However, analysis of families where this seems to have occurred has yielded positive results on 1q.

Most recently, a genome scan was performed on families originating from the old Choggia population - a town on an island of the Venetian lagoon - with either schizophrenia or BP (Vazza *et al.* 2007). A single affection model classifying individuals with schizophrenia, BP or schizoaffective disorder as affected was employed. Non-parametric analysis yielded a NPL score of 1.51 on 1q43 but this did not reach the threshold for suggestive linkage based on their own simulations.

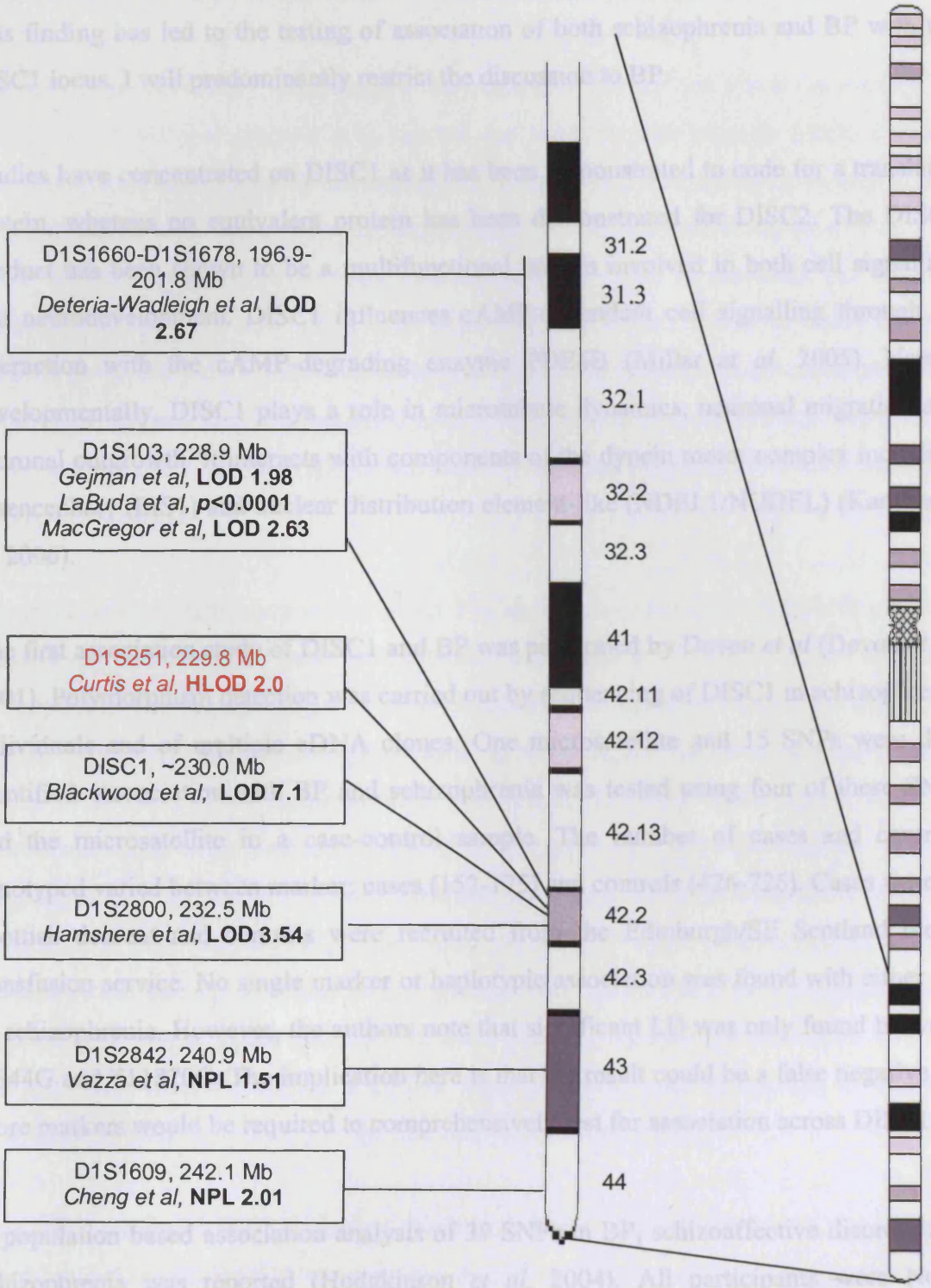
Linkage was performed on 22, predominately extended, Scottish families affected by BP or schizophrenia (Macgregor *et al.* 2004). Analysis of the 'BP' families yielded a maximum two-point LOD of 2.63 at the marker D1S103 (~229 Mb) under a recessive model with broad phenotypic definition. 'BP' families in which cases of schizophrenia occurred were included in this analysis. Multipoint variance component linkage analysis was also carried out for the chromosome 1 markers. This gave a maximum LOD at a slightly more centromeric location near D1S419 (~213 Mb). Interestingly, no evidence for linkage was found in the 'schizophrenic' families. However, the effect of analysing linkage data using more lax diagnostic boundaries was dramatic.

Pedigrees containing at least one member with a diagnosis of schizoaffective disorder (manic type) were selected from a collection of pedigrees ascertained for schizophrenia and BP linkage scans (Hamshere *et al.* 2005). An affection status model which included schizophrenia and BPI as well as schizoaffective disorder (manic type) was then applied to these pedigrees. Genomewide significant linkage, LOD=3.54, was obtained on 1q42 near the marker D1S2800 (~233 Mb). Further analysis of LOD scores from individual families suggested that both BP and schizophrenic families were contributing to this significant finding. Sex covariate linkage analysis was also performed which led to a significant increase in the LOD score with increased allele sharing in same sex sib-pairs (either male or female).

In light of these findings, it is worth noting the consequence of varying the affection models on the linkage results from the Scottish (1;11) (q42;q14.3) translocation family. The highest LOD, 7.1, was obtained with the broadest model, including individuals with schizophrenia, BP and recurrent major depression as affected (Blackwood *et al.* 2001). More restricted models: schizophrenia only and affective disorders only gave LODs of 3.6 and 4.5 respectively.

The linkage results are summarised in figure 5.4 below. Considering only those linkage studies which have respected traditional diagnostic boundaries, there is reasonable evidence for a 'pure' BP susceptibility locus in the region. However, with reference to Hamshere and Blackwood's findings it is difficult to escape the conclusion that a susceptibility locus for a non-traditional phenotype, encompassing the schizophrenia, schizoaffective, BP and recurrent depression phenotypes exists in this region. It is also plausible that a linkage signal is being detected from more than one locus in the region. It is of significance that linkage analyses of schizophrenia performed at the UCL laboratory, in which an identical set of markers to those used by Curtis *et al* was utilised, gave a peak LOD on 1q in a more proximal position with no overlap with the positive LODs for BP (Gurling *et al.* 2001). In other words, in the UCL data, at least, there is evidence for separate susceptibility loci for schizophrenia and BP on 1q.

Figure 5.4. Summary of linkage results compatible with UCL signal on 1q31-44



5.4.2 Association results for DISC1

Analysis of the breakpoint on chromosome 1q42 of the Scottish translocation family revealed direct disruption of two genes, denoted DISC1 and DISC2 (Millar *et al.* 2000b). This finding has led to the testing of association of both schizophrenia and BP with the DISC1 locus. I will predominantly restrict the discussion to BP.

Studies have concentrated on DISC1 as it has been demonstrated to code for a translated protein, whereas no equivalent protein has been demonstrated for DISC2. The DISC1 product has been shown to be a multifunctional protein involved in both cell signalling and neurodevelopment. DISC1 influences cAMP-dependent cell signalling through its interaction with the cAMP-degrading enzyme PDE4B (Millar *et al.* 2005). Neuro-developmentally, DISC1 plays a role in microtubule dynamics, neuronal migration and neuronal outgrowth. It interacts with components of the dynein motor complex including lissencephaly (LIS1) and nuclear distribution element-like (NDEL1/NUDEL) (Kamiya *et al.* 2006).

The first association study of DISC1 and BP was performed by Devon *et al* (Devon *et al.* 2001). Polymorphism detection was carried out by sequencing of DISC1 in schizophrenic individuals and of multiple cDNA clones. One microsatellite and 15 SNPs were thus identified. Association with BP and schizophrenia was tested using four of these SNPs and the microsatellite in a case-control sample. The number of cases and controls genotyped varied between marker: cases (152-175) and controls (426-726). Cases were of Scottish descent and controls were recruited from the Edinburgh/SE Scotland blood transfusion service. No single marker or haplotypic association was found with either BP or schizophrenia. However, the authors note that significant LD was only found between A844G and T11870C. The implication here is that the result could be a false negative, as more markers would be required to comprehensively test for association across DISC1.

A population based association analysis of 39 SNPs in BP, schizoaffective disorder and schizophrenia was reported (Hodgkinson *et al.* 2004). All participants were North

American Caucasians of European descent. The BP group consisted of 82 individuals, which was compared against a control group of 217. None of the markers were individually associated with BP. However, allelic association was observed with schizophrenia (five markers) and schizoaffective disorder (one marker).

The authors proceeded by identifying seven 'haplotype blocks' for which the average *D'* was >0.9. Haplotype analysis was carried out with respect to each block. Disease association was found detected for haplotypes from the first four consecutive blocks. For BP, association was found with haplotypes from block one and block four. Two rare block one haplotypes, denoted hap8 and hap14, were found in cases but not controls (both p=0.0192). One rare block four haplotype, hap7, was found in cases but not controls (p=0.0048) while the common hap1 was underrepresented in BP cases (p=0.027). None of the BP associations remained after Bonferroni correction, although the authors argue that its application is overly conservative. Haplotypic association was also detected with schizophrenia.

A comprehensive association analysis of a 510 kb region, which included DISC1 and the immediately centromeric gene translin-associated factor X (TRAX), in BP and schizophrenia was attempted (Thomson *et al.* 2005). The region of study was extended to include the TRAX gene as intergenic splicing with DISC1 has been shown to occur (Millar *et al.* 2000a). To optimise SNP selection an LD map of the region was first constructed using the information from 45 SNPs genotyped in 29 Scottish BP parent–offspring trios. Tagging SNPs thus identified were supplemented with SNPs showing association in previous studies, giving a total of 30 SNPs which were carried forward to the association testing. The case-control groups were drawn from the same populations and numbered 381 BP cases and 478 controls. Following the observation in schizophrenia that the under-transmission of a DISC1 haplotype was only significant in affected females (Hennah *et al.* 2003), sex specific analysis was undertaken.

Only one marker, rs1030711, showed allelic association with BP (p=0.032). This marker is not located within DISC1 but maps to intron 5 of TRAX. No single markers showed

association with schizophrenia. Analysing the sexes separately yielded one associated marker, rs821663, with BP males (p=0.047). By contrast, rs821663 is located towards the 3' end of DISC1 within exon 10.

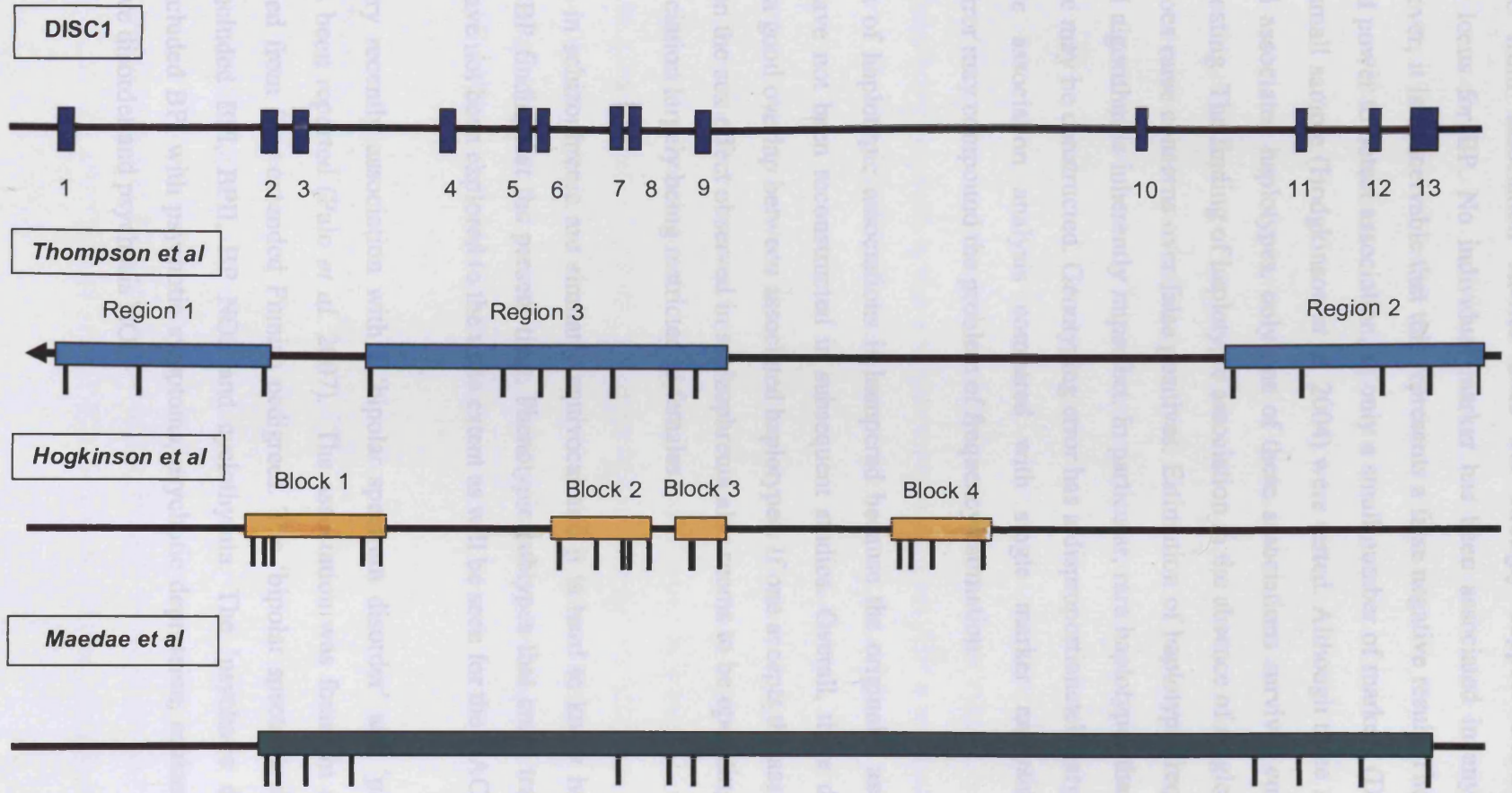
Several haplotypes across the TRAX/DISC1 showed global and/or individual association with BP. Interpretation of these haplotypic associations is not straightforward. The strongest association was found with haplotypes from what can loosely be described as the central portion of DISC1 spanning rs1954175 to rs6675281 (Thompson *et al*'s region 3). Analysis by sex revealed that association is restricted to affected females. This region was also the only region to show association with schizophrenia.

However, association with BP was also found at both the 'centromeric' region 1 and 'telomeric' region 2. Thompson *et al*'s region 1 refers to the interval spanning TRAX and exons 1 and 2 of DISC1, rs1630250-rs3738401. The global and individual haplotypic associations indicate that this region is associated with BP males. Thompson *et al*'s region 2 refers to the interval spanning rs821663-rs11122396. The gender effect observed with the individual marker rs821663 was not supported by the haplotype analysis. The only haplotypic associations found in this region were for a number of individual haplotypes which occurred in the female BP group and not the control group. Thompson was unable to test the haplotypes associated with BP in the Hodkingson *et al* study, as the constituent SNPs were not typed. However, it can be seen from figure 5.5 (after Maeda *et al*) below that the regions do not overlap.

A family based association study of DISC1 was conducted (Maeda *et al.* 2006). Fifty-seven pedigrees from the Hopkins/Dana collection were selected and 12 SNPs genotyped in a total of 297 subjects. The SNP set included one novel SNP which had been discovered by direct sequencing of ten BP subjects prior to the association phase of the study. These subjects were identified from pedigrees with the strongest evidence for linkage to 1q42. Tests of allelic and haplotypic association were performed using FBAT. Sex specific analysis was also carried out. No individual marker showed association with BP. 'Various' haplotypes were tested. Two 'common' 12 marker haplotypes, denoted

HP1 and HP2, showed association. HP1 was over-transmitted to affected females (p=0.004) while HP2 was under-transmitted. Maeda *et al* proceeded to examine the relationship between the 12 SNP haplotypes and DISC1 expression in human lymphoblasts. Consistent with the association results, DISC1 mRNA in females was found to be significantly lower in affected individuals carrying the HP1 haplotype as compared with unaffected individuals with the 'protective' HP2 haplotype (p=0.001). The correlation of mRNA levels with a number of clinical variables was then investigated. A statistically significant inverse correlation was found with a number of manic symptoms

Figure 5.5: Summary of haplotypic associations with DISC1.



The evidence from association studies does not strongly support DISC1 as a susceptibility locus for BP. No individual marker has been associated in any of the studies. However, it is conceivable that this represents a false negative result. The earlier studies lacked power to detect association, as only a small number of markers (Devon *et al.* 2001) or small sample (Hodgkinson *et al.* 2004) were tested. Although three separate studies found associated haplotypes, only one of these associations survives correction for multiple testing. The finding of haplotypic association in the absence of single marker association does raise concerns over false positives. Estimation of haplotype frequencies using the EM algorithm is inherently imperfect. In particular, rare haplotypes that do not exist in nature may be constructed. Genotyping error has a disproportionately large effect of haplotypic association analysis compared with single marker analysis. Such genotyping error may compound the problem of frequency estimation.

Interpretation of haplotypic associations is hampered because the originally associated haplotypes have not been reconstructed in subsequent studies. Overall, there does not appear to be a good overlap between associated haplotypes. If one accepts the association to be real then the sex effect observed in schizophrenia also seems to be operating in BP, with the association largely being restricted to females.

The findings in schizophrenia are similarly equivocal and it is hard to know how they relate to the BP findings at the present time. Phenotypic subtypes that cross traditional boundaries have not been explored to the same extent as will be seen for the DAOA locus.

However, very recently association with a 'bipolar spectrum disorder' and 'psychotic disorder' has been reported (Palo *et al.* 2007). The association was found in a family sample derived from four extended Finnish pedigrees. The 'bipolar spectrum disorder' phenotype included BPI, BPII, BP NOS and cyclothymia. The 'psychotic disorder' phenotype included BPI with psychotic symptoms, psychotic depression, schizophrenia, schizoaffective disorder and psychosis NOS.

One marker, rs1655285, showed individual association (p=0.025). SNP rs1655285 maps to intron 5 of TRAX (referred to as TSNAX in the paper). Analysis of 2-SNP haplotypes supported the association of TRAX with 'bipolar spectrum disorder'. Haplotypic association was also found with 2-SNP and 3-SNP haplotypes at the 3' end of DISC1. This pattern of association is similar to that obsevered by Thomson *et al.*

No individual marker association was found with 'psychotic disorder'. Furthermore analysis of 2-SNP haplotypes did not yield association in the whole dataset. However, analysis by gender revealed association with 2-SNP and 3-SNP haplotypes at the 5' end of DISC1. This male specific association in the 5' region of DISC1 was also observed by Thomson *et al.*

The question of which phenotype is of primary relevance at the DISC1 locus is further thrown into question by the finding that a single non-synonymous SNP is associated with major depressive disorder (MDD) (Hashimoto *et al.* 2006).

While DISC1 has received most attention in this region other genes that are compatible with the UCL linkage have been investigated by association analysis. Eleven polymorphisms across the adenosine A1 receptor were genotyped in a very small case-control sample (Deckert *et al.* 1998). No association was found. However, allelic association was found with a functional polymorphism in the angiotensinogen gene in a case-control sample from Brazil, p=0.009. The angiotensinogen gene maps very close to DISC1 at 229 Mb (Meira-Lima *et al.* 2000).

5.5 Other evidence for a bipolar locus on chromosome 11p

5.5.1 Linkage findings on 11p

The first significant linkage to BP was found with the HRAS1 locus which maps to the teleomere of 11p, at ~0.5 Mb (Egeland *et al.* 1987). While this finding has subsequently

been brought into question, other independent studies have reported supportive linkage in this region.

Egeland *et al* performed linkage analysis of a large composite pedigree (family 110) from the Old Order Amish community of Pennsylvania using RFLP markers. Comprehensive genealogical records were available for the community and indicated that the entire community of 12,000 was descended from 30 founders who emigrated from Europe in the 18th Century. Analysis under a dominant model with the penetrance value set to 0.63, in accordance with the segregation analysis, gave a two-point LOD of 3.34 at HRAS1. Three point analysis, including data from the nearby insulin locus, yielded a peak LOD score of 4.90 at HRAS1. Subsequent analysis of linkage to these markers using a three-allele model resulted in an increase in the LOD scores (Sandkuyl and Ott 1989).

Diagnostic revision, additional genotyping and extension of the original 'core' pedigree would seem to have negated the evidence for linkage to the 11p15.5 region. The 'core' pedigree was re-analysed using updated information (Kelsoe *et al.* 1989). Missing genotypes for ten unaffected individuals were obtained, two new unaffected individuals were added to the pedigree and two individuals reclassified as affected, having developed affective disorder in the interim. Evidence for linkage to HRAS1 and INS was markedly reduced following these amendments. Extension of the 'core' pedigree, right and left lateral extensions, further reduced the evidence for linkage. In fact, linkage to HRAS1 or INS was excluded in the large right lateral extension. Linkage to 11p was also excluded in a second large lateral extension of the 'core' pedigree (Pauls *et al.* 1991).

Theoretically, genotyping error could contribute to the discrepancy between results. The genotyping of individuals from the 'core' pedigree was repeated (Law *et al.* 1992). Due to the difficulty of resolving some of the alleles of the INS marker an alternative assay was designed. Scoring of INS alleles and diagnostic models differed somewhat from those of Kelsoe *et al* but the linkage findings were in general agreement.

Under a heterogeneity model, suggestive linkage was found - maximum LOD 2.89 - with a marker pj4.7/*taq*1 which flanks the tyrosine hydroxylase locus (Malafosse *et al.* 1997). The tyrosine hydroxylase gene is located at approximately 2 Mb. Four markers spanning tyrosine hydroxylase, INS and HRAS1 loci were typed in 11 French pedigrees. Analysis was performed using a broad disease definition. Two of the families yielded positive LOD scores when analysed individually.

Analysis of the second wave of the NIMH collection provided evidence of linkage to the 11p15 region (Zandi *et al.* 2003). This sample comprised of 56 families from the USA. NPL scores of more than two were found for the region extending from D11S1984 to D11S2362 using an intermediate phenotypic model. The peak score, 2.96 (p=0.002), occurred at D11S1923. The results of a parametric analysis, allowing for heterogeneity, were consistent with this finding.

Subsequent analysis of the combined NIMH wave 1 and wave 2 datasets, comprising of a total of 153 families, supported the 11p15 finding (McInnis *et al.* 2003a). The peak NPL score was again found at D11S1923, but was somewhat reduced at 2.2 (p=0.01). Possible heterogeneity or epistatic effects were tested but the analyses did not increase the evidence for linkage on 11p15.

Other investigators have implicated a slightly more centromeric region, although it is possible they have detected the same linkage signal. Suggestive linkage, two-point LOD 1.95, was obtained with the 11p14 marker D11S1312, which maps to approximately 30 Mb (McInnes *et al.* 1996). This result emanated from a genomewide linkage scan of two large Costa Rican pedigrees, CR001 and CR004, ascertained from the genetically homogenous population of the Central Valley. A narrow definition of affection, including only those individuals with BPI or schizoaffective disorder (manic type) was employed.

Positive LODs were also obtained in the more centromeric region under a dominant model and narrow phenotype definition (Detera-Wadleigh *et al.* 1999). However, the LODs of 1.89 and 1.62 at D11S915 and D11S904 respectively cannot be considered as

confirming linkage in light of the re-evaluation of Egeland *et al*'s findings. This genome scan of the CNG families has been described in the section 5.3.1. It is of note that the pedigrees included the 'right extension' of the Old order Amish family 110.

In summary, the linkage evidence for the 11p14-15 region is not as compelling as for the other regions discussed so far. Suggestive rather than significant linkage has been found, with a telomeric position being favoured.

5.5.2 Association findings on 11p14-15

Association studies in the 11p14-15 region have focused on four promising functional candidates: tyrosine hydroxylase (TH), dopamine D4 receptor (DRD4), tryptophan hydroxylase (TPH) and brain-derived neurotrophic factor (BDNF). The association literature relating to these loci, particulary TH, is very large. I will discuss the BDNF findings in the most detail, as I tested this association in the UCL sample. Thus for the other loci I will, where appropriate, restrict the discussion to meta-analyses or reviews rather than individual studies.

Tyrosine hydroxylase

TH catalyses the hydroxylation of tyrosine to L-Dopa. It is the first, and importantly, the rate-limiting step in the synthetic pathway of dopamine and noradrenaline. Catehcolamines have widely been implicated in the aetiology of BP. For instance, anti-dopaminergic drugs are effective against mania and dopaminergic drugs, such as amphetamine, cause behavioural changes which, to some extent, mimic those seen in mania. Hence, TH has been thought to be an excellent functional candidate for BP.

Association with TH was first reported by Leboyer *et al* (Leboyer *et al.* 1990), who tested two RFLPs - a 5' *Taq*1 site and 3' *Bgl*II site - in a small French sample. The same group subsequently reported genotypic association with a tetranucleotide VNTR in intron one (Meloni *et al.* 1995). Several association studies of the TH locus have now been performed. The *Taq*1 and tetranucleotide (now referred to as HUMTH01) have been the

most commonly studied polymorphisms. The tetranucleotide polymorphism is thought to be functional. It was demonstrated that allelic variation had a quantative silencing effect on TH expression through differential binding of the zinc finger protein ZNF 191 (Albanese *et al.* 2001).

Three meta-analyses have now been performed. The most recent is that of Furlong *et al* (Furlong *et al.* 1999b). The meta-analysis was restricted to the intron one tetranucleotide VNTR. The group searched the Bath Information and Data Services database in March 1998. They identified 12 studies, including their own findings, of which 11 were included in the analysis. One Japanese study (Kawada *et al.* 1995b) was omitted due to uncertainty over how the alleles they reported corresponded to other published alleles. The rare allele (termed 6 or F), found only by Furlong *et al* and Todd (Todd and O'Malley 1989), was also omitted from the analysis. Both case-control and family based studies were included. In total, the meta-analysis comprised of 583 BP(I or II) cases and 745 controls. Meta-analysis was performed using unconditional logistic regression as implemented in the SPSS. The authors claimed such analysis was robust to differences to control allele frequencies in the different studies. No evidence of association was found. A further meta-analysis, including UP cases as affected, also failed to show association.

A meta-analysis had previously been performed (Turecki *et al.* 1997), but a number of methodological issues were subsequently highlighted (Bellivier *et al.* 1998b). Turecki *et al*'s primary analysis was of association of either the *Taq*1 RFLP or intron 1 VNTR with BP. Separate analyses of the two markers were also performed. The group carried out a search which encompassed published data up to January 1996. They included eight studies, totalling 547 cases and 522 controls. There was some variation in the studies included in this meta-analysis compared to that of Furlong *et al.* For example, the study by Kawada *et al* (Kawada *et al.* 1995b) was not excluded.

On the basis of a homogeneity test, Turecki *et al* considered the analysis of pooled data to be valid. No evidence for association was found when the markers were considered together or separately. However, Belliever *et al* objected to the pooling of studies using

different markers, as the LD between the marker and putative aetiological change may differ. Turecki *et al* argued that it was valid to pool because meta-analysis combines a "relative measure of the observed effect towards or against a hypothesis". Furthermore, the studies were testing the same hypothesis, i.e. that variation at the TH locus influences susceptibility to BP.

Belliever *et al* (Bellivier *et al.* 1998b) performed their own meta-analysis of the *Taq*I and *Bgl*II polymorphisms. However, they found significant heterogeneity of allele frequencies between centres for the *Taq*I marker, and between control groups for the *Bgl*II marker. The group did not perform meta-analysis for the intron 1 VNTR, arguing that the variation in tests of association applied in the constituent studies rendered this "unfeasible". In line with Turecki *et al*, they found no association with *Taq*I. They found a significant association with *Bgl*II (p=0.02).

Since these meta-analyses, a number of individual studies of TH have been published, all of which have been negative: (McQuillin *et al.* 1999a), (Muglia *et al.* 2002) and (Serretti *et al.* 2003).

The sizes of these meta-analyses are not great and they may not have sufficient power to detect small effects. As discussed in chapter 4, the issue of variation in studies included and, importantly, heterogeneity in allele frequencies is highlighted by the TH example. The findings of Belliever *et al* indicate that such heterogeneity can occur even if the analysis is restricted to samples from populations assumed to be broadly comparable. The implications of such heterogeneity are unclear, but do raise concerns regarding the possibility of population stratification inducing error.

Dopamine D4 receptor

The DRD4 gene maps to 11p15.5 at approximately 0.6 Mb. It encodes the D4 subtype of the dopamine receptor, which is highly expressed in the limbic system. The most commonly studied polymorphism in this gene has been the exon 3 VNTR. The polymorphism is an imperfect 48 bp tandem repeat. Repeat numbers of 2-10 have been

observed. Certain repeat numbers have been shown to influence the pharmacological properties of the receptor (Muglia *et al.* 2002).

Association studies have provided little evidence for DRD4 being involved in the aetiology of BP. Eight studies have failed to find association: (Lim *et al.* 1994); (Perez de Castro *et al.* 1994); (Kawada *et al.* 1995b); (Di Bella *et al.* 1996); (Oruc *et al.* 1997); (Li *et al.* 1999); (Bocchetta *et al.* 1999); (Serretti *et al.* 2001); (Leszczynska-Rodziewicz *et al.* 2005). However, sample sizes were mostly small.

Only one positive association has been reported with BP (Muglia *et al.* 2002). Muglia *et al* tested for association with the exon 3 VNTR in a sample of 145 trios from Canada. Transmission distortion was observed for two alleles. The two-repeat allele was under-transmitted to probands (p=0.0114), while the four-repeat allele was over-transmitted (p=0.0278). Separate analysis of maternal and paternal transmissions revealed that the transmission was restricted to the maternal meiosis. The authors note that the DRD4 gene is a region where genomic imprinting is known to operate.

One other study has reported association with alleles at the exon 3 VNTR (Manki *et al.* 1996). However, the case group included UPD. When the BP cases were analysed separately no association was found.

While Muglia *et al*'s result should not be ignored, the weight of evidence does not favour a role for the exon three polymorphism in BP. I was unable to find any meta-analyses of association results for this polymorphism. However, one cannot exclude DRD4 in the aetiology of BP as the study of this marker alone will not capture all the haplotypic variability at the locus.

Tryptophan hydroxylase

The tryptophan hydroxylase 1 gene (TPH1) maps to 11p15.1 at approximately 18 Mb. TPH1 catalyses the conversion of tryptophan to 5-hydroxytryptophan, which is the rate-

limiting step in the synthesis of 5HT. The most commonly studied marker has been the intron 7 A218C RFLP.

Belliever *et al* (Bellivier *et al.* 1998a) initially reported allelic association between the A allele of the A218C polymorphism and BP ($p<0.001$) in a small case-control sample from France. However, attempts to replicate this finding have not been successful: (Furlong *et al.* 1998a); (Kirov *et al.* 1999b); (McQuillin *et al.* 1999b); (Vincent *et al.* 1999); (Kunugi *et al.* 1999); (Rietschel *et al.* 2000); (Souery *et al.* 2001); (Rotondo *et al.* 2002). A more systematic association analysis of the TPH1 locus was attempted (Lai *et al.* 2005). Ten SNPs were tested in a relatively small Taiwanese case-control sample. No allelic or genotypic association was found for individual markers. Haplotype analysis revealed one 10-marker haplotype which occurred in cases but not controls. However, the significance of this finding did not survive correction for multiple testing.

It is clear that there is scant evidence to suggest a role for TPH1 in BP. However, despite Lai *et al*'s more extensive coverage of the TPH1 locus, association cannot be confidently excluded. More recently attention has shifted to the study of the other TPH isoform, TPH2, which maps to chromosome 12. TPH2 is solely brain expressed and some promising association results have emerged, e.g. (Van Den Bogaert *et al.* 2006a).

Brain-derived neurotrophic factor

The BDNF gene maps to 11p14.1 at approximately 27.6 Mb. BDNF promotes survival, growth and differentiation of neurons. A number of lines of evidence suggest it is an excellent functional candidate for BP. Antidepressants and mood stabilizers have been shown to increase BDNF transcription. In animal models of depression, decreased BDNF expression and hippocampal atrophy have been demonstrated. In drug naïve patients with MDD, serum BDNF has been shown to be lowered (Hashimoto *et al.* 2004).

The most intensively studied marker at the BDNF locus is the Val66Met polymorphism. This amino acid is present in the pro-peptide but cleaved in the formation of the mature

peptide. However, variation at this position affects BDNF secretion in vitro and memory in human subjects (Egan *et al.* 2003).

Association of BDNF with BP was first described by Sklar *et al* (Sklar *et al.* 2002). This association was identified in the course of a screen of 76 candiadate genes using 90 SNPs. The test sample consisted of 136 American trios. Only the Val66Met polymorphism was tested at the BDNF locus. Association with the valine (G) allele was found at the 0.042 level. Replication of the association was attempted in two independent trio samples, NIMH (n=189) and UK (n=145). A trend for over-transmission was observed. Resequencing of the gene in cases and controls identified 44 SNPs. Eight of these were successfully genotyped. Three additional SNPs showed transmission distortion: a40 (p=0.0047), a20 (p=0.0021) and a13 (p=0.0124). a40 is 5' while a20 and a13 are both 3' of the coding sequence. Haplotype analysis demonstrated significant global transmission distortion of the six eight-marker haplotypes with frequencies above 2% (p=0.034). With regard to individual haplotypes there was significant under-transmission of haplotype 3 (p=0.03013) and over-transmission of haplotype 6 (p=0.00096).

In the same year, association was reported in an independent family based sample (Neves-Pereira *et al.* 2002). A 5' dinucleotide repeat, -1040bp (GT)n, and the Val66Met were tested in 283 Canadian trios. The Val66Met valine allele was again over-transmitted, p=0.00064. Preferential transmission of allele 3 (A3) of the -1040bp (GT)n polymorphism was also observed, p=0.042. Haplotype analysis supported this association with over-transmission of the A3-G haplotype, p=0.000394.

This sample was extended and further markers tested (Muller *et al.* 2006). A small number of families were added and siblings included, giving a total of 350 BP cases and 693 non-affected relatives. The association with the valine allele at Val66Met and allele 3 of the -1040bp (GT)n repeat remained, p=0.001 and p=0.008 respectively. Four additional SNPs were typed of which two showed significant association: hCV11592756 and rs2049045. Muller *et al* proceeded to demonstrate that the associations were only observed in patients with the rapid cycling form of the disorder.

This finding of association with rapid cycling BP was consistent with the results from a very large UK case-control sample (Green *et al.* 2006). The Val66Met polymorphism was typed in 962 cases and 2,100 controls. No association was found in the primary analysis. Six subtypes of BP were defined according to phenotypic variables, e.g rapid cycling, on the basis that these variables may equate to distinct genetic subtypes. Testing of the rapid cycling subtype n=131 revealed association with the valine allele at the 0.004 level (uncorrected). Green *et al* therefore re-analysed the data for the Cardiff trio sample used for replication in the original Sklar *et al* paper. By restricting the analysis to only those trios with rapid cycling probands, significant over-transmission of the valine allele was found, p<0.03 (one tailed).

