

From THE DEPARTMENT OF CLINICAL
NEUROSCIENCE
Karolinska Institutet, Stockholm, Sweden

GENETIC STUDIES IN MOOD DISORDERS

Magnus Lekman



**Karolinska
Institutet**

Stockholm 2014

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by Åtta.45 Tryckeri AB

© Magnus Lekman, 2014

ISBN 978-91-7549-570-5

Genetic studies in mood disorders

THESIS FOR DOCTORAL DEGREE (Ph.D.)

by

Magnus Lekman

Principal Supervisor:

Associate Professor Ingrid Kockum
Karolinska Institutet
Department of Clinical Neuroscience
Division of Neuroimmunology Unit

Opponent:

Professor Thomas Schulze
University of Goettingen, Germany
Department of Psychiatry and Psychotherapy

Co-supervisor(s):

Professor Ola Hössjer
Stockholm University
Department of Mathematics
Division of Mathematical Statistics

Examination Board:

Professor Nancy Pedersen
Karolinska Institutet
Department of Medical Epidemiology and
Biostatistics

Professor Maria Anvret
University of Gothenburg
The Sahlgrenska Academy

Associate Professor Catharina Lavebratt
Karolinska Institutet
Department of Molecular Medicine and
Surgery
Division of Neurogenetics

Associate Professor Lars Feuk
Uppsala University
Department of Immunology, Genetics and
Pathology

ABSTRACT

Mood disorders, including bipolar disorder (BPD) and major depressive disorder (MDD), are highly complex psychiatric disorders. Decades of genetic studies have generated a large number of putative genetic susceptibility variants. However, with exception of *CACNA1C*, *SYNE1* and *ANK3* in BPD no robust association has as yet been identified. In this thesis my aim was to find predisposing genetic risk factors for mood disorders.

In paper I, my hypotheses were that the *FKBP5* gene is a risk gene for MDD, contributes to severity and is involved in treatment response to an antidepressant, Citalopram. We tested for association of three markers using the STAR*D cohort (Level 1). Rs1360780 was significantly associated with MDD. Rs4713916 was significantly associated with remission following treatment with Citalopram when two study populations were analyzed together. We determined that there is a stratification issue and no correlation can be made for treatment response or severity of MDD. **In paper II**, my hypothesis was that candidate genes for MDD are acting synergistically to contribute to risk for MDD. We applied three different algorithms to evaluate SNP-SNP interactions. Although none of the interactions survived corrections for multiple comparisons, our results contribute valuable information to future genetic studies in MDD. The logistic regression methods identified large interaction effects. None of the top interactions explain a large proportion of MDD in the general population. Among the top interactions, none of the three algorithms identify identical pairs of markers for risk of MDD. Moreover, none of the top ranked interactions has previously been implicated to act synergistically in MDD-susceptibility. We also noted that markers selected for predicted interaction effects were not among the top interactions. **In paper III**, my hypothesis was that rare and highly penetrant large structural genomic variations (CNVs) increase the risk for BPD. We searched for CNVs across diagnostic boundaries and included individuals with BPD, schizophrenia (SZ) or schizoaffective disorder (SA). To increase the possibility that the CNV should be highly penetrant we searched for CNVs in affected individuals in BP-pedigrees and identified CNVs in the *MAGII* gene in two families and showed that it was more frequent in individuals with BPD, SZ and SA than controls. **In paper IV**, my hypothesis was that inherited CNVs contribute with risk to BPD irrespective of their frequency. We developed an algorithm that combines linkage-data with the CNV content within and across families. We identified one significant region with a CNV that maps to 19q13, and stretches over the *PSG* genes. The *PSG* proteins has been shown to activate TGF- β . Moreover, two CNV SNPs are reported as likely eQTL's for regulation of NF κ B. Thus, this CNV involves several putative molecular targets in BPD-etiology.

In conclusion, the work conducted in this thesis has contributed to our knowledge of the etiology of mood disorders. For BPD we found two new susceptibility loci. In the analysis of MDD we increased the knowledge of the genetic interacting landscape in 63 candidate genes. We also showed that *FKBP5* is associated with risk for MDD.

LIST OF SCIENTIFIC PAPERS

- I. **Magnus Lekman**, Gonzalo Laje, Dennis Charney, Augustus J. Rush, Alexander F. Wilson, Alexa J. M. Sorant, Robert Lipsky, Stephen R. Wisniewski, Husseini Manji, Francis J. McMahon, and Silvia Paddock

The FKBP5-Gene in Depression and Treatment Response – an Association Study in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Cohort.
Biological Psychiatry, 2008;63:1103-1110

- II. **Magnus Lekman**, Ola Hössjer, Peter Andrews, Henrik Källberg, Daniel Uvehag, Dennis Charney, Husseini Manji, Augustus J. Rush, Francis J. McMahon, Jason H. Moore, and Ingrid Kockum

The Genetic Interacting Landscape of 63 Candidate Genes in Major Depressive Disorder: An explorative Study.
Manuscript

- III. Robert Karlsson, Lisette Graae, **Magnus Lekman**, Dai Wang, Reyna Favis, Tomas Axelsson, Dagmar Galter, Andrea Carmine Belin, and Silvia Paddock

MAGI1 Copy Number Variation in Bipolar Affective Disorder and Schizophrenia.
Biological Psychiatry, 2012;71(10):922-930

- IV. **Magnus Lekman**, Robert Karlsson, Lisette Graae, Ola Hössjer, and Ingrid Kockum

A Significant Risk Locus on 19q13 for Bipolar Disorder Identified Using a Combined Genome-wide Linkage and Copy Number Variation Analysis.
Manuscript

Additional publication

Magnus Lekman, Silvia Paddock, Francis J. McMahon

Pharmacogenetics of Major Depression: Insights from Level 1 of the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial.
Mol Diagn Ther. 2008;12(5):321-30. Review.

CONTENTS

1	Introduction	1
1.1	Mood disorders – Origin, evolution and current diagnostic system .	2
2	Theories for the underlying etiology of mood disorders.....	4
2.1	Overview	4
2.2	Limbic system	4
2.3	HPA-axis	6
2.4	Glutamate hypothesis	7
2.5	Neurotrophin hypothesis	7
2.6	Monoamine hypothesis	7
2.7	Other molecular mechanisms.....	8
3	The genetic landscape of mood disorder	9
4	Copy number variation.....	14
4.1	Background.....	14
4.2	Formation of CNVs.....	15
4.3	Implications of CNVs on human phenotype and disease	18
5	Finding genetic risk factors for mood disorders.....	20
6	Hypothesis	23
7	Aims.....	24
8	Materials	25
8.1	Study population I; STAR*D (papers I & II).....	25
8.2	Study population II; GAIN MDD (paper II)	25
8.3	Study population III; NIMH BP-pedigrees (papers III & IV)	26
8.4	Replication samples (paper III).....	27
9	Methods	28
9.1	Candidate genes, genotyping and marker selection	28
9.1.1	STAR*D (papers I & II)	28
9.1.2	GAIN (paper II).....	28
9.1.3	NIMH BP-pedigrees (papers III & IV)	28
9.1.4	Quality control	29
9.2	Statistical models to map risk loci to mood disorders.....	29
9.2.1	Background	29