Replication of the Val66Met association has now been reported in both family based and case-control samples. Excess transmission of the valine allele, p=0.001, was found in a sample of 212 nuclear families from the historically isolated Antioquina region of North West Colombia (Kremeyer *et al.* 2006). The -1040bp (GT)n repeat, which was denoted BDNF-CA in this study, was also tested. Association at this marker was not replicated, but transmission of the same two-marker haplotype as Neves-Pereira *et al* was observed, (p=0.025). Of note, Kremeyer *et al* found the microsatellite alleles to be 57 bp larger than reported in previous studies. However, they were confident that the alleles corresponded, e.g. Kremeyer *et al*'s 227 bp allele=Neves-Pereira *et al*'s 170 bp allele.

A case-control design was employed to test association at the Val66Met polymorphism (Lohoff *et al.* 2005). 621 unrelated BPI cases of European descent were identified from the NIMHGI collection. By definition, all had a family history of BP. The control group consisted of 998 individuals. Association with the valine allele was found at the 0.028 level.

Schumacher *et al* (Schumacher *et al.* 2005a) also used a case-control design to test for association with BP, MDD and schizophrenia. The Val66Met, -1040bp (GT)n repeat and additional SNP, rs988748, were genotyped. The BP sample numbered 281 and the control

sample, 1,097. Although no single marker association was found for BP, a significant difference in the global frequencies of the three marker haplotypes was observed between BP cases and controls, p=0.0057. However, the most robust findings of association with BDNF were for MDD.

Association results have suggested that variation at the BDNF locus confers susceptibility specifically for BP of early onset. Children and adolescents, mean age 10.7, with a "bipolar disorder phenotype" were studied (Geller *et al.* 2004). Case attribution included the requirement that euphoria or grandiosity had been present. One could question how exactly this "bipolar disorder phenotype" relates to the traditional adult BP construct. Certainly, the mean duration of baseline manic episodes of 3.2 years seems very long considering the mean age of the sample group. However, longitudinal validation of the phenotype was claimed. The Val66Met polymorphism was tested in 53 American trios. Over-transmission of the valine allele was found, p=0.014.

Some support has been provided for this association between BDNF and early onset BP (Strauss *et al.* 2004). However, there were important differences between the studies. Strauss employed a much broader phenotypic definition – childhood onset mood disorders (COMD). In fact, only 23 of the 99 subjects had a diagnosis of BP. Furthermore, subjects were recruited as adults – some had been diagnosed in childhood while the others were diagnosed retrospectively. A case-control design was used. The cases had different ethnic backgrounds. To reduce the potential for population stratification the controls were ethnically matched and genomic control was performed. The dinucleotide repeat -1040bp (GT)n and Val66Met polymorphism were tested. A significant difference in the distribution of -1040bp (GT)n alleles was observed, p=0.0032. The 168 bp allele was overrepresented in the cases. The Val66Met polymorphism did not individually show association but haplotypic analysis supported the implication of the BDNF locus.

Subsequently, Strauss *et al* was able to demonstrate the association of COMD with BDNF in an independent sample of 258 Hungarian trios. However, the authors comment

in the conclusion that they "were not able to test for association with bipolar disorder". This was due to the very small number of subjects with a diagnosis of BP in the sample.

It is important to recognize, however, that there have been a number of negative findings: (Hong *et al.* 2003); (Oswald *et al.* 2004); (Kunugi *et al.* 2004); (Skibinska *et al.* 2004); (Nakata *et al.* 2003). Furthermore, meta-analysis of the Val66Met polymorphism published since the commencement of this work failed to find association (Kanazawa *et al.* 2007). The numbers involved in the meta-analysis were large – 3,143 cases and 6,347 controls. However, of the 11p candidate loci BDNF would still seem worthy of further study given the association in multiple independent samples using different study designs and populations.

5.6 Loci implicated by meta-analysis of the linkage data

There is a natural desire to follow up ones own linkage results in the hope that they will lead to the identification of novel susceptibility loci. However, independent replication of findings is crucial to the advancement of the field. Therefore, the UCL laboratory has tried to contribute to replication studies. It would be logistically impossible to follow up all the positive findings reported for BP. Hence, the laboratory has concentrated on attempting to replicate association findings in regions where there is strong *a priori* evidence for a susceptibility locus from the linkage data. At the time my work commenced, there had been only one meta-analysis of the linkage data for BP which found 13q and 22q to be the regions with the best evidence for harbouring a BP susceptibility locus.

5.6.1 Evidence for 22q bipolar locus

Both cytogenetic abnormalities and linkage studies have provided evidence that an aetiological gene maps to 22q. While I will essentially confine the discussion to BP, it is

important to state that these lines of evidence have also implicated a 22q schizophrenia susceptibility locus.

Cytogenetic evidence

Velo-cardio facial syndrome, also known as 22q11.2 deletion syndrome, is an autosomal domimantly inherited condition which is characterised by cardiac malformations, cleft palate and learning disabilities. Investigators have variously reported an association with schizophrenia or BP.

A very high rate of early onset BP was found amongst VCFS patients (Papolos *et al.* 1996). Twenty-five unselected patients diagnosed with VCFS were recruited and underwent diagnostic psychiatric interviews. Ages ranged from 5 to 34. 64% were diagnosed with BP and all had onset in late childhood or early adolescence.

The characteristics of the deletions seen in VCFS with psychiatric disorders were investigated (Carlson *et al.* 1997). Loss of heterozygosity analysis revealed that a 1.5 Mb deletion bounded by the markers D22S427 (~17.0 Mb) – D22S264 (~19.1 Mb) was common to all the patients in which there was evidence of a deletion.

Therefore, it is hypothesised that both the physical and psychiatric manifestations of VCFS are due to haplo-insufficiency of gene(s) that map to this minimally deleted interval.

Linkage

Many of the studies referred to below have been discussed in previous sections. I will therefore try not to repeat descriptions of these studies. I have indicated such studies with an asterisk.

The first report of linkage compatible with a susceptibility locus in the VCFS deletion region was that of Lachmann *et al* (Lachman *et al.* 1997). 22q microsatellites were typed in 17 BP families. 13 of the families were derived from the University of California, San

Diego/University of British Columbia (UCSD/UBC), 2 from New York and the other was a branch of the Old Order Amish pedigree 110. The strongest evidence for linkage was obtained when the UCSD/UBC families were analysed separately. Under a dominant model a maximum LOD of 2.51 was found at the 22q11 marker D22S303 (at ~21.6 Mb).

Subsequently, a genomewide survey of 20 North American familes was reported which included the 13 UCSD/UBC families above (Kelsoe *et al.* 2001). Significant linkage, maximum LOD 3.8, to the slightly more telomeric marker D22S278 (at ~34.7 Mb) was found. However, the authors noted a broad linkage region, LODs >1.0, spanning 32 cM with a possible second linkage peak at D22S419 (~24.3 Mb). A maximum LOD of 2.19 was found at this marker under a recessive model.

Linkage to the 22q11-12 region is supported by results from the NIMH (both NIMHGI and CNG) pedigrees. Analysis of the first wave of 97 NIHMGI families revealed increased allele sharing at the marker D22S533, at ~24.2 Mb (Edenberg *et al.* 1997). Under a broad affection model a maximum LOD of 2.46 was obtained. Using the same chromosome 22 markers and analysis employed in the genome survey, 57 of these families were examined (Kelsoe *et al.* 2001). The results were consistent with those reported by Edenberg *et al*, with the peak LOD 2.72 occurring at D22S419 which maps very closely to D22S533.

The CNG families are entirely separate from the NIMHGI pedigrees but do include the right extension of the Old Order Amish pedigree 110. Under a broad affection status model, excess allele sharing was observed in a region delineated by D22S689 and D22S685, LOD 2.1 (Detera-Wadleigh *et al.* 1999).

Implication of the 22q region by these individual studies is strengthened by meta-analysis of the linkage data (Badner and Gershon 2002a). Signifcant linkage was found at 36 cM on 22q, $p<1 \times 10^{-5}$. This represented the region with the second strongest evidence for linkage after 13q. Of note, significant evidence for linkage to schizophrenia was also found on 22q (at 32 cM).

Since the publication of the meta-analysis a number of studies have supported 22q linkage. Linkage has been reported in two independent large multiply affected pedigrees. Two-point parametric analysis of the entire Costa Rican kindred cr201 did not yield evidence for linkage to 22q (Service *et al.* 2006)*. However, division of the pedigree to enable multipoint analysis revealed a linkage peak, p=0.00732, at approximately 40 cM on 22q in one section of the family.

A genomewide scan on a kindred from Eastern Cuba with 28 affecetd members was performed (Marcheco-Teruel *et al.* 2006)*. Nonparametric analysis yielded suggestive linkage in both the entire family and one individual branch under a broad phenotypic model. The peak NPL score was observed at 46.4 cM for the complete family, p=0.015 and 51.56 cM, p=0.005, in section 1a of the family.

Linkage analysis of potential subtypes of BP has produced some promising results on 22q, particularly with regard to a psychotic subtype. Suggestive linkage, p=0.005 at D22S277 (34.6 Mb), was obtained when the linkage analysis was restricted to families with at least 3 members with psychotic mood disorder (Potash *et al.* 2003)*. Linkage analysis of pedigrees containing at least one member with schizoaffective disorder (BP type) gave a maximum LOD of 1.96 at D22S420, at ~16.2 Mb (Hamshere *et al.* 2005)*.

Turecki *et al* (Turecki *et al.* 2001) postulated that a lithium response subtype of BP exists and attempted mapping by linkage analysis. A cohort of patients with BP attending an affective disorder clinic in Canada was followed prospectively. Families were ascertained through patients that showed an excellent response to lithium. In total, 31 families were included in the linkage analysis. Again, suggestive linkage was found at the marker D22S420, LOD 1.91 under a recessive model in which only lithium responders were considered affected.

An attempt was made to identify chromosomal segments shared by distantly related BP subjects in the isolated population of the Faroe Islands (Jorgensen *et al.* 2002). This

follows the approach of Degn *et al* in mapping of the 12q23-24 region, referred to in section 5.2. Thirty-five microsatellite markers across 22q were typed in 14 BP patients, 10 schizophrenia patients and 44 unrelated controls. Where possible, parents of both patients and controls were genotyped so as to reconstruct haplotypes. The most significant result was for the telomeric two-marker haplotype, D22S1161-D22S922. Haplotype frequencies were significantly different at the 0.0081 level.

Overall, the linkage of BP to 22q is convincing. Given that 22q has also been implicated in schizophrenia, the recent findings of Potash *et al* and Hamshere *et al* are consistent with the hypothesis of a common susceptibility locus. However, the results could alternatively be explained by the presence of separate susceptibly loci in this region.

Association studies of catechol-o-methyltransferase

Association studies have implicated multiple BP susceptibility loci along chromosome 22 at positions which would be consistent with the linkage evidence. However, within the VCFS deleted region only catechol-o-methyltransferase (COMT) has shown association. Therefore, I will limit the discussion to this locus.

COMT maps to 22q11.21 at approximately 18.3 Mb. COMT is considered an excellent functional candidate due to its role in catecholamine metabolism. The first step of catecholamine neurotransmitter degradation proceeds either by oxidation or methylation. MAO catalyses the oxidation while COMT catalyses the methylation. The relative importance of theses pathways varies between brain regions but COMT degradation is pronounced in the frontal cortex.

COMT exists in two isoforms – a membrane bound form, MB-COMT and a shorter soluble form, S-COMT. It has been demonstrated that the expression of each isoform is under the control of separate promotors, P2 and P1 respectively (Tenhunen *et al.* 1994). While overall expression of S-COMT predominates, the MB-COMT is the more common isoform in the CNS.

The most extensively studied COMT polymorphism has been the exon 4 Val158Met substitution. I will refer to it as rs4680 from now on. The presence of the methionine residue results in heat lability of the protein and a dramatic decrease in enzymatic activity (Lachman *et al.* 1996b). It is of relevance to genetic association studies that allele frequencies at this polymorphism have been shown to vary between populations (McLeod *et al.* 1998).

A gender difference in COMT activity has also been demonstrated, with lower activity in females (Floderus *et al.* 1981). It appears that this effect is mediated by oestrogen response elements and C/EBP binding sites within the promotor regions. A region between -1405 and -1083, with respect to the MB-COMT transcription start site, was found to be critical for the oestrogen-dependent repression of the P2 promotor (Xie *et al.* 1999). A *Hind*III RSP was identified within this critical region, at -1217 bp (DeMille *et al.* 2002).

Association of the rs4680 methionine allele with BP was initially found in a sample of VCFS patients with psychiatric disorders (Lachman *et al.* 1996a). The rs4680 polymorphism was typed in 23 patients and a screened control group, n=89. Association of the methionine allele with rapid cycling BP was found, although no correction was made for multiple testing of diagnostic categories. Since this original report, multiple studies have tested the association between BP and COMT.

Craddock *et al* (Craddock *et al.* 2001) performed a meta-analysis for the rs4680 polymorphism which included the following studies: (1997); (Gutierrez *et al.* 1997); (Lachman *et al.* 1997); (Kunugi *et al.* 1997); (Mynett-Johnson *et al.* 1998); (Li *et al.* 1997); (Ohara *et al.* 1998). These studies contributed a total of 910 BP cases and 1,069 controls to the analysis. Association with the methionine allele was confirmed, albeit with a very modest pooled OR of 1.18 (CI 1.02–1.35).

Three subsequent studies have supported association with the COMT locus. Analysis of a population based sample from Italy, cases=111, controls=127, gave association with the

methionine allele at the p=0.004 level (Rotondo *et al.* 2002). Subdivision of the case sample by comorbid panic disorder revealed that the association was only seen in the group without comorbid panic disorder. This finding is somewhat unexpected since the met allele has been implicated in the aetiology of panic disorder itself, e.g. (Woo *et al.* 2002).

Association with three COMT SNPs - rs737865, rs4680 (referred to as rs165688 in the article), and rs165599 - was tested in an Ashkenazi Jewish sample (Shifman *et al.* 2004). Allele frequencies of 217 BP cases were compared with over 4,000 controls. No association was found with rs4680 alleles. However, both of the other SNPs showed association, p=0.049 and p=0.0015. Haplotypic analysis produced two intriguing results. Firstly, haplotypic association was restricted to BP women. Secondly, the G-G-G haplotype was found to be associated. The G allele at rs4680 corresponds to the high activity valine residue, which is at odds with the hypothesis that low COMT activity predisposes to BP.

Further evidence of allele switching comes from the study of Funke *et al* (Funke *et al.* 2005), although the results are rather difficult to interpret. In addition to the 3 SNPs tested by Shifman *et al*, a SNP located in the P2 promotor region, rs2097603 was genotyped. Variation at this SNP had previously been shown to have a small effect on enzyme activity. Individual marker association was found at rs2097603, rs4680 and rs165599 in a US Caucasian sample of 394 cases and 467 controls. Again the valine allele of rs4680 was associated. However, the case group included patients with diagnoses of BP, schizophrenia, major depression, depressive disorder, NOS and psychotic disorder NOS. When the BP cases were analysed separately n=82 no single marker associations remained.

Most recently association of BPI with SNP rs165599 was supported, p=0.02 in a small case control study from the US (Burdick *et al.* 2007). The risk allele (G) was subsequently shown to predict poorer performance on tests of verbal memory in both cases and controls. Association with rs4680 was not, however, replicated.

Some additional evidence that the methionine allele at rs4680 increases susceptibly to the rapid cycling (RC) form of BP has emerged. However, this association has only been observed when BP cases with RC are compared to BP cases without RC, rather than controls (Papolos *et al.* 1998); (Kirov *et al.* 1998). Furthermore, the association was restricted to a subgroup of RC, ultra-ultra RC, in the Papolos *et al* study.

It needs to be recorded that there have also been multiple negative studies since Craddock *et al*'s meta-analysis: (Kirov *et al.* 1999a); (Geller and Cook 2000); (Serretti *et al.* 2003); (Massat *et al.* 2005); (Van Den Bogaert *et al.* 2006b); (Prata *et al.* 2006; Serretti *et al.* 2006).

Revised meta-analysis is awaited, but the methionine allele at rs4680 does seem to confer a small increase in risk of BP. Although this is a functional polymorphism it is not necessarily responsible for the increased risk, as the methionine allele may be in LD with the true aetiological variant. It has been suggested that variation at the COMT locus alters risk of psychiatric illness in some generic way as, in addition to panic disorder, COMT has also been implicated in schizophrenia and alcoholism.

5.6.2 Evidence for bipolar loci on 13q

Linkage

Several studies have found linkage between BP and markers on the long arm of chromosome 13. The distribution of these linked markers would suggest at least two distinct regions of linkage: 13q11-13q21 and 13q22-13q32.

The more telomeric 13q22-13q32 region is robustly supported by the linkage data. As discussed in chapter 3, the first meta-analysis of linkage scans in BP found the strongest evidence for a susceptibility locus on 13q at 79 cM (Badner and Gershon 2002a). Subsequently, significant linkage with markers mapping to this region has been reported (Shaw *et al.* 2003). A parametric multipoint analysis using a BP-only, autosomal

dominant, medium penetrance model yielded a peak LOD of 3.4, approximately 7 cM centomeric to D13S154. The relative position of this marker is indicated on figure 5.6 below. In addition, there have been a number of independent reports which support linkage to the 13q22-13q32 region: (Detera-Wadleigh *et al.* 1999); (Liu *et al.* 2003); (Potash *et al.* 2003); (Cheng *et al.* 2006).

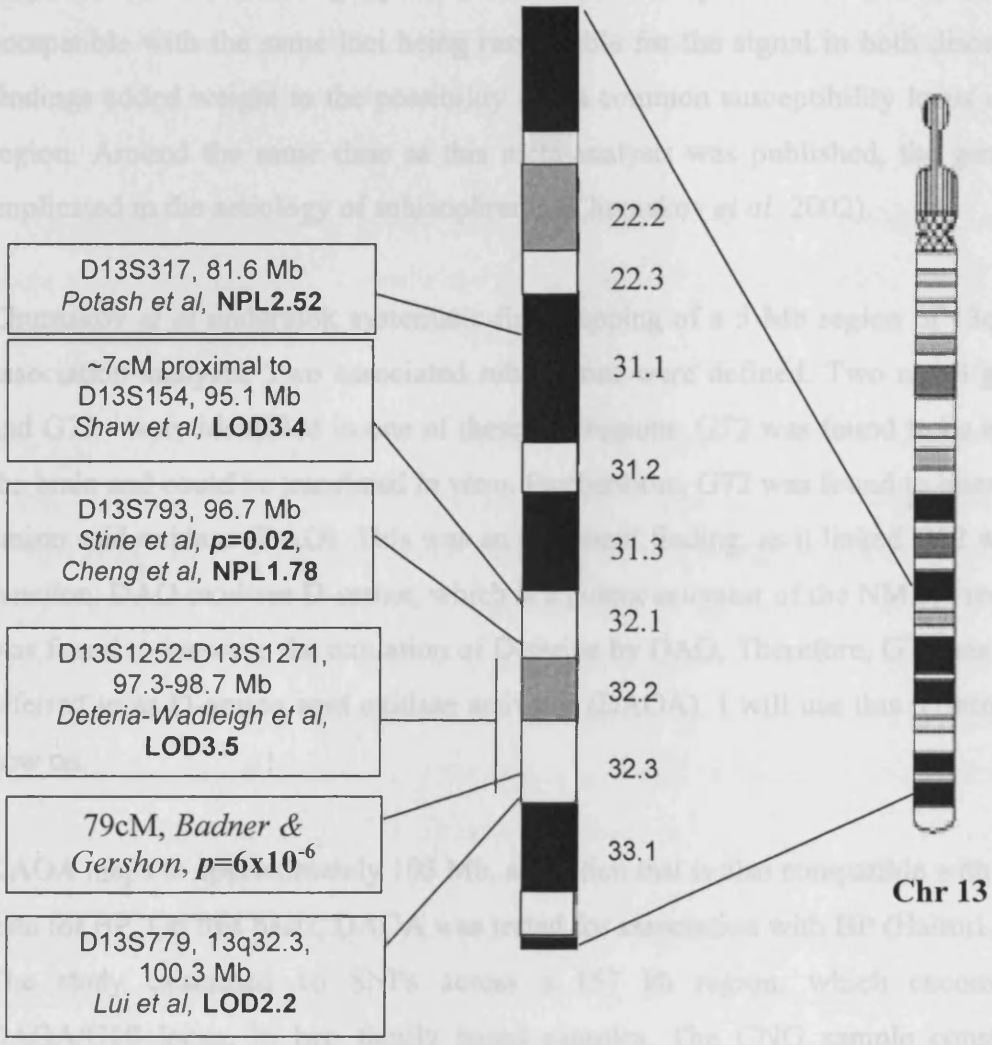
Under multipoint nonparametric analysis suggestive linkage at the D13S1252-D13S1271 interval, p=0.000028, with a broad phenotypic definition was obtained (Detera-Wadleigh *et al.* 1999). Subsequently, additional microsatellite markers were genotyped in this sample and a slight telomeric shift in the interval showing strongest evidence for linkage (D13S779-D13S225) was found (Liu *et al.* 2001a).

Multipoint ASP analysis yielded a suggestive linkage signal near D13S779, LOD score of 2.2, in a large sample of American and Israeli pedigrees (Liu *et al.* 2003). When this dataset was reclassified to include only individuals exhibiting psychotic symptoms as affected, suggestive evidence for linkage persisted at this marker, multipoint, LOD=1.96 under a dominant model (Park *et al.* 2004). Attempts have also been made to identify susceptibility loci for 'psychotic' BP by sub-dividing BP pedigrees according to the number of members with psychotic symptoms (Potash *et al.* 2003). In families with three or more members exhibiting psychotic mood disorder a nonparametric LOD score of 2.52 at D13S317 was obtained using ASPEX SIBIBD.

More recently, the effect of different phenotypic definitions on the linkage results has been further explored (Cheng *et al.* 2006). No evidence was found to support linkage of 'psychotic' BP to this region, but a nonparametric LOD of 1.78 at D13S793 was obtained when a standard narrow affection status model was applied. Although this LOD did not reach the Lander and Krugylak threshold for suggestive linkage, the authors argued that it should be considered as such, as it exceeded the threshold derived from simulation analysis of their dataset. Since approximately 16% of people with BP develop panic disorder at some point in their lives (Henry *et al.* 2003), it is interesting to note that

significant linkage of panic disorder to this marker (LOD 3.57) has been published (Hamilton *et al.* 2003).

Figure 5.6: Summary of linakge results on 13q31-33



Three independent genomewide linkage scans have provided suggestive linkage evidence for the more centromeric 13q11-13q21 region: (Badenhop *et al.* 2002); (Cichon *et al.* 2001); (McInnis *et al.* 2003b). These studies defined a broad region - spanning approximately 30 Mb - but one that could still be compatible with a single locus.

Association findings for DAOA/G30

Multiple studies have also demonstrated linkage of schizophrenia to the 13q22-34 region. Indeed, Badner and Gershon's (Badner and Gershon 2002a) meta-analysis of schizophrenia linkage scans yielded remarkably similar results to that of their BP meta-analysis. Again, the strongest evidence for linkage of a susceptibility locus was found on 13q. The location was slightly more distal for schizophrenia at 85 cM, but is entirely compatible with the same loci being responsible for the signal in both disorders. These findings added weight to the possibility that a common susceptibility locus exists in the region. Around the same time as this meta-analysis was published, the gene G72 was implicated in the aetiology of schizophrenia (Chumakov *et al.* 2002).

Chumakov *et al* undertook systematic fine mapping of a 5 Mb region of 13q34 by SNP association analysis. Two associated sub-regions were defined. Two novel genes - G30 and G72 - were identified in one of these sub-regions. G72 was found to be expressed in the brain and could be translated *in vitro*. Furthermore, G72 was found to interact with D-amino acid oxidase (DAO). This was an important finding, as it linked G72 with NMDA function. DAO oxidises D-serine, which is a potent activator of the NMDA receptor. G72 was found to increase the oxidation of D-serine by DAO. Therefore, G72 has come to be referred to as D-amino acid oxidase activator (DAOA). I will use this terminology from now on.

DAOA maps to approximately 105 Mb, a position that is also compatible with the linkage data for BP. On this basis, DAOA was tested for association with BP (Hattori *et al.* 2003). The study examined 16 SNPs across a 157 kb region, which encompassed the DAOA/G30 locus, in two family based samples. The CNG sample consisted of 83 individuals from pedigrees in which linkage to 13q32 had previously been found. The NIMH samples comprised of 474 individuals. TDT and haplotype analysis was performed. It was stated that both tests controlled for the presence of linkage which could potentially confound association.

Five SNPs showed individual association with BP in the CNG sample: rs1815686 (p=0.041), rs1935058 (p=0.00077), rs1341402 (p=0.0075), rs1935062 (p=0.0078) and rs778294 (p=0.018). No individual association was found in the NIMH sample, although rs1935058 approached significance at p=0.055. Haplotype analysis was performed for the seven SNPs in LD with one another. Significant perturbations from the expected transmission frequencies were found for the two most common haplotypes in the CNG sample. Furthermore, significant results were obtained for the global test of transmission disequilibrium in both datasets.

Subsequently, a number of groups have attempted to replicate the association of DAOA with schizophrenia, BP, a psychosis phenotype and panic disorder. I will restrict my consideration to those that have studied BP.

Six SNPs - rs2027446, rs1935058, rs2391191, rs1935062, rs947267 and rs954581 - were tested in a population based sample of 139 cases and 113 ethnically matched controls (Chen *et al.* 2004). Only two of the five associated SNPs in the Hattori *et al* study were tested. Attempts were made to genotype two more of the Hattori *et al* associated SNPs - rs1341402 and rs778294 - but the assays failed quality control. Significant association was found with two adjacent SNPs: rs1935062 (p=0.007) and rs947267 (p=0.04). However, here the A allele of rs1935062was found to be associated, whereas it was under-transmitted to cases in the Hattori *et al* study.

Because of the assay failure, Chen *et al* were unable to reconstruct the Hattori *et al* haplotypes. However, analysis of five SNP haplotypes provided further evidence for association, with significant results for both individual haplotypes and the global distribution of haplotypes (p=0.004). The associated haplotypes did not correspond between the two studies. Of the three individually associated haplotypes in the study of Chen *et al*, two were not reported in that of Hattori *et al.* It is possible that these rare haplotypes did not exist in the Hattori *et al* samples, or that they were an artifact of haplotype estimation.

Seven SNPs in schizophrenia and BP case-control samples from Germany were genotyped (Schumacher *et al.* 2004). 300 BP cases and 300 controls were included. As can be seen from table 5.1 below, not all of the SNPs which Hattori *et al* found to be associated were typed and it was not possible to reconstruct the original seven SNP haplotypes. One SNP, rs3918342 (M23), was associated at the 0.013 level. A four-marker haplotype was tested. No difference in the overall distribution of haplotypes between cases and controls was observed. However, one individual haplotype was found to be significantly less frequent in the cases (p=0.032).

On the premise that the DAOA/G30 locus confers susceptibility to both disorders, the hypothesis that a psychotic subtype of BP is accounting for the association observed in the above dataset was examined (Schulze *et al.* 2005). Logistic regression was performed to identify variables from the 21 OPCRIT symptoms of psychosis, which influenced M23 genotype. Lifetime history of persecutory delusions was found to be the only significant variable. A case group of BP with persecutory delusion was thus defined (n=90) and the association data re-analysed for the four markers that made up the associated haplotype in the Schumacher *et al* study. Two of these markers showed association - rs3918342 (M23) and rs1421292 (M24), with p-values of <0.0004 and <0.007 respectively. No association was found at any of the markers for BP without persecutory delusions. Haplotype analysis supported the association with BP with persecutory delusion, yielding significant results for global and individual haplotypes.

Replication of this finding was attempted in a Polish sample of 294 BP cases and 311 controls. Only the positively associated rs3918342 (M23) and rs1421292 (M24) markers were genotyped in this sample. Individually neither marker was found to be significantly associated, but a modest association was observed with one haplotype (p<0.04). This haplotype was consistent with one of the associated four SNP haplotypes in the German sample.

As described in relation to DAO, a screen of 64 candidate genes in Ashkenazi Jewish trios was conducted (Fallin *et al.* 2005). Fourteen SNPs across the DAOA/G30 locus

were tested and the suggestive association with BP reported. However, I was unable to glean from the report any details of the association.

A meta-analysis of the association findings in schizophrenia and BP was performed (Detera-Wadleigh and McMahon 2006). For BP it included the studies of Hattori *et al*, Chen *et al* and Schumacher *et al*. Three markers - rs1935058, rs1935062 and rs3918342 (M23) - were found to be significantly associated. However, the results were combined "without regard to the allele identity".

An attempt was made to systematically test for association of the DAOA/G30 locus with BP and schizophrenia (Williams *et al.* 2006). Initially, mutation screening of 14 unrelated schizophrenic individuals was performed. Exonic 5' and 3' untranslated regions and the flanking intronic regions were sequenced. Nineteen sequence variants were identified. Thirteen of these were not analysed further because they captured the same information or had minor allele frequencies of <10%. The remaining six SNPs and three intronic SNPs chosen from dbSNP were used to define the LD structure using a sample of 96 unrelated individuals. From this, six tag SNPs were identified. These were supplemented with three SNPs that had previously been associated with schizophrenia for the association testing stage. Case-control samples were used comprising of 706 BP cases, 709 schizophrenics and 1,409 controls. Three SNPs showed association with BP: rs3916965 (M12), DAOA_3'UTR_SNP12 and rs1341402. The results of haplotype analysis were not shown.

Table 5.1: Summary of published association studies for DAOA in BP

	Hattori *et al*	Chen *et al*	Schumacher *et al*	Williams *et al*
rs3916965 (M12)	-	-	0.225	**0.047**
rs1815686	**0.041**	-	-	-
rs2027446	-	NS	-	-
rs1935058	**0.00077**	NS	-	0.30
rs1341402	**0.0075**	-	0.437	**0.03**
rs2391191 (M15)	0.08	NS	0.222	0.09
rs1935062	**0.0078 A (controls)**	**0.0072 A (cases)**	0.715	-
rs947267	0.17	**0.0403 A**	-	-
DAOA_3'UTR_SNP12				
rs778294 (M19)	**0.018**	-	0.337	0.35
rs954581	0.62	NS	-	0.88
rs778293 (M22)	-	-	-	0.36
rs3918342 (M23)	-	-	**0.013**	-
rs1421292 (M24)	-	-	0.064	0.59

Abbreviations: NS=not statistically significant, A=associated allele. Results that are statistically significant are indicated in red and SNPs used as part of Hattori *et al's* 7-SNP haplotype are shown in blue.

It is difficult to know how to interpret the DAOA/G30 results. The difficulty emanates from two sources: firstly, the variation in markers tested between individual studies and secondly, the fact that different alleles at the same marker were found to be associated.

No study tested all the markers that were found to be associated by Hattori *et al* (Hattori *et al.* 2003). Consequently, it is particularly difficult to interpret the haplotype findings for subsequent studies. The example of DAOA association studies raises the more general point of what constitutes a formal replication. It could be argued that if the LD structure of the region is first delineated then it may not be necessary to type the exact same SNPs, as tagging SNPs will be able to capture the information. This is broadly the approach taken by Williams *et al* (Williams *et al.* 2006).

Individual association was found with the marker rs1935062 by Hattori *et al* and then by Chen *et al.* The p-values were similar and impressive, but unfortunately for different alleles. However, such findings are not necessarily incompatible with DAOA being a susceptibility gene for BP. One explanation is that there were a number of aetiological

mutational events which occurred on different haplotypic backgrounds. Hence, patterns of association could differ between populations. This argument has been used to explain the findings for Dysbidin in schizophrenia (Schwab *et al.* 2003).

While I have avoided detailed discussion of the findings in schizophrenia, I think it is pertinent to note that there have been similar inconsistencies in the associated alleles but overall the evidence for association is much stronger (Detera-Wadleigh and McMahon 2006). Detera-Wadleigh and McMahon also performed a meta-analysis where schizophrenia and BP studies were combined, which in fact yielded the strongest evidence for association. Susequently Williams *et al* performed analysis across traditional diagnostic boundaries (Williams *et al.* 2006). No association was found when cases were defined by the presence of psychosis, but the evidence for association was increased when the definition was according to the presence of major mood disorder.

Although association studies on the telomeric region of 13q have focused on the DAOA locus, it should be remembered that the linkage results implicate a wide region. Furthermore, the possibility exists that there is more than one susceptibility locus in this region. The parametric multipoint linkage analysis of Shaw *et al* (Shaw *et al.* 2003) showed two peaks. Of these, the distal peak is more compatible with the DAOA locus, while it was the proximal peak that showed the greater evidence for linkage.

To follow up the more proximal findings at D13S280, three genes - KDELC1, BIVM, ERCC5 - and the hypothetical gene LOC93081 were screened (Ferraren *et al.* 2005). No convincing evidence for association was found, but nonetheless this study serves as a reminder that the search may need to extend beyond the DAOA locus.

Chapter 6: Aims

The overall intention of the work was to identify novel, and further investigate previously implicated, BP susceptibility loci by LD mapping. Within this framework the aims were three-fold:

1) To follow up the preliminary association findings on the 12q24 and 21q22 regions identified through linkage studies at UCL in a larger case-control sample.

2) To follow up the UCL linkage findings on 1q42.

3) To attempt to replicate association findings at loci within regions either implicated by linkage studies at UCL or by meta-analysis of the linkage data (Badner and Gershon 2002a). Specifically to:
    a) replicate the association of BDNF with BP reported by Sklar *et al*/Neves-Pereira *et al* (Sklar *et al.* 2002)/(Neves-Pereira *et al.* 2002)
    b) replicate the association of P2RX7 with BP reported by Barden *et al* (Barden *et al.* 2006)
    c) replicate the association of the methionine allele of the COMT functional exon 4 polymorphism, rs4680 with BP
    d) replicate the association of DAOA with BP reported by Hattori *et al* (Hattori *et al.* 2003).

For BDNF, P2RX7, COMT and DAOA additional polymorphisms were tested for association to further investigate the role of these genes in the aetiology of BP.

Chapter 7: Materials and Methods

LD between alleles at DNA markers and BP was tested for using a population based case-control design. I will first describe the case-control sample, genotyping/sequencing techniques and statistical analysis in a generic sense. I will then proceed to discuss the specific details for each locus studied.

7.1 Case-control samples

7.1.1 Power considerations

It is not possible to calculate *a priori* the size of sample needed to detect association with a given allele, since the relative risk conferred by that allele in the population from which our sample is drawn is not known. This is evidently true for novel loci but also applies to loci in which replication of association is being attempted. The estimates of the relative risk conferred by a particular allele from previous studies will only apply to the populations which have been studied. Overall, association studies in BP suggest that considerable variability in risk alleles exists across populations so it would seem dangerous to rely on power calculations based on such assumptions.