9.2.2	Linkage analysis.....	31
9.2.3	Association analysis.....	32
9.2.4	Gene-gene interaction analysis.....	32
9.2.5	Approach to find inherited CNVs with risk to BPD.....	37
10	Results	39
10.1	Paper I.....	39
10.2	Paper II.....	42
10.3	Paper III	46
10.4	Paper IV	47
11	Discussion	50
11.1	Paper I.....	50
11.2	Paper II.....	51
11.3	Papers III & IV	53
12	Future perspectives.....	56
13	Acknowledgements	58
14	References	62

LIST OF ABBREVIATIONS

ASM	Affection status model
BDNF	Brain-derived neurotrophic factor
BPD	Bipolar disorder
BP-I	Bipolar disorder type 1
BP-II	Bipolar disorder type 2
CV-CD	Common variant common disease
CD-RV	Common variant rare disease
CIDI	Composite International Diagnostic Interview
CNV	Copy number variation
CRH	Corticotrophin releasing hormone
CVC	Cross-validation consistency
GR	Glucocorticoid receptor
GRE	Glucocorticoid response elements
DSM	Diagnostic and Statistical Manual of Mental Disorders
DZ	Dizygotic twins
eQTL	Expression quantitative trait locus
FISH	Fluorescence <i>in-situ</i> hybridization
fMRI	Functional magnetic resonance imaging
FoSTeS	Fork stalling template switching
FWER	Family-wise error rate
GAIN	Genetic Association Information Network
GWAS	Genome-wide association study
HLOD	Heterogeneity logarithm of odds
HPA-axis	Hypothalamic pituitary adrenaline axis
HRSD ₁₇	Hamilton Rating Scale for Depression 17-item
IBD	Identical-by-descent
IBS	Identical-by-state
LCR	Low copy repeats
LD	Linkage disequilibrium
LOD	Logarithm of odds
MDD	Major depressive disorder
MZ	Monozygotic twins
MRI	Magnetic resonance imaging
MAOI	Monoamino oxidase inhibitor

NAHR	Non-allelic homologous recombination
NHEJ	Non-homologous end joining
NIMH	National Institutes of Mental Health
NMI	Never mentally ill
NPL	Non-parametric linkage
OR	Odds ratio
PAF	Population attributable fraction
PCR	Polymerase chain reaction
PET	Positron emission tomography
PFC	Prefrontal cortex
QC	Quality control
QIDS-C ₁₆	Quick Inventory of Depressive Symptomatology Clinician 16-item
RDC	Research Diagnostic Criteria
ROH	Runs of homozygosity
RUDD	Recurrent unipolar depressive disorder
SA	Schizoaffective disorder
SABP	Schizoaffective bipolar type
SZ	Schizophrenia
TF	Transcription factor

1 INTRODUCTION

Mood disorders (also known as affective disorders) are psychiatric disorders that cause an abnormal and uncontrollable shift in a person's mood. Mood disorders are divided into *major depressive disorder* (MDD) and *bipolar disorder* (BPD). Both MDD and BPD are diverse groups of disorders and several subcategories are used for their classification (1).

Major depressive disorder, also known as *unipolar disorder*, is manifest by an abnormally low mood and periods of persistent anhedonia with a sad mood and pessimism, loss of interest, low energy, problems with concentration and decision-making, feelings of worthlessness or emptiness, sleeping and eating disturbances. Thoughts of death and suicide are common. Psychomotor disturbances frequently occur which often consists of slowing down of locomotor expression. Agitation is occasionally the locomotory sign (1, 2). Bipolar disorder manifests by shifts in mood that alternate from abnormally elevated mood (mania) to periods of depressed mood. The manic phase involve inflated self-image and increased self-esteem, irritability, racing thoughts, decreased need for sleep, acting impulsively and having high-risk behavior. The psychomotor activity is characterized by accelerated activity both in speech and locomotion. The symptoms during the depressed phase are similar to those observed in MDD (1, 2). Persons who are affected by BPD rarely experience a single episode of symptoms. The most common course of the disorder is a recurrent or chronic abnormal variation in mood which can be depressed (MDD) or cycling, or a mixed state of depressed and elevated mood (BPD) (1, 3).

Mood disorders are highly prevalent, affecting people of all ages, from children to the elderly, of both sexes as well as across race and ethnicity (4-7). MDD and BPD are conditions that have a severe impact on a person's normal daily life activities and the World Health Organization (WHO) ranks MDD and BPD among the top leading causes of impairments on quality of life (8, 9). Clinical observations and biological research unambiguously demonstrate that mood disorders are brain disorders with disrupted brain function as the under lying cause. This means that all facets of a persons' life will be involved. Dysregulation of cognitive processes, autonomic and endocrine functions affect the person's physical and mental health (3). Research has demonstrated that there is an accumulated effect from an unexpectedly large number of genetic and environmental risk factors that are likely to interact with each other to impact on vulnerability to disease, rather than a single causative factor. MDD and BPD are thus both highly complex disorders. In parallel to their complex nature, mood disorders are highly heterogeneous whereby different risk factors have central roles in different subtypes of these disorders (1, 4). Although research has increased understanding of the disease causing mechanisms and why people are at risk the exact etiology is still poorly understood. A variety of treatment options (both pharmacological and cognitive therapies) are available for MDD and BPD. Although the acute efficacy of treatments tends to be satisfactory, with both high response- and remission rates the long-term effects are largely unknown. A recognized problem is the large variation in inter-

individual response effects to certain treatment options. Since the hallmark of MDD and BPD is a cycling feature with recurrent symptoms, there is a need to evaluate the efficacy of available treatments that results in short and long-term remission effects, respectively (10-12).

Taken together, MDD and BPD constitute a spectrum of psychiatric disorders that have a high prevalence, morbidity, and mortality (13-17), resulting in a substantial burden, for both for the individual as well as society in general (18, 19). It is therefore of a critical concern, for both those who are affected and for the public in general to improve the understanding of the underlying causes of these disorders. Improvements may finally lead to better prevention, patient management and ultimately curative treatments. The scientific community as a whole has therefore made attempts to accelerate research in these disorders by organizing and working in large collaborative projects and consortia (20, 21).