It is clear that the larger the sample, the greater the power to detect allelic association. Thus the UCL laboratory has tried to maximise the size of the sample. By necessity, the collection of the UCL sample has run parallel to the association testing. **Therefore, there is considerable variability in the total number of individuals genotyped for different markers, reflecting the point in time at which the genotyping was performed. However, no selection of cases or controls occurred for the association testing. At the point at which a particular marker was genotyped, the full sample that was available was used.** The only exception to this was where genotyping was outsourced to Kbiosciences, as the stock of DNA held by them was, for logistic reasons, updated in batches. At the time I began the work, the sample consisted of approximately 300 cases

and 300 controls. The recruitment of cases preceeded that of controls. Therefore in all but the most recent tests of association the number of cases is greater than controls. The DAOA locus is an exception to this pattern, as approximately 300 cases and 300 controls were initially tested. At a later date further controls were genotyped in relation to an association study in schizophrenia and the control genotypes added into the bipolar analysis. The maximum number of cases and controls available for testing was 613 and 600 respectively.

Power calculation shows that a sample of 600 cases and 600 controls has 89% (2-sided) power to detect LD at the $p<0.05$ significance level when there is an allele frequency difference of 5% in the context of less common alleles - alleles present in 5% of controls and 10% of cases. To detect a 5% difference in a common marker allele which has a frequency of 45% in controls and 50% in cases, the power is only 38% (2-sided).

7.1.2 UCL Sample

U.K. National Health Service (NHS) multicentre and Local Research Ethics Committee approval was obtained. Each volunteer was given an information sheet and description of the study. All volunteers signed an approved consent form.

The cases were recruited from London NHS psychiatric services and from volunteers who were members of the Manic Depression Fellowship, a support organisation for sufferers of BP. The control subjects were recruited from London branches of the National Blood Service, from local NHS family doctor clinics and from university student/staff volunteers.

As discussed in chapter 4, it is crucially important for association studies that the case and control sample are derived from the same population. Ancestry criteria for participation in the study had been set out at the very start of the study. Volunteers had to be of predominately British or Irish ancestry. This was defined as having at least three of four grandparents of such ancestry. Volunteers with a grandparent of non-western

European ancestry were excluded. All volunteers were screened with an ancestry questionnaire. The validity of using this questionnaire had previously been assessed. Twenty-six microsatellite markers were genotyped at loci not thought to be associated with affective disorders in the first 300 BP cases and 300 controls to test for population stratification, as suggested by Pritchard and Rosenberg (Pritchard and Rosenberg 1999). No difference in allele frequencies between the groups was observed (McQuillin *et al.* 2006).

All volunteers were interviewed using the SADS-L (Endicott and Spitzer 1978). This information was supplemented by material from case-notes where appropriate. Diagnoses were always assigned prior to genotyping and were made at the "probable" level of the RDC. A second psychiatrist reviewed the diagnoses. Most of the cases had BPI (88%). All the BP cases were rated with the 90 item OPCRIT checklist (McGuffin *et al.* 1991). Information regarding family history of mental disorder and drug treatment response was also collected. Phenotypic information was collated in a Microsoft Access database.

Only 'supernormal' controls were recruited. Control volunteers with a personal history of any RDC defined mental disorder or a family history of BP, schizophrenia or alcoholism were excluded. It has been argued that it is more efficient to use such controls (Morton and Collins 1998).

7.2 Genotyping

7.2.1 DNA extraction

DNA was extracted from whole frozen venous blood using a cell lysis, proteinase K, phenol/chloroform method (Sambrook *et al.* 1989). The DNA was quantified by fluorimetry with Picogreen according to manufacturer's instructions (Invitrogen).

7.2.2 PCR reactions

Genotyping was performed using DNA amplified by the polymerase chain reaction (PCR). All PCR reactions were made up to a volume of 12 ul. 50 ng of genomic DNA (2 ul of 25 ng/ul) was used in each reaction. The concentration of 0.2 mM dNTPs (0.096 ul of 25 mM dNTPs. Oligonucleotide primer concentrations varied depending on the marker.

PCR conditions were optimised by varying the magnesium concentration (2.0-2.5 mM), the addition of 1 mM betaine and the PCR programme. The PCR reactions for each marker were carried out separately in 96-well PCR plates. Reaction mixes were plated out with a MWG Biotech RoboAmp 4200. PCR was performed in MWG Biotech Primus Thermocyclers (MWG Biotech, DE). Details of the PCR programmes after given in appendix 1.

7.2.3 Microsatellites

An infrared fluorescence-based method of genotyping most of the microsatellite markers was used (Oetting *et al.* 1995).

Generally, established microsatellite markers were genotyped. Where necessary, sequence databases were interrogated for repeat structures. Primers were designed for repeats thus identified using the program Primer 3 (Rozen and Skaletsky 2000).

A 19 bp M13 tail (CACGACGTTGTAAAACGA) was added to the 5' end of the primers *in silico*. The four primers thus created were analysed using NetPrimer, (www.premierbiosoft.com/netprimer), for self-complimentary and likely formation of secondary structures. The primer pair with the highest combined score, i.e. lowest likelihood of developing these structures, was chosen. The oligonucleotides were synthesised by MWG (MWG Biotech, DE). 2-4 pmoles (0.167 uM - 0.333 uM) of each digonuclectide was used in the 12 ul reaction.

In order to increase throughput, a strategy of running pooled PCR products on the polyacrilamide gels was employed. Consequently, primers were selected or re-designed so that, for a given group of markers, there was no overlap of product sizes.

A third universal M13 sequence primer labeled with IRD 700 or 800 was included in the PCR reaction so as to directly label the product. Post-PCR Fuschin (50mgs of Basic Fuchsin in 50mls formamide 0.05%) loading buffer was added to each reaction. Pooling of aliquots from each marker was then carried out.

Labelled PCR products were denatured (three minutes at 96°C) and manually loaded on 6% polyacrylamide gels. Separation and visualisation was achieved by electrophoresis on LiCor DNA 4200L sequencers (Li-Cor, NE, US). Molecular weight markers, microSTEP-20a IRD700/800 (Microzone Limited, UK), were run in multiple lanes in accordance with the genotyping software requirements.

Allele scoring was carried out with either GeneImagIR (Scanalytics, VA, USA) or SAGA-GT (LiCor) genotyping software.

A few of the early markers were genotyped using PCR with a ^{32}P-labelled oligonucleotide primer followed by polyacrylamide gel electrophoresis and autoradiography. One of the oligonucleotide primers was end-labelled with γ^{32}P-ATP using T4 polynucleotide kinase (Promega, UK). Once amplification had been completed the products were denatured in formamide loading buffer and separated on 6% polyacrylamide gels under denaturing conditions. Products were detected by autoradiography.

7.2.4 SNPS

The SNPs were genotyped by one or more of the following methods: KASPar (Kbiosciences, UK), Pyrosequencing (Pyrosequencing, Sweden), Taqman (Applied Biosystems, CA, US), Amplifluor (Serologicals, GA, US) or PCR/restriction digest/

agarose gel electrophoresis. Some of the genotyping was outsourced to Kbiosciences (Hoddeson, UK) who employed the KASPar and Amplifluor methods. 'In house' assays were performed according to manufacturers' instructions. For amplifluor PCR reactions Titanium Taq (Clontech) was used. The fluorescence of Taqman and Ampliflour assay products was end-read on an ABI700 or Roche LightCycler 480.

RFLPs were typed by PCR, restriction enzyme digest and agarose gel electrophoresis. Digests were performed overnight at 37°C. Products were run on 1.5% agarose gels alongside molecular weight markers (Hyperladder, Bioline). DNA detection was achieved by incorporation of ethidium bromide and visualisation under UV light. Digital photographs of the gel were taken and the genotypes scored manually.

7.2.5 Sequencing

Sequencing was carried out using the Li-Cor infrared fluorescent sequencing system. The method involved direct sequencing of PCR products amplified from genomic DNA using M13 tagged PCR primers. Amplification products were purified with Microclean reagent according to the manufacturers' instructions (Microzone, UK). PCR products were then sequenced simultaneously in both directions with the SequiTherm Excel II DNA sequencing kit (Epicentre, WI, US) using two M13 primers fluorescently labelled with IRD700 and IRD800 (MWG Biotech, DE). Sequencing products were analysed on Li-Cor 4200 automated sequencers (Li-Cor, NE, US). The samples were arranged such that the guanosine dideoxy nucleotide terminated tracks from each of the samples were loaded together followed by the adenosine, thymidine and cytidine. This facilitated visual scoring of potential mutations and allowed recognition of polymorphisms in their heterozygote state. The validity of potential mutations was assessed by comparison with the gel image produced from simultaneous sequencing reactions primed from the opposite strand.

7.3 Statistical analysis

7.3.1 Error detection

Genotypes were read by two independent assessors, each blind to both the diagnosis and the allele scoring of the other. Genotypes that could not be agreed upon were repeated.

To assess the general comparability of results between in house and Kbiosciences genotyping 621 individuals were genotyped for the SNP rs6265 by Taqman in house and Amplifluor at Kbiosciences. 98.7% of genotypes correlated.

For markers genotyped by Kbiosciences, 17% of samples within each 96 well plate were duplicated in order to detect error and confirm the reproducibility of genotypes.

The data for each SNP was analysed to confirm the presence of HWE. The SNP data was also tested for significant differences in haplotype frequencies between 96 well plates using the SCANGROUP support program of the GENECOUNTING software (Zhao *et al.* 2002). This allows plates with atypical haplotype frequencies to be identified and scrutinised. If genotyping errors occur on one particular plate, overall allele frequencies and thus allelic association may be minimally affected. However, such errors may produce rare or non-existent haplotypes and create strong haplotypic association.

7.3.2 Testing of allelic association

Allelic association between each SNP marker and BP was tested using the chi-squared test. Allelic association between BP and each microsatellite marker was tested using the CLUMP programme (Sham and Curtis 1995). This performs a Monte Carlo evaluation of significance levels to produce an empirical p-value for the observed chi-squared statistics, since the asymptotic p-value may be inaccurate for markers having large numbers of

alleles. Four different chi-squared statistics are calculated, labelled T1-T4: T1 is the standard Pearson chi-squared statistic; T2 is the statistic obtained by collapsing together rare alleles; the T3 statistic is the maximum obtained by considering each allele separately against all the rest; the T4 statistic is obtained by grouping together all the alleles positively and negatively associated with disease.

Association testing of subsets of the case group was performed only where there was prior evidence that the chromosome region might harbour a susceptibility allele to a specific subtype of BP. The subanalysis performed is summarised in table 7.1 below. This evidence may have originated from linkage, chromosomal aberration or association studies. Subdivision of the case group was achieved using queries of the Microsoft Access phenotypic database. Only markers associated in the primary analysis were tested in the subanalysis, except for the COMT locus where all markers were tested.

Table 7.1: Summary of allelic association subanalysis performed

Subtype analysis	Loci						
	12q UCL	21q UCL	DISC1	BDNF	P2RX7	COMT	DAOA
Age of onset	Y	Y	Y	Y	Y	Y	Y
Gender	-	-	Y	-	-	Y	-
Psychotic symptoms	-	-	Y	-	-	Y	Y
Persecutory delusions	-	-	-	-	-	-	Y
Rapid cycling	-	-	-	Y	-	Y	-
Comorbid alcoholism	-	-	-	-	-	Y	-
Lithium response	-	-	-	-	-	Y	-

Linkage disequilibrium between pairs of markers was determined using LDPAIRS, another option within the GENECOUNTING package. Markers found to be in linkage disequilibrium with one another were used for subsequent haplotype analysis.

7.3.3 Haplotype analysis

Haplotype analysis was carried out using the above quoted program, GENECOUNTING. This method computes maximum likelihood estimates of haplotype frequencies from phase unknown case-control data. It uses an algorithm capable of analysing both SNPs and microsatellites. It also has the advantage of being able to deal with missing data. The significance of any overall haplotypic association with BP was computed using permutation testing of the observed haplotype distributions. Where empirical significance was high, greater numbers of permutations were employed to achieve better estimates of significance. Analysis of haplotype block structure was performed using HAPLOVIEW (Barrett *et al.* 2005).

The results of both the single marker and haplotype analysis are presented without correction for multiple testing and therefore these significance values should be interpreted with caution. Correction for multiple testing has not been applied because the appropriate method for correcting these analyses is unclear.

7.4 UCL 12q24 linked region

LD fine mapping of the 12q24 linkage region was initially carried out using microsatellite markers.

Fifteen microsatellite marker loci were selected and oligonucleotide sequence information obtained from the Marshfield Centre for Medical Genetics database (http://research.marshfieldclinic.org/genetics) and from the Genome database (http://www.gdb.org). In addition, publicly available genomic sequence was examined for potential microsatellite repeat structures within the region. These sequences were amplified in 32 samples and the products were investigated for potential polymorphic markers. Seven novel microsatellites were identified using this approach and subsequently genotyped. The oligonucleotide sequences for these microsatellites are given in appendix 2. A total of 21 microsatellite markers were typed.

Analyses of these microsatellite markers implicated a subregion, which was further investigated using SNP markers. The SNP genotyping was performed in stages. In the first stage, SNPs were either identified by re-sequencing of predicted genes in the sub-region or from the databases. Re-sequencing of exons and flanking intronic sequence was performed in 32 individuals. In the second stage, further SNPs were selected using Tagger (http://www.broad.mit.edu/mpg/tagger) to more systematically fine map the region. Oligonucleotide sequence information was obtained from the databases for known SNPs. The oligonucleotide sequences for SNPs identified by sequencing are shown in appendix 2.

In total 22 SNPs were tested. SNPs rs4765108, 47783-CT, rs1194031 and 29818-insT were typed using Pyrosequencing. SNPs rs2292446, rs7133178 and rs10773323 were typed using Taqman MGB probes. SNPs Mette_1, rs1212337, rs7308299, rs4765448, rs10773322, rs7311054, rs1194050, rs10847208, rs11058793, rs1194029, rs1616006, rs4765451, rs4765449 and rs10773324 were typed by Kbiosciences using the Amplifluor SNP genotyping method. The rs3830490 polymorphism was genotyped as per the methods for microsatellite markers above. The positions of all the markers tested are shown in figure 8.1 in the results section.

7.5 UCL 21q22.3 linked region

In a similar manner to that described in section 7.4, LD fine mapping of the 21q22.3 linkage region was initially carried out using microsatellite markers and then extended by SNP genotyping.

10 microsatellites (D21S266, D21S1260, D21S212, White-PolyT, D21S1411, PKNOX1-CA, PFKL-CA, D21S171, D21S1903, and D21S1897) were selected and sequence information obtained from the Marshfield Genetics and Genome databases. These were supplemented by a further nine novel microsatellites (22A5.GT124a, 218C10.CA44a,

C21orf2, AGAT10, ATTT11, 314n7CT18, 314n7CA28, 314n7CA23 and Col6a1) identified from the sequence databases. The oligonucletide sequences for these microsatellites are given in appendix 3. C21orf2 was genotyped using ^{32}P-labelled primers. The Col6a1 VNTR was genotyped by PCR and agarose gel electrophoresis.

Microsatellite analysis implicated the genes TSPEAR and TRPM2. Therefore these genes were screened for polymorphisms in BP subjects selected for having inherited polymorphisms associated with or linked to the chromosome 21q22.3 BP locus. In total, 28 BP subjects were chosen for sequencing, including 16 affected members from seven English BP families and seven NIMH BP families that had produced LOD scores >1 on 21q. A further 12 BP patients were selected from the association sample based on the fact that they carried alleles associated at the D21S171 and AGAT10 loci. Four control subjects were also screened. If nucleotide changes were detected, a further 32 controls were sequenced to verify whether the detected nucleotide changes were likely to be pathogenic or simply common polymorphisms. Controls were chosen without regard to their genotypes at the chromosome 21 loci. Primers were designed to sequence all the exons and 100 bases into each intron.

Again, SNPs for association testing were either identified through re-sequencing or from the databases. 11 SNPs were genotyped in total. Oligonucleotide sequence information was obtained from GDB and dbSNP for TMEM1_IN/DEL, ITGB2-*Aci*I, S100B-*Hae*III, rs1556314, rs1785437, rs1618355, rs933151, rs1785467, rs458178 and rs233283. Oligonucleotide sequence information for the PCNT2-*Alu*I SNP identified by sequencing is shown in appendix 3. SNPs rs1556314 and rs1785437 were typed using Pyrosequencing. SNPs rs1618355 and rs458178 were typed using Taqman. SNPs rs933151, rs1785467 and rs233283 were typed using Amplifluor. TMEM1, ITGB2-*Aci*I and S100B-*Hae*III were genotyped using ^{32}P-labelled primers. The positions of all markers tested are shown in figure 8.2 in the results section.

7.6 UCL 1q42 linked region

Initial fine mapping of the 1q42 linked region was performed with microsatellite markers as for the 12q24 and 21q22 regions. Four microsatellites in the region were selected from UCSC Genome Browser: D1S103, D1S3462, D1S235 and D1S251. Association was found with D1S3462, which maps to intron 8 of DISC1. Therefore, two further database microsatellites mapping in or near DISC1 were typed: DISC_5'_microsatellite and D1S2709. The positions of all markers tested are shown in figure 8.3 in chapter 8. The oligonucleotide sequences and PCR conditions are set out in appendix 4.

7.7 BDNF

In order to attempt to replicate the findings of Neves-Pereira *et al* (Neves-Pereira *et al.* 2002), the BDNF –1040bp (GT)n repeat was genotyped. To further investigate LD at the locus an intragenic microsatellite, AFMA275YB9, was selected using the UCSC Genome Browser.

The four SNPs which were found to be individually associated with BP in the Sklar *et al* (Sklar *et al.* 2002) study were selected for genotyping by Kiobioscience: the Val66Met SNP rs6265 (a39); rs11030096 (a13); rs16917204 (a20) and rs11030101 (a40). Unfortunately, the last two of these - rs16917204 (a20) and rs11030101 (a40) - failed assay design and were abandoned. The Val66Met SNP rs6265 was also genotyped in-house in 621 individuals by Taqman. The positions of all markers tested are shown in figure 8.5 in chapter 8. The oligonucleotide sequences and PCR conditions are set out in appendix 5.

Two subanalyses were performed. The association of rapid cycling BP was tested after Muller *et al* (Muller *et al.* 2006) and Green *et al* (Green *et al.* 2006). The association of early onset BP was tested after Geller *et al* (Geller *et al.* 2004) and Strauss *et al* (Strauss *et al.* 2004).

7.8 P2RX7

An attempt was made to replicate the finding of association with the microsatellite marker NBG6 (Shink *et al.* 2005a). To further test association at the P2RX7 locus, the UCSC Genome Browser sequence for P2RX7 was interrogated for repeat structures. Three potential markers were identified and optimisation attempted. However, only the marker denoted P2RX73 gave products which could be adequately resolved. Therefore a total of two microsatellites were genotyped. The oligonucleotide sequences and PCR conditions are set out in appendix 6.

Assays were designed for nine SNPs from the original sudy by Barden *et al* (Barden *et al.* 2006): rs208293 (P2RX7-I04B), rs208294 (P2RX7-E05A), rs504677 (P2RX7-I07E), rs7958311 (P2RX7-E08A), rs1718119 (P2RX7-E11B), rs2230911 (P2XR7E11C), rs2230912 (P2RX7-E13A), rs11065501 (P2XR4-UTR3A) and rs3817190 (CAMKK2-E01B). Of note, rs2230911 is referred to as rs6489795 by Barden *et al.* These SNPs were chosen so that the replication of the individual allelic association and 7-SNP sliding window haplotypic association reported in the study could be attempted. Unfortunately, two of these SNPs - rs7958311 (P2RX7-E08A) and rs2230911 (P2XR7E11C) - failed to give reliable genotypes and were abandoned.

Genotyping was carried out by Kbiosciences using the Kaspar method. Genotyping was initially carried out on the 604 BP samples and 450 controls available. Subsequently, 110 additional control samples were recruited and these samples were typed for the two markers showing association in the preliminary analysis - rs208294 (P2RX7-E05A) and rs2230912 (P2RX7-E13A).

The entire sample was screened for the exon 13 deletion described by Barden *et al* (Barden *et al.* 2006). Primers were designed to flank the deletion (see appendix 6). Alleles were separated and visualised on 1% agarose gel using standard techniques.

The positions of all markers tested are shown in figure 8.6 in chapter 8. Analysis of individual haplotype association was performed using HAPLOVIEW.

A combined association analysis for marker rs2230912 was performed using the data from this thesis, the Barden *et al* (Barden *et al.* 2006) and Lucae *et al* (Lucae *et al.* 2006) studies. Lucae *et al* tested for association with UP rather than BP. It was considered valid to combine the results on the basis that the conditions may be genetically related. With regard to the 12q24 region specifically, the linkage findings of Abkevich *et al* (Abkevich *et al.* 2003) could suggest the presence of a common susceptibility locus.

7.9 COMT

The Val158Met polymorphism (rs4680) was genotyped in house using the Amplifluor system and genotypes were detected on a Roche LightCycler 480.

To extend the testing of association of COMT with BP a further SNP (rs2075507) and microsatellite (intron 1 tetranucleotide) were also genotyped.

The *Hind*III restriction site polymorphism rs2075507, identified by DeMille *et al* (DeMille *et al.* 2002), maps to the critical region for the oestrogen-dependent repression of the P2 promotor. DeMille *et al* also identified the tetranucleotide repeat that maps to intron 1 of MB-COMT isoform (Alfred site identifier - SI000269R). The quoted primers were redesigned to give PCR products of a smaller size which would accord with our pooling strategy.

The UCSC Genome Browser sequence for COMT was interrogated for other repeat structures and three potential markers were selected for optimisation. The three potential markers were either found to be monomorphic or the alleles could not be accurately scored.

Primer sequences and cycling conditions are show in appendix appendix 7. The positions of all markers tested are shown in figure 8.8 and 8.9 in chapter 8.

7.10 DAOA/G30

An attempt was made to replicate the individual marker and haplotypic association reported by Hattori *et al* (Hattori *et al.* 2003). Therefore, the eight relevant SNPs were selected for genotyping: rs1815686, rs1935058, rs1341402, rs2391191 (M15), rs1935062, rs947267 (M18), rs778294 (M19) and rs954581. Unfortunately, rs1815686 and rs1935058 failed assay design.

To extend the testing of association of DAOA with BP, rs3918342 (M23) and rs1421292 (M24) were also genotyped.

Genotyping was performed by Kbiosciences using the Amplifluor method. The positions of all markers tested are shown in figure 8.10 in chapter 8.

Chapter 8: Results

8.1 UCL 12q24 region

All SNPs were in HWE. The association results are summarised in figure 8.1, table 8.1 and table 8.2 below.

In total seven markers showed allelic association with BP. Four of the associated markers were microsatellites: 1634tet, 307GT4, D12S307 and D12SDK1. Three were SNPs: rs4765449, rs7133178 and 29818-insT.

The results for 1634tet, 307GT4, D12S307, 29818-insT and D12SDK1 are the same as reported by our group previously (Kalsi *et al.* 2006). Typing of additional individuals resulted in rs7133178 becoming slightly less significant. The pufu in/del (rs3830490) was previously associated at the ***p*=0.034** level but was no longer formally significant, p=0.079, in the enlarged dataset. From the most recent set of SNPs tested one further SNP, rs4765449, was found to be associated, ***p*=0.039**.

The region defined by the most proximal and distal associated markers - 1634tet at 125,621,931 bp and D12SDK1 at 125,802,874 bp respectively - spans approximately 180 kb.

Analysis of BP cases with early onset illness against the control group generally increased the evidence for association. The most significant results were obtained with D12S307 (***p*=0.0001**) and rs7133178 (***p*=0.0003**). The minor allele (A) of rs7133178 was increased in cases compared to controls (15% versus 9%). The direction of the association was the same as observed in the whole dataset.

Haplotype analysis based on the associated markers was consistent with the allelic association results for the whole dataset and early onset subset. However, it's usefulness for identifying haplotypes likely to harbour aetiological mutations has proved limited.

Figure 8.1: LocusView output of 12q24 markers typed (May 2004 Assembly)

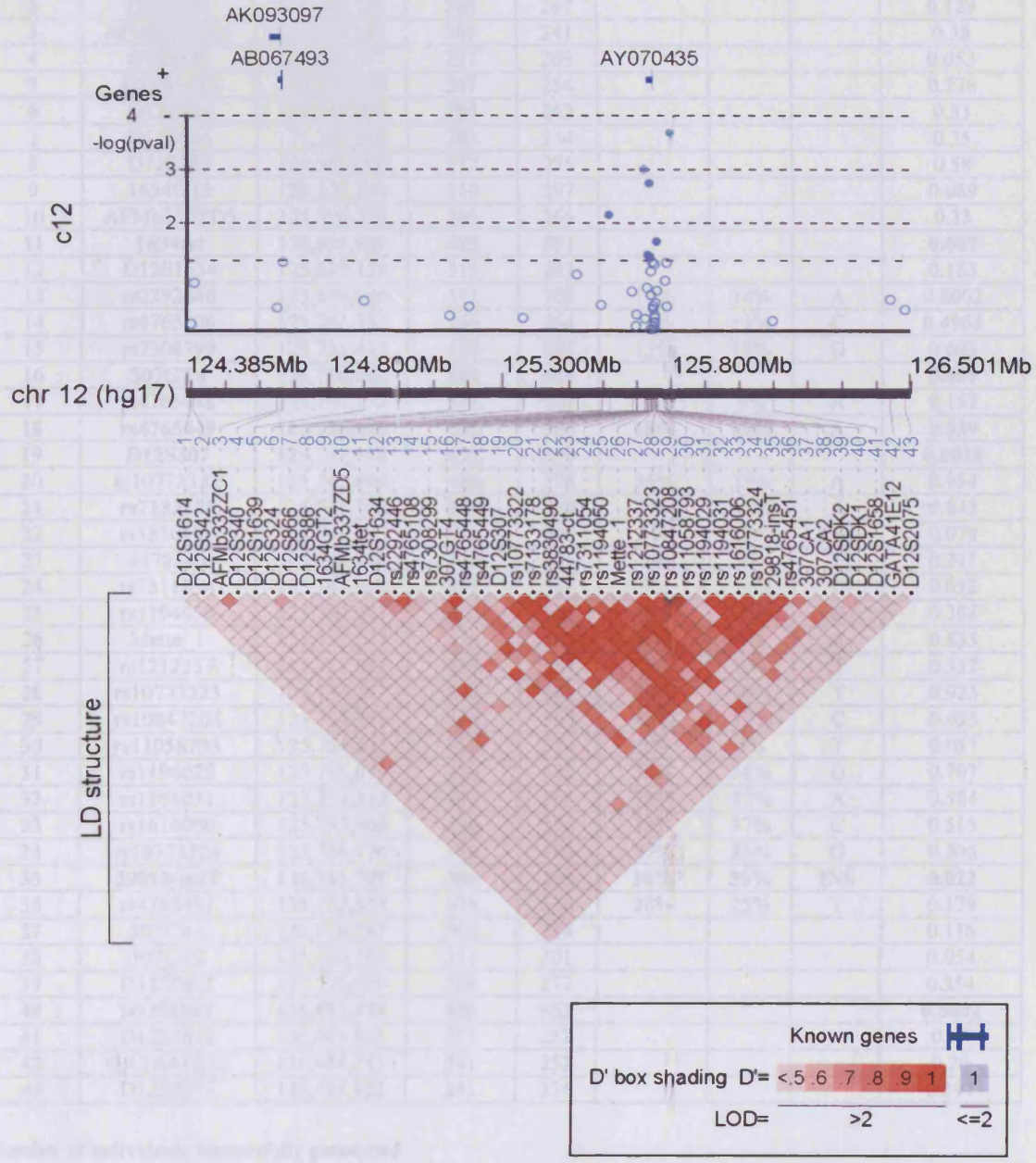


Table 8.1: Single marker association results for the UCL 12q24 region

Marker		Location (May 04 Assembly)	Number of Samples*		MAF***		Minor Allele	Allelic Association
			Controls	BP	Controls	BP		p-value**
1	D12S1614	124,400,331	295	271				0.736
2	D12S342	124,408,720	298	267				0.129
3	AFMb332ZC1	124,650,140	288	241				0.38
4	D12S340	124,666,617	297	263				0.053
5	D12S1639	124,903,049	247	254				0.276
6	D12S324	125,149,966	294	262				0.53
7	D12S866	125,207,206	293	234				0.35
8	D12S386	125,366,006	277	275				0.58
9	1634GT2	125,526,244	310	297				0.089
10	AFMb337ZD5	125,596,235	283	265				0.33
11	**1634tet**	**125,621,931**	**412**	**561**				**0.007**
12	D12S1634	125,689,128	315	282				0.183
13	rs2292446	125,699,946	311	300	14%	14%	A	0.8002
14	rs4765108	125,701,352	266	264	15%	13%	C	0.4968
15	rs7308299	125,725,413	424	587	13%	13%	G	0.603
16	**307GT4**	**125,726,085**	**394**	**544**				**0.001**
17	rs4765448	125,732,172	551	581	8%	9%	A	0.157
18	**rs4765449**	**125,735,688**	**531**	**586**	**18%**	**15%**	**A**	**0.039**
19	**D12S307**	**125,741,191**	**431**	**590**				**0.0018**
20	rs10773322	125,741,404	409	576	35%	35%	A	0.954
21	**rs7133178**	**125,741,788**	**595**	**589**	**9%**	**12%**	**A**	**0.042**
22	rs3830490	125,743,175	576	600	9%	12%	DEL	0.079
23	44783-ct	125,743,735	290	296	3%	2%	T	0.241
24	rs7311054	125,747,922	537	568	9%	11%	C	0.052
25	rs1194050	125,750,661	412	555	3%	2%	G	0.388
26	Mette_1	125,751,572	435	588	6%	6%	A	0.833
27	rs1212337	125,751,852	435	587	2%	2%	T	0.337
28	rs10773323	125,754,207	442	581	49%	50%	T	0.923
29	rs10847208	125,754,386	426	565	46%	47%	C	0.693
30	rs11058793	125,754,644	538	574	3%	5%	T	0.057
31	rs1194029	125,755,075	432	585	38%	38%	G	0.707
32	rs1194031	125,755,312	443	591	38%	37%	A	0.584
33	rs1616006	125,755,566	426	556	39%	37%	C	0.515
34	rs10773324	125,756,776	436	586	27%	25%	G	0.306
35	**29818-insT**	**125,761,701**	**306**	**293**	**36%**	**30%**	**INS**	**0.022**
36	rs4765451	125,762,344	435	586	28%	25%	T	0.179
37	307CA1	125,786,142	306	284				0.116
38	307CA2	125,790,180	311	301				0.054
39	D12SDK2	125,790,929	308	272				0.354
40	**D12SDK1**	**125,802,874**	**426**	**553**				**0.0002**
41	D12S1658	126,099,303	267	252				0.66
42	GATA41E12	126,444,213	291	252				0.26
43	D12S2075	126,485,922	241	274				0.397

* Number of individuals successfully genotyped
** Most significant p-value quoted from CLUMP results
*** MAF = minor allele frequency

Table 8.2: Distribution of alleles for associated microsatellite markers in the UCL 12q24 region

Marker	Location (May 2004 Assembly)	Sample (Number)*	Allelic association															
			Distribution of alleles														Chi^2	Empirical p-value
1634tet	125,621,931		217	221	225	229	233	237	241	245	249	253	257	261	265	269	273	
		Controls (412)	0	8	41	175	219	169	56	14	28	34	46	26	6	1	1	
			0%	1%	5%	21%	27%	21%	7%	2%	3%	4%	6%	3%	1%	0%	0%	18.33^{T4}
		Bipolars (561)	4	16	90	263	242	190	85	14	49	38	69	41	16	4	1	0.007
			0%	1%	8%	23%	22%	17%	8%	1%	4%	3%	6%	4%	1%	0%	0%	
307GT4	125,726,085		352	354	355	356	358	360	362	363	364	366	368	370	372	374	376	
		Controls (394)	1	2	0	1	8	7	61	0	289	244	99	25	49	1	1	
			0%	0%	0%	0%	1%	1%	8%	0%	37%	31%	13%	3%	6%	0%	0%	25.75^{T2}
		Bipolars (544)	1	1	1	8	13	37	58	1	369	367	146	35	45	4	2	0.001
			0%	0%	0%	1%	1%	3%	5%	0%	34%	34%	13%	3%	4%	0%	0%	
D12S307	125,741,191		148	152	154	156	158	160	162	164	166	168	170					
		Controls (431)	64	1	5	700	47	24	1	0	3	15	2					
			7%	0%	1%	81%	5%	3%	0%	0%	0%	2%	0%					20.94^{T2}
		Bipolars (590)	133	4	4	891	80	36	5	3	6	10	8					0.0018
			11%	0%	0%	76%	7%	3%	0%	0%	1%	1%	1%					
D12SDK1	125,802,874		158	160	164	166	168	170	172	174	176	178						
		Controls (426)	0	6	515	9	0	137	118	54	12	1						
			0%	1%	60%	1%	0%	16%	14%	6%	1%	0%						27.16^{T1}
		Bipolars (553)	1	3	650	9	1	179	157	106	0	0						0.0002
			0%	0%	59%	1%	0%	16%	14%	10%	0%	0%						

For each marker the repeat number is shown in the top row. Allele frequencies and percentages are shown below.
*Number of individuals successfully genotyped.
T^1,T^2,T^4=T1,T2,T4 statistic from CLUMP

8.2 UCL 21q22 region

All SNPs were in HWE. The results of the association testing are summarised in figure 8.2 and tables 8.3 and 8.4 below.

Two microsatellites - D21S171 and D21S1260 - provided significant evidence for association with empirical p-values of **0.016** and **0.031** respectively, both with the T3 test of CLUMP. The 90 kb region between AGAT10 and D21S171 was targeted for further analysis. Although AGAT10 was not associated using the empirical tests of significance an uncorrected p-value of 0.026 was obtained when each allele was compared with the sum of the remaining alleles in a 2x2 contingency table. At that time there were only two candidate genes - TRPM2 and C21orf29 - within the region.

Seven SNPs were typed across these genes. SNPs rs1556314 and rs1785467 showed allelic association with BP with p-values of **0.008** and **0.025** respectively. The minor allele (G) of rs1556314 was increased in controls (23%) compared to BP cases (18%). SNP rs1556314 is in exon 11 of TRPM2 and causes an amino-acid change from aspartic acid to glutamic acid at position 543. The minor allele (C) of rs1785467 was also increased in controls (17%) compared to cases (12%). SNP rs1785467 is in intron 23 of TRPM2. There is strong LD between these two markers, D'=0.876.

Five other SNPs in TRPM2 and C21orf29 were identified by sequencing (see table 8.5). The three TRPM2 SNPs are now referred to as rs36007753, rs34214578 and rs35288229. The minor allele (A) of the exon 6 SNP rs36007753 was observed in one of 28 BP patients and not in any of 36 controls. However, it was not predicted to alter an amino acid. From an evolutionary standpoint it should be noted that there is conservation of the major allele (G) across human, chimp and rat sequences, but the minor allele (A) is observed in the mouse. The minor allele of the intron 8 SNP rs34214578 was detected in one of four controls examined and not in any of 28 BP patients. This polymorphism occurs 78 nucleotides before exon 9. SNP rs35288229 is located in exon 15 and is predicted to alter the amino acid at position 755 from arginine to cysteine. This amino

acid is located in part of one of the protein's predicted transmembrane domains. However analysis of the amino acid change with transmembrane structure prediction programmes TMHMM (http://www.cbs.dtu.dk/services/TMHMM-2.0/) and TMpred (http://www.ch.embnet.org/software/TMPRED_form.html) did not predict significant alterations in TRPM2 transmembrane structures.

The two C21orf29 SNPs are now referred to as rs35449326 and rs35028190. The intron 7 substitution, rs35449326, was detected 26 nucleotides before exon 8 in one of 28 BP cases but in none of 48 controls. SNP rs35028190 is located in exon 8. It is predicted to alter amino acid 408 from histidine to glutamine. However, the minor allele was observed in 5 of 28 BP cases and in 3 of 48 controls that were sequenced.