1.1 MOOD DISORDERS – ORIGIN, EVOLUTION AND CURRENT DIAGNOSTIC SYSTEM

MDD and BPD have been established as psychiatric diagnoses through a process that can be traced back to a couple of thousand years. Mania and melancholia were already described as two different mental disorders by Hippocrates (460-337 BC), although these disorders had also been described by Greek physicians even before this time period. During the 1st century CE Aretaeus of Cappadocia described for the first time the possible connection between mania and melancholia and suggested them as two different images of the same disorder (22). Since then many physicians and psychiatrists have described these conditions (21, 22). However, it was not until the end of the 19th century that schizophrenia (SZ) and affective disorder became two separated entities, when Emil Kraepelin established a new paradigm that dichotomized insanity into “dementia praecox” and “manic-depressive insanity”. The Kraepelin categorization led to a unification of affective disorders into one group (manic-depressive insanity). Contemporary leading psychiatrists opposed Kraepelin, however, and argued that different but subtly affective syndromes occur, including unipolar depression (22). The classification of mental disorders into different diagnoses continued, which in the middle of twentieth century culminated in a classification of *manic-depressive reaction* and *depressive reaction* in the American Psychiatric Association’s first edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-I) 1952. In the second edition of DSM (1968) depressive neurosis was described, and in the 3rd edition, DSM-III, (1980) diagnoses of major depressive disorder and bipolar disorder were described for the first time (23). In early 2013, the 5th edition of the DSM classification system was published.

The DSM was developed by the American Psychiatric Association [24] and the International Statistical Classification of Diseases and Related Health Problems (ICD),

was developed by the World Health Organization (WHO). Of note, although these diagnostic systems have evolved and accentuated under a long period of time, throughout subsequent editions in two independent manuals, there is yet no established protocol for a molecular, genetic- or other laboratory tests that can be performed to verify a diagnosis. The only diagnostic tool available today is interview-based and focuses on the persons' own experiences. It is often laborious and difficult in the clinical situation to determine a correct diagnosis due to the relatively discrete boundaries between the different subtypes of mood disorders. In mild cases there is also an ambiguous distinction between diseased and non-diseased (4). In some literature mood disorder is also known as *affective disorder*. This classification is replaced with *mood disorder* in the ICD-10 and DSM-IV since in some European classification affective disorder refers to include anxiety states (1).

Taken together, endpoint diagnoses of the various subtypes of MDD and BPD are a final consensus entity of the contemporarily divergent and evolving field of psychopathology. A diagnosis of mood disorders should hence not only be seen in perspective of current knowledge and standards of the medical discipline but also in the perspective of social, ideological, political, cultural and the moral standards of its time (24).

2 THEORIES FOR THE UNDERLYING ETIOLOGY OF MOOD DISORDERS

2.1 OVERVIEW

It has become clear that both psychosocial and neurobiological factors underlie the etiology of mood disorders. These factors are not mutually exclusive but rather are interdependent and likely to act together to manifest into a mood disorder (25). Our current knowledge is nonetheless rudimentary in understanding the exact mechanisms for how these risk factors connect into different networks and predispose to mood disorders. Although our current knowledge is limited we have gained a lot of insight into the disease causing mechanisms through the results from clinical observations, animal models, pharmacological studies, post mortem studies, advances in brain imaging and from genetic studies. These lines of evidence have converged into theories which suggest that abnormalities of the brain nerve circuits, dysregulation of neurotransmitters, neurotrophin systems and immunological mechanisms are the underlying causes of mood disorders (25-28). These theories will be summarized below to be able to understand the scope of the problem and to motivate the two main genetically approaches, candidate gene analyses and genome-wide scans, which were both applied in this thesis.

2.2 LIMBIC SYSTEM

Since the limbic system was identified to have a role in controlling emotions there have been a plethora of studies focusing on the limbic structures, i.e. amygdala, hippocampus, thalamus and hypothalamus, with trajectories to the prefrontal cortex and brain stem regions, for understanding the basis of mood disorders (29). The prefrontal cortex (PFC) integrates sensory-motoric input with motivation and regulates emotion, mood, locomotory systems, neuroendocrine and autonomic functions through nerve circuits to both the limbic system and brain stem region (see Figure 1 for illustration) (30-32). The Striatum, nucleus accumbens and amygdala mediates rewarding to emotional stimuli, and thus, may thus explain the loss of motivation and anhedonia that is one dominating characteristic feature of depressive disorder. Activation of the limbic system will also activate regions in the brain stem and midbrain that in turn mediates pain modulation and locomotor functions.

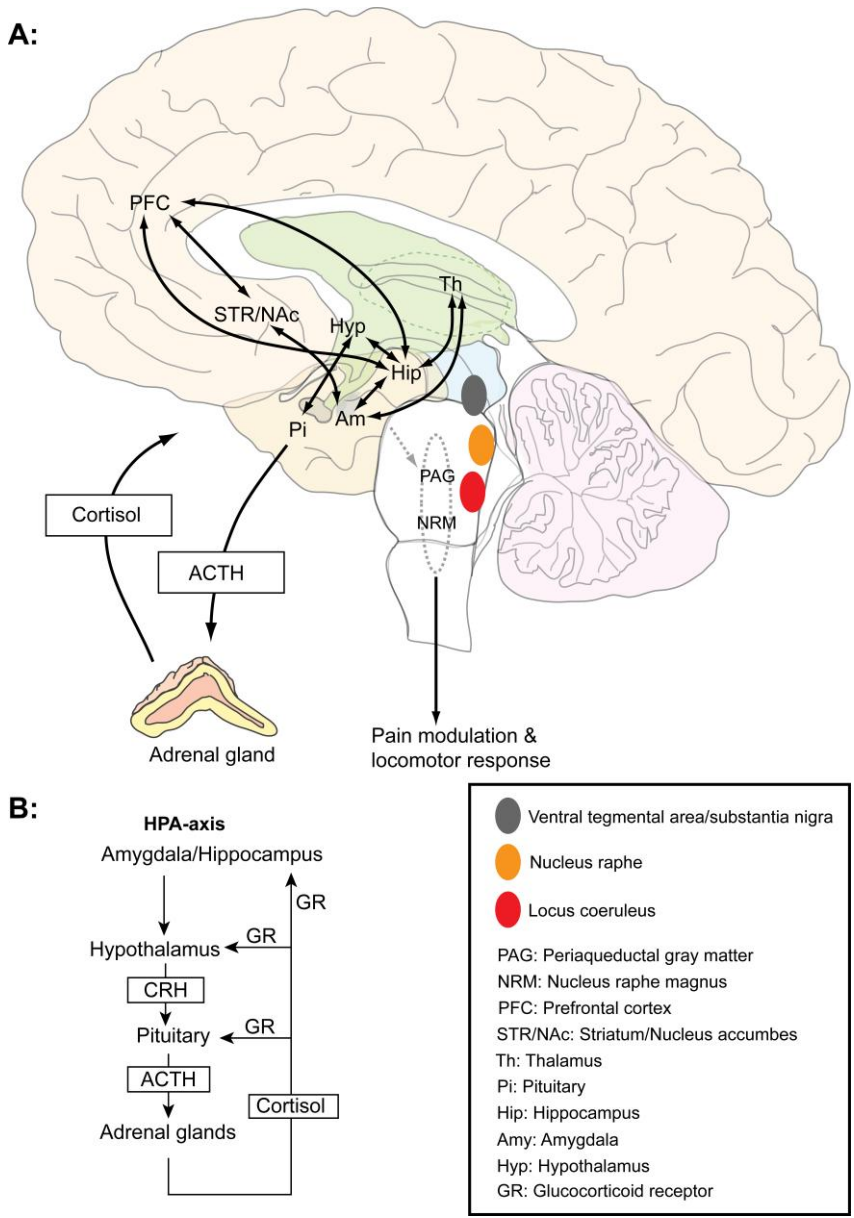


Figure 1. Brain circuits in mood disorders.

A: Connecting nerve circuits between various brain regions that have been suggested to be involved in regulation of mood disorders.

B: Feed-forward and feed-backward regulation of the HPA-axis.

These circuits explain the dysregulation in pain modulation and psychomotor impairments that are observed in patients with mood disorders. Brain imaging studies, using magnetic resonance imaging (MRI and fMRI) and positron emission tomography (PET) technologies, have demonstrated strong correlation between mood disorders and structural and functional abnormalities in PFC, the limbic system and their connections (33-37). The structural and functional abnormalities that are connected with mood disorders consist of a decrease in brain volume and in decrease of blood flow in the limbic-PFC circuits. These abnormalities can be reversed by treatment with pharmacological compounds (38). These findings are not consistent, however, and many questions remain unsolved. In fact, some studies contradict the change of brain volume in patients with mood disorder (28, 39).

2.3 HPA-AXIS

Dysregulation of the hypothalamic-pituitary-adrenal (HPA)-axis is a well established finding in mood disorders (40). The theory of the HPA-axis incorporates psychosocial stress into neurobiology and reactivity to mood disorder. Nerve circuits from PFC to the limbic system activate the HPA-axis as a response to stress. Studies in both humans and in animal models have shown that experiences of stressful life events will activate a feed-forward cascade in which the hypothalamus releases corticotrophin releasing hormone (CRH) which in turn activates the pituitary gland to release adrenocorticotrophic hormone (ACTH) (41). The adrenal glands are activated by ACTH to produce the stress hormone cortisol which prepares the entire body, including the brain, to adapt to the stressful event (Figure 1). Cortisol will act on glucocorticoid receptors located in the amygdala, hippocampus, hypothalamus and pituitary. The lipophilic properties of the glucocorticoids (cortisol) allows for diffusion over the plasma membrane into the cytosol where it binds to the glucocorticoid receptors (GR). This ligand-receptor complex translocates into the nucleus where it acts as a transcription factor. Upon binding to glucocorticoid response elements (GRE) it induces transcription of a variety of genes necessary for stress adaptation (41). A negative feedback loop is activated to ensure that, the active state of the HPA-axis successively normalizes after the stressful event has ceased. Chronic stress, or dysregulation due to mutation of key regulatory genes within the HPA-axis, will cause a disinhibition of the HPA-axis. This leads to pathological levels of cortisol and to decreased levels of glucocorticoid receptors (GR) (42-44), resulting in pathologically increased activation of the HPA-axis (28). Multiple brain regions convey afferent stimuli to the structures of the HPA-axis. Stress induced activation of the hippocampus and other limbic regions lead to expression of serotonin and noradrenaline neurons that in turn activates the hypothalamus and feed the forward cascade of the HPA-axis that progressively leads to increased level of cortisol (28). The stress hormone response pathway described as the HPA-axis is thus more comprehensive than the nerve circuits between hypothalamus-pituitary and the adrenal gland. The hypothalamus also innervates the locus coeruleus with activation of the autonomic nervous system in preparation for adaptive stress

response and maintenance of body homeostasis (45). Based on these observations several genes with key regulatory functions in the HPA-axis have become interesting targets for mood disorders (41, 46).

2.4 GLUTAMATE HYPOTHESIS

An accumulation of results indicates that mood disorders are influenced by dysregulation of the glutamatergic system (47). Glutamate transmission regulates synaptic plasticity and neuronal survival. Exactly how altered glutamate levels contribute to the pathophysiology underlying mood disorders is still unclear although there is growing evidence that glutamate mediate cognitive-emotional functions related to stress response through cell atrophy and cell death (48). Stress-induced atrophy in neurons of the hippocampus (49) has been attributed to dysfunction in glucocorticoid receptors (50). Moreover, glutamate has also been reported to be involved in stress mediated dysfunction of the HPA-axis (28, 51). Treatment with antidepressants has been reported to reverse structural and functional alterations that are observed in glutamate receptors in chronic stress models (48). Of note, the mood stabilizer compound lithium targets the NMDA receptor signaling cascade, underlining the significance of glutamate in regulation of mood disorders.

2.5 NEUROTROPHIN HYPOTHESIS

The neurotrophin hypothesis refers to reduced levels of neurotrophin levels in the limbic system which give rise to a reduction in cell survival, growth and differentiation. Such morphological changes have been a hallmark of mood disorders (28, 52). The structural variations in the brain observed in post mortem studies and neuroimaging studies of mood depressed individuals have been attributed to decreased levels of neurotrophic factors, e.g. the brain-derived neurotrophic factor (BDNF) (25, 53). Genes coding for neurotrophins, such as BDNF, its receptors and the signal transduction cascades that are involved in the regulation of neuroplastic events such as neural cell growth, axonal sprouting and synaptogenesis, involves a large number of plausible susceptibility candidates of mood disorders (25).

2.6 MONOAMINE HYPOTHESIS

A milestone in the understanding of the neurobiology of mood disorders is the *monoamine hypothesis*. This theory proposed that deficiency of monoamines forms the basis of depressed mood (54) and has been a prevailing hypothesis for understanding the etiology of mood disorders for the last 40 years (55). This theory has indisputable paved

the way for the development of several new drugs with better efficacy and fewer side effects. In a broad sense, antidepressant drugs used routinely in the clinic act either by inhibiting the reuptake (i.e. reuptake inhibitors, SSRIs) or by inhibiting the degradation (MAOIs) to increase synaptic levels of monoamine transmitters (52, 56). Pre-clinical results from human and animal models combined with the results from clinical studies have produced evidence for abnormal (i) synthesis (ii) receptor binding and (iii) degradation of the monoamines as a source of information to understand the disease mechanisms of mood disorders (56). How significant the role of monoamines is in the etiology of mood disorders is still intensively discussed. Given the complex etiology of mood disorders and the large variation in the efficacy of compounds based on the monoamine theory, there is an ongoing debate which questions whether this group of neurotransmitters will actually help in understanding the nature of mood disorders (52, 57).

2.7 OTHER MOLECULAR MECHANISMS

Additional theories should have merited a detailed description but will only be briefly mentioned. First, the innate immune system has recently been shown to associate with mood disorders and is a good candidate for further studies to unravel one of the many sources leading to vulnerabilities of MDD and BPD. The background is the observation that stress influence the function of pro-inflammatory cytokines (IL-1 and IL-6) and tumor necrosis factor alpha (TNF-alpha) (58). These cytokines have a role in neurotransmission and thus become putative candidates in dysregulation of neurotransmission in the brain (58). Other candidates that have been proposed to be involved in mood disorders are molecules that mediate functions in cell adhesion, DNA repair, epigenetic regulation, transcriptional modifiers and neurotrophic signaling (25, 52, 55). In summary, although the different theories described above are well established there are still many controversies and there is no consensus for the exact mechanisms of how these risk factors may affect brain nerve circuits and their reactivity to stress (28, 54). Models that are currently in use to study stress-induced variation regarding the central nervous system have major limitations. Clinical studies of depressed individuals have still not unraveled cause from effect. We should also keep in mind that not all persons who are exposed to stress will develop a mood disorder.