These polymorphisms have not yet been typed in the full sample.

Analysis of BP cases with early onset illness against the control group increased the evidence for association with TRPM2. Both rs1556314 and rs1785467 became more significant with p-values of **0.0023** and **0.0173** respectively. The direction of association remained the same.

Haplotype analysis using the two associated SNPs supported the single marker association results for the whole dataset and early onset subset, but again did not help identify haplotypes for sequencing.

Figure 8.2: LocusView output of 21q22 markers typed (May 2004 Assembly)

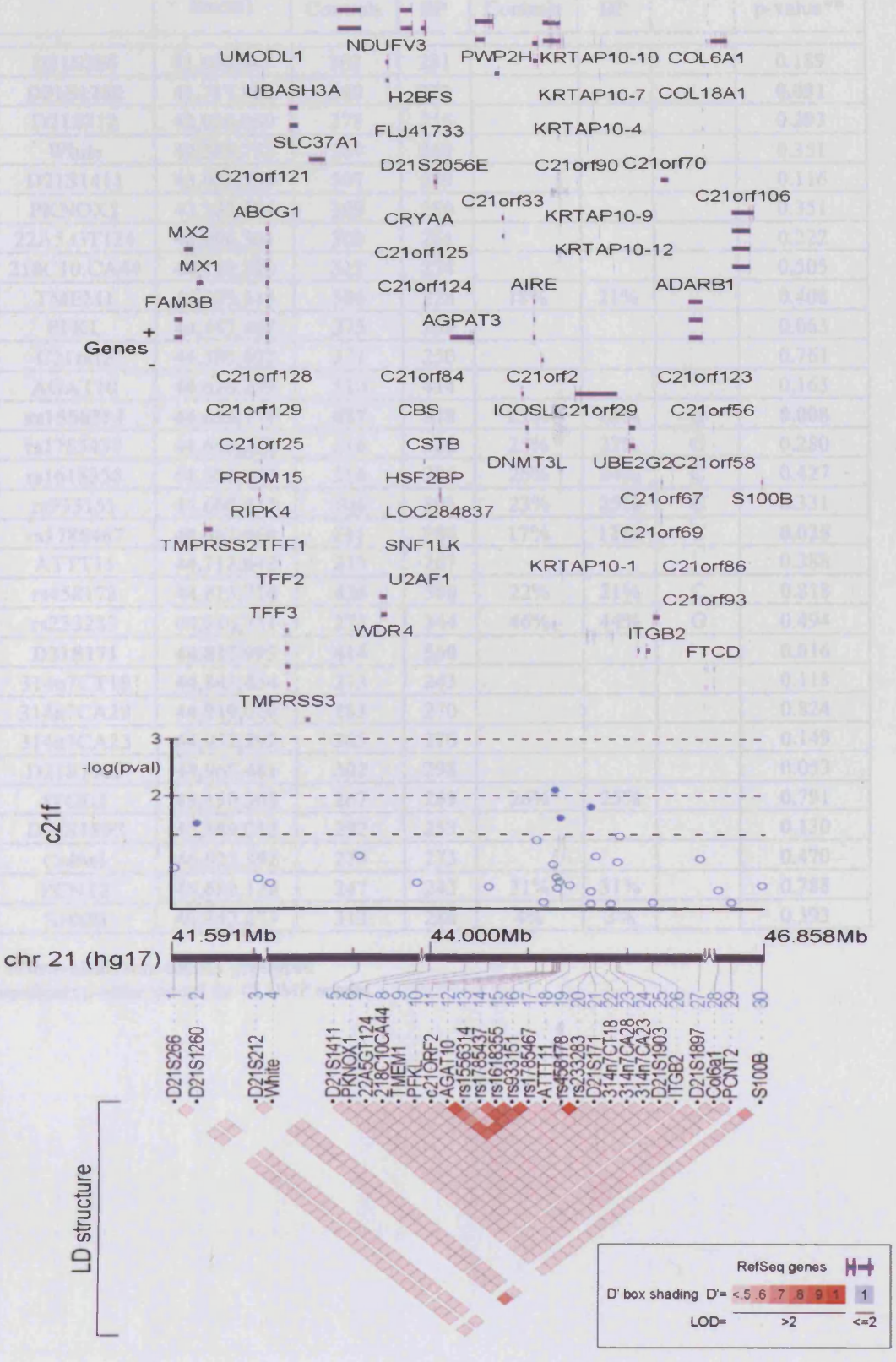


Table 8.3: Single marker association results for UCL 21q22 region

Marker		Location (May 04 freeze)	Number of Samples *		MAF		Minor Allele	Allelic Association
			Controls	BP	Controls	BP		p-value**
1	D21S266	41,606,447	302	281				0.189
2	**D21S1260**	**41,717,912**	**290**	**272**				**0.031**
3	D21S212	42,026,060	278	256				0.293
4	White	42,589,753	289	269				0.351
5	D21S1411	43,033,739	307	279				0.116
6	PKNOX1	43,323,394	309	280				0.351
7	22A5.GT124	43,996,361	300	261				0.222
8	218C10.CA44	44,182,320	312	274				0.505
9	TMEM1	44,303,314	306	278	18%	21%		0.408
10	PFKL	44,543,467	275	259				0.063
11	C21orf2	44,580,622	271	250				0.761
12	AGAT10	44,626,239	510	414				0.163
13	**rs1556314**	**44,635,771**	**427**	**578**	**23%**	**18%**	**G**	**0.008**
14	rs1785437	44,646,222	316	306	25%	23%	G	0.280
15	rs1618355	44,650,890	316	306	26%	24%	C	0.427
16	rs933151	44,660,433	306	303	23%	25%	C	0.331
17	**rs1785467**	**44,663,060**	**311**	**286**	**17%**	**12%**	**C**	**0.025**
18	ATTT11	44,712,642	275	267				0.388
19	rs458178	44,815,716	438	580	22%	21%	C	0.818
20	rs233283	44,816,731	273	244	46%	44%	G	0.494
21	**D21S171**	**44,817,095**	**414**	**560**				**0.016**
22	314n7CT18	44,843,434	273	243				0.118
23	314n7CA28	44,919,010	283	270				0.824
24	314n7CA23	44,952,292	305	270				0.149
25	D21S1903	44,965,481	302	298				0.053
26	ITGB2	45,130,368	267	269	26%	25%		0.791
27	D21S1897	45,369,052	292	253				0.130
28	Col6a1	46,623,392	274	273				0.470
29	PCNT2	46,689,126	247	243	31%	31%		0.788
30	S100B	46,842,673	313	286	4%	3%		0.393

* Number of individuals successfully genetyped
** Most significant p-value quoted for CLUMP results

Table 8.4: Distribution of alleles for associated microsatellite markers in the UCL 21q22 region

Marker	Location (May 2004 Assembly)	Sample (Number)*	Allelic association														
			Distribution of alleles													Chi^2	Empirical p-value
D21S1260	41,717,912		204	206	208	210	212	214	216	218	220	224	226	228		8.128^{T3}	0.031
		Controls (290)	5	35	15	161	29	217	84	20	3	5	4	2			
			1%	6%	3%	28%	5%	37%	14%	3%	1%	1%	1%	0%			
		Bipolars (272)	4	27	7	156	51	179	80	25	2	3	7	3			
			1%	5%	1%	29%	9%	33%	15%	5%	0%	1%	1%	1%			
D21S171	44,817,095		213	215	217	219	223	225	227	229	231	233	235	237	239	8.953^{T3}	0.016
		Controls (414)	1	0	23	7	4	84	385	122	141	55	6	0	0		
			0%	0%	3%	1%	0%	10%	46%	15%	17%	7%	1%	0%	0%		
		Bipolars (560)	1	1	23	9	1	117	571	172	137	72	9	5	2		
			0%	0%	2%	1%	0%	10%	51%	15%	12%	6%	1%	0%	0%		

For each marker the repeat number is shown in the top row. Allele frequencies and percentages are shown below.
*Number of individuals successfully genotyped.
T3=T3 statistic from CLUMP

Table 8.5: Results of mutation screening of TRPM2 and C21orf29

SNP	Gene	Position	Sequence around change [Major allele/Minor Allele]	Predicted Amino Acid Change	MAF (number individuals tested)	
					Controls	Bipolar
rs36007753	TRPM2	Exon 6	GGACCAGGCTGGAGAAGTTCATATC[G/A] GAGCAGACCAAGGAAAGAGGAGGTA	None	0.00 (36)	0.02 (28)
rs34214578	TRPM2	Intron 8	ACCCAAGGCTCTGGCTGTGTCCGGG[G/T] CCTCAATTCTATGTGTCCTGGAGGA	None	0.13 (4)	0.00 (28)
rs35288229	TRPM2	Exon 15	GCTCTCCGTGGACAATGGGCTGTGG[C/T] GTGTGACCCTGTGCATGCTGGCCTT	R755C	0.13 (4)	0.02 (28)
rs9974831	TRPM2	Intron 15	GTGGAGGTCAGGACCCTGGCCCTGG[C/T] GGTCACTGTCTCCTCTCTGCTGTGA	None	0.07 (36)	0.05 (28)
rs35449326	C21orf29	Intron 7	CTGGCACCGACCCAAGCAGAACCA[C/G] GTGGTCCCGGTTTACTCTCTTTCAG	None	0.00 (48)	0.02 (28)
rs35028190	C21orf29	Exon 8	TCTCTGTCATTTACAAATGGAGCCA[C/G] AGAAAGCTGAAGTTTACCCCATATC	H408Q	0.03 (48)	0.05 (28)

8.3 DISC1

The positions of the typed markers are shown in figures 8.3 and 8.4. The association results are summarised in table 8.6.

The marker D1S3462, mapping to intron 8 of DISC1, was found to be significantly associated (**p=0.0160**). Analysis of D1S3462 by gender revealed that significant association was observed with BP males (**p=0.0126**) but not BP females (p=0.2005). Analysis of BP cases with OPCRIT defined psychotic symptoms against controls resulted in the association becoming more significant, (**p=0.0089**). No association was found with non-psychotic BP cases. On this occasion the evidence for association was not increased by selecting out the early onset cases (p=0.0516).

The marker DISC1_5'_Microsatellite, which was observed to be an imperfect repeat, was also found to be associated (**p=0.0189**). A gender effect was not observed for this marker. Evidence for association was neither increased by subanalysis according to the presence of psychotic symptoms nor by age of onset.

Low intermarker LD, as measured by D', was observed except between D1S251 and DISC1_5'_Microsatellite (D'=0.803). Thus haplotype analysis was only performed with these two markers. Association was not supported by this analysis.

Figure 8.3: UCSC multiple BLAT output for 1q42 markers genotyped

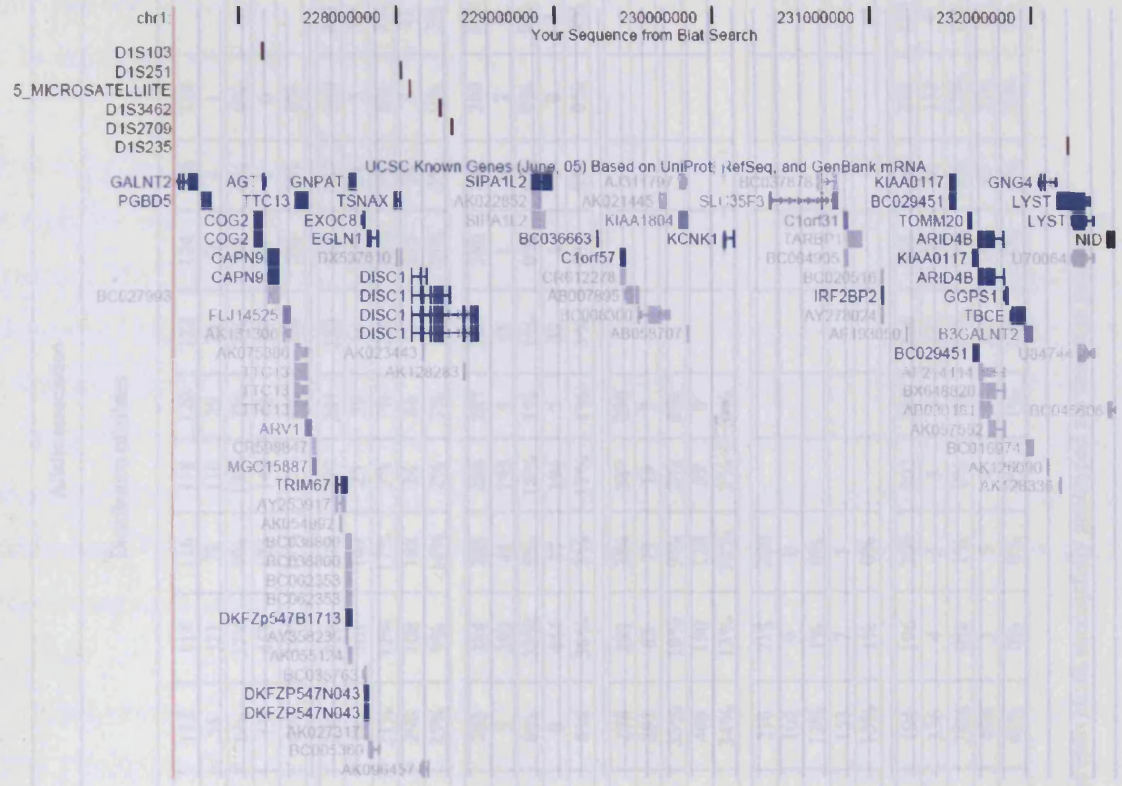


Figure 8.4: Positions of the four genotyped markers which map in or near DISC1

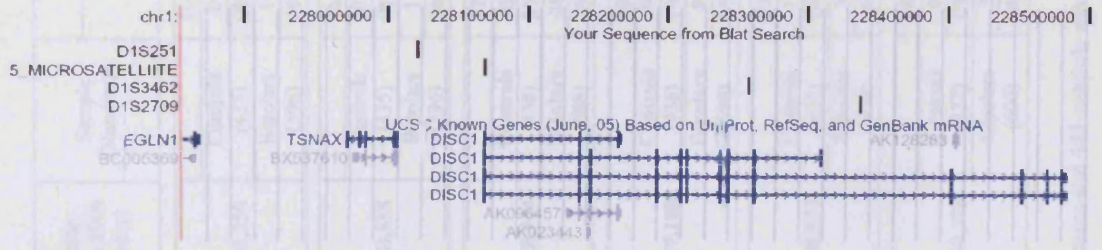


Table 8.6: Single marker association results for 1q42 region

Marker	Location (March 2006 Assembly)	Sample (Number)*	Allelic association															
			Distribution of alleles															Empirical p-value
D1S103	228,903,359		104	106	108	110	112	114	116	118	120	122	124	126	128				0.718^{T1}
		Controls (435)	0	38	7	149	54	321	78	110	29	21	58	4	1				
			0%	4%	1%	17%	6%	37%	9%	13%	3%	2%	7%	0%	0%				
		Bipolars (596)	1	41	9	200	73	457	125	148	43	29	65	1	0				
			0%	3%	1%	17%	6%	38%	10%	12%	4%	2%	5%	0%	0%				
D1S251	229,780,888		263	265	269	271	273	275	277	279	281	283	285	287	289	291	295		0.235^{T2}
		Controls (435)	0	0	48	10	185	101	147	25	59	14	225	24	3	29	0		
			0%	0%	6%	1%	21%	12%	17%	3%	7%	2%	26%	3%	0%	3%	0%		
		Bipolars (596)	1	1	73	14	298	108	188	24	86	15	315	20	5	41	3		
			0%	0%	6%	1%	25%	9%	16%	2%	7%	1%	26%	2%	0%	3%	0%		
DISC1_5'_MICRO	**229,828,914**		**276**	**278**	**280**	**282**	**283**	**284**	**285**	**286**	**287**	**288**	**290**	**292**	**300**	**302**	**304**	**306**	**0.018^{T2}**
		Controls (438)	**19**	**289**	**1**	**1**	**3**	**293**	**41**	**155**	**5**	**11**	**1**	**0**	**2**	**0**	**52**	**3**	
			2%	**33%**	**0%**	**0%**	**0%**	**33%**	**5%**	**18%**	**1%**	**1%**	**0%**	**0%**	**0%**	**0%**	**6%**	**0%**	
		Bipolars (588)	**28**	**350**	**1**	**4**	**0**	**444**	**37**	**196**	**7**	**33**	**1**	**1**	**0**	**2**	**69**	**3**	
			2%	**30%**	**0%**	**0%**	**0%**	**38%**	**3%**	**17%**	**1%**	**3%**	**0%**	**0%**	**0%**	**0%**	**6%**	**0%**	
D1S3462	**230,017,109**		**266**	**269**	**272**	**275**	**278**	**281**	**284**	**287**	**290**								**0.016^{T1}**
		Controls (436)	**9**	**2**	**112**	**277**	**291**	**83**	**77**	**19**	**2**								
			1%	**0%**	**13%**	**32%**	**33%**	**10%**	**9%**	**2%**	**0%**								
		Bipolars (598)	**8**	**0**	**119**	**354**	**406**	**150**	**120**	**39**	**0**								
			1%	**0%**	**10%**	**30%**	**34%**	**13%**	**10%**	**3%**	**0%**								
D1S2709	230,096,973		208	210	212	214	216	218	220										0.701^{T3}
		Controls (423)	1	103	233	398	102	9	0										
			0%	12%	28%	47%	12%	1%	0%										
		Bipolars (600)	3	147	336	553	153	7	1										
			0%	12%	28%	46%	13%	1%	0%										
D1S235	233,960,378		182	184	188	192	194	196	200	202	204	206	208	210	212	214	216		0.161^{T3}
		Controls (437)	0	0	12	1	336	4	5	3	5	222	129	40	112	4	1		
			0%	0%	1%	0%	38%	0%	1%	0%	1%	25%	15%	5%	13%	0%	0%		
		Bipolars (600)	2	1	9	4	498	3	5	4	9	265	164	43	176	15	2		
			0%	0%	1%	0%	42%	0%	0%	0%	1%	22%	14%	4%	15%	1%	0%		

* For each marker 613 cases and 443 controls were genotyped. The numbers of individuals successfully genotyped are shown in brackets

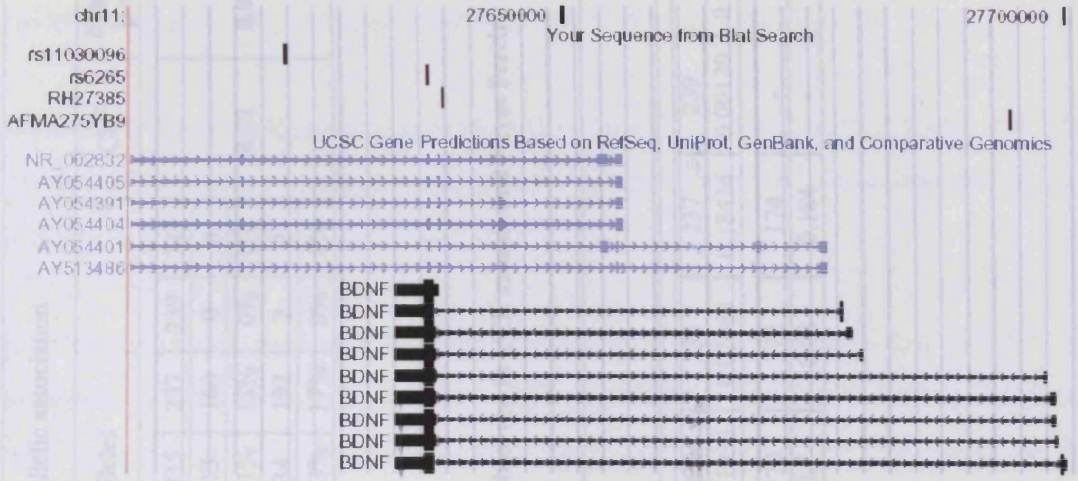
8.4 BDNF

Both SNPs were in HWE. A 98.7% correlation between in-house and Kbiosciences genotypes was obtained. The markers genotyped are shown in figure 8.5. The association results are summarised in tables 8.7 and 8.8.

Statistically significant association was found with the marker –1040bp (GT)n repeat, although the association was modest: ***p*=0.048**. As can be seen in table 8.4.2, a modest increase in the frequency of the 235 bp allele in cases compared to controls is observed (3% versus 1%). The alignment and frequency of the alleles reported by Neves-Periera *et al* (Neves-Pereira *et al.* 2002) and those observed here is shown in table 8.9. Analysis of BP cases with rapid cycling against controls led to an increase in the significance of the association ***p*=0.0028**. The overrepresentation of the 235 bp allele in the case group was accentuated (7% versus 1%). No association was found when early onset BP cases were compared against controls.

Moderately strong LD, *D*'=0.841, was observed between –1040bp (GT)n repeat and the rs6265 (Val66Met) polymorphism. Haplotype analyisis incorporating these two markers provided additional evidence for association, empirical ***p*=0.0021**. The LRT contribution suggested an overrepresentation of the Valine-235 and Valine-237 haplotypes in the case sample (3% versus 1% and 5 % versus 2% respectively). The Methionine-237 haplotype was by contrast underrepresented in the case group, 12% versus 16%.

Figure 8.5: UCSC multiple BLAT output for markers genotyped at BDNF locus



* RH27385 is a synonym for the -1040bp (GT)n repeat

Table 8.7: Single marker association results for genotyped BDNF markers

	Marker	Location (March 06 Assembly)	Number of Samples Tested*		MAF		Minor Allele	Allelic Association
			Controls (443)	BP (613)	Controls	BP		p-value
1	rs11030096	27,622,119	436	578	49%	49%	C	0.786
2	rs6265(Val66Met)	27,636,492	441	582	17%	15%	A	0.347
3	**-1040bp (GT)n repeat**	**27,637,640**	**436**	**561**				**0.048** [T4]
4	AFMA275YB9	27,694,422	419	600				0.223 [T3]

* Total number of individuals genotyped shown in brackets

Table 8.8: Distribution of alleles for BDNF -1040bp (GT)n repeat

Marker	Location (March 2006 Assembly)	Sample Number*	Allelic association										
			Distribution of alleles									Chi^2	Empirical p-value
-1040bp (GT)n repeat	27,637,640		225	227	229	231	233	235	237	239	243	8.72	**0.048** [T4]
		Controls (436/443)	0	3	8	56	632	13	160	0	0		
			0%	0%	1%	6%	72%	1%	18%	0%	0%		
		Bipolars (561/613)	3	1	14	71	803	34	192	2	2		
			0%	0%	1%	6%	72%	3%	17%	0%	0%		

*(Number of individuals successfully genotyped/total number of individuals genotyped)

Table 8.9: Comparison between -1040bp (GT)n repeat allele fragment sizes and frequencies observed in UCL sample and Neves Pereira et al (Neves-Pereira et al. 2002)

Sample	Repeat number Frequencies											
UCL	-	-	-	225	227	229	231	233	235	237	239	243
	-	-	-	0.00043	0.00602	0.00473	0.07134	0.71036	0.02063	0.18436	0.00129	0.00086
Neves-Pereira *et al*	156	158	160	162	164	166	168	170	172	174	-	-
	0.002	0.004	0.004	0.003	-	0.009	0.065	0.7	0.03	0.184	-	-

8.5 P2RX7

The 21 bp deletion in exon thirteen13 of P2RX7 was not observed in any of the BP individuals. The data for six of the seven SNPs were in HWE for the control samples. SNP rs11065501 showed a modest departure from HWE (***p*=0.027**).

The positions of the genotyped markers in relation to P2RX7 are shown in figure 8.6. The LD between SNP markers is shown in figure 8.7. The association results are summarised in and tables 8.10 and 8.11.

Statistically significant evidence for allelic association with BP was obtained from the microsatellite NBG6 (***p*=0.01**) and the SNP rs2230912 (***p*=0.043**). Strong LD was observed between markers NBG6 and rs2230912 (*D'*=0.976). The frequency of the minor allele (G) of rs2230912 was increased in cases compared to controls (18% versus 15%).

Relatively low LD was observed between the seven SNP markers. Specifying a spine of LD with a *D'* of >0.8, two small haplotype blocks were generated using HAPLOVIEW. Block one consisted of two SNPs: rs208293 (P2RX7-I04B) and rs208294 (P2RX7-E05A). Block two consisted of three SNPs: rs504677 (P2RX7-I07E), rs1718119 (P2RX7-E11B) and rs2230912 (P2RX7-E13A). This block covers the region that includes the microsatellite marker NBG6.

It is not possible to directly compare these results with LD patterns observed in the study of Barden *et al* (Barden *et al.* 2006) where a greater number of markers spanning a broader region were typed. However, block 2 sits within Barden et al's wider block 2. Analysis of this seemingly important block two revealed significant association using data from rs504677 (P2RX7-I07E) and rs2230912 (P2RX7-E13A), global empirical ***p*=0.0079**. Testing of individual haplotypes from this analysis using HAPLOVIEW revealed two significantly associated haplotypes. The frequency of haplotype TG was estimated to be 14% in controls and 18% in cases (empirical ***p*=0.03**), while haplotype

TA was estimated to be found in 29% of controls and 24% of cases (empirical ***p*=0.027**). It was not possible to reconstruct the 7-SNP haplotype which Barden *et al* reported to be associated. However, the haplotype analysis provided further evidence for the association of P2RX7 with BP, see table 8.12.

The present study found association of the mutated G (arginine) allele of rs2230912 with BP. This is in line with the findings of both Barden *et al* and Lucae *et al.* A total of 3,586 individuals were included in the combined analysis of this marker. This revealed significant allelic association (***p*=0.0025**) and genotypic association (***p*=0.0005**). The allelic and genotypic associations are explained by an increase of heterozygotes rather than homozygotes which suggests a dominant mode of transmission, odds ratio of 1.302 (CI 1.129-1.503).

Analysis of UCL BP cases with early onset illness against the control group increased the evidence for association of the G allele, ***p*=0.0062**. The G allele occurred in 21% of cases compared with 15% of controls. However the evidence for association of NBG6 decreased slightly, ***p*=0.0181**, with this analysis.

Figure 8.6: UCSC multiple BLAT output for markers genotyped in the P2RX7 region

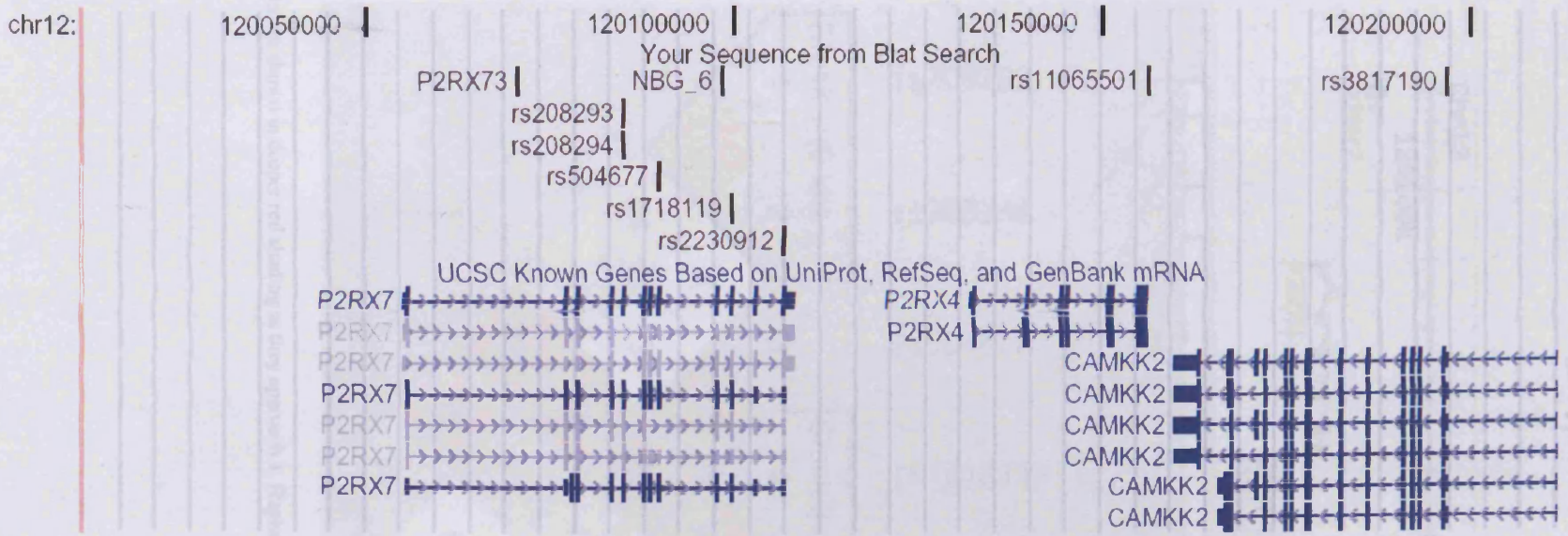
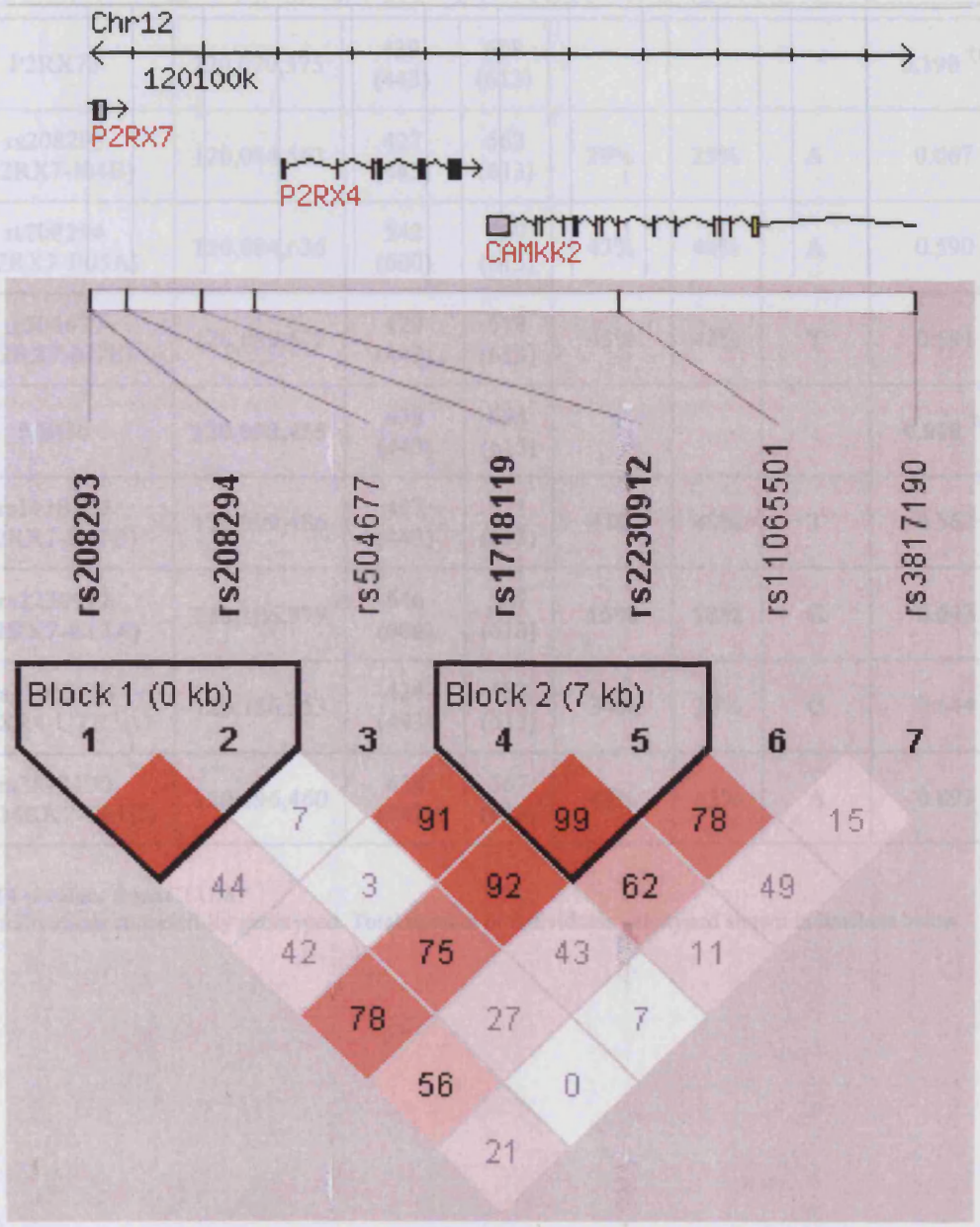


Figure 8.7: Marker to marker LD between genotyped SNPs at the P2RX7 locus.



D' statistics are shown in deeper red shading as they approach 1. Haplotype blocks are defined by a solid spine of LD *D'*>0.8.

Table 8.10: Marker positions, allele frequencies and tests of allelic association in or adjacent to the P2RX7 locus

Marker		Location (March 06 Assembly)	Number of Samples*		MAF		Minor Allele	Allelic Association
			Controls	BP	Controls	BP		p-value
1	P2RX73	120,070,375	439 (443)	608 (613)				0.190 [T1]
2	rs208293 (P2RX7-I04B)	120,084,563	427 (443)	562 (613)	29%	25%	A	0.067
3	rs208294 (P2RX7-E05A)	120,084,636	542 (600)	569 (613)	43%	44%	A	0.590
4	rs504677 (P2RX7-I07E)	120,089,572	429 (443)	578 (613)	43%	42%	T	0.591
5	**NBG6**	**120,098,415**	**438 (443)**	**604 (613)**				**0.010** [T4]
6	rs1718119 (P2RX7-E11B)	120,099,486	427 (443)	571 (613)	41%	40%	T	0.562
7	**rs2230912 (P2RX7-E13A)**	**120,106,579**	**546 (600)**	**587 (613)**	**15%**	**18%**	**G**	**0.043**
8	rs11065501 (P2XR4-UTR3A)	120,156,353	424 (443)	583 (613)	34%	33%	G	0.644
9	rs3817190 (CAMKK2-E01B)	120,196,460	424 (443)	567 (613)	43%	43%	A	0.893

[T1,T4] T1 and T4 p-values from CLUMP

* Number of individuals successfully genotyped. Total number of individuals genotyped shown in brackets below

Table 8.11: Distribution of alleles for the microsatellite NBG6

Marker	Marker location (May 2004 Assembly)	Sample (Number)*	Allelic association														
			Distribution of alleles**													Chi^2	Empirical p-value
NBG6	120,098,415		211	215	217	219	221	223	225	227	229	231	233	235	237	239	15.2	0.010^{T4}
		Controls (438/443)	1	14	7	364	98	148	17	72	48	30	17	27	12	21		
			0%	2%	1%	42%	11%	17%	2%	8%	5%	3%	2%	3%	1%	2%		
		Bipolars (604/613)	0	9	8	528	138	231	36	88	46	55	12	26	7	24		
			0%	1%	1%	44%	11%	19%	3%	7%	4%	5%	1%	2%	1%	2%		

*(Number of individuals successfully genotyped/total number of individuals genotyped) ** For each marker the repeat number is shown in the top row. Allele frequencies and percentages are shown below.