3 THE GENETIC LANDSCAPE OF MOOD DISORDER

Genetic epidemiological research of family and twin studies has consistently supported the genetic contribution to susceptibility to mood disorders. Caution should be made when interpreting risk estimates, due to the relatively large variability of inclusion criteria, diagnosis definitions, study designs and the number of included study individuals that have occurred over time. These estimates are substantially larger for BPD than what is reported for MDD (59, 60). A survey of the literature for familial aggregation reveals that in MDD the first degree relatives of depressed probands are 3-10 times more likely than controls to be affected by MDD (61-63), while first degree relatives to BPD-diagnosed probands are 7 times more likely to develop bipolar disorder (64, 65). Although measurements of concordance rates in monozygotic twins compared to dizygotic twins show a large variability they support the contribution of a substantial genetic background for BPD (8 times higher in MZ compared to DZ twins) and to a lesser extent MDD (3 times higher in MZ compared to DZ) (61) (64). The concordance rate for MZ twins is less than 100% (< 70% in BPD and < 50% in MDD) which underlines the additional influences of non-genetic risk factors (61, 64). Although the heritability rates (derived from concordance rate estimates in twin studies) vary between studies they are higher for BPD (71-93%) (66-68) than for MDD (~37%) (61). Genetic epidemiological studies further reveal that there is a shared genetic background between BPD and MDD. A large Danish study reported that relatives to BPD affected individuals have an increased risk for developing depressive disorder, schizophrenia as well as other mental disorders (69). From other population survey studies there is evidence for an increased frequency of occurrence of mood disorders in close relatives to probands with BPD or SZ (65, 70-72). Another observation is that the relative risk of mood disorders declines substantially when genetic relatedness diminishes (64, 73, 74). Attempts to understand the mode of inheritance have largely been ineffective in establishing a simple mode of inheritance. Results from a large number of linkage studies indicates that mood disorders have a complex mode of inheritance and leaves no room for a model representative of few genes predisposing for either MDD or BPD.

Although the extensive reports from family and twin studies provides strong evidence that genetic factors confer risk to mood disorders, linkage and association approaches have not delivered any robust and convincingly replicated predisposing gene or risk locus. Those genes that have emerged in single studies, both in candidate gene and genome-wide association studies (GWAS) approaches, only confer risk with a relative small effect sizes with odds ratios (OR) generally ranging between 1.1 and 1.4. Linkage analyses face the same reality, and no region has yet yielded strong evidence for linkage in genome-wide scans. Figure 2 summarizes association and linkage studies performed up to the present time (73, 75-81). These results show that an exceptional number of regions have been reported to influence susceptibility to mood disorders, but with a general absence of overlap between independent studies. In spite of this complicated and puzzling picture there are some interesting observations. Firstly, in none of the eight GWAS in MDD that has been published, did any locus reach GWAS significance level

(82-89). Neither was a recent meta-analysis successful (90). However, when an approach was applied to increase the spectrum of a shared genetic background using a combined dataset of MDD and BPD samples (known as a cross-disorder meta-analysis), the GWAS yielded several significant markers in 3p21.1 (90). Secondly, genetic studies in BPD have been somewhat more rewarding than for MDD (8, 91). The first waves of BPD GWAS failed to reach conventional GWAS significance levels for the top candidate signals. The following waves with larger sample sizes have been more successful, where four, *ANK3*, *CACNA1*, *NCAN* and *ODZA*, reach genome-wide significance. Additionally, a few sub-threshold candidates have been reported which overlap with 3p21.1 in the MDD-BPD cross-disorder analysis (8, 90). These results together with familial studies, support clinical and genetic overlap between mood disorders. A shared genetic background has also been reported between BPD and SZ (92, 93). Accumulating data thus demonstrates that MDD and BPD are both clinically and genetically heterogeneous disorders with different risk factors being important in different subgroups. These results also reveal that these disorders are exceptionally complex with multiple genetic and environmental risk factors predisposing to vulnerability. Such risk factors are presumably not acting on their own, but are rather part of networks with interacting effects between multiple genes and environmental exposures.

A recent discovery that may explain a part of the elusive genetic risk to psychiatric disorders is represented by a variety of genomic polymorphisms. These genomic changes are not only restricted to variation at individual nucleotide bases such as SNPs in the DNA sequence (93-96).

Figure 2A:

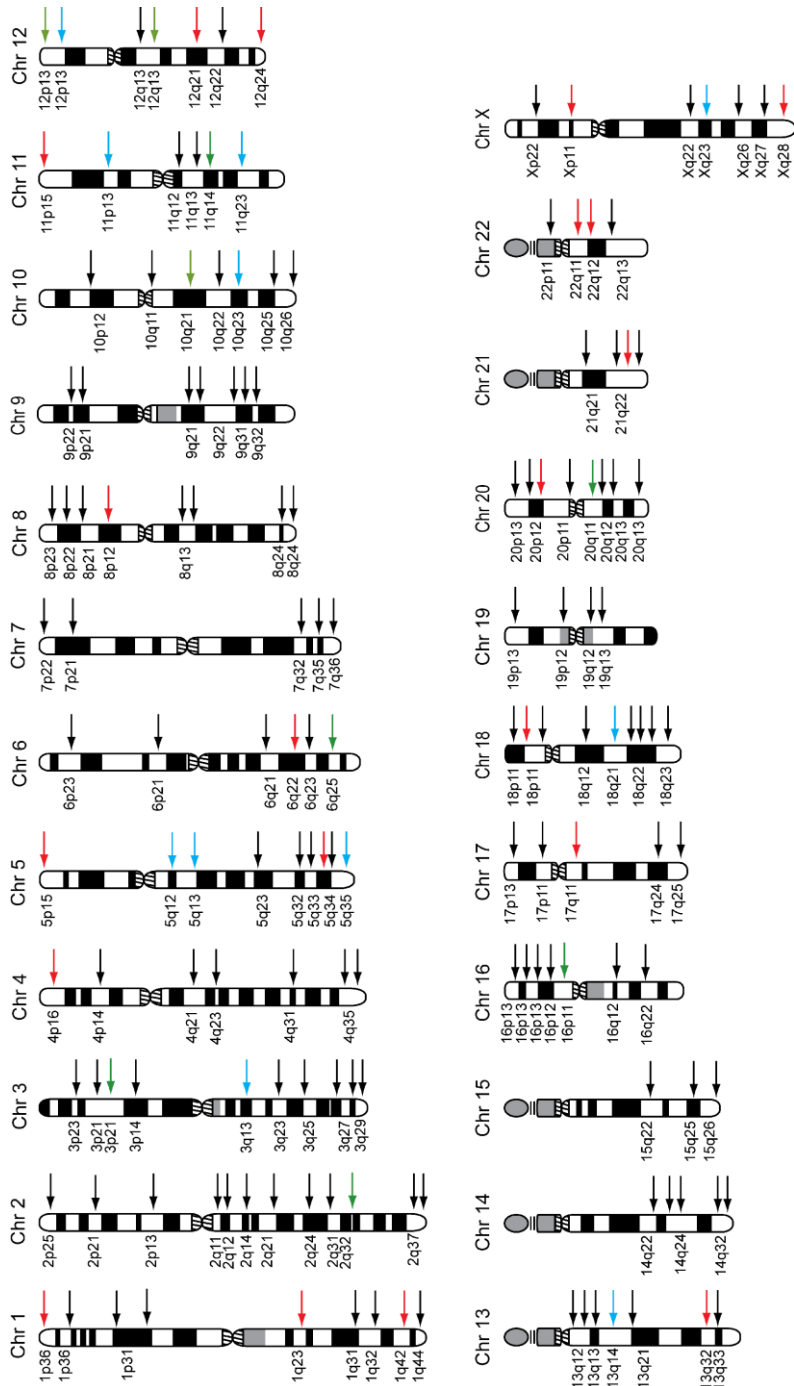


Figure 2B:

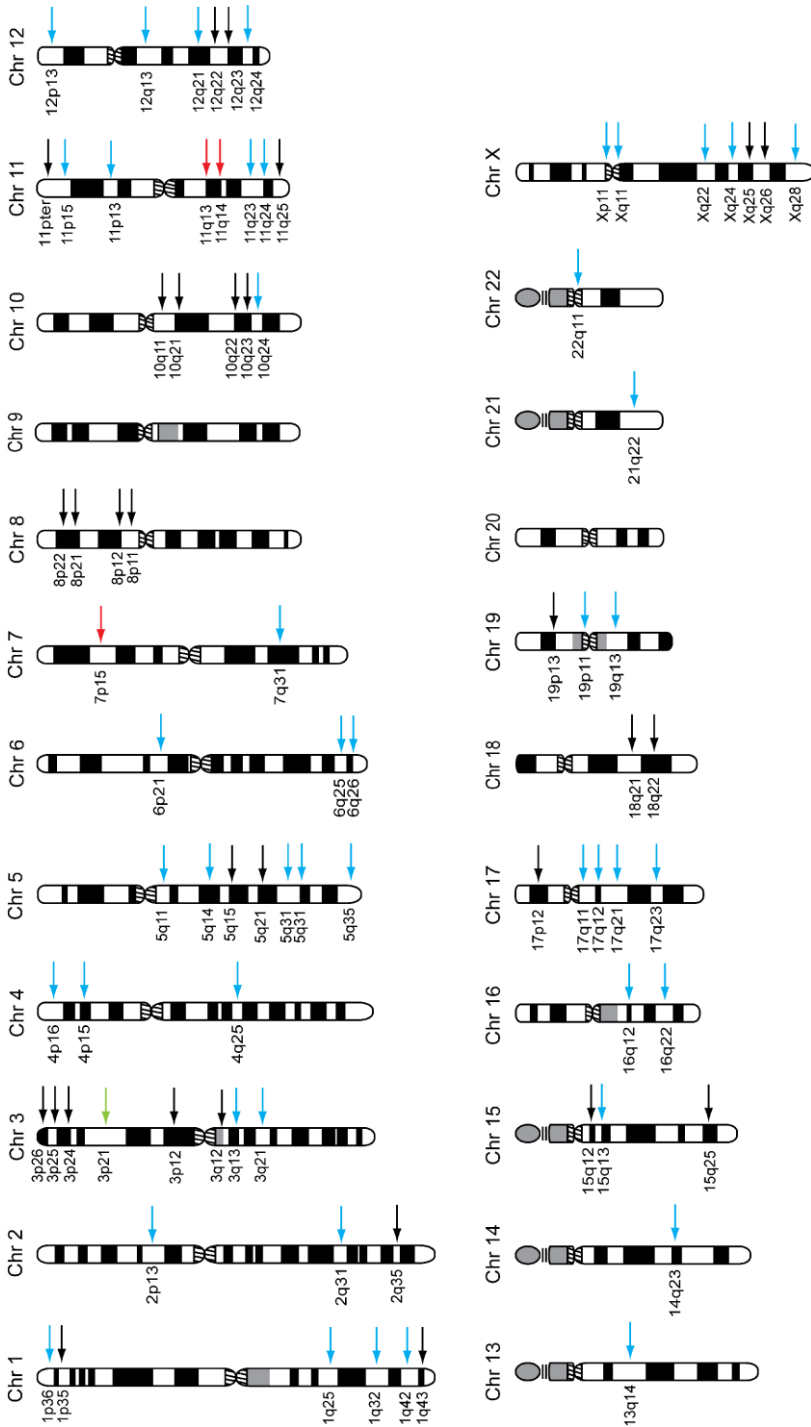


Figure 2. Karyotype illustrating genome regions for which linkage and association have been reported in BPD and MDD.

Black arrows: region with linkage.

Blue arrows: region with gene from association study surviving significance level.

Green arrows: region with gene at GWAS significant level.

Red arrows: region with associated gene reported in region with linkage.

A: Results of analysis in BPD reported since 1990. Arrows represent regions with at least suggestive linkage from genome-wide linkage scans and from candidate gene studies and GWAS reaching study-wise significant levels.

B: Results of analysis in MDD with at least suggestive linkage from genome-wide linkage scans reported in the literature since 2001, and from candidate gene studies and GWAS reaching study-wise significant levels, as reviewed in 2007.

Ideogram has been derived from David Adler and with permission for publication. (<http://www.pathology.washington.edu/research/cytopages/idiograms/human/>)

4 COPY NUMBER VARIATION

4.1 BACKGROUND

Due to rapid advances in the development of molecular biology techniques we now have a firm understanding that variation in the human genome is far more complex and diverse than microscopic visible aberrations. The diversity of such variations comprises aberrations involving gain or loss of entire chromosomes (aneuploidy) to changes at the level of single nucleotide base pairs within the DNA. Several lines of evidence clearly illustrate that DNA rearrangements involve a diverse form of variable DNA segments that causes the genomic architecture to differ between any two individuals (97, 98). Historically, large structural abnormalities were detected using microscopically visualized stained and non-stained chromosomes. In contrast, the techniques currently used, such as PCR-based sequencing, can identify variations at the level of a single base-pair and repetitive segments such as micro- and minisatellites, ranging from a few base pairs up to a couple of hundreds of base pairs in size (99). With the advent of technologies such as fluorescence *in-situ* hybridization (FISH) and more recently of high-throughput arrays (clone-based and SNP genotyping-based), the capability to interrogate the genome at a high resolution dramatically improved and larger segments, mega-bases in size, were further identified (99). Within this group, the low copy repeats (LCR) (defined as short repetitive sequences that occur a few times in one diploid genome compared to another) and copy number variants (CNVs) (defined as deletions, duplications, inversions and translocations ranging between 1 kb to 3 Mb in one diploid genome compare to another) have attracted a great deal of interest in understanding the diversity of human traits since these forms of variants contains a substantial part of DNA, including genes and regulatory regions (100-102).