Table 8.12: Haplotypic association analysis results for 2-SNP haplotypes from block 2 of the P2RX7 locus

Marker	Alleles						Estimated frequency (cases)	Estimated frequency (controls)	LRT	LRT	d.f.	p-value	Empirical p-value
	1	2	3	4	5	6							
3 rs504677 5 rs2230912			C		A		58%	56%	0.26	11.88	3	0.0078	0.0079
			C		G		0%	1%	2.93				
			T		A		24%	29%	4.15				
			T		G		18%	14%	4.11				
4 rs1718119 5 rs2230912				C	A		60%	58%	0.19	9.52	3	0.023	0.019
				T	A		21%	26%	4.14				
				T	G		18%	15%	2.36				

LRT: Likelihood ratio test, cells with estimated frequencies of zero in both cases and controls omitted

Table 8.13: Combined analysis of SNP rs2230912 data from Barden *et al*, Lucae *et al* and UCL study

Study	Sample	Genotype		
		AA	GA	GG
Barden *et al* 2006	Controls	150	54	8
	BP	134	73	6
	Totals	284	127	14
Lucae *et al* 2006	Controls	760	235	34
	Unipolar depression	684	293	22
	Totals	1444	528	56
UCL study	Controls	389	147	10
	BP	390	177	20
	Totals	779	324	30
Combined analysis	Controls	1299	436	52
	BP and unipolar depression	1208	543	48
	Totals	2507	979	100

Allelic association Chi^2=9.171, ***p*=0.002**
Allelic odds ratio=1.213 confidence interval=1.070-1.375
Genotypic association Chi^2=15.118, ***p*=0.0005**
Genotypic odds ratio (GG+GA vs AA)=1.302 confidence interval=1.129-1.503

8.6 COMT

Both SNPs were in HWE. The positions of the markers typed are shown in figures 8.8 and 8.9. None of the markers showed allelic or genotypic association with BP, (table 8.14). Highest LD was found between the intron 1 tetranucleotide repeat and rs4680, D'=0.77. No haplotypic association was found using these two markers.

Analysis of BP cases with comorbid panic disorder revealed modest association with the valine allele of rs4680 uncorrected for multiple testing, ***p*=0.0485**. No association of rs2075507 or the intron 1 tetranucleotide was found with this subgroup.

Further analysis of rs4680 was performed. BP cases with psychotic symptoms, comorbid alcoholism, early onset disease, lithium response and rapid cycling were tested against controls. Analysis by gender was also performed. No significant associations were found.

It has been recognised that the intron 1 tetranucleotide is an imperfect repeat (Palmatier *et al.* 2004). In addition to the TAAA ladder there is a secondary ladder with a final TAA repeat sequence. The imperfect 285 bp allele has generally been found to be the most common across different populations (http://alfred.med.yale.edu/alfred/). Table 8.15 shows how the alleles observed correspond with those quoted in the Alfred database. The allele frequencies observed in the UCL sample were consistent with those previously found in European populations.

Figure 8.8: UCSC multiple BLAT output of COMT markers genotyped

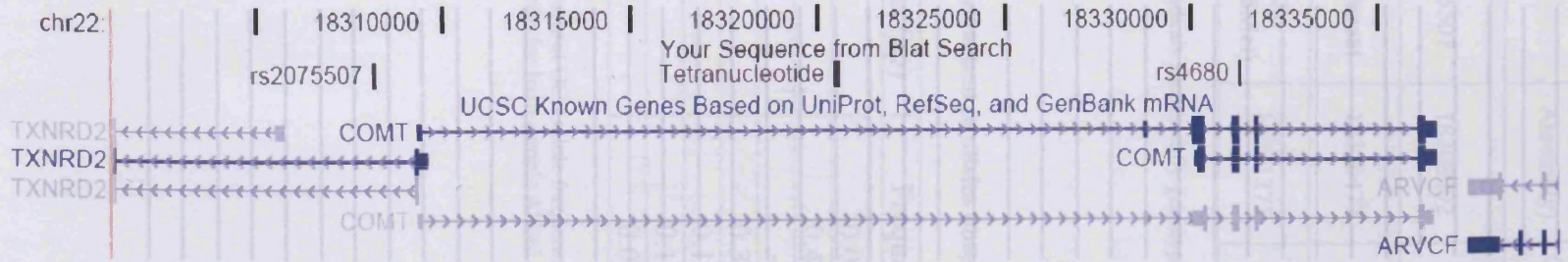


Figure 8.9: Structure of COMT gene and position of markers genotyped

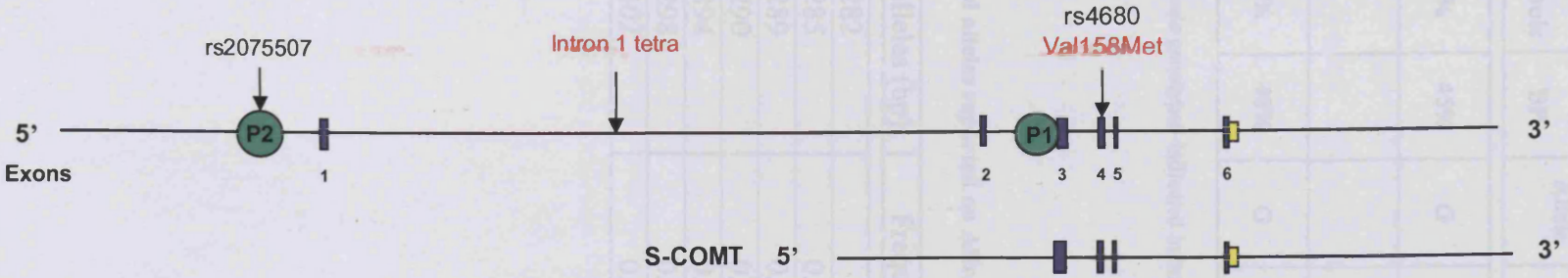


Table 8.14: Single marker association results for COMT markers

Marker		Location (March 06 Assembly)	Number of Samples*		MAF		Minor Allele	Allelic Association
			Controls (443)	BP (613)	Controls	BP		p-value
1	rs2075507	18308092	439	599	45%	45%	G	0.76
2	Tetra repeat	18320477	423	600				0.77 [T2T3]
3	rs4680 (Val158Met)	18331271	431	585	47%	48%	G	0.56

*Number of individuals successfully genotyped. Total number of individuals genotyped indicated in brackets.

Table 8.15: Comparison of alleles observed in UCL sample and alleles reported on Alfred database

UCL alleles (bp)	Frequency*	Alfred alleles (bp)	Frequency**
114	0.000	282	-
117	0.458	285	0.447
121	-	289	0.013
122	0.320	290	0.292
126	0.112	294	0.124
130	0.103	298	0.115
134	0.007	302	0.009

* combined case/control UCL allele frequencies
** Allele frequencies for Irish sample Alfred ID SA000057M

8.7 DAOA/G30

All SNPs were found to be in HWE. The positions of the genotyped markers in relation to DAOA are shown in figure 8.10.The LD between SNP markers is shown in figure 8.11. The association results are summarised in table 8.16.

Allelic association was detected with rs3918342 (M23), ***p*=0.038**. The T allele was increased in cases compared with controls (54% versus 48%). A trend towards association was observed with rs954581, p=0.073.

Markers rs1341402 to rs778294 (M19) form an LD block defined with a solid spine of LD with a D' >0.8. Strong LD was also observed between rs3918342 (M23) and rs1421292 (M24). Haplotype analysis within the LD blocks did not show evidence for association.

Analysis of the data for rs3918342 (M23) using phenotypic subdivisions that were previously found to be important in DAOA/G30 studies did not provide additional evidence for association. BP cases with psychotic symptoms showed association at the p=0.04 level. No significant association was observed when BP cases with persecutory delusions were compared against controls. However there were only 45 such cases. There was an insufficient number of BP cases with panic disorder to perform a meaningful analysis. Selection of cases by early onset again increased the evidence for association, ***p*=0.0247**, with overrepresentation of the T allele.

Figure 8.10: UCSC multiple BLAT output of DAOA markers genotyped

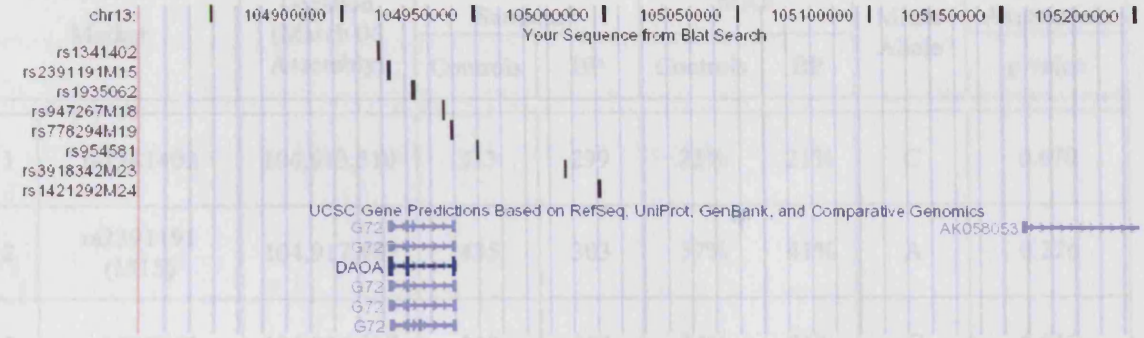


Figure 8.11: Haploview of LD between SNP markers across DAOA locus

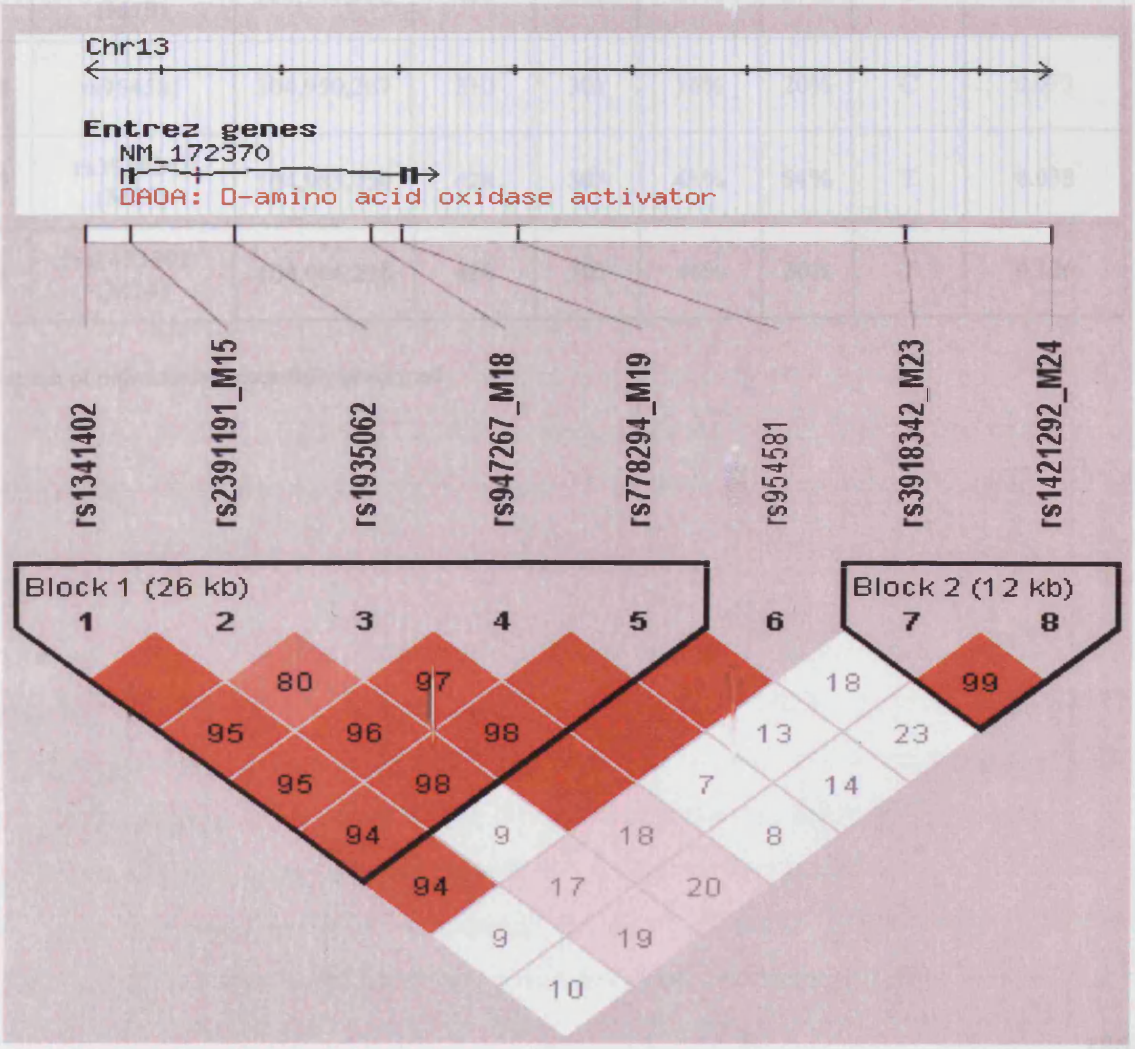


Table 8.16: Single marker association results for DAOA locus

Marker		Location (March 06 Assembly)	Number of Samples*		MAF		Minor Allele	Allelic Association
			Controls	BP	Controls	BP		p-value
1	rs1341402	104,913,510	313	299	22%	21%	C	0.679
2	rs2391191 (M15)	104,917,447	435	303	37%	41%	A	0.226
3	rs1935062	104,926,137	313	303	34%	34%	C	0.917
4	rs947267 (M18)	104,937,663	435	302	42%	41%	C	0.548
5	rs778294 (M19)	104,940,236	438	302	27%	27%	T	0.856
6	rs954581	104,950,267	310	301	16%	20%	C	0.073
7	**rs3918342 (M23)**	**104,983,750**	**428**	**303**	**48%**	**54%**	**T**	**0.038**
8	rs1421292 (M24)	104,996,236	430	303	46%	50%	T	0.126

* Number of individuals successfully genotyped

Chapter 9: Conclusions

In this chapter I will consider each locus individually and then make some overall comments. I intend to be limited and specific in the inferences I make from the results. I will indicate the future direction of enquiry for each locus. I will defer discussion of the broader issues raised by these association studies until chapter 10.

12q24 UCL region

The results presented here provide evidence for the presence of a novel BP susceptibility locus on 12q24.

As was discussed in section 5.2, there is evidence for multiple BP susceptibility loci in the 12q23-q24 region. P2RX7 is the nearest of the previously implicated genes to the associated markers in this study. However the distance between P2RX7 and these markers is still not compatible with a single locus producing the evidence for association that was detected. LD over such a large distance would be unprecedented.

Three of the four markers originally found to be associated in the Danish cohort (Kalsi *et al.* 2006) lie within the 180 kb region defined by the most proximal and distal markers associated in the UCL sample. However, these particular markers were not found to be significantly associated in the UCL sample. Therefore only partial replication for this region can be claimed.

There are five ACEVIEW (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly) gene predictions within the 180 kb region. These are: LOC338825, Skoynar, LOC387895, Slynar and Zorgawby (see figure 9.1). The region is not well characterized, but of these the most promising candidate is Slynar. Slynar is expressed in the human brain and so far five splice variants have been demonstrated. No known function has been attributed to this gene, although predicted translation products of some splice variants contain PSORT-predicted membrane localization motifs (Nakai and Horton 1999).

Another candidate gene termed LOC387895 is antisense to Slynar and so may have a role in the control of Slynar expression (Yelin *et al.* 2003). LOC387895 is also brain expressed, although to a lesser extent than Slynar (http://www.ncbi.nlm.nih.gov/UniGene/ESTProfileViewer). 13 alternative splice variants have been isolated from various tissues.

The analysis by age of onset suggests that the region harbours a susceptibility locus for an early onset form of BP. The most significant allelic associations with early onset BP were found with D12S307 and rs7133178. Both of these markers map to within the LOC387895/Slynar predicted genes, see figure 9.2. D12S307 is intronic. Relative to Slynar rs7133178 is intronic but is located in the predicted exon4 of LOC387895 (www.ncbi.nlm.nih.gov/SNP). SNP rs7133178 is synonymous and hence unlikely to confer susceptibility in itself. LD between these markers and the aetiological mutations is thus presumed.

On the basis of these results, mutation screening of LOC387895 and Slynar has now begun. Given the significance of the association of rs7133178, sequencing of exon 4 of LOC387895 must be considered a priority. Since sequencing of all individuals is unfeasible, careful consideration must be given to the selection of cases for analysis. Sequencing of early onset cases with the minor, A, allele of rs7133178 (n=85), in the first instance, would seem to represent the most parsimonious approach.

Replication of the genetic association findings presented here in multiple samples is now needed to confirm that a novel susceptibility locus has been identified. Testing of the hypothesis that the risk is specific to early onset BP may help alleviate the problem of non-replication in the context of genetic heterogeneity.

Figure 9.1: UCSC multiple BLAT output of associated markers in the UCL 12q24 region, annotated with gene prediction tracks

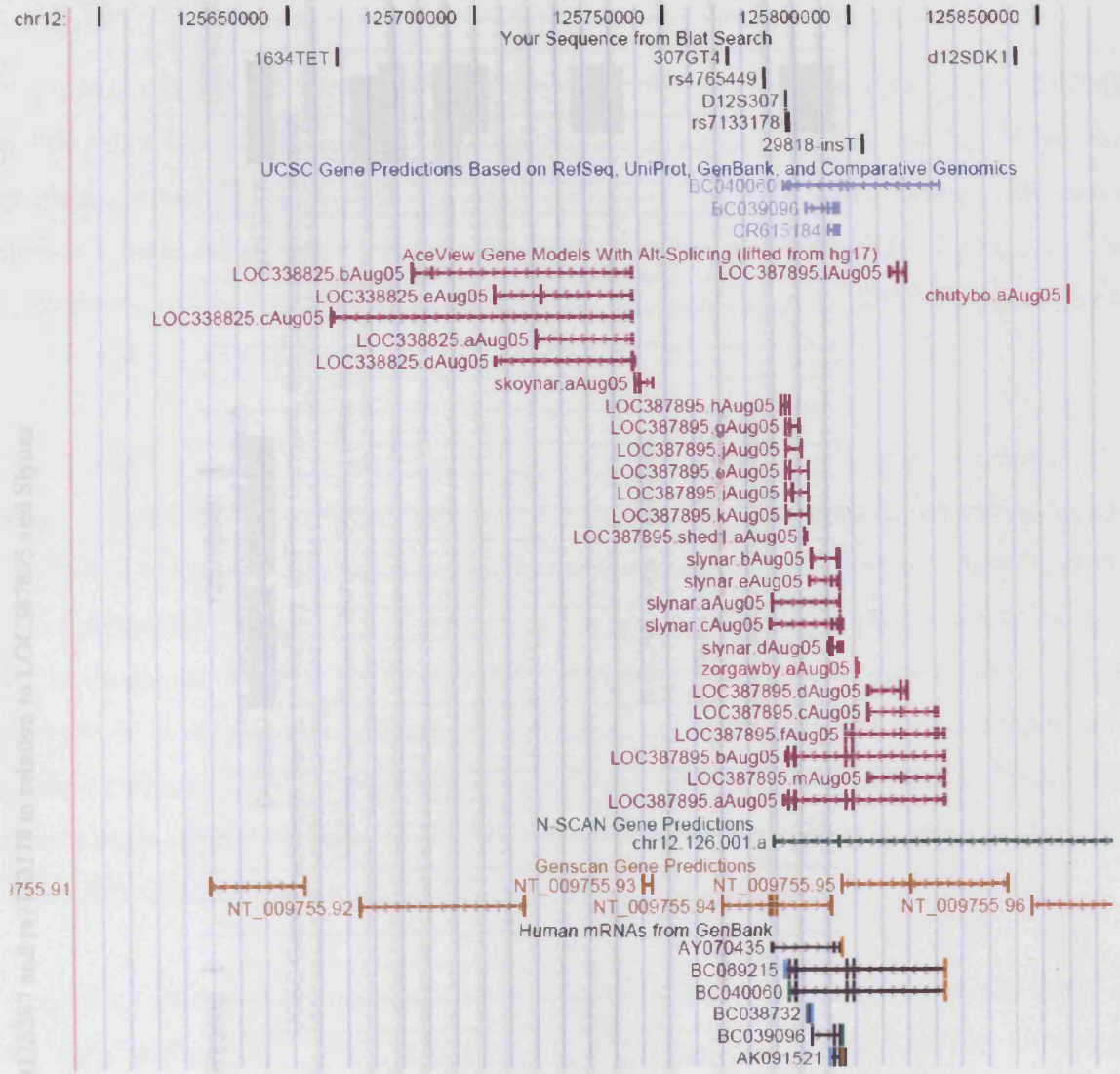
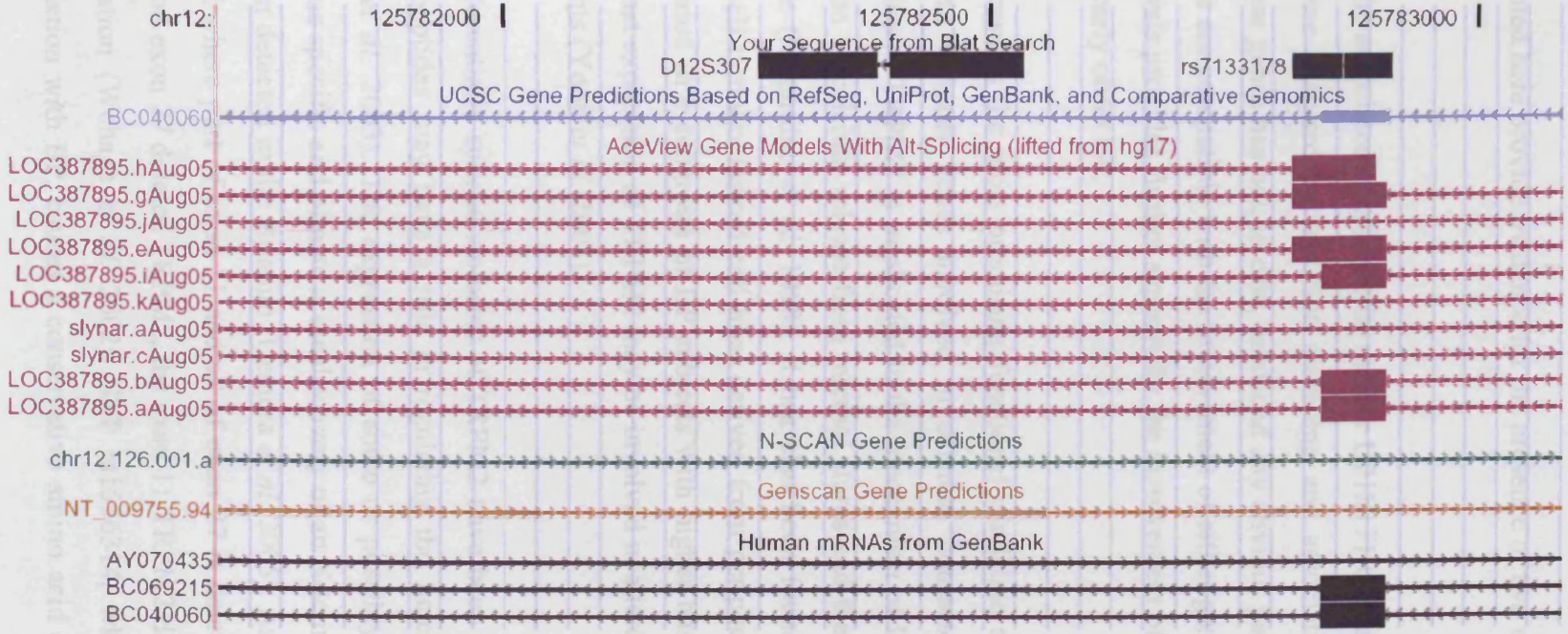


Figure 9.2: Position of D12S307 and rs7133178 in relation to LOC387895 and Slynar



UCL 21q22 region

The results presented here provide evidence for the presence of two novel susceptibility BP loci on 21q22.3.

In the region of the associated microsatellite marker D21S171 the candidate genes TRPM2 and C21orf29 were analysed by mutation screening and association testing. While the sequencing of these genes has not, to date, produced any obvious pathogenic mutations the association results are compatible with the involvement of either gene (see figure 9.3). The age of onset analysis provides further support for the involvement of TRPM2, and suggests it confers risk to early onset BP.

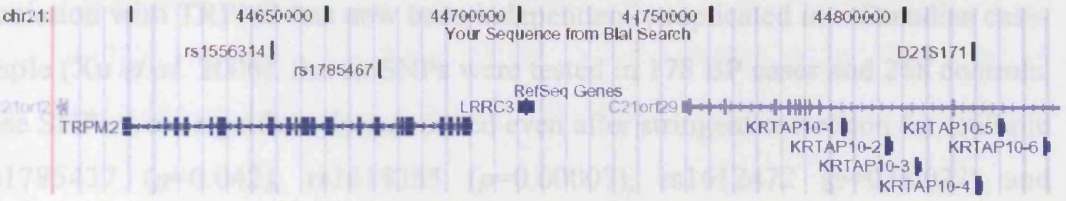
TRPM2 also represents the most promising functional candidate as it is a member of a family of calcium channel genes involved in calcium homeostasis which is highly expressed in the brain. TRPM2 is implicated in the mechanism of lithium action by virtue of the coregulation of calcium release from internal stores and the production of inositol 1,4,5 triphosphate (Nagamine *et al.* 1998). It has also been found that TRPM2 mRNA transcripts in B cell lymphoblastoid cell lines derived from peripheral blood lymphocytes were down-regulated in a subgroup of BP subjects with higher basal calcium levels. The authors suggest that expression of TRPM2 may be involved in altered calcium homeostasis in some BP patients (Yoon *et al.* 2001).

A number of differentially spliced isoforms of TRPM2 have been identified and some of these truncated peptides may have a role in regulating the function of the full-length isoform (Zhang *et al.* 2003). The expression of some or possibly all of these isoforms appears to be tissue specific and of note is the discovery of an N-terminal truncated TRPM2 that has only been detected in the striatum (Uemura *et al.* 2005). Another TRPM2 isoform has been isolated where part of exon 11 and all of exon 27 is deleted. In contrast to full-length TRPM2 and exon 27 deleted TRPM2, the exon 11 TRPM2 deletion did not respond to H2O2 stimulation (Wehage *et al.* 2002). SNP rs1556314, which showed the most significant association with BP, causes a conservative amino acid change in the exon 11

deleted region. However there is no evidence that this change has any role in either the dysregulation of cellular calcium homeostasis in response to oxidative stress or in BP susceptibility.

There is no database support for C21orf29 brain expression. While this has not been empirically tested in the UCL laboratory, C21orf29 cannot be considered a strong candidate gene.

Figure 9.3: Associated markers in relation to TRPM2 and C21orf29



The UCSC RefSeq annotation of the interval between D21S171 and the associated SNPs now includes the gene leucine-rich repeat-containing 3 precursor (LRRC3) and a cluster of Keratin-associated protein (KRTAP) genes. Although LRRC3 is predominately expressed in the cervix, brain expression has also been demonstrated and thus it must also be considered a candidate. The KRTAP genes encode keratin imtermediate filaments which are a major component of hair fibers. Multiple intronless KRTAPs map to the introns of C21orf29. Expression is restricted to the hair root and hence the involvement of this family of genes in BP susceptibility would seem unlikely.

Allelic association was also detected with the more proximal marker D21S1260. SNP analysis of this region has not yet been performed. This marker is over 3 Mb from D21S171 so association cannot be explained by a single mutation. However, the evidence for association at D21S1260 is very modest even without correction for multiple tests. The issue of multiple testing will be revisited at the end of this chapter.

While the results argue for the further analysis of TRPM2 the optimal strategy for this is not immediately apparent. An increased frequency of the major alleles at both the TRPM2 associated SNPs in BP cases was observed. Hence the majority of the BP cases would have to be sequenced to be confident that an aetiological change is not missed. So far mutation screening has only been carried out in a small proportion of the BP cases.

In contrast to the 12q24 region, it seems that further association testing is indicated. It should be possible to reduce the sequencing burden by employing the systematic approach of genotyping haplotype tagging SNPs across TRPM2. Study of the polymorphisms already identified by sequencing (table 8.5) should also be included in future association testing.

Allelic association with TRPM2 has now been independently replicated in a Canadian case-control sample (Xu *et al.* 2006). Seven SNPs were tested in 178 BP cases and 268 controls. Four of these SNPs were significantly associated even after stringent correction for multiple testing: rs1785437 (p=0.042), rs1618355 (p=0.00007), rs1612472 (p=0.00022) and rs933151 (p=0.011). For all of these SNPs the minor allele was increased in the cases compared to controls. Highly significant haplotypic association was found with haplotypes incorporating all of the SNPs tested. One particular haplotype, TTTCTTA, was estimated to be present in 12% of cases but only 0.9% of controls ($p=2.3\times10^{-12}$).

The two SNPs associated in the UCL sample were not tested. SNPs rs1785437, rs1618355 and rs933151 were tested in the UCL sample but not found to be associated, suggesting that the aetiological mutation(s) occurred on a different haplotypic background in the two samples. This is somewhat surprising as approximately 25% of the Canadian sample has British ancestry (both case and control groups). It is of note that the MAF of rs1785437, rs1618355 and rs933151 in the control was similar in the Canadian and UCL samples.

More recently, Roche *et al* failed to replicate TRPM2 association in a sample of 125 BPI trios from Ireland (Roche *et al.* 2007). 14 SNPs were tested across the TRPM2 gene. No evidence of individual marker or haplotypic association was found. Of the six previously

associated markers two were tested by Roche *et al*, rs1556314 and rs933151. By contrast they implicated the neutrophic factor S100B. Two SNPs were found to be associated: rs2839350 and rs3788266. The evidence for association was increased in the subset of trios with psychosis. In the UCL sample one SNP was tested in S100B but no association found. Since the loci are over 2 Mb apart it does not seem feasible that the same signal is being detected.

Kostyrko *et al* (Kostyrko *et al.* 2006) performed mutation screening of TRPM2 in 60 unselected BP cases by SSCP and sequencing. Seven polymorphisms were thus identified which were then genotyped in a control group. No significant difference in allele frequencies was observed. This study would not seem to add greatly to the data on TRPM2. The sample size was very small and little information is given about the control group other than that it consisted of 66 "healthy persons". Furthermore, only about 30% of the coding sequence of TRPM2 was screened. Interestingly, a synonymous exon 19 change, at rs35698354, was found in the BP group but not in the controls.

DISC1

The results add to the evidence that DISC1 influences susceptibility to BP. For the first time individual markers rather than haplotypes have been found to be associated. Furthermore, the results suggest that DISC1 specifically confers risk to BP patients with psychotic symptoms. Given that association has also been reported with schizophrenia and schizoaffective disorder, it does raise the possibility that DISC1 association defines a psychotic illness that crosses traditional diagnostic boundaries. I will return to the subject of common susceptibility loci in chapter 10.

It is difficult to relate the UCL microsatellite results to previous findings (section 5.4.2) where SNPs were typed. The associated microsatellite D1S3462, mapping to intron 8, is located within the associated region 3 of Thompson *et al* (Thomson *et al.* 2005) and the broader associated region of Maeda *et al* (Maeda *et al.* 2006). The DISC1_5'_Microsatellite is within the associated region 1 of Thompson *et al.* Thus some

consistency of results is discernible. However, the relationship of the microsatellite alleles to the individually associated haplotypes in these regions is not known.

Despite the equivalence of BP and schizophrenia rates in males and females there is accumulating evidence of a gender effect at the DISC1 locus. However, the male specific association described here is generally at odds with previous findings, although Thompson *et al* did observe male specific association in the more centromeric region 1. Epigenetic mechanisms or hormonal influences have to be invoked to explain sex effects for autosomal markers.

It certainly cannot be claimed that systematicic LD fine mapping of the UCL 1q42 linked region was performed or that DISC1 is unequivocally responsible for the linkage peak. However, given that allelelic association was observed with a previously implicated locus in the first set of markers tested *a priori* probability suggests efforts should focus on DISC1. Since the SNP and microsatellite results cannot be directly compared and the distribution of microsatellite allele frequencies are not helpful in selecting individuals for sequencing, systematic LD mapping with SNPs is again indicated. This is now underway as part of a multicenter European consortium. However, this does not preclude more extensive fine mapping of 1q42.

BDNF

Recent work on the -1040bp (GT)n microsatellite has called into question the validity of using this marker in association studies. Sequencing has demonstrated that it is a highly polymorphic marker with a complex structure rather than a simple dinucleotide repeat (Okada *et al.* 2006). A total of 23 allelic variants were identified, arising from a combination of three separate dinucleotide repeats, an insertion/deletion and a substitution. Since some of the variants are of the same size, the variants cannot be unequivocally identified by fragment size based genotyping methodologies. This is not only relevant to the association findings presented here but also to the original finding of association at the

marker (Neves-Pereira *et al.* 2002) and the subsequent finding of haplotypic association (Kremeyer *et al.* 2006).

The fragment sizes obtained in the UCL sample were at variance from those reported by Neves-Pereira *et al.* Kreyemer *et al* also noticed variance with the Neves-Pereira *et al* allele and concluded that the fragments were 57 bp larger than previously described. However, the allele frequencies of Kreyemer *et al* did not correspond well with those reported by Neves-Pereira *et al.* Allele frequencies in the UCL sample corresponded closely with Neves-Pereira *et al* frequencies if it is assumed that the fragments are 53 bp larger. It is not clear that these discrepancies can be accounted for by the underlying complexity of the polymorphism.

In the UCL sample, association was only found with the -1040bp (GT)n microsatellite or with haplotypes incorporating this marker. Association was not replicated with either of the SNPs. It is intriguing that the association became more significant when cases with rapid cycling were compared against controls, in line with the findings of Green *et al* (Green *et al.* 2006), and that the frequency of haplotypes containing the valine allele of rs6265 was estimated to be increased in the case group. However I do not think it would be right to interpret these results in light of the new information regarding the marker. Thus, it has to be simply concluded that neither the valine allele of rs6265 nor alleles at rs11030096 influence susceptibility to BP in the UCL sample. This by no means excludes the involvement of BDNF in BP.

Okada *et al* (Okada *et al.* 2006) suggests that the -1040bp (GT)n microsatellite is a functional polymorphism affecting transcriptional activity. While it would be of interest to perform a more detailed analysis of this marker in the UCL sample, sequencing of this region in each individual would represent a considerable undertaking.

P2RX7

The results presented here independently replicate the association of P2RX7 with BP. Two markers intragenic to P2RX7, the intron 9 microsatellite NBG6 and the non-synonmous exon 13 SNP rs2230912, had previously shown association. Both of these associations were replicated in the UCL dataset. It was not possible to test the 7-SNP haplotypic association reported by Barden *et al* (Barden *et al.* 2006) due to SNP assay failures. However, 2-SNP haplotype analysis of the haplotype block covering the region to which the associated markers mapped supported the individual marker associations.

Association of P2RX4 and CAMKK2 was not supported, although only one marker relating to each gene was tested. Very recently association of P2RX4 and CAMKK2 with anxiety disorders has been reported (Erhardt *et al.* 2007). The case group consisted of inviduals with anxiety disorders and syndromal panic attacks. Association with illness thus defined was found for one SNP in the 5'UTR of P2RX4. Severity of panic and agrophobic symptoms was associated with rs3817190 mapping to CAMKK2. Further analysis according co-morbid anxiety disorder would therefore seem to be justified.

The initial association of rs2230912 was found in a family based sample, derived from SLSJ pedigrees in which the linkage to 12q24 had originally been found. Hence, replication in a population based case-control sample suggests that initial association results were not confounded by linkage. The consistent association of the G allele of rs2230912 is particularly interesting because this allele alters a highly conserved glutamine residue in an SH-3 domain to arginine. The arginine substitution has been found to lead to a functional decrease in Ca^{2+} influx in transfected recombinant human embryonic kidney cells and in lymphocytes from patients with chronic lymphocytic leukaemia (Cabrini *et al.* 2005). The residue is located in an intracellular domain of P2RX7 and may have effects on P2RX7-mediated signalling (Denlinger *et al.* 2001). However, the G allele occurs in controls in

both the heterozygous and homozygous state. Thus more complex aetiological mechanisms would have to be invoked.

COMT

The results do not generally support the involvement of COMT in the aetiology of BP. The previous association of the low activity methionine allele of the Val158Met functional polymorphism with the disorder was not replicated in the primary analysis. The modest association of BP cases with comorbid panic disorder implies that there may be a subtype of BP disorder in which COMT variation is relevant. This analysis should only be considered exploratory, yet provides a cogent hypothesis that could be tested in other samples.

Risk of BP was not affected by variation at a SNP within the region of the promoter critical for oestrogen-mediated repression of COMT expression. No gender affect was seen with this marker. Furthermore, no association was found between the intron 1 tetranucleotide repeat and BP.

COMT cannot be excluded as an aetiological gene in BP, even in the UCL sample, as much more extensive genotyping of markers spanning the COMT gene and associated promoter regions would be required to capture all the haplotypic diversity.

DAOA/G30

Only very modest evidence to support a role for DAOA was found in the UCL sample. None of the individual marker associations reported by Hattori *et al* (Hattori *et al.* 2003) were replicated. However the results do not constitute a formal non-replication, as two of the individually associated markers were not tested due to assay failures. Consequently, the haplotypic association could not be tested either. Of note, these failed SNPs included the most strongly associated marker in the Hattori *et al* study, rs1935058.

Evidence for the involvement of DAOA came from the SNP rs3918342, (M23), which was associated originally with schizophrenia by Chumakov *et al* (Chumakov *et al.* 2002). The association of this marker with BP replicates the finding of Schumacher *et al* (Schumacher *et al.* 2004). However, the results did not support the hypothesis that DAOA specifically confers risk to a psychotic subtype of BP. The ramifications of association in both BP and schizophrenia will be explored further in the next chapter.

It should be borne in mind that rs3918342 does not actually map to DAOA but is located approximately 40 kb 3' to the gene. According to UCSC annotation the SNP is intergenic. The closest gene is the anti-sense G30 to which rs3918342 lies approximately 30 kb 5'. Although translation of G30 has not been demonstrated, the genetic evidence argues that it should not be overlooked.

Summary

Two novel BP susceptibility loci have been implicated by LD fine mapping of linked regions of chromosomes 12q and 21q. The identification of novel risk loci represents a significant step in attempts to understand the aetiology of BP. The fact that this was derived from a case-control sample implies that the loci are relevant at a population level.

No clear aetiological base pair changes have been identified, although neither locus has yet been comprehensively analysed. Hence, the possibility that the observed associations are false positive results must at least be considered. The most likely source of type I error is multiple testing. Empirical testing of significance negates the need for correction for multiple alleles or multiple haplotypes. However, it is difficult to know how to correct for testing of multiple markers. A simple Bonferroni correction would be overly conservative, as results from different markers are likely to be correlated to some extent because of LD and therefore represent non-independent tests of association.

I would argue for the robustness of the results presented in this thesis on a number of counts. Firstly there was *a priori* probability of susceptibility loci in these regions on the

basis of confirmed linkage. Furthermore, multiple associated markers were found at each locus. Association was not confined to one type of marker, with SNPs and microsatellite markers both yielding association. This assertion is supported by the replication of TRPM2 in an entirely independent sample.

Independent replication is important as it addresses the problem of false positive findings. Generally this has been hard to achieve for genes implicated in BP, but is necessary if research efforts are to be well directed. In this study association of the gene P2RX7 with BP was independently replicated. Hence further investigation of this gene is advocated.

With regard to the DAOA/G30 I would be more cautious in my claims. Since only one marker was modestly associated with BP and no haplotypic association was found, I would suggest that the UCL sample provides some supporting evidence for the involvement of this locus. The inability to reconstruct the originally associated haplotypes is clearly a weakness.

It is here, in the interpretation of negative or equivocal replication results, that the limitations of the current study are demonstrated. Only very circumscribed inferences can be made. For example, COMT could not be excluded as a susceptibly locus for BP. All that can be concluded is that the previously associated SNP, rs4680, showed no association in an outbred population derived from the UK and Ireland. Despite the UCL sample being relatively large, the power to detect association at this marker would not have been high and therefore may represent a type II error. Thus the result should, strictly speaking, be considered preliminary and be added to as more volunteers become available. When this study began it was only realistic to replicate a limited number of markers. However, recent technological advances have revolutionised the field and much more systematic allelic association studies can now be undertaken.

Phenotypic subanalysis yielded some intriguing findings. Most notably, the results suggest that selection of cases by early onset enriches the genetic liability. This not only generates a specific hypothesis to test in relation to future association studies of Slynar/ LOC387895,

TRPM2 and P2RX7 but also has more general repercussions for the design of genetic association studies of BP.

Chapter 10: Discussion

In this final chapter I will briefly comment on some of the broader issues which have been raised in the course of this thesis.

What exactly are we mapping?

The historical paradigm of genetic disease mapping - identification of a gene followed by isolation of aetiological coding changes - has not, to date, revealed unequivocal BP risk alleles. The nature of the variation that we are attempting to map is a fundamental question at both the population and molecular level, with implications for association studies, subsequent mutation screening and functional studies.

From a population genetics perspective, the common-disease/common variant (CDCV) and multiple rare variant (MRV) models have been hypothesised to account for common disorders exhibiting complex inheritance (Pritchard 2001). In relation to association studies, the disease allele frequency is one of the determinants of the observable marker effect size and thus has major implications for the power of a study to map novel loci or replicate previous association (Zondervan and Cardon 2004). For instance, if there are multiple rare variants with only individually small effect sizes, extremely large case-control samples will be required.

Under the CDCV hypothesis invocation of an interactional model would seem to be required given the heterogeneity of linkage results. Gene-environment interaction, gene-gene interaction and allelic interaction at a particular locus have been postulated in other complex conditions. An interaction between childhood maltreatment and MAOA genotype was reported (Caspi *et al.* 2002). Exposure to severe childhood maltreatment and a genotype conferring low MAOA activity greatly elevated the risk of developing both conduct disorder and antisocial personality disorder.

Bardet-Biedl syndrome provides an example of a condition in which gene-gene interaction can be necessary for expression of the clinical phenotype. On the assumption of autosomal recessive inheritance, two genes were identified: BBS2 and BBS6. Subsequently, affected individuals were shown to possess three mutant alleles in a number of pedigrees. In addition, unaffected individuals with two mutant copies of the BBS2 gene were found. Hence a triallelic disease model was proposed (Katsanis *et al.* 2001). It is feasible that more than three alleles are involved and that the number of relevant alleles varies between sets of susceptibility loci. However, the interaction of risk alleles may not necessarily operate across different genes but occur at one particular locus. Potentially, mutation synergy could involve one copy of a homologous gene pair, both copies or operate in a combined manner.

Mutation screening has, to date, focused on exonic changes which would affect amino acid residues and, in turn, protein characteristics. A MRV interpretation would be that this has not yielded convincing aetiological variants. However, from a CDCV perspective one could argue that, for example, the mutant G allele of the P2RX7 exon 13 SNP rs2230912 could be causative. With regard to the novel UCL loci, exon screening has not been fruitful for TRPM2, although only a subset of individuals carrying the associated alleles has been sequenced. It is also now usual to perform sequencing of the promoter region. Again, interpretation of associations with promoter variants, e.g. the 5HTTLPR polymorphism at the 5-HTT gene, is dependent on one's theoretical stance.

An increasingly complex anatomy of genes and their regulatory mechanisms is emerging. Thus, comprehensive mutation screening needs to extend beyond the exons and promoters. Intronic and 3' regulatory sequences are now well-established sites for disease mutations, e.g.(Yokoi *et al.* 1991; Qian *et al.* 2004). Intronic changes may alter splice sites and thus affect mRNA species. The 3'UTR's of some genes contain microRNA recogniton sites.

The issue is, perhaps, even more fudamental than conceptualizing the boundaries of a 'gene'. A radical shift in mindset away from a blinkered search for exon containing genomic sequence is now being indicated. Preliminary results from the ENCODE project indicate that, contrary to previous assumptions, there is in fact very little 'junk DNA' in the

genome. In the 1% of the genome studied, the majority of sequence was found to be transcribed (2007b). Since non-translated RNA species such as microRNAs are known to perform important cellular functions the inference is that a disease mutation could occur at virtually any base.

Recently, large duplications, insertions and deletions have been shown to occur frequently in the genome (Sebat *et al.* 2004). Such copy number variation (CNV) is a plausible mechanism for genetic disease. There is now some evidence that the CNV is relevant to BP and schizophrenia. In a small post-mortem study, CNV at three loci containing genes involved in glutamate signalling was only found in the case group (Wilson *et al.* 2006). CNV may not have previously been detected using standard association and screening techniques.

Association findings can only be considered the starting point to the convincing implication of variant in BP risk. In general, expression and functional studies will be necessary to achieve this goal. There is a very real dilemma as to when research efforts should make the transition from the genome to the transcriptome and proteome. It seems likely that this will be made earlier in an era when the non-synonymous base change is no longer thought of as the 'holy grail' of the gene-hunter.

The future of LD mapping in bipolar disorder

The enthusiasm for LD mapping in BP seems to be ever increasing. As discussed in section 4.5, two GWA studies have recently been published and others are in progress. Indeed the UCL sample forms part of one of these studies in progress. However, there remain serious methodological considerations for both the design and interpretation of association studies in BP.

Study design

The cross-sectional nature of association studies limits the ability to test for gene-environment interactions. However, the prevalence of BP means that truly prospective studies would be unfeasible. Information regarding psychosocial stressors could be more extensively gathered in cross-sectional studies but would be affected by retrospective bias. The recording of accurate phenotypic data is important if multivariate analysis is to be performed. This has not generally been done in loci specific association studies to date. However, on the assumption that at least some genetic subtypes correlate with phenotypic subtypes it would seem a missed opportunity if the GWA results were not subjected to this form of analysis

Without wishing to recapitulate the family based versus population based LD mapping debate, I think a few points need to be stressed, especially as the current large GWA studies have employed a case-control design. In my opinion the problem of population stratification remains a concern. Allele frequency differences between control groups have been observed at markers associated with BP (Bellivier *et al.* 1998b). This highlights the need to ensure that the case and control groups are derived from the same population. The trend towards multicentre studies and the use of shared control groups would seem to increase the risk of cryptic differences. However, this can be addressed by 'genomic control' using ancestrally informative markers. Testing of association using both designs is clearly desirable and this has been carried out at some loci, e.g. P2RX7 (Barden *et al.* 2006). However, obtaining a suitably sized trio sample for GWA represents a financial and logistical challenge.

The overarching aim of testing multiple markers in an implicated region is to identify specific haplotypes which harbour the aetiological mutation(s) and thus refine cases for further study. This is rendered more difficult by unphased case-control data. In the case-control data the significance in the overall distribution of estimated haplotype frequencies was, generally, tested rather than the difference in individual haplotype frequencies.

Association studies in BP have now reached an industrial scale. Although consumable costs, such as SNP genotyping, are concomitantly reducing, large studies require significant research funding. Given finite research resources it is beholden on the research community to strive to achieve mapping in the most efficient way without compromising scientific rigour. For example, pooling strategies have been attempted (Baum *et al.* 2007). While such pooling techniques have not been without their detractors, strategic and practical collaboration in association studies would seem to be dictated.

Power

As alluded to above, the marker effect size will be a function of a number of factors: the effect size of the aetiological variant, the allele frequencies of marker and aetiological variant and the LD relationship between them (Zondervan and Cardon 2004). Since some of these parameters are unknown it cannot be established *a priori* what power is required. In the course of reviewing the association results the lack of adequately powered replication studies became apparent. According to Zondervan *et al*'s simulations the full UCL sample of 600 cases and controls - which represents one of the larger international single centre samples - would still be insufficient under a variety of scenarios. This should caution against over-interpretation of negative replications presented here and elsewhere and argues for continued meta-analysis.

These concerns regarding power have clearly come to the fore in the design of the GWA studies. Large sample sizes have been achieved by pooling samples from multiple centres. Potential sources of heterogeneity between centres, e.g. due to method of recruitment, must be considered with such study designs. However, a coordinated multicentre study should be less affected by heterogeneity than a meta-analysis of individual studies. Certain aspects of the experiment can be standardized, for example genotyping may be performed on a single platform in one designated centre. Nevertheless, even with GWA studies of in excess of 1,000 subjects, meta-analysis of GWA results will be required.

The UCL sample now consists of approximately equal numbers of cases and controls. Unfortunately at the time some markers were genotyped there was an excess of cases. In fact, there is an argument that controls should be in excess. The population of mentally healthy controls is larger than the population of BP cases. Consequently a larger sample is required to accurately represent the control population.

Subjects

Genetic heterogeneity of BP has implications for the optimal population to use for LD fine mapping based on linkage results. Sampling from the same population in which linkage was obtained should maximise the probability of finding association. This issue is not directly relevant to GWA, although variation in results between studies conducted in different populations is to be expected under a heterogeneity model. The likelihood of inconsistent results is further increased by the possible lack of sensitivity of case-control studies to allelic heterogeneity.

Case definition remains a critical issue and is intimately related to study power. The identification of endophenotypes, representing more genetically homogenous subtypes of BP, continues to be pursued by investigators. This subject has recently been reviewed (MacQueen *et al.* 2005). Subsequently, the familiality of phenotypic features in 172 multiplex BP pedigrees has been studied using mixed-regression analysis (Schulze *et al.* 2006). DSM criteria continue to prevail in ascertaining BP 'caseness' for association studies. However, *post hoc* analysis using putative endophenotypes is becoming common practice. Such analysis in the UCL dataset yielded increased significance of association at four loci, most notably with respect to early onset disorder. These results could be interpreted as evidence for an early onset subtype or that the genetic liability is increased if the case group is defined by early onset. The latter interpretation would be consistent with the age of onset effect seen in both segregation and linkage analyses. While it may be parsimonious to continue to ascertain cases by DSM criteria, I believe there is a compelling rationale to test specific hypotheses based on putative endophenotypes in the primary analysis of future studies.

Recently the argument that BP and schizophrenia share some genetic liability has gained momentum. Proponents of 'common susceptibility loci' have cited family, twin, linkage and association findings as evidence (Craddock and Owen 2005). In my view the association findings provide the most persuasive evidence that there are at least some loci where variation confers susceptibility to both disorders or genetically define a clinical entity which crosses traditional diagnostic boundaries. In medical genetics there is certainly a precedent for different mutations at a given gene to give rise to differing clinical phenotypes, e.g. the dystrophin gene (Muntoni *et al.* 2003).Yet it is also plausible that the same mutations have pleitropic effects.

Of the loci tested in the UCL sample, DISC1, COMT and DAOA have been implicated in the aetiology of schizophrenia by association studies. In addition, the genes neuregulin 1 and dysbidin which were originally implicated in schizophrenia have subsequently be associated with BP (Breen *et al.* 2006; Blackwood *et al.* 2007). Ideally these loci should therefore be tested in both BP and schizophrenia samples. However, this has obvious resource implications. The alternative approach of recruiting both individuals with schizophrenia and BP to a combined case group does not follow from these findings.

Markers

A study's power to detect association is dependent on marker selection. Two properties of the marker influence power - LD with the aetiological mutation and the marker allele frequency. Unfortunately it is not possible to optimise these properties *a priori.*

SNP markers are now used in the vast majority of association studies. This in itself confers some benefit, as it facilitates comparison and meta-analysis of studies. Generally SNPs with a MAF>10% are selected. However, the optimal SNP MAF for detecting association is unknown. This will depend on the allele frequency of the aetiological variant with which it is in LD. The allele frequencies of aetiological variants are likely to vary between loci and possibly at a given locus. In some scenarios a very large case-control sample - of more than

1,000 cases and 1,000 controls - would be required to detect association (Zondervan and Cardon 2004).

The HapMap project is progressively elucidating the LD relationships between SNPs in different human populations. However, the process is not complete and variants aetiological to BP may not be database SNPs. Increasingly haplotype 'tagging' SNPs have been selected for systematic LD fine mapping. Implicit in this approach is the assumption that non-database aetiological variants will conform with the LD relationships of SNPs within a given LD block. While this may not be an entirely safe assumption, haplotype tagging represents the best available method for optimising marker/unknown aetiological variant LD in a systematic way.

The HapMap project also has a bearing on replication studies. Traditionally, replication has referred to the testing of an associated marker or markers. This has been the general approach taken in the thesis, so as to test simple hypotheses and minimise multiple testing. However, some investigators have started to consider the gene as the unit of replication, and test a tagging SNP set designed to screen the gene. Such an approach may mitigate the effect of locus heterogeneity but does normally increase the number of tests carried out. Furthermore, it may create problems comparing results between centres as LD patterns and therefore the tagging SNP set may vary between samples. Care in the interpretation of results is needed with both approaches.

No such database of LD relationships between microsatellite markers exists. Consequently, they must be derived empirically in each sample. However, LD mapping using microsatellite markers did prove successful in the UCL sample with reference to the chromosome 12q and 21q loci. In neither case could it be claimed that the linkage regions were systematically mapped according to measures of LD. It has been suggested that microsatellite markers are more informative than SNP markers (Varilo *et al.* 2003). It may be that multiallelic microsatellite markers are less likely than biallelic markers to give false negative results when aetiological variants occur on different haplotypic backgrounds. With a bialleic marker an aetiological variant may be in LD with a particular allele in some

individuals and with the opposite allele in others. This would diminish any observable differences between the case and control groups. With multiallelic markers the probability of this cancelling out of association is reduced. It was interesting that the strongest evidence for association was generally obtained with microsatellite markers.

This thesis did, however, highlight some of the disadvantages of using microsatellite markers. Established highly polymorphic microsatellite markers may have cryptic complexity which renders interpretation of allele frequencies between case-control groups difficult. Using the microsatellite genotyping methodology employed in this thesis, PCR fragment size differences of 1 bp cannot be reliably resolved. I would suggest that this may account for some of the allele frequency variation between populations reported for the COMT microsatellite (http://alfred.med.yale.edu/). Consequently it is recommended that only perfect repeats are tested.

It is clear that the gene chip technology currently being employed for GWA studies is elevating the examination of the genome for variation aetiological to BP to a new level. However, the combination of uncertainties over the amount of haplotypic diversity being captured, power of the studies to detect association and nature of variation being sought argues against such studies being perceived as definitive.

Analysis

There is a dilemma as to the extent to which association data should be analyzed. Statistical rigour argues for the minimisation of multiple testing. However, there appears to be a danger of overlooking covert associations which have been masked by overinclusiveness of the diagnostic criteria for BP or complex mechanisms of inheritance.

In this study I chose to test a limited set of putative endophenotypes for which there was *a priori* evidence in the literature. No claim is made that this was exhaustive as some legitimate endophenotype, e.g. lithium responsiveness, were not readily available from the database for all individuals. Alternatively, association found in the primary analysis could

have been explored by multivariate analysis. Certainly application of these statistics to the GWA data may help to define genetic subtypes.

No analysis of genetic interaction was attempted in this thesis. Interaction between genes, for example belonging to a neurotransmitter system, or even interaction between alleles at a given gene is highly plausible. Recent evidence from GWA data suggests an interaction between the established late onset Alzheimer's susceptibility allele APOE4 and novel loci GAB2 (Reiman *et al.* 2007). It could be argued that interaction analysis should be deferred until definite aetiological variants have been identified, but this is certainly of great interest.

A case-control monopoly?

One could question the relative cost-effectiveness of large scale GWA studies as compared to sequence-based approaches targeted at genomic regions implicated by linkage and/or association studies. Given the uncertainties alluded to above direct sequencing is attractive in its simplicity.

It may be that genetic susceptibility to BP will be difficult to interpret without investigation of gene environment interactions. While technically possible within a LD mapping framework, employing prospective cohort designs would represent massive practical challenges. However, there are alternatives to this.

One classical approach to studying environmental effects has employed MZ twins discordant for the trait of interest. Recently a molecular perspective has been brought to bare on this paradigm (Kato *et al.* 2005). Using micoarrays, gene expression was assayed in the discordant twins. While differences in expression cannot be definitively attributed to the specific environment - as, for example, lyonisation in female twins could cause differential expression - they could give a strong indication of the molecular sequelae of environmental effects. The integration of human data with expression studies in animals, where the environment can be manipulated in a more controlled manner, is another possible avenue for investigation.

Final remarks

It is disappointing that despite the great investment in genetic association studies no clinical benefit for patients suffering from BP has yet been realised. The identification of BP susceptibility genes continues to challenge our understanding of the mechanisms of inheritance operating in complex disorders and the nature of sequence variation that can result in disease. More positively, a number of novel therapeutic targets have been identified and drug development has begun. Furthermore, the coalescence of advances in SNP genotyping technology, expanding case-control samples and multicentre collaboration is yielding unprecedented power to examine BP susceptibility with genetic association studies.

References

(1997). "No association between bipolar disorder and alleles at a functional polymorphism in the COMT gene. Biomed European Bipolar Collaborative Group." Br J Psychiatry **170**: 526-8.

(2003). "The International HapMap Project." Nature **426**(6968): 789-96.

(2007a). "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls." Nature **447**(7145): 661-78.

(2007b). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." Nature **447**(7146): 799-816.

Abkevich, V., N. J. Camp, et al. (2003). "Predisposition locus for major depression at chromosome 12q22-12q23.2." Am J Hum Genet **73**(6): 1271-81.

Ahearn, E. P., M. C. Speer, et al. (2002). "Investigation of Notch3 as a candidate gene for bipolar disorder using brain hyperintensities as an endophenotype." Am J Med Genet **114**(6): 652-8.

Aita, V. M., J. Liu, et al. (1999). "A comprehensive linkage analysis of chromosome 21q22 supports prior evidence for a putative bipolar affective disorder locus." Am J Hum Genet **64**(1): 210-7.

Albanese, V., N. F. Biguet, et al. (2001). "Quantitative effects on gene silencing by allelic variation at a tetranucleotide microsatellite." Hum Mol Genet **10**(17): 1785-92.

Althoff, R. R., S. V. Faraone, et al. (2005). "Family, twin, adoption, and molecular genetic studies of juvenile bipolar disorder." Bipolar Disord **7**(6): 598-609.

Angst, J. and A. Marneros (2001). "Bipolarity from ancient to modern times: conception, birth and rebirth." J Affect Disord **67**(1-3): 3-19.

Angst, J. and R. Sellaro (2000). "Historical perspectives and natural history of bipolar disorder." Biol Psychiatry **48**(6): 445-57.

Anguelova, M., C. Benkelfat, et al. (2003). "A systematic review of association studies investigating genes coding for serotonin receptors and the serotonin transporter: I. Affective disorders." Mol Psychiatry **8**(6): 574-91.

Antonarakis, S. E. and J. S. Beckmann (2006). "Mendelian disorders deserve more attention." Nat Rev Genet **7**(4): 277-82.

APA (1994). Diagnostic and Statistical Manual of Mental Disorders DSM-IV, American Psychiatric Association; 4th edition (January 15, 1994).

Badenhop, R. F., M. J. Moses, et al. (2003). "Genetic refinement and physical mapping of a 2.3 Mb probable disease region associated with a bipolar affective disorder susceptibility locus on chromosome 4q35." Am J Med Genet B Neuropsychiatr Genet **117**(1): 23-32.

Badenhop, R. F., M. J. Moses, et al. (2002). "A genome screen of 13 bipolar affective disorder pedigrees provides evidence for susceptibility loci on chromosome 3 as well as chromosomes 9, 13 and 19." Mol Psychiatry **7**(8): 851-9.

Badner, J. A. and E. S. Gershon (2002a). "Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia." Mol Psychiatry **7**(4): 405-11.

Badner, J. A. and E. S. Gershon (2002b). "Regional meta-analysis of published data supports linkage of autism with markers on chromosome 7." Mol Psychiatry **7**(1): 56-66.

Barden, N., M. Harvey, et al. (2006). "Analysis of single nucleotide polymorphisms in genes in the chromosome 12Q24.31 region points to P2RX7 as a susceptibility gene to bipolar affective disorder." Am J Med Genet B Neuropsychiatr Genet **141**(4): 374-82.

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

Baum, A. E., N. Akula, et al. (2007). "A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder." Mol Psychiatry.

Baum, M., L. Ortiz, et al. (2003). "Fetal origins of cardiovascular disease." Curr Opin Pediatr **15**(2): 166-70.

Beck, A. T., C. H. Ward, et al. (1962). "Reliability of psychiatric diagnosis. 2. A study of consistency of clinical judgments and ratings." Am J Psychiatry **119**: 351-7.

Bell, J. T., C. Wallace, et al. (2006). "Two-dimensional genome-scan identifies novel epistatic loci for essential hypertension." Hum Mol Genet **15**(8): 1365-74.

Bellivier, F., M. Leboyer, et al. (1998a). "Association between the tryptophan hydroxylase gene and manic-depressive illness." Arch Gen Psychiatry **55**(1): 33-7.

Bellivier, F., F. Schurhoff, et al. (1998b). "Methodological problems in meta-analysis of association studies between bipolar affective disorders and the tyrosine hydroxylase gene." Am J Med Genet **81**(4): 349-52.

Berrios, G. and R. Porter (1995). A History of Clinical Psychiatry: The Origin and History of Psychiatric Disorders. London, The Athlone Press.

Bertelsen, A., B. Harvald, et al. (1977). "A Danish twin study of manic-depressive disorders." Br J Psychiatry **130**: 330-51.

Blackwood, D. H., A. Fordyce, et al. (2001). "Schizophrenia and affective disorders--cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family." Am J Hum Genet **69**(2): 428-33.

Blackwood, D. H., B. J. Pickard, et al. (2007). "Are some genetic risk factors common to schizophrenia, bipolar disorder and depression? Evidence from DISC1, GRIK4 and NRG1." Neurotox Res **11**(1): 73-83.

Bocchetta, A., M. P. Piccardi, et al. (1999). "Family-based association study between bipolar disorder and DRD2, DRD4, DAT, and SERT in Sardinia." Am J Med Genet **88**(5): 522-6.

Bohm, D., K. Hoffmann, et al. (2006). "Association of Jacobsen syndrome and bipolar affective disorder in a patient with a de novo 11q terminal deletion." Am J Med Genet A **140**(4): 378-82.

Botstein, D. and N. Risch (2003). "Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease." Nat Genet **33 Suppl**: 228-37.

Botstein, D., R. L. White, et al. (1980). "Construction of a genetic linkage map in man using restriction fragment length polymorphisms." Am J Hum Genet **32**(3): 314-31.

Breen, G. (2002). "Novel and alternate SNP and genetic technologies." Psychiatr Genet **12**(2): 83-8.

Breen, G., D. Prata, et al. (2006). "Association of the dysbindin gene with bipolar affective disorder." Am J Psychiatry **163**(9): 1636-8.

Brown, S. and N. Bass (2004). "The psychiatric intensive care unit (PICU): Patient characteristics, treatment and outcome." Journal of mental health **13**(6): 601-609.
Brunn, T. G. and H. Ewald (1999). "Selection bias of susceptibility genes possible when using parent-offspring trios in genetic association studies." Mol Psychiatry **4**(5): 415-6.
Brunner, H. G., M. Nelen, et al. (1993). "Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A." Science **262**(5133): 578-80.
Bucher, K. D., R. C. Elston, et al. (1981). "The transmission of manic depressive illness--II. Segregation analysis of three sets of family data." J Psychiatr Res **16**(1): 65-78.
Burdick, K. E., B. Funke, et al. (2007). "COMT genotype increases risk for bipolar I disorder and influences neurocognitive performance." Bipolar Disord **9**(4): 370-6.
Cabrini, G., S. Falzoni, et al. (2005). "A His-155 to Tyr polymorphism confers gain-of-function to the human P2X7 receptor of human leukemic lymphocytes." J Immunol **175**(1): 82-9.
Carlson, C., D. Papolos, et al. (1997). "Molecular analysis of velo-cardio-facial syndrome patients with psychiatric disorders." Am J Hum Genet **60**(4): 851-9.
Cases, O., I. Seif, et al. (1995). "Aggressive behavior and altered amounts of brain serotonin and norepinephrine in mice lacking MAOA." Science **268**(5218): 1763-6.
Caspi, A., J. McClay, et al. (2002). "Role of genotype in the cycle of violence in maltreated children." Science **297**(5582): 851-4.
Cassidy, F., C. Zhao, et al. (2007). "Genome-wide scan of bipolar disorder and investigation of population stratification effects on linkage: Support for susceptibility loci at 4q21, 7q36, 9p21, 12q24, 14q24, and 16p13." Am J Med Genet B Neuropsychiatr Genet.
Chen, Y. S., N. Akula, et al. (2004). "Findings in an independent sample support an association between bipolar affective disorder and the G72/G30 locus on chromosome 13q33." Mol Psychiatry **9**(1): 87-92; image 5.
Cheng, R., S. H. Juo, et al. (2006). "Genome-wide linkage scan in a large bipolar disorder sample from the National Institute of Mental Health genetics initiative suggests putative loci for bipolar disorder, psychosis, suicide, and panic disorder." Mol Psychiatry **11**(3): 252-60.
Cho, H. J., I. Meira-Lima, et al. (2005). "Population-based and family-based studies on the serotonin transporter gene polymorphisms and bipolar disorder: a systematic review and meta-analysis." Mol Psychiatry **10**(8): 771-81.
Chumakov, I., M. Blumenfeld, et al. (2002). "Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia." Proc Natl Acad Sci U S A **99**(21): 13675-80.
Cichon, S., J. Schumacher, et al. (2001). "A genome screen for genes predisposing to bipolar affective disorder detects a new susceptibility locus on 8q." Hum Mol Genet **10**(25): 2933-44.
Clerget-Darpoux, F., C. Bonaiti-Pellie, et al. (1986). "Effects of misspecifying genetic parameters in lod score analysis." Biometrics **42**(2): 393-9.
Collier, D. A., M. J. Arranz, et al. (1996a). "The serotonin transporter is a potential susceptibility factor for bipolar affective disorder." Neuroreport **7**(10): 1675-9.

Collier, D. A., G. Stober, et al. (1996b). "A novel functional polymorphism within the promoter of the serotonin transporter gene: possible role in susceptibility to affective disorders." Mol Psychiatry **1**(6): 453-60.
Collins, A., C. Lonjou, et al. (1999). "Genetic epidemiology of single-nucleotide polymorphisms." Proc Natl Acad Sci U S A **96**(26): 15173-7.
Coyle, N., I. Jones, et al. (2000). "Variation at the serotonin transporter gene influences susceptibility to bipolar affective puerperal psychosis." Lancet **356**(9240): 1490-1.
Craddock, N., S. Dave, et al. (2001). "Association studies of bipolar disorder." Bipolar Disord **3**(6): 284-98.
Craddock, N. and I. Jones (1999). "Genetics of bipolar disorder." J Med Genet **36**(8): 585-94.
Craddock, N., V. Khodel, et al. (1995). "Mathematical limits of multilocus models: the genetic transmission of bipolar disorder." Am J Hum Genet **57**(3): 690-702.
Craddock, N. and M. Owen (1994). "Chromosomal aberrations and bipolar affective disorder." Br J Psychiatry **164**(4): 507-12.
Craddock, N., M. Owen, et al. (1994). "Familial cosegregation of major affective disorder and Darier's disease (keratosis follicularis)." Br J Psychiatry **164**(3): 355-8.
Craddock, N. and M. J. Owen (2005). "The beginning of the end for the Kraepelinian dichotomy." Br J Psychiatry **186**: 364-6.
Craddock, N., P. Van Eerdewegh, et al. (1997). "Single major locus models for bipolar disorder are implausible." Am J Med Genet **74**(1): 18-20.
Curtis, D., G. Kalsi, 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): 77-84.
Curtis, D. and P. C. Sham (1995). "Model-free linkage analysis using likelihoods." Am J Hum Genet **57**(3): 703-16.
Dalton, E. J., T. D. Cate-Carter, et al. (2003). "Suicide risk in bipolar patients: the role of co-morbid substance use disorders." Bipolar Disord **5**(1): 58-61.
Daly, M. J., J. D. Rioux, et al. (2001). "High-resolution haplotype structure in the human genome." Nat Genet **29**(2): 229-32.
Das Gupta, R. and J. F. Guest (2002). "Annual cost of bipolar disorder to UK society." Br J Psychiatry **180**: 227-33.
Dawson, E., M. Gill, et al. (1995). "Genetic association between alleles of pancreatic phospholipase A2 gene and bipolar affective disorder." Psychiatr Genet **5**(4): 177-80.
Deckert, J., M. M. Nothen, et al. (1998). "Adenosine A1 receptor and bipolar affective disorder: systematic screening of the gene and association studies." Am J Med Genet **81**(1): 18-23.
Degn, B., M. D. Lundorf, et al. (2001). "Further evidence for a bipolar risk gene on chromosome 12q24 suggested by investigation of haplotype sharing and allelic association in patients from the Faroe Islands." Mol Psychiatry **6**(4): 450-5.
Del Zompo, M., V. De Luca, et al. (2007). "Haplotype association study between DRD1 gene and bipolar type I affective disorder in two samples from Canada and Sardinia." Am J Med Genet B Neuropsychiatr Genet **144**(2): 237-41.

DeMille, M. M., J. R. Kidd, et al. (2002). "Population variation in linkage disequilibrium across the COMT gene considering promoter region and coding region variation." Hum Genet **111**(6): 521-37.

Denlinger, L. C., P. L. Fisette, et al. (2001). "Cutting edge: the nucleotide receptor P2X7 contains multiple protein- and lipid-interaction motifs including a potential binding site for bacterial lipopolysaccharide." J Immunol **167**(4): 1871-6.

Derby, I. (1933). "Manic-depressive "Exhaustion" deaths: An Analysis of "Exhaustion" case histories." Psychiatric Quarterly **7**(3): 436-449.

Detera-Wadleigh, S. D., J. A. Badner, et al. (1999). "A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11.2." Proc Natl Acad Sci U S A **96**(10): 5604-9.

Detera-Wadleigh, S. D., J. A. Badner, et al. (1996). "Affected-sib-pair analyses reveal support of prior evidence for a susceptibility locus for bipolar disorder, on 21q." Am J Hum Genet **58**(6): 1279-85.

Detera-Wadleigh, S. D. and F. J. McMahon (2006). "G72/G30 in schizophrenia and bipolar disorder: review and meta-analysis." Biol Psychiatry **60**(2): 106-14.

Devon, R. S., S. Anderson, 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): 71-8.

Di Bella, D., M. Catalano, et al. (1996). "Association study of a null mutation in the dopamine D4 receptor gene in Italian patients with obsessive-compulsive disorder, bipolar mood disorder and schizophrenia." Psychiatr Genet **6**(3): 119-21.

Donahue, R. P., W. B. Bias, et al. (1968). "Probable assignment of the Duffy blood group locus to chromosome 1 in man." Proc Natl Acad Sci U S A **61**(3): 949-55.