An overview of different forms of human genomic variations, mutation rate and technologies used for their detection is provided in Figure 3. Investigation of the CNV content in non-diseased individuals reveals that CNVs contain a large proportion of the human genome (~12%), frequently overlap with genes (~13%) and frequently alter the gene transcripts (101-103). CNVs are likely to be an important source of genomic variations with influences on survival and fitness since reported CNVs often map to genes responsible for cell adhesion and immune responses as well as responses to environmental stimuli (101). Cross-population (102, 103) and cross-species CNV-based analyses (104, 105) demonstrate that CNVs are highly preserved across populations and species, thereby suggesting that CNVs are have a role in evolution and adaptation. Of particular interest, several independent studies have recently linked CNVs to a variety of complex disorders (103, 106) including psychiatric disorders (94, 96, 107).

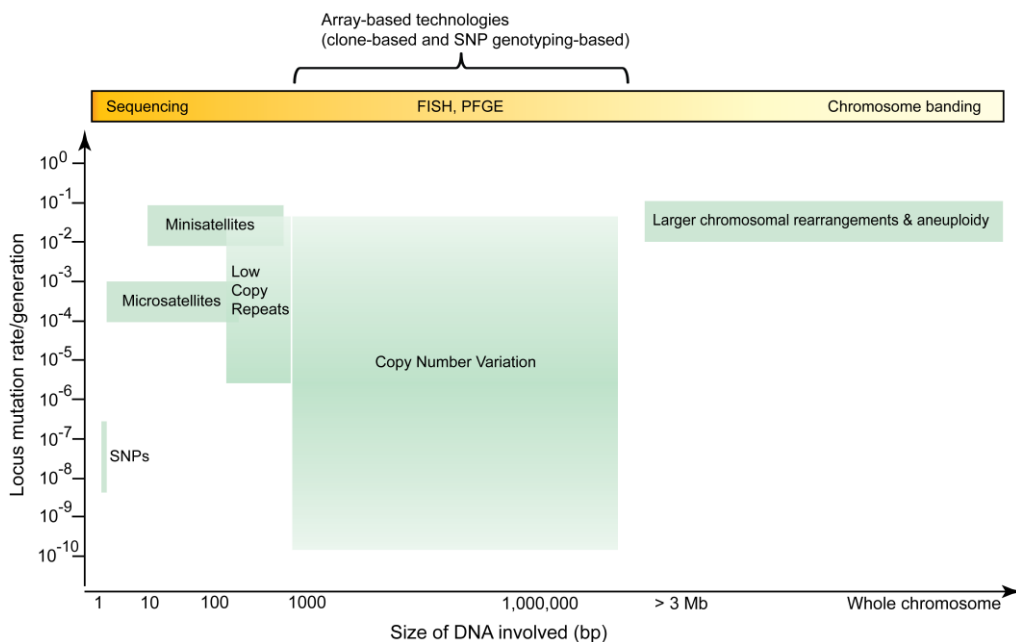


Figure 3. Genomic variations, mutation rates and resolution of detection methods.

Different classes of genomic variation in the human genome with corresponding mutation rates in correlation to size of the DNA involved. The top panel illustrates the detection methods. Variation in the DNA sequence occurs from the level of one single base pair to the entire chromosome. The figure shows that CNVs are highly variable, comprise a substantial part of DNA and that detection methods have an optimum range to find various forms of genomic variations. Abbreviations: FISH; fluorescence *in-situ* hybridization, PFGE; pulsed-field gel-electrophoresis.

The figure is adapted from ref. 103 and 111. Data included in figure are compiled from ref. 101, 103, 104, 108 and 110.

4.2 FORMATION OF CNVS

Various models of both recombination- and replication based mechanisms are widely accepted for formation of CNVs (108, 109). A well documented mechanism underlying rearrangements, studied in different model systems, has revealed that genomic regions enriched for low-copy repeats (LCR), mediates formation of non-allelic homologous recombination (NAHR) (106). During cross-over, the presence of large segmental duplications located within a close distance, possessing a high degree of sequence

similarity, enables misalignment of homologous chromosomes and formation of NAHR. Mismatching may also occur between sister chromatids or within the same chromatid arm (Figure 4 for illustration) (100).

Another form of recombination-based mechanisms for generation of CNVs is non-homologous end joining (NHEJ). The initiator of this form of chromosomal rearrangements is the repair of a DNA double-strand break without the need for LCR. Double strand repair mechanisms are present in virtually all organisms as DNA strand breaks are deleterious for all cells. Evolutionary preserved proteins are responsible for bringing the DNA ends together, bridging the gap between the non-matching ends and finally their ligation without the need for homologous sequences (Figure 4 for illustration). The repair region is prone to insertion of additional nucleotides resulting in a variable segment with insertions, deletions of various lengths (108, 110).

Experimental evidence for replication-error based genomic rearrangements has recently been proposed, termed *fork stalling and template switching* (FoSTeS) (108, 111, 112). The initial step of FoSTeS consists of an arrest of the DNA polymerase. The interruption in propagation of the replication fork allows for switching of the lagging strand between the original stalled and adjacent replication fork. The lagging strand disengages from its template and attach to another DNA polymerase. The template switch may occur with a DNA polymerase located in the physical proximity up/down stream of the original, or a DNA polymerase located on a homologous or non-homologous chromosome. Depending on the direction of the invading fragment and the direction of the replication fork different forms of genomic rearrangements are generated (111) (illustrated in Figure 4). Although CNVs are inherited they also occur as somatic mutations that results in differences in CNVs between different tissues in the same individual and between identical twins (113, 114).

Figure 4:

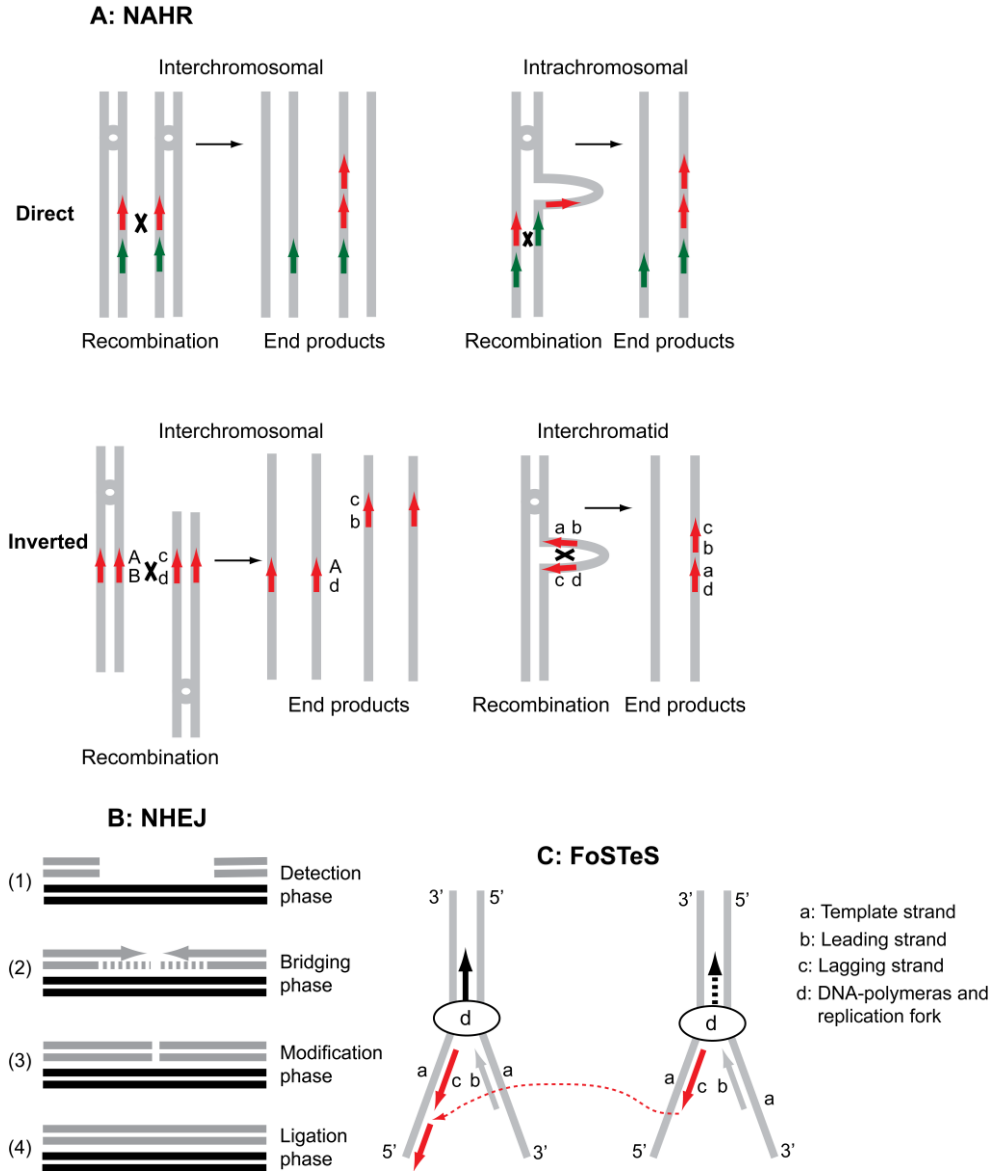


Figure 4. Formation of CNVs.

A: Non-allelic homologous recombination (NAHR). Chromosomes (grey), with centromere (grey circle), are illustrated with regions of low copy repeats (LCR) (colored arrows). Black crosses indicate recombination. The figure illustrates the principles for recombination-based genomic rearrangements, leading to formation of deletions, duplications, inversions and translocations.

B: Non-homologous end-joining (NHEJ). NHEJ does not require a homologous template. The successive steps contains; (1) detection, (2) bridging the two DNA-ends, (3) modification and (4) ligation. The repaired region may contain additional DNA sequences, and thus formation of extra segments and CNV.

C: Replication error, fork stalling and template switching (FoSTeS). Fork stalling (right) allows the lagging strand (c) to disengage from the template strand (a) and to anneal to an adjacent replication fork (template switch) at the 3' end. This event can lead to priming of continued DNA synthesis and formation of CNV. Figure is adapted from ref. 102, 112 and 114.

4.3 IMPLICATIONS OF CNVS ON HUMAN PHENOTYPE AND DISEASE

Exactly how CNVs affect gene expression and susceptibility to disease is not known in detail. Common CNVs identified in blood donors often overlap with established disease genes and it therefore seems that not all CNVs result in phenotypic change. However, CNVs have in numerous studies been reported to cause genetic imbalance and convey severe consequence to human health, leading to a variety of complex disorders through various mechanisms (106, 115).

CNVs represent variants of gain or loss and may alter the gene-dosage resulting in increased or decreased gene expression. Alteration in transcription levels may be the consequence of deletions or insertions in regulatory sequences. Transcription level modifications may also be induced from variations occurring in intronic or exonic sequences. These forms of variants may also cause splice variants. Inversions that flip an exonic region can lead to change of amino acid sequence, whereas inversions that flip a regulatory sequence may silence (or activate) a gene. Conversely, translocations will give rise to a multitude of different changes depending on the region involved. For example, a translocation of a gene into a different regulatory region of an active gene may lead to increased transcription, whereas translocations that involve part of the gene may cause truncated gene product (98, 116, 117).

There are many more potentially mechanisms through which CNVs may disrupt gene function including indirect transcriptional regulation through cis- and trans-regulatory elements, collectively known as 'positional effects'. Of interest, emerging evidence shows that CNVs may give rise to positional effects through gene regulatory elements of distances up to ~2 Mb up/down stream of the causative gene (106, 118, 119). A probable mechanism is alterations in DNA sequence which induces chromatin structural changes and epigenetic modifications that indirectly influence gene function (119). A fundamental mechanism, albeit often overlooked, concerns alterations in binding sites for transcription factors (TF), which will influence on transcription levels (120, 121). These observations indicate the uncertainty in predicting the various direct and indirect benign and pathological effects of the presence of CNV. It also apparent that unraveling the functional impact on CNVs on human diseases will be of a significant importance but also a significant challenge for future studies. In this respect several caveats must be taken into consideration. The most obvious problem is the lack of reference sample. Results from CNV should therefore be taken with carefully interpreted. In terms of computational strategies, the different CNV-calling algorithms differ widely in sensitivity of detecting CNVs (122). Finally, the conceptually different detection methods do not coincide in identification of CNVs and none of the methods seems to be adequate enough to discover all forms of genetic variations (101). For these reasons caution is advised before functional implications for certain CNVs can be established.

5 FINDING GENETIC RISK FACTORS FOR MOOD DISORDERS

As described in section 1.3, it is unlikely that a few causative risk factors (genetic or environmental) will be sufficient to lead to the development of a mood disorder. The extensive heterogeneity and the exceptional complex nature of mood disorders have made it difficult to pinpoint risk genes. The observation that detected genetic loci explain only a very small proportion of the genetic variance has been described as *the missing heritability* (123). Fundamentally different hypotheses, ranging from rare to common allele frequencies, have been used to explain the genetic susceptibility and may serve to explain the missing heritability (Figure 5). The two most well known hypotheses, that have contributed with valuable insights into the nature of the disorders, are the *common-disease common-variant* (CD-CD) hypothesis and the *common-disease rare-variant* (CD-RV) hypothesis. However, neither has been demonstrated to be perfectly adequate (further discussed in 5.4.1).

Of interest, promising knowledge has emerged. Firstly, genetic studies indicate that there is an overlap with shared genetic and clinical signatures between the subtypes of BPD and MDD. Severe forms of BPD (bipolar disorder type I) seem to share genetic and clinical manifestations with SZ whereas subtypes of MDD seem to overlap