Dupont, B., S. E. Oberfield, et al. (1977). "Close genetic linkage between HLA and congenital adrenal hyperplasia (21-hydroxylase deficiency)." Lancet **2**(8052-8053): 1309-12.

Edenberg, H. J., T. Foroud, et al. (1997). "Initial genomic scan of the NIMH genetics initiative bipolar pedigrees: chromosomes 3, 5, 15, 16, 17, and 22." Am J Med Genet **74**(3): 238-46.

Egan, M. F., M. Kojima, et al. (2003). "The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function." Cell **112**(2): 257-69.

Egeland, J. A., D. S. Gerhard, et al. (1987). "Bipolar affective disorders linked to DNA markers on chromosome 11." Nature **325**(6107): 783-7.

Endicott, J. and R. L. Spitzer (1978). "A diagnostic interview: the schedule for affective disorders and schizophrenia." Arch Gen Psychiatry **35**(7): 837-44.

Erhardt, A., S. Lucae, et al. (2007). "Association of polymorphisms in P2RX7 and CaMKKb with anxiety disorders." J Affect Disord **101**(1-3): 159-68.

Etain, B., F. Mathieu, et al. (2006). "Genome-wide scan for genes involved in bipolar affective disorder in 70 European families ascertained through a bipolar type I early-onset proband: supportive evidence for linkage at 3p14." Mol Psychiatry **11**(7): 685-94.

Ewald, H., B. Degn, et al. (1998). "Significant linkage between bipolar affective disorder and chromosome 12q24." Psychiatr Genet **8**(3): 131-40.

Ewald, H., T. Flint, et al. (2002). "A genome-wide scan shows significant linkage between bipolar disorder and chromosome 12q24.3 and suggestive linkage to chromosomes 1p22-21, 4p16, 6q14-22, 10q26 and 16p13.3." Mol Psychiatry **7**(7): 734-44.

Ewald, H., T. A. Kruse, et al. (2003). "Genome wide scan using homozygosity mapping and linkage analyses of a single pedigree with affective disorder suggests oligogenic inheritance." Am J Med Genet B Neuropsychiatr Genet **120**(1): 63-71.

Ewald, H., F. P. Wikman, 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): 25-30.

Falconer, D. S. (1965). "The inheritance of liability to certain diseases, estimated from the incidence among relatives." Ann. Hum. Genet. **29**: 51-76.

Falk, C. T. and P. Rubinstein (1987). "Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations." Ann Hum Genet **51**(Pt 3): 227-33.

Fallin, M. D., V. K. Lasseter, et al. (2005). "Bipolar I disorder and schizophrenia: a 440-single-nucleotide polymorphism screen of 64 candidate genes among Ashkenazi Jewish case-parent trios." Am J Hum Genet **77**(6): 918-36.

Faraone, S. V., W. S. Kremen, et al. (1990). "Genetic transmission of major affective disorders: quantitative models and linkage analyses." Psychol Bull **108**(1): 109-27.

Faraone, S. V., J. Lasky-Su, et al. (2006). "Early onset bipolar disorder: possible linkage to chromosome 9q34." Bipolar Disord **8**(2): 144-51.

Farmer, A., P. Cosyns, et al. (1993). "A SCAN-SADS comparison study of psychotic subjects and their first-degree relatives." Eur Arch Psychiatry Clin Neurosci **242**(6): 352-6.

Feighner, J. P., E. Robins, et al. (1972). "Diagnostic criteria for use in psychiatric research." Arch Gen Psychiatry **26**(1): 57-63.

Ferraren, D. O., C. Liu, et al. (2005). "Linkage disequilibrium analysis in the LOC93081-KDELC1-BIVM region on 13q in bipolar disorder." Am J Med Genet B Neuropsychiatr Genet **133**(1): 12-7.

Floderus, Y., S. B. Ross, et al. (1981). "Erythrocyte catechol-O-methyltransferase activity in a Swedish population." Clin Genet **19**(5): 389-92.

Freeman, M. P. and S. A. Freeman (2006). "Lithium: clinical considerations in internal medicine." Am J Med **119**(6): 478-81.

Funke, B., A. K. Malhotra, et al. (2005). "COMT genetic variation confers risk for psychotic and affective disorders: a case control study." Behav Brain Funct **1**: 19.

Furlong, R. A., L. Ho, et al. (1998a). "No association of the tryptophan hydroxylase gene with bipolar affective disorder, unipolar affective disorder, or suicidal behaviour in major affective disorder." Am J Med Genet **81**(3): 245-7.

Furlong, R. A., L. Ho, et al. (1999a). "Analysis of the monoamine oxidase A (MAOA) gene in bipolar affective disorder by association studies, meta-analyses, and sequencing of the promoter." Am J Med Genet **88**(4): 398-406.

Furlong, R. A., L. Ho, et al. (1998b). "Analysis and meta-analysis of two serotonin transporter gene polymorphisms in bipolar and unipolar affective disorders." Am J Med Genet **81**(1): 58-63.

Furlong, R. A., J. S. Rubinsztein, et al. (1999b). "Analysis and metaanalysis of two polymorphisms within the tyrosine hydroxylase gene in bipolar and unipolar affective disorders." Am J Med Genet **88**(1): 88-94.
Gejman, P. V., M. Martinez, et al. (1993). "Linkage analysis of fifty-seven microsatellite loci to bipolar disorder." Neuropsychopharmacology **9**(1): 31-40.
Gelder, M., R. Mayou, et al. (2001). Shorter Oxford Textbook of Psychiatry. Oxford, Oxford University Press.
Gelehrter, T. D. and F. S. Collins (1990). Principles of Medical Genetics. Baltimore, Williams & Wilkins.
Geller, B., J. A. Badner, et al. (2004). "Linkage disequilibrium of the brain-derived neurotrophic factor Val66Met polymorphism in children with a prepubertal and early adolescent bipolar disorder phenotype." Am J Psychiatry **161**(9): 1698-700.
Geller, B. and E. H. Cook, Jr. (2000). "Ultradian rapid cycling in prepubertal and early adolescent bipolarity is not in transmission disequilibrium with val/met COMT alleles." Biol Psychiatry **47**(7): 605-9.
Gershon, E. S., J. Hamovit, et al. (1982). "A family study of schizoaffective, bipolar I, bipolar II, unipolar, and normal control probands." Arch Gen Psychiatry **39**(10): 1157-67.
Glaser, B., G. Kirov, et al. (2005a). "Identification of a potential bipolar risk haplotype in the gene encoding the winged-helix transcription factor RFX4." Mol Psychiatry **10**(10): 920-7.
Glaser, B., G. Kirov, et al. (2005b). "Linkage disequilibrium mapping of bipolar affective disorder at 12q23-q24 provides evidence for association at CUX2 and FLJ32356." Am J Med Genet B Neuropsychiatr Genet **132**(1): 38-45.
Goes, F. S., P. P. Zandi, et al. (2007). "Mood-incongruent psychotic features in bipolar disorder: familial aggregation and suggestive linkage to 2p11-q14 and 13q21-33." Am J Psychiatry **164**(2): 236-47.
Goldin, L. R., E. S. Gershon, et al. (1983). "Segregation and linkage analyses in families of patients with bipolar, unipolar, and schizoaffective mood disorders." Am J Hum Genet **35**(2): 274-87.
Goldman, D., G. L. Brown, et al. (1993). "DRD2 dopamine receptor genotype, linkage disequilibrium, and alcoholism in American Indians and other populations." Alcohol Clin Exp Res **17**(2): 199-204.
Goldstein, D. B. and M. E. Weale (2001). "Population genomics: linkage disequilibrium holds the key." Curr Biol **11**(14): R576-9.
Gonzalez-Pinto, A., J. van Os, et al. (2003). "Age-dependence of Schneiderian psychotic symptoms in bipolar patients." Schizophr Res **61**(2-3): 157-62.
Goodwin, F. and K. Jamison (1990). Manic-depressive illness. Oxford, Oxford University Press.
Gourovitch, M. L., E. F. Torrey, et al. (1999). "Neuropsychological performance of monozygotic twins discordant for bipolar disorder." Biol Psychiatry **45**(5): 639-46.
Green, E., G. Elvidge, et al. (2005). "Localization of bipolar susceptibility locus by molecular genetic analysis of the chromosome 12q23-q24 region in two pedigrees with bipolar disorder and Darier's disease." Am J Psychiatry **162**(1): 35-42.

Green, E. K., R. Raybould, et al. (2006). "Genetic variation of brain-derived neurotrophic factor (BDNF) in bipolar disorder: case-control study of over 3000 individuals from the UK." Br J Psychiatry **188**: 21-5.

Grigoroiu-Serbanescu, M., M. Martinez, et al. (2001). "Different familial transmission patterns in bipolar I disorder with onset before and after age 25." Am J Med Genet **105**(8): 765-73.

Gurling, H. M., G. Kalsi, et al. (2001). "Genomewide genetic linkage analysis confirms the presence of susceptibility loci for schizophrenia, on chromosomes 1q32.2, 5q33.2, and 8p21-22 and provides support for linkage to schizophrenia, on chromosomes 11q23.3-24 and 20q12.1-11.23." Am J Hum Genet **68**(3): 661-73.

Gutierrez, B., J. Bertranpetit, et al. (1997). "Association analysis of the catechol O-methyltransferase gene and bipolar affective disorder." Am J Psychiatry **154**(1): 113-5.

Hageman, G. S., D. H. Anderson, et al. (2005). "A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration." Proc Natl Acad Sci U S A **102**(20): 7227-32.

Hall, J. M., M. K. Lee, et al. (1990). "Linkage of early-onset familial breast cancer to chromosome 17q21." Science **250**(4988): 1684-9.

Hamilton, S. P., A. J. Fyer, et al. (2003). "Further genetic evidence for a panic disorder syndrome mapping to chromosome 13q." Proc Natl Acad Sci U S A **100**(5): 2550-5.

Hamshere, M. L., P. Bennett, et al. (2005). "Genomewide linkage scan in schizoaffective disorder: significant evidence for linkage at 1q42 close to DISC1, and suggestive evidence at 22q11 and 19p13." Arch Gen Psychiatry **62**(10): 1081-8.

Harris, E. C. and B. Barraclough (1997). "Suicide as an outcome for mental disorders. A meta-analysis." Br J Psychiatry **170**: 205-28.

Hashimoto, K., E. Shimizu, et al. (2004). "Critical role of brain-derived neurotrophic factor in mood disorders." Brain Res Brain Res Rev **45**(2): 104-14.

Hashimoto, R., T. Numakawa, et al. (2006). "Impact of the DISC1 Ser704Cys polymorphism on risk for major depression, brain morphology and ERK signaling." Hum Mol Genet **15**(20): 3024-33.

Hattori, E., C. Liu, et al. (2003). "Polymorphisms at the G72/G30 gene locus, on 13q33, are associated with bipolar disorder in two independent pedigree series." Am J Hum Genet **72**(5): 1131-40.

Hawkes, C. H. (1997). "Twin studies in medicine--what do they tell us?" Qjm **90**(5): 311-21.

Helzer, J. E., L. N. Robins, et al. (1977). "Reliability of psychiatric diagnosis. I. A methodological review." Arch Gen Psychiatry **34**(2): 129-33.

Hennah, W., T. Varilo, et al. (2003). "Haplotype transmission analysis provides evidence of association for DISC1 to schizophrenia and suggests sex-dependent effects." Hum Mol Genet **12**(23): 3151-9.

Henry, C., D. Van den Bulke, et al. (2003). "Anxiety disorders in 318 bipolar patients: prevalence and impact on illness severity and response to mood stabilizer." J Clin Psychiatry **64**(3): 331-5.

Herzberg, I., A. Jasinska, et al. (2006). "Convergent linkage evidence from two Latin-American population isolates supports the presence of a susceptibility locus for bipolar disorder in 5q31-34." Hum Mol Genet **15**(21): 3146-53.

Hesselbrock, V., J. Stabenau, et al. (1982). "A comparison of two interview schedules. The Schedule for Affective Disorders and Schizophrenia-Lifetime and the National Institute for Mental Health Diagnostic Interview Schedule." Arch Gen Psychiatry **39**(6): 674-7.

Heun, R. and W. Maier (1993). "The distinction of bipolar II disorder from bipolar I and recurrent unipolar depression: results of a controlled family study." Acta Psychiatr Scand **87**(4): 279-84.

Hodgkinson, C. A., D. Goldman, et al. (2004). "Disrupted in schizophrenia 1 (DISC1): association with schizophrenia, schizoaffective disorder, and bipolar disorder." Am J Hum Genet **75**(5): 862-72.

Hodgkinson, S., R. Sherrington, et al. (1987). "Molecular genetic evidence for heterogeneity in manic depression." Nature **325**(6107): 805-6.

Hong, C. J., S. J. Huo, et al. (2003). "Association study of a brain-derived neurotrophic-factor genetic polymorphism and mood disorders, age of onset and suicidal behavior." Neuropsychobiology **48**(4): 186-9.

Hotamisligil, G. S. and X. O. Breakefield (1991). "Human monoamine oxidase A gene determines levels of enzyme activity." Am J Hum Genet **49**(2): 383-92.

Ioannidis, J. P., E. E. Ntzani, et al. (2001). "Replication validity of genetic association studies." Nat Genet **29**(3): 306-9.

Jacobsen, N., J. Daniels, et al. (1996). "Association study of bipolar disorder at the phospholipase A2 gene (PLA2A) in the Darier's disease (DAR) region of chromosome 12q23-q24.1." Psychiatr Genet **6**(4): 195-9.

Jones, I., M. Hamshere, et al. (2007). "Bipolar affective puerperal psychosis: genome-wide significant evidence for linkage to chromosome 16." Am J Psychiatry **164**(7): 1099-104.

Jones, I., N. Jacobsen, et al. (2002). "Evidence for familial cosegregation of major affective disorder and genetic markers flanking the gene for Darier's disease." Mol Psychiatry **7**(4): 424-7.

Jorde, L. B. (2000). "Linkage disequilibrium and the search for complex disease genes." Genome Res **10**(10): 1435-44.

Jorgensen, T. H., A. D. Borglum, et al. (2002). "Search for common haplotypes on chromosome 22q in patients with schizophrenia or bipolar disorder from the Faroe Islands." Am J Med Genet **114**(2): 245-52.

Kakiuchi, C., K. Iwamoto, et al. (2003). "Impaired feedback regulation of XBP1 as a genetic risk factor for bipolar disorder." Nat Genet **35**(2): 171-5.

Kalsi, G., A. McQuillin, et al. (2006). "Identification of the Slynar gene (AY070435) and related brain expressed sequences as a candidate gene for susceptibility to affective disorders through allelic and haplotypic association with bipolar disorder on chromosome 12q24." Am J Psychiatry **163**(10): 1767-76.

Kamiya, A., T. Tomoda, et al. (2006). "DISC1-NDEL1/NUDEL protein interaction, an essential component for neurite outgrowth, is modulated by genetic variations of DISC1." Hum Mol Genet **15**(22): 3313-23.

Kanazawa, T., S. J. Glatt, et al. (2007). "Meta-analysis reveals no association of the Val66Met polymorphism of brain-derived neurotrophic factor with either schizophrenia or bipolar disorder." Psychiatr Genet **17**(3): 165-70.

Kaneva, R. P., V. M. Chorbov, et al. (2004). "Linkage analysis in bipolar pedigrees adds support for a susceptibility locus on 21q22." Psychiatr Genet **14**(2): 101-6.
Kato, T., K. Iwamoto, et al. (2005). "Genetic or epigenetic difference causing discordance between monozygotic twins as a clue to molecular basis of mental disorders." Mol Psychiatry **10**(7): 622-30.
Katsanis, N., S. J. Ansley, et al. (2001). "Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder." Science **293**(5538): 2256-9.
Kawada, Y., M. Hattori, et al. (1995a). "Possible association between monoamine oxidase A gene and bipolar affective disorder." Am J Hum Genet **56**(1): 335-6.
Kawada, Y., M. Hattori, et al. (1995b). "No evidence of linkage or association between tyrosine hydroxylase gene and affective disorder." J Affect Disord **34**(2): 89-94.
Kelsoe, J. R., E. I. Ginns, et al. (1989). "Re-evaluation of the linkage relationship between chromosome 11p loci and the gene for bipolar affective disorder in the Old Order Amish." Nature **342**(6247): 238-43.
Kelsoe, J. R., M. A. Spence, et al. (2001). "A genome survey indicates a possible susceptibility locus for bipolar disorder on chromosome 22." Proc Natl Acad Sci U S A **98**(2): 585-90.
Kerem, B., J. M. Rommens, et al. (1989). "Identification of the cystic fibrosis gene: genetic analysis." Science **245**(4922): 1073-80.
Kerner, B., D. L. Brugman, et al. (2007). "Evidence of linkage to psychosis on chromosome 5q33-34 in pedigrees ascertained for bipolar disorder." Am J Med Genet B Neuropsychiatr Genet **144**(1): 74-8.
Kieseppa, T., T. Partonen, et al. (2004). "High concordance of bipolar I disorder in a nationwide sample of twins." Am J Psychiatry **161**(10): 1814-21.
Kirov, G., I. Jones, et al. (1999a). "Family-based association studies of bipolar disorder with candidate genes involved in dopamine neurotransmission: DBH, DAT1, COMT, DRD2, DRD3 and DRD5." Mol Psychiatry **4**(6): 558-65.
Kirov, G., C. A. Lowry, et al. (2001). "Screening ABCG1, the human homologue of the Drosophila white gene, for polymorphisms and association with bipolar affective disorder." Mol Psychiatry **6**(6): 671-7.
Kirov, G., K. C. Murphy, et al. (1998). "Low activity allele of catechol-O-methyltransferase gene associated with rapid cycling bipolar disorder." Mol Psychiatry **3**(4): 342-5.
Kirov, G., M. J. Owen, et al. (1999b). "Tryptophan hydroxylase gene and manic-depressive illness." Arch Gen Psychiatry **56**(1): 98-9.
Klar, A. J. (2002). "The chromosome 1;11 translocation provides the best evidence supporting genetic etiology for schizophrenia and bipolar affective disorders." Genetics **160**(4): 1745-7.
Kornberg, J. R., J. L. Brown, et al. (2000). "Evaluating the parent-of-origin effect in bipolar affective disorder. Is a more penetrant subtype transmitted paternally?" J Affect Disord **59**(3): 183-92.
Kostyrko, A., J. Hauser, et al. (2006). "Screening of chromosomal region 21q22.3 for mutations in genes associated with neuronal Ca2+ signalling in bipolar affective disorder." Acta Biochim Pol **53**(2): 317-20.

Kremeyer, B., I. Herzberg, et al. (2006). "Transmission distortion of BDNF variants to bipolar disorder type I patients from a South American population isolate." Am J Med Genet B Neuropsychiatr Genet **141**(5): 435-9.
Krishnan, K. R. (2005). "Psychiatric and medical comorbidities of bipolar disorder." Psychosom Med **67**(1): 1-8.
Kruglyak, L. (1999a). "Genetic isolates: separate but equal?" Proc Natl Acad Sci U S A **96**(4): 1170-2.
Kruglyak, L. (1999b). "Prospects for whole-genome linkage disequilibrium mapping of common disease genes." Nat Genet **22**(2): 139-44.
Kunugi, H., Y. Iijima, et al. (2004). "No association between the Val66Met polymorphism of the brain-derived neurotrophic factor gene and bipolar disorder in a Japanese population: a multicenter study." Biol Psychiatry **56**(5): 376-8.
Kunugi, H., S. Ishida, et al. (1999). "No evidence for an association of polymorphisms of the tryptophan hydroxylase gene with affective disorders or attempted suicide among Japanese patients." Am J Psychiatry **156**(5): 774-6.
Kunugi, H., H. P. Vallada, et al. (1997). "No evidence for an association of affective disorders with high- or low-activity allele of catechol-o-methyltransferase gene." Biol Psychiatry **42**(4): 282-5.
Kwok, J. B., L. J. Adams, et al. (1999). "Nonparametric simulation-based statistical analyses for bipolar affective disorder locus on chromosome 21q22.3." Am J Med Genet **88**(1): 99-102.
LaBuda, M. C., M. Maldonado, et al. (1996). "A follow-up report of a genome search for affective disorder predisposition loci in the Old Order Amish." Am J Hum Genet **59**(6): 1343-62.
Lachman, H. M., J. R. Kelsoe, et al. (1997). "Linkage studies suggest a possible locus for bipolar disorder near the velo-cardio-facial syndrome region on chromosome 22." Am J Med Genet **74**(2): 121-8.
Lachman, H. M., B. Morrow, et al. (1996a). "Association of codon 108/158 catechol-O-methyltransferase gene polymorphism with the psychiatric manifestations of velo-cardio-facial syndrome." Am J Med Genet **67**(5): 468-72.
Lachman, H. M., D. F. Papolos, et al. (1996b). "Human catechol-O-methyltransferase pharmacogenetics: description of a functional polymorphism and its potential application to neuropsychiatric disorders." Pharmacogenetics **6**(3): 243-50.
Lai, T. J., C. Y. Wu, et al. (2005). "Polymorphism screening and haplotype analysis of the tryptophan hydroxylase gene (TPH1) and association with bipolar affective disorder in Taiwan." BMC Med Genet **6**: 14.
Lander, E. and L. Kruglyak (1995). "Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results." Nat Genet **11**(3): 241-7.
Lander, E. S. and N. J. Schork (1994). "Genetic dissection of complex traits." Science **265**(5181): 2037-48.
Lasky-Su, J. A., S. V. Faraone, et al. (2005). "Meta-analysis of the association between two polymorphisms in the serotonin transporter gene and affective disorders." Am J Med Genet B Neuropsychiatr Genet **133**(1): 110-5.
Law, A., C. W. Richard, 3rd, et al. (1992). "Genetic linkage analysis of bipolar affective disorder in an Old Order Amish pedigree." Hum Genet **88**(5): 562-8.

Leboyer, M., A. Malafosse, et al. (1990). "Tyrosine hydroxylase polymorphisms associated with manic-depressive illness." Lancet **335**(8699): 1219.
LeGuern, E., A. Guilbot, et al. (1996). "Homozygosity mapping of an autosomal recessive form of demyelinating Charcot-Marie-Tooth disease to chromosome 5q23-q33." Hum Mol Genet **5**(10): 1685-8.
Leszczynska-Rodziewicz, A., J. Hauser, et al. (2005). "Lack of association between polymorphisms of dopamine receptors, type D2, and bipolar affective illness in a Polish population." Med Sci Monit **11**(6): CR289-295.
Levinson, D. F., M. D. Levinson, et al. (2003). "Genome scan meta-analysis of schizophrenia and bipolar disorder, part I: Methods and power analysis." Am J Hum Genet **73**(1): 17-33.
Lewontin, R. C. (1964). "The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models." Genetics **49**(1): 49-67.
Li, Q., X. Luo, et al. (2005). "Regulation of the P2X7 receptor permeability to large molecules by extracellular Cl- and Na+." J Biol Chem **280**(29): 26922-7.
Li, T., X. Liu, et al. (1999). "Association analysis between dopamine receptor genes and bipolar affective disorder." Psychiatry Res **86**(3): 193-201.
Li, T., H. Vallada, et al. (1997). "Catechol-O-methyltransferase Val158Met polymorphism: frequency analysis in Han Chinese subjects and allelic association of the low activity allele with bipolar affective disorder." Pharmacogenetics **7**(5): 349-53.
Lim, L. C., M. M. Nothen, et al. (1994). "No evidence of association between dopamine D4 receptor variants and bipolar affective disorder." Am J Med Genet **54**(3): 259-63.
Lim, L. C., J. Powell, et al. (1995). "Evidence for a genetic association between alleles of monoamine oxidase A gene and bipolar affective disorder." Am J Med Genet **60**(4): 325-31.
Lin, P. I., M. G. McInnis, et al. (2005). "Assessment of the effect of age at onset on linkage to bipolar disorder: evidence on chromosomes 18p and 21q." Am J Hum Genet **77**(4): 545-55.
Liu, C., J. A. Badner, et al. (2001a). "Fine mapping supports previous linkage evidence for a bipolar disorder susceptibility locus on 13q32." Am J Med Genet **105**(4): 375-80.
Liu, J., S. H. Juo, et al. (2003). "Evidence for a putative bipolar disorder locus on 2p13-16 and other potential loci on 4q31, 7q34, 8q13, 9q31, 10q21-24, 13q32, 14q21 and 17q11-12." Mol Psychiatry **8**(3): 333-42.
Liu, J., S. H. Juo, et al. (2001b). "A follow-up linkage study supports evidence for a bipolar affective disorder locus on chromosome 21q22." Am J Med Genet **105**(2): 189-94.
Lohmueller, K. E., C. L. Pearce, et al. (2003). "Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease." Nat Genet **33**(2): 177-82.
Lohoff, F. W., T. Sander, et al. (2005). "Confirmation of association between the Val66Met polymorphism in the brain-derived neurotrophic factor (BDNF) gene and bipolar I disorder." Am J Med Genet B Neuropsychiatr Genet **139**(1): 51-3.
Lolak, S., E. Dannemiller, et al. (2005). "48,XXYY syndrome, mood disorder, and aggression." Am J Psychiatry **162**(7): 1384.
Lonjou, C., A. Collins, et al. (1999). "Allelic association between marker loci." Proc Natl Acad Sci U S A **96**(4): 1621-6.

Lotrich, F. E. and B. G. Pollock (2004). "Meta-analysis of serotonin transporter polymorphisms and affective disorders." Psychiatr Genet **14**(3): 121-9.
Lucae, S., D. Salyakina, et al. (2006). "P2RX7, a gene coding for a purinergic ligand-gated ion channel, is associated with major depressive disorder." Hum Mol Genet **15**(16): 2438-45.
Lykken, D. T., M. McGue, et al. (1987). "Recruitment bias in twin research: the rule of two-thirds reconsidered." Behav Genet **17**(4): 343-62.
Lyons-Warren, A., J. J. Chang, et al. (2005). "Evidence of association between bipolar disorder and Citron on chromosome 12q24." Mol Psychiatry **10**(9): 807-9.
Macayran, J. F., S. G. Brodie, et al. (2006). "Duplication 8q22.1-q24.1 associated with bipolar disorder and speech delay." Bipolar Disord **8**(3): 294-8.
Macgregor, S., P. M. Visscher, et al. (2004). "A genome scan and follow-up study identify a bipolar disorder susceptibility locus on chromosome 1q42." Mol Psychiatry **9**(12): 1083-90.
MacIntyre, D. J., D. H. Blackwood, et al. (2003). "Chromosomal abnormalities and mental illness." Mol Psychiatry **8**(3): 275-87.
MacKenzie, A. and J. Quinn (1999). "A serotonin transporter gene intron 2 polymorphic region, correlated with affective disorders, has allele-dependent differential enhancer-like properties in the mouse embryo." Proc Natl Acad Sci U S A **96**(26): 15251-5.
MacQueen, G. M., T. Hajek, et al. (2005). "The phenotypes of bipolar disorder: relevance for genetic investigations." Mol Psychiatry **10**(9): 811-26.
Maeda, K., E. Nwulia, et al. (2006). "Differential expression of disrupted-in-schizophrenia (DISC1) in bipolar disorder." Biol Psychiatry **60**(9): 929-35.
Maj, M., R. Pirozzi, et al. (2002). "The prognostic significance of "switching" in patients with bipolar disorder: a 10-year prospective follow-up study." Am J Psychiatry **159**(10): 1711-7.
Malafosse, A., M. Leboyer, et al. (1997). "Manic depressive illness and tyrosine hydroxylase gene: linkage heterogeneity and association." Neurobiol Dis **4**(5): 337-49.
Manki, H., S. Kanba, et al. (1996). "Dopamine D2, D3 and D4 receptor and transporter gene polymorphisms and mood disorders." J Affect Disord **40**(1-2): 7-13.
Mansell, W., F. Colom, et al. (2005). "The nature and treatment of depression in bipolar disorder: a review and implications for future psychological investigation." Clin Psychol Rev **25**(8): 1076-100.
Marcheco-Teruel, B., T. J. Flint, et al. (2006). "A genome-wide linkage search for bipolar disorder susceptibility loci in a large and complex pedigree from the eastern part of Cuba." Am J Med Genet B Neuropsychiatr Genet **141**(8): 833-43.
Massat, I., D. Souery, et al. (2005). "Association between COMT (Val158Met) functional polymorphism and early onset in patients with major depressive disorder in a European multicenter genetic association study." Mol Psychiatry **10**(6): 598-605.
Maziade, M., M. A. Roy, et al. (2001). "A search for specific and common susceptibility loci for schizophrenia and bipolar disorder: a linkage study in 13 target chromosomes." Mol Psychiatry **6**(6): 684-93.

McGuffin, P., A. Farmer, et al. (1991). "A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system." Arch Gen Psychiatry **48**(8): 764-70.
McGuffin, P. and R. Katz (1989). "The genetics of depression and manic-depressive disorder." Br J Psychiatry **155**: 294-304.
McGuffin, P., F. Rijsdijk, et al. (2003). "The heritability of bipolar affective disorder and the genetic relationship to unipolar depression." Arch Gen Psychiatry **60**(5): 497-502.
McInnes, L. A., M. A. Escamilla, et al. (1996). "A complete genome screen for genes predisposing to severe bipolar disorder in two Costa Rican pedigrees." Proc Natl Acad Sci U S A **93**(23): 13060-5.
McInnis, M. G., D. M. Dick, et al. (2003a). "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): 1265-73.
McInnis, M. G., T. H. Lan, et al. (2003b). "Genome-wide scan of bipolar disorder in 65 pedigrees: supportive evidence for linkage at 8q24, 18q22, 4q32, 2p12, and 13q12." Mol Psychiatry **8**(3): 288-98.
McInnis, M. G., F. J. McMahon, et al. (1993). "Anticipation in bipolar affective disorder." Am J Hum Genet **53**(2): 385-90.
McLeod, H. L., A. C. Syvanen, et al. (1998). "Ethnic differences in catechol O-methyltransferase pharmacogenetics: frequency of the codon 108/158 low activity allele is lower in Kenyan than Caucasian or South-west Asian individuals." Pharmacogenetics **8**(3): 195-9.
McMahon, F. J., O. C. Stine, et al. (1995). "Patterns of maternal transmission in bipolar affective disorder." Am J Hum Genet **56**(6): 1277-86.
McQuay, H. and R. Moore (1997). "Using Numerical Results from Systematic Reviews in Clinical Practice." Annals of Internal Medicine **126**(9): 712-720.
McQueen, M. B., B. Devlin, et al. (2005). "Combined analysis from eleven linkage studies of bipolar disorder provides strong evidence of susceptibility loci on chromosomes 6q and 8q." Am J Hum Genet **77**(4): 582-95.
McQuillin, A., N. J. Bass, et al. (2006). "Fine mapping of a susceptibility locus for bipolar and genetically related unipolar affective disorders, to a region containing the C21ORF29 and TRPM2 genes on chromosome 21q22.3." Mol Psychiatry **11**(2): 134-42.
McQuillin, A., J. Lawrence, et al. (1999a). "Adjacent genetic markers on chromosome 11p15.5 at or near the tyrosine hydroxylase locus that show population linkage disequilibrium with each other do not show allelic association with bipolar affective disorder." Psychol Med **29**(6): 1449-54.
McQuillin, A., J. Lawrence, et al. (1999b). "No allelic association between bipolar affective disorder and the tryptophan hydroxylase gene." Arch Gen Psychiatry **56**(1): 99-101.
Meira-Lima, I. V., A. C. Pereira, et al. (2000). "Angiotensinogen and angiotensin converting enzyme gene polymorphisms and the risk of bipolar affective disorder in humans." Neurosci Lett **293**(2): 103-6.
Meloni, R., M. Leboyer, et al. (1995). "Association of manic-depressive illness with tyrosine hydroxylase microsatellite marker." Lancet **345**(8954): 932.

Mendes de Oliveira, J. R., P. A. Otto, et al. (1998). "Analysis of a novel functional polymorphism within the promoter region of the serotonin transporter gene (5-HTT) in Brazilian patients affected by bipolar disorder and schizophrenia." Am J Med Genet **81**(3): 225-7.

Mendlewicz, J. and J. D. Rainer (1977). "Adoption study supporting genetic transmission in manic--depressive illness." Nature **268**(5618): 327-9.

Metherell, L. A., J. P. Chapple, et al. (2005). "Mutations in MRAP, encoding a new interacting partner of the ACTH receptor, cause familial glucocorticoid deficiency type 2." Nat Genet **37**(2): 166-70.

Millar, J. K., S. Christie, et al. (2000a). "Chromosomal location and genomic structure of the human translin-associated factor X gene (TRAX; TSNAX) revealed by intergenic splicing to DISC1, a gene disrupted by a translocation segregating with schizophrenia." Genomics **67**(1): 69-77.

Millar, J. K., B. S. Pickard, et al. (2005). "DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling." Science **310**(5751): 1187-91.

Millar, J. K., J. C. Wilson-Annan, et al. (2000b). "Disruption of two novel genes by a translocation co-segregating with schizophrenia." Hum Mol Genet **9**(9): 1415-23.

Morissette, J., A. Villeneuve, et al. (1999). "Genome-wide search for linkage of bipolar affective disorders in a very large pedigree derived from a homogeneous population in quebec points to a locus of major effect on chromosome 12q23-q24." Am J Med Genet **88**(5): 567-87.

Morton, L. A. and K. K. Kidd (1980). "The effects of variable age-of-onset and diagnostic criteria on the estimates of linkage: an example using manic-depressive illness and color blindness." Soc Biol **27**(1): 1-10.

Morton, N. E. (1955). "Sequential tests for the detection of linkage." Am J Hum Genet **7**(3): 277-318.

Morton, N. E. (1956). "The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type." Am J Hum Genet **8**(2): 80-96.

Morton, N. E. and A. Collins (1998). "Tests and estimates of allelic association in complex inheritance." Proc Natl Acad Sci U S A **95**(19): 11389-93.

Moskvina, V., N. Craddock, et al. (2006). "Effects of differential genotyping error rate on the type I error probability of case-control studies." Hum Hered **61**(1): 55-64.

Mouaffak, F., T. Gallarda, et al. (2007). "Gender identity disorders and bipolar disorder associated with the ring Y chromosome." Am J Psychiatry **164**(7): 1122-3.

Muglia, P., A. Petronis, et al. (2002). "Dopamine D4 receptor and tyrosine hydroxylase genes in bipolar disorder: evidence for a role of DRD4." Mol Psychiatry **7**(8): 860-6.

Muller, D. J., V. de Luca, et al. (2006). "Brain-derived neurotrophic factor (BDNF) gene and rapid-cycling bipolar disorder: family-based association study." Br J Psychiatry **189**: 317-23.

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

Munafo, M. R. and J. Flint (2004). "Meta-analysis of genetic association studies." Trends Genet **20**(9): 439-44.

Mundo, E., M. Walker, et al. (2001). "The role of serotonin transporter protein gene in antidepressant-induced mania in bipolar disorder: preliminary findings." Arch Gen Psychiatry **58**(6): 539-44.
Muntoni, F., S. Torelli, et al. (2003). "Dystrophin and mutations: one gene, several proteins, multiple phenotypes." Lancet Neurol **2**(12): 731-40.
Murphy, V. E., L. A. Mynett-Johnson, et al. (2000). "Search for bipolar disorder susceptibility loci: the application of a modified genome scan concentrating on gene-rich regions." Am J Med Genet **96**(6): 728-32.
Murray, C. J. and A. D. Lopez (1997). "Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study." Lancet **349**(9063): 1436-42.
Mynett-Johnson, L. A., V. E. Murphy, et al. (1998). "Preliminary evidence of an association between bipolar disorder in females and the catechol-O-methyltransferase gene." Psychiatr Genet **8**(4): 221-5.
Nagamine, K., J. Kudoh, et al. (1998). "Molecular cloning of a novel putative Ca2+ channel protein (TRPC7) highly expressed in brain." Genomics **54**(1): 124-31.
Nakai, K. and P. Horton (1999). "PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization." Trends Biochem Sci **24**(1): 34-6.
Nakata, K., H. Ujike, et al. (2003). "Association study of the brain-derived neurotrophic factor (BDNF) gene with bipolar disorder." Neurosci Lett **337**(1): 17-20.
Neuman, R. J., B. Geller, et al. (1997). "Increased prevalence and earlier onset of mood disorders among relatives of prepubertal versus adult probands." J Am Acad Child Adolesc Psychiatry **36**(4): 466-73.
Neves-Pereira, M., E. Mundo, et al. (2002). "The brain-derived neurotrophic factor gene confers susceptibility to bipolar disorder: evidence from a family-based association study." Am J Hum Genet **71**(3): 651-5.
Nickerson, D. A., S. L. Taylor, et al. (1998). "DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene." Nat Genet **19**(3): 233-40.
Noga, J. T., K. Vladar, et al. (2001). "A volumetric magnetic resonance imaging study of monozygotic twins discordant for bipolar disorder." Psychiatry Res **106**(1): 25-34.
Nothen, M. M., K. Eggermann, et al. (1995). "Association analysis of the monoamine oxidase A gene in bipolar affective disorder by using family-based internal controls." Am J Hum Genet **57**(4): 975-8.
Oetting, W. S., H. K. Lee, et al. (1995). "Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers." Genomics **30**(3): 450-8.
Ohara, K., M. Nagai, et al. (1998). "Low activity allele of catechol-o-methyltransferase gene and Japanese unipolar depression." Neuroreport **9**(7): 1305-8.
Okada, T., R. Hashimoto, et al. (2006). "A complex polymorphic region in the brain-derived neurotrophic factor (BDNF) gene confers susceptibility to bipolar disorder and affects transcriptional activity." Mol Psychiatry **11**(7): 695-703.
Olajossy, M., L. Olajossy-Hilkesberger, et al. (2005). "Bipolar affective disorder in a male with a deletion of Y chromosome -- a case report." Bipolar Disord **7**(3): 298-301.
Oliveira, J. R., D. R. Carvalho, et al. (2000). "Analysis of the serotonin transporter polymorphism (5-HTTLPR) in Brazilian patients affected by dysthymia, major depression and bipolar disorder." Mol Psychiatry **5**(4): 348-9.

Olson, M., L. Hood, et al. (1989). "A common language for physical mapping of the human genome." Science **245**(4925): 1434-5.
Oruc, L., G. R. Verheyen, et al. (1997). "Analysis of the tyrosine hydroxylase and dopamine D4 receptor genes in a Croatian sample of bipolar I and unipolar patients." Am J Med Genet **74**(2): 176-8.
Osby, U., L. Brandt, et al. (2001). "Excess mortality in bipolar and unipolar disorder in Sweden." Arch Gen Psychiatry **58**(9): 844-50.
Oswald, P., J. Del-Favero, et al. (2004). "Non-replication of the brain-derived neurotrophic factor (BDNF) association in bipolar affective disorder: a Belgian patient-control study." Am J Med Genet B Neuropsychiatr Genet **129**(1): 34-5.
Ott, J. (1974). "Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies." Am J Hum Genet **26**(5): 588-97.
Owen, M. J., P. Holmans, et al. (1997). "Association studies in psychiatric genetics." Mol Psychiatry **2**(4): 270-3.
Palmatier, M. A., A. J. Pakstis, et al. (2004). "COMT haplotypes suggest P2 promoter region relevance for schizophrenia." Mol Psychiatry **9**(9): 859-70.
Palo, O. M., M. Antila, et al. (2007). "Association of distinct allelic haplotypes of DISC1 with psychotic and bipolar spectrum disorders and with underlying cognitive impairments." Hum Mol Genet.
Papolos, D. F., G. L. Faedda, et al. (1996). "Bipolar spectrum disorders in patients diagnosed with velo-cardio-facial syndrome: does a hemizygous deletion of chromosome 22q11 result in bipolar affective disorder?" Am J Psychiatry **153**(12): 1541-7.
Papolos, D. F., S. Veit, et al. (1998). "Ultra-ultra rapid cycling bipolar disorder is associated with the low activity catecholamine-O-methyltransferase allele." Mol Psychiatry **3**(4): 346-9.
Papp, L., E. S. Vizi, et al. (2004). "Lack of ATP-evoked GABA and glutamate release in the hippocampus of P2X7 receptor-/- mice." Neuroreport **15**(15): 2387-91.
Park, N., S. H. Juo, et al. (2004). "Linkage analysis of psychosis in bipolar pedigrees suggests novel putative loci for bipolar disorder and shared susceptibility with schizophrenia." Mol Psychiatry **9**(12): 1091-9.
Pauls, D. L., J. N. Bailey, et al. (1995). "Complex segregation analyses of old order Amish families ascertained through bipolar I individuals." Am J Med Genet **60**(4): 290-7.
Pauls, D. L., D. S. Gerhard, et al. (1991). "Linkage of bipolar affective disorders to markers on chromosome 11p is excluded in a second lateral extension of Amish pedigree 110." Genomics **11**(3): 730-6.
Pauls, D. L., L. A. Morton, et al. (1992). "Risks of affective illness among first-degree relatives of bipolar I old-order Amish probands." Arch Gen Psychiatry **49**(9): 703-8.
Peltonen, L., A. Palotie, et al. (2000). "Use of population isolates for mapping complex traits." Nat Rev Genet **1**(3): 182-90.
Perez de Castro, I., P. Torres, et al. (1994). "No association between dopamine D4 receptor polymorphism and manic depressive illness." J Med Genet **31**(11): 897-8.
Perlis, R. H., J. A. Welge, et al. (2006). "Atypical antipsychotics in the treatment of mania: a meta-analysis of randomized, placebo-controlled trials." J Clin Psychiatry **67**(4): 509-16.

Pickard, B. S., E. J. Hollox, et al. (2004). "A 4q35.2 subtelomeric deletion identified in a screen of patients with co-morbid psychiatric illness and mental retardation." BMC Med Genet **5**: 21.
Poolsup, N., A. Li Wan Po, et al. (2000). "Systematic overview of lithium treatment in acute mania." J Clin Pharm Ther **25**(2): 139-56.
Potash, J. B., P. P. Zandi, et al. (2003). "Suggestive linkage to chromosomal regions 13q31 and 22q12 in families with psychotic bipolar disorder." Am J Psychiatry **160**(4): 680-6.
Prata, D. P., G. Breen, et al. (2006). "Bipolar 1 disorder is not associated with the RGS4, PRODH, COMT and GRK3 genes." Psychiatr Genet **16**(6): 229-30.
Preisig, M., F. Bellivier, et al. (2000). "Association between bipolar disorder and monoamine oxidase A gene polymorphisms: results of a multicenter study." Am J Psychiatry **157**(6): 948-55.
Pritchard, J. K. (2001). "Are rare variants responsible for susceptibility to complex diseases?" Am J Hum Genet **69**(1): 124-37.
Pritchard, J. K. and N. A. Rosenberg (1999). "Use of unlinked genetic markers to detect population stratification in association studies." Am J Hum Genet **65**(1): 220-8.
Qian, X., A. Shu, et al. (2004). "A novel insertion mutation in the FOXL2 gene is detected in a big Chinese family with blepharophimosis-ptosis-epicanthus inversus." Mutat Res **554**(1-2): 19-22.
Reich, D. E., M. Cargill, et al. (2001). "Linkage disequilibrium in the human genome." Nature **411**(6834): 199-204.
Reich, T., P. J. Clayton, et al. (1969). "Family history studies: V. The genetics of mania." Am J Psychiatry **125**(10): 1358-69.
Reif, A., W. Kress, et al. (2004). "Duplication 15q14 --> pter: a rare chromosomal abnormality underlying bipolar affective disorder." Eur Psychiatry **19**(3): 179-81.
Reiman, E. M., J. A. Webster, et al. (2007). "GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers." Neuron **54**(5): 713-20.
Rice, J., T. Reich, et al. (1987). "The familial transmission of bipolar illness." Arch Gen Psychiatry **44**(5): 441-7.
Rietschel, M., A. Schorr, et al. (2000). "Association study of the tryptophan hydroxylase gene and bipolar affective disorder using family-based internal controls." Am J Med Genet **96**(3): 310-1.
Risch, N. (1988). "A new statistical test for linkage heterogeneity." Am J Hum Genet **42**(2): 353-64.
Risch, N. and K. Merikangas (1996). "The future of genetic studies of complex human diseases." Science **273**(5281): 1516-7.
Robinson, L. J. and I. N. Ferrier (2006). "Evolution of cognitive impairment in bipolar disorder: a systematic review of cross-sectional evidence." Bipolar Disord **8**(2): 103-16.
Roche, S., F. Cassidy, et al. (2007). "Candidate gene analysis of 21q22: Support for S100B as a susceptibility gene for bipolar affective disorder with psychosis." Am J Med Genet B Neuropsychiatr Genet.
Rotondo, A., C. Mazzanti, et al. (2002). "Catechol o-methyltransferase, serotonin transporter, and tryptophan hydroxylase gene polymorphisms in bipolar disorder patients with and without comorbid panic disorder." Am J Psychiatry **159**(1): 23-9.

Rozen, S. and H. J. Skaletsky (2000). Methods in Molecular Biology. Totowa, NJ, Humana Press.

Rubinsztein, D. C., J. Leggo, et al. (1996). "Genetic association between monoamine oxidase A microsatellite and RFLP alleles and bipolar affective disorder: analysis and meta-analysis." Hum Mol Genet **5**(6): 779-82.

Saito, T., F. Guan, et al. (2001). "Mutation analysis of SYNJ1: a possible candidate gene for chromosome 21q22-linked bipolar disorder." Mol Psychiatry **6**(4): 387-95.

Sajatovic, M. (2005). "Bipolar disorder: disease burden." Am J Manag Care **11**(3 Suppl): S80-4.

Sambrook, J., E. F. Fritsch, et al. (1989). Molecular Cloning: A laboratory manual. New York, Cold Harbour Spring Laboratory Press.

Sandkuyl, L. A. and J. Ott (1989). "Affective disorders: evaluation of a three-allele model accounting for clinical heterogeneity." Genet Epidemiol **6**(1): 265-9.

Santosa, C. M., C. M. Strong, et al. (2006). "Enhanced creativity in bipolar disorder patients: A controlled study." J Affect Disord.

Schulze, T. G., D. Hedeker, et al. (2006). "What is familial about familial bipolar disorder? Resemblance among relatives across a broad spectrum of phenotypic characteristics." Arch Gen Psychiatry **63**(12): 1368-76.

Schulze, T. G., S. Ohlraun, et al. (2005). "Genotype-phenotype studies in bipolar disorder showing association between the DAOA/G30 locus and persecutory delusions: a first step toward a molecular genetic classification of psychiatric phenotypes." Am J Psychiatry **162**(11): 2101-8.

Schumacher, J., R. A. Jamra, et al. (2005a). "Evidence for a relationship between genetic variants at the brain-derived neurotrophic factor (BDNF) locus and major depression." Biol Psychiatry **58**(4): 307-14.

Schumacher, J., R. A. Jamra, et al. (2004). "Examination of G72 and D-amino-acid oxidase as genetic risk factors for schizophrenia and bipolar affective disorder." Mol Psychiatry **9**(2): 203-7.

Schumacher, J., R. Kaneva, et al. (2005b). "Genomewide scan and fine-mapping linkage studies in four European samples with bipolar affective disorder suggest a new susceptibility locus on chromosome 1p35-p36 and provides further evidence of loci on chromosome 4q31 and 6q24." Am J Hum Genet **77**(6): 1102-11.

Schwab, S. G., M. Knapp, 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): 185-90.

Sebat, J., B. Lakshmi, et al. (2004). "Large-scale copy number polymorphism in the human genome." Science **305**(5683): 525-8.

Segurado, R., S. D. Detera-Wadleigh, et al. (2003). "Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: Bipolar disorder." Am J Hum Genet **73**(1): 49-62.

Serretti, A., C. Cusin, et al. (2003). "Multicentre Italian family-based association study on tyrosine hydroxylase, catechol-O-methyl transferase and Wolfram syndrome 1 polymorphisms in mood disorders." Psychiatr Genet **13**(2): 121-6.

Serretti, A., R. Lilli, et al. (2001). "DRD4 exon 3 variants associated with delusional symptomatology in major psychoses: a study on 2,011 affected subjects." Am J Med Genet **105**(3): 283-90.

Serretti, A., A. Rotondo, et al. (2006). "Catechol-O-methyltransferase gene variants in mood disorders in the Italian population." Psychiatr Genet **16**(5): 181-2.
Service, S., J. Molina, et al. (2006). "Results of a SNP genome screen in a large Costa Rican pedigree segregating for severe bipolar disorder." Am J Med Genet B Neuropsychiatr Genet **141**(4): 367-73.
Sham, P. (1998). Statistics in Human Genetics. London, Arnold.
Sham, P. C. and D. Curtis (1995). "Monte Carlo tests for associations between disease and alleles at highly polymorphic loci." Ann Hum Genet **59**(Pt 1): 97-105.
Shaw, S. H., Z. Mroczkowski-Parker, et al. (2003). "Linkage of a bipolar disorder susceptibility locus to human chromosome 13q32 in a new pedigree series." Mol Psychiatry **8**(5): 558-64.
Shifman, S., M. Bronstein, et al. (2004). "COMT: a common susceptibility gene in bipolar disorder and schizophrenia." Am J Med Genet B Neuropsychiatr Genet **128**(1): 61-4.
Shink, E., M. Harvey, et al. (2005a). "Analysis of microsatellite markers and single nucleotide polymorphisms in candidate genes for susceptibility to bipolar affective disorder in the chromosome 12Q24.31 region." Am J Med Genet B Neuropsychiatr Genet **135**(1): 50-8.
Shink, E., J. Morissette, et al. (2005b). "A genome-wide scan points to a susceptibility locus for bipolar disorder on chromosome 12." Mol Psychiatry **10**(6): 545-52.
Skibinska, M., J. Hauser, et al. (2004). "Association analysis of brain-derived neurotrophic factor (BDNF) gene Val66Met polymorphism in schizophrenia and bipolar affective disorder." World J Biol Psychiatry **5**(4): 215-20.
Sklar, P., S. B. Gabriel, et al. (2002). "Family-based association study of 76 candidate genes in bipolar disorder: BDNF is a potential risk locus. Brain-derived neutrophic factor." Mol Psychiatry **7**(6): 579-93.
Smeraldi, E., F. Negri, et al. (1977). "A genetic study of affective disorders." Acta Psychiatr Scand **56**(5): 382-98.
Smith, C. A. (1963). "Testing for Heterogeneity of Recombination Fraction Values in Human Genetics." Ann Hum Genet **27**: 175-82.
Smyth, C., G. Kalsi, et al. (1996). "Further tests for linkage of bipolar affective disorder to the tyrosine hydroxylase gene locus on chromosome 11p15 in a new series of multiplex British affective disorder pedigrees." Am J Psychiatry **153**(2): 271-4.
Smyth, C., G. Kalsi, et al. (1997). "Two-locus admixture linkage analysis of bipolar and unipolar affective disorder supports the presence of susceptibility loci on chromosomes 11p15 and 21q22." Genomics **39**(3): 271-8.
Sobell, J. L., L. L. Heston, et al. (1992). "Delineation of genetic predisposition to multifactorial disease: a general approach on the threshold of feasibility." Genomics **12**(1): 1-6.
Somanath, C. P., S. Jain, et al. (2002). "A family study of early-onset bipolar I disorder." J Affect Disord **70**(1): 91-4.
Souery, D., S. Van Gestel, et al. (2001). "Tryptophan hydroxylase polymorphism and suicidality in unipolar and bipolar affective disorders: a multicenter association study." Biol Psychiatry **49**(5): 405-9.
Spence, M. A., P. L. Flodman, et al. (1995). "Bipolar disorder: evidence for a major locus." Am J Med Genet **60**(5): 370-6.

Spielman, R. S., R. E. McGinnis, et al. (1993). "Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM)." Am J Hum Genet **52**(3): 506-16.
Spitzer, R. L., J. Endicott, et al. (1978). "Research diagnostic criteria: rationale and reliability." Arch Gen Psychiatry **35**(6): 773-82.
St Clair, D., D. Blackwood, et al. (1990). "Association within a family of a balanced autosomal translocation with major mental illness." Lancet **336**(8706): 13-6.
Sterling, T. D. (1959). "Publication decisions and their possible effects on inferences drawn from tests of significance - or vice versa." Journal of the American Statistical Association **54**: 30-4.
Stopkova, P., J. Vevera, et al. (2004). "Analysis of SYNJ1, a candidate gene for 21q22 linked bipolar disorder: a replication study." Psychiatry Res **127**(1-2): 157-61.
Straub, R. E., T. Lehner, et al. (1994). "A possible vulnerability locus for bipolar affective disorder on chromosome 21q22.3." Nat Genet **8**(3): 291-6.
Strauss, J., C. L. Barr, et al. (2004). "Association study of brain-derived neurotrophic factor in adults with a history of childhood onset mood disorder." Am J Med Genet B Neuropsychiatr Genet **131**(1): 16-9.
Strober, M., W. Morrell, et al. (1988). "A family study of bipolar I disorder in adolescence. Early onset of symptoms linked to increased familial loading and lithium resistance." J Affect Disord **15**(3): 255-68.
Suarez, B. K., C. L. Hampe, et al. (1994). Problems of replicating linkage claims in psychiatry. Genetic approaches to mental disorders. C. R. Coninger. Washington DC, Am Psychiatric Press.
Tenhunen, J., M. Salminen, et al. (1994). "Genomic organization of the human catechol O-methyltransferase gene and its expression from two distinct promoters." Eur J Biochem **223**(3): 1049-59.
Thomson, P. A., N. R. Wray, et al. (2005). "Association between the TRAX/DISC locus and both bipolar disorder and schizophrenia in the Scottish population." Mol Psychiatry **10**(7): 657-68, 616.
Tienari, P. J., J. D. Terwilliger, et al. (1994). "Two-locus linkage analysis in multiple sclerosis (MS)." Genomics **19**(2): 320-5.
Todd, R. D. and K. L. O'Malley (1989). "Population frequencies of tyrosine hydroxylase restriction fragment length polymorphisms in bipolar affective disorder." Biol Psychiatry **25**(5): 626-30.
Tomas, C., F. Canellas, 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): 145-51.
Turecki, G., P. Grof, et al. (2001). "Mapping susceptibility genes for bipolar disorder: a pharmacogenetic approach based on excellent response to lithium." Mol Psychiatry **6**(5): 570-8.
Turecki, G., G. A. Rouleau, et al. (1997). "Lack of association between bipolar disorder and tyrosine hydroxylase: a meta-analysis." Am J Med Genet **74**(4): 348-52.
Uemura, T., J. Kudoh, et al. (2005). "Characterization of human and mouse TRPM2 genes: identification of a novel N-terminal truncated protein specifically expressed in human striatum." Biochem Biophys Res Commun **328**(4): 1232-43.

Van Den Bogaert, A., K. Sleegers, et al. (2006a). "Association of brain-specific tryptophan hydroxylase, TPH2, with unipolar and bipolar disorder in a Northern Swedish, isolated population." Arch Gen Psychiatry **63**(10): 1103-10.
Van Den Bogaert, A., K. Sleegers, et al. (2006b). "No allelic association or interaction of three known functional polymorphisms with bipolar disorder in a northern Swedish isolated population." Psychiatr Genet **16**(5): 209-12.
Van Eerdewegh, P., R. D. Little, et al. (2002). "Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness." Nature **418**(6896): 426-30.
Varilo, T., T. Paunio, et al. (2003). "The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories." Hum Mol Genet **12**(1): 51-9.
Vazza, G., C. Bertolin, et al. (2007). "Genome-wide scan supports the existence of a susceptibility locus for schizophrenia and bipolar disorder on chromosome 15q26." Mol Psychiatry **12**(1): 87-93.
Venken, T., M. Alaerts, et al. (2007). "Chromosome 10q harbors a susceptibility locus for bipolar disorder in Ashkenazi Jewish families." Mol Psychiatry.
Vincent, J. B., M. Masellis, et al. (1999). "Genetic association analysis of serotonin system genes in bipolar affective disorder." Am J Psychiatry **156**(1): 136-8.
Vitzthum, V. J. (2003). "A number no greater than the sum of its parts: the use and abuse of heritability." Hum Biol **75**(4): 539-58.
Wainwright, B. J., P. J. Scambler, et al. (1985). "Localization of cystic fibrosis locus to human chromosome 7cen-q22." Nature **318**(6044): 384-5.
Weeks, D. E. and K. Lange (1988). "The affected-pedigree-member method of linkage analysis." Am J Hum Genet **42**(2): 315-26.
Weeks, D. E. and K. Lange (1992). "A multilocus extension of the affected-pedigree-member method of linkage analysis." Am J Hum Genet **50**(4): 859-68.
Weeks, D. E., T. Lehner, et al. (1990). "Measuring the inflation of the lod score due to its maximization over model parameter values in human linkage analysis." Genet Epidemiol **7**(4): 237-43.
Wehage, E., J. Eisfeld, et al. (2002). "Activation of the cation channel long transient receptor potential channel 2 (LTRPC2) by hydrogen peroxide. A splice variant reveals a mode of activation independent of ADP-ribose." J Biol Chem **277**(26): 23150-6.
Weiss, K. M. (1998). "In search of human variation." Genome Res **8**(7): 691-7.
Weissman, M. M., R. C. Bland, et al. (1996). "Cross-national epidemiology of major depression and bipolar disorder." Jama **276**(4): 293-9.
Wender, P. H., S. S. Kety, et al. (1986). "Psychiatric disorders in the biological and adoptive families of adopted individuals with affective disorders." Arch Gen Psychiatry **43**(10): 923-9.
WHO (1993). The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research. Geneva, World Health Organization.
Williams, N. M., E. K. Green, 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): 366-73.

Wilson, G. M., S. Flibotte, et al. (2006). "DNA copy-number analysis in bipolar disorder and schizophrenia reveals aberrations in genes involved in glutamate signaling." Hum Mol Genet **15**(5): 743-9.
Wilson, S. R. (1997). "On extending the transmission/disequilibrium test (TDT)." Ann Hum Genet **61**(Pt 2): 151-61.
Wise, L. H., J. S. Lanchbury, et al. (1999). "Meta-analysis of genome searches." Ann Hum Genet **63**(Pt 3): 263-72.
Woo, J. M., K. S. Yoon, et al. (2002). "Catechol O-methyltransferase genetic polymorphism in panic disorder." Am J Psychiatry **159**(10): 1785-7.
Wright, A. F., A. D. Carothers, et al. (1999). "Population choice in mapping genes for complex diseases." Nat Genet **23**(4): 397-404.
Xie, T., S. L. Ho, et al. (1999). "Characterization and implications of estrogenic down-regulation of human catechol-O-methyltransferase gene transcription." Mol Pharmacol **56**(1): 31-8.
Xu, C., F. Macciardi, et al. (2006). "Association of the putative susceptibility gene, transient receptor potential protein melastatin type 2, with bipolar disorder." Am J Med Genet B Neuropsychiatr Genet **141**(1): 36-43.
Yelin, R., D. Dahary, et al. (2003). "Widespread occurrence of antisense transcription in the human genome." Nat Biotechnol **21**(4): 379-86.
Yokoi, T., K. Shinoda, et al. (1991). "A 3' splice site consensus sequence mutation in the intron 3 of the alpha-galactosidase A gene in a patient with Fabry disease." Jinrui Idengaku Zasshi **36**(3): 245-50.
Yoon, I. S., P. P. Li, et al. (2001). "Altered TRPC7 gene expression in bipolar-I disorder." Biol Psychiatry **50**(8): 620-6.
Zandi, P. P., J. A. Badner, et al. (2007). "Genome-wide linkage scan of 98 bipolar pedigrees and analysis of clinical covariates." Mol Psychiatry **12**(7): 630-9.
Zandi, P. P., V. L. Willour, et al. (2003). "Genome scan of a second wave of NIMH genetics initiative bipolar pedigrees: chromosomes 2, 11, 13, 14, and X." Am J Med Genet B Neuropsychiatr Genet **119**(1): 69-76.
Zhang, W., X. Chu, et al. (2003). "A novel TRPM2 isoform inhibits calcium influx and susceptibility to cell death." J Biol Chem **278**(18): 16222-9.
Zhang, X., M. Zhang, et al. (2005). "Stimulation of P2X7 receptors elevates Ca2+ and kills retinal ganglion cells." Invest Ophthalmol Vis Sci **46**(6): 2183-91.
Zhao, J. H., S. Lissarrague, et al. (2002). "GENECOUNTING: haplotype analysis with missing genotypes." Bioinformatics **18**(12): 1694-5.
Zondervan, K. T. and L. R. Cardon (2004). "The complex interplay among factors that influence allelic association." Nat Rev Genet **5**(2): 89-100.

Appendix 1: PCR cycling conditions

PCR Programme

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

* standard 55 °C and standard 57 °C cycles are the same standard 60 °C cycle but with different annealing temperature. Variation in cycle number is indicated in brackets in subsequent appendices.

Appendix 2: Novel microsatellite and SNP markers at the 12q24.3 locus generated for fine mapping

Marker	Type of marker	Size range of alleles (bp) or sequence around SNP	Oligonucleotide name	Oligonucleotide Sequence
307CA2	Dinucleotide	122-166	307CA2F	TTGCCATCATATTCCAACACA
			307CA2R	CAGGTTCCGTTTTCAAATGC
1634GT2	Dinucleotide	140-150	1634GT2F	GGCATACAATGTGGTGTTTCG
			1634GT2R	GGAGGGCTTCAATGTTTTCA
1634TET	Tetranucleotide	212-268	1634TETF	GCTGGACTTTAGAATCCCACA
			1634TETR	GCATGAGGTACAGGCATCAA
307GT4	Dinucleotide	350-376	307GT4F	TTCCAGATGTCCATTGCAGA
			307GT4R	CACCACTATCTTCCCCCAAT
307CA1	Dinucleotide	296-308	307CA1F	AGCTGTGCTAAGGAGCAAAAA
			307CA1R	TTGTCTACCAACCAGATAATGGA
D12SDK1	Dinucleotide	134-152	D12SDK1F	TTGGGATCACACCACTGC
			D12SDK1R	CCTCCTTCAGCCAACAATTC
D12SDK2	Dinucleotide	115-164	D12SDK2F	AAATGGGCAAAGCTTGGC
			D12SDK2R	GCCCAGATATTGCTCAGTACAG
47783-CT	SNP	aagtctatagatcG/Actggcacaagcttcca	47783-CT-F	GCAAAGAGAACAGGCTCACAGA
			47783-CT-R	TGCTGAACACTTGCTTCCTT
pufu in/del	SNP	gaagaagaaaagtGATGAACACAT/-ttggattttataa	pufu-indel-F	CATGGGGCATATGGAAAAAC
			pufu-indel-R	TCCAGAAAGATGGATGTCTTCA
29818-insT	SNP	agcatatttttttT/-acagaagaatctgcagt	29818-insT-F	TTGCAAAATGAAACCAAGTCC
			29818-insT-R	ACCCTTTGCACCTGTGTTG
rs10773323	SNP	agaaaaggaagaaA/ Tcaacactttcagagat	Forward:	CAAACAAACGATATTGATTCCTGTTT
			Reverse:	GGCTCAACAGGCTCCATCTC
Mette_1	SNP	gctggcccattggaaG/ Agagacctgagcttgg	Forward:	CATGGACCCAAGAGGGTATG
			Reverse:	AGGATTCATGGCATTTCCAC

Appendix 3: Novel microsatellite and SNP markers at 21q22.3 generated for fine mapping

Marker	Type	Oligo name	Oligo sequence
22A5.GT124a	Dinucl.	22A5.GT124-1	GCAGGGTCCAGATTCTGCTA
		22A5.GT124-2	AGAAACTGTCTCCACCGCAC
218C10.CA44a	Dinucl.	218C10.CA44-1	AGAGGGAGGAGGAAGCAGCT
		218C10.CA44-2	ATCCTGTTAGCAGCCCAAGA
C21orf2c	In/del	C21orf2-1	GTAATGAATGCTCTGTCTCCACTG
		C21orf2-2	AGCGAGGTACCTGAGCGTG
AGAT10	Tetranucl.	AGAT10-1	AATGAATTGATATGATAGATAGATGGA
		AGAT10-2	ACACACACCCACCTACCAAC
ATTT11	Tetranucl.	ATTT11-1	GTGCCGTAGCTTCAGCATTT
		ATTT11-2	CTGGGCAACAGAGTGAGACA
314n7CT18	Dinucl.	314n7CT18-1	GTCCCATATGGAAGCAGAGG
		314n7CT18-2	CGGTGGGATGAATTGTCTTT
314n7CA28	Dinucl.	314n7CA28-1	CATGACCACACCACGACT G
		314n7CA28-2	CCAGTTTTTAGAGCATAGTCTCTG G
314n7CA23	Dinucl.	314n7CA23-1	ATCCCTCCTGGCTGCCTA
		314n7CA23-2	TCTCTCGCTCGCTCTCTCTC
Col6a1b	VNTR	Col6a1-1	CACCGTCTCCTCCTGTGTGTTCCAGGGAAGAC
		Col6a1-2	CTGGGAAGACCGGGGGAGTCACACACGCTG
PCNT2-AluI	RFLP (AluI)	PCNT-1	GAAGTGATCCAGCAAAGATTGG
		PCNT-2	CAAAACAACAGGAAGCCATTCT

Appendix 4: Oligonucleotide sequences and PCR conditions for 1q42 markers

Marker	Oligonucletide sequence*	[$MgCl_2$] mM	Betaine	Cycling Conditions
D1S235	Forward- **CACGACGTTGTAAAACGAC**CAGCAAGAGTTCATGGGA	2.0	Y	Touch Down
	Reverse- AACAGTCAATTACAAAATATGTGTG			
D1S3462	Forward- **CACGACGTTGTAAAACGAC**TTTCTCACCTTTAAATGTCATCA	2.5	N	Touch Down
	Reverse- CCAGTACGCAGATGGTCCTA			
D1S251	Forward- GTCTCCAGCCTGCCAC	2.5	Y	Standard 60 (30)
	Reverse- **CACGACGTTGTAAAACGAC**GACCAAGCAACTTCACTCC			
D1S103	Forward- ACGAACATTCTACAAGTTAC	2.5	N	Standard 57 (27)
	Reverse- **CACGACGTTGTAAAACGAC**TTTCAGAGAAACTGACCTGT			
DISC_5'_ microsatellite	Forward- **CACGACGTTGTAAAACGAC**TTAAATCATGCGTCTGGAAGG	2.0	Y	Standard 55
	Reverse- TTCTTCTCCTCAGCGAGTCC			
D1S2709	Forward-**CACGACGTTGTAAAACGAC**TCATACCACATATCAGAATGTC	2.5	N	Touch Down
	Reverse- ATCAATCAGTATCTAATAGCATCA			

* M13 tail indicated in bold

Appendix 5: Oligonucleotide sequences and PCR conditions for BDNF microsatellite markers

Marker	Oligonucletide sequence*	[$MgCl_2$] mM	Betaine	Cycling Conditions
-1040bp (GT)n repeat	Forward- **CACGACGTTGTAAAACGAC**AGCACTAGCTGCCTATT	2.0	Y	Touch Down
	Reverse- GCCACTTTATCTCCTCCAGT			
AFMA275YB9	Forward- **CACGACGTTGTAAAACGAC**TGGCATGTAAATTTAAG	2.0	Y	Standard 55
	Reverse- CTGCTACCTCAGAAGTATCTCAA			

*M13 tail indicated in bold

Appendix 6: Oligonucleotide sequences and PCR conditions for P2RX7 microsatellite markers

Marker	Oligonucletide sequence	[$MgCl_2$] mM	Betaine	Cycling Conditions
NBG6	Forward- **CACGACGTTGTAAAACGAC**CGAACCAGCCCTAACCT	2.0	N	Touch Down
	Reverse- GTTTCTTGGACAGGATGTCGTGGG			
P2RX73	Forward- TGCAGCCACGTTTGTAACAT	2.5	Y	Standard 60
	Reverse- **CACGACGTTGTAAAACGAC**CTACTTGGGAGGCTGAG			
Exon 13 deletion	Forward- TTCCTGGACAACCAGAGGAG	2.5	Y	Touch Down
	Reverse- GAACAGCTCTGAGGTGGTGA			
P2RX71*	Forward- AGGGGCAAGATAGAGGGAAG			
	Reverse- **CACGACGTTGTAAAACGAC**CTCCAGCCTGGGTGATA			
P2RX72*	Forward- **CACGACGTTGTAAAACGAC**TTGTAAGGTAAGCCCCA			
	Reverse- GCCCTATTGGACAGCACAGA			

*Markers that failed optimization

Appendix 7: Oligonucleotide sequences and PCR conditions for COMT markers

Marker	Oligonucleotide sequence	[$MgCl_2$] mM	Betaine	Cycling Conditions
rs4680	Forward red– GAAGGTCGGAGTCAACGGATTCATGCACACCTTGTCCTTCAC			Standard 57
	Forward green– GAAGGTGACCAAGTTCATGCTCATGCACACCTTGTCCTTCAT			
	Reverse– ACCCAGCGGATGGTGGATTT			
rs2075507	Forward– CTCTGGCGGAAAGGAAT	2.5	N	Touch Down
	Reverse- TCGGCATCAAAAGGAGGAAAA			
Intron 1 tetranucleotide	Forward- **CACGACGTTGTAAAACGAC**CAGAGTGACAGCCCGTCTC	2.0	Y	Touch Down
	Reverse –GGCCTTGAACAAACACTAACA			
COMT2*	Forward- CCAATATGCATGTAACTCAAGGA			
	Reverse-.**CACGACGTTGTAAAACGAC**AGATCGTGCCACTGCACTC			
COMT3*	Forward- CAGATCACACCATGGCACTC			
	Reverse- **CACGACGTTGTAAAACGAC**TGAGGTCAAAGCTGGCTAGAA			
COMT4*	Forward- **CACGACGTTGTAAAACGAC**CTTCCTGAGCAGGAGGATGT			
	Reverse- GCTGAGGCAGGAGAATCACT			

*Markers that failed optimization

Appendix 8: Publications related to this thesis

1. Kalsi, G., McQuillin A., et al. (2006). "Identification of the Slynar gene (AY070435) and related brain expressed sequences as a candidate gene for susceptibility to affective disorders through allelic and haplotypic association with bipolar disorder on chromosome 12q24." Am J Psychiatry **163**(10): 1767-76.

2. McQuillin, A., Bass N.J., et al. (2006). "Fine mapping of a susceptibility locus for bipolar and genetically related unipolar affective disorders, to a region containing the C21ORF29 and TRPM2 genes on chromosome 21q22.3." Mol Psychiatry **11**(2): 134-42.

3. McQuillin A, Bass N.J., (2008). "Case-control studies show that a non-conservative amino-acid change from a glutamine to arginine in the P2RX7 purinergic receptor protein is associated with both bipolar- and unipolar-affective disorders." Mol Psychiatry. Feb 12 [Epub ahead of print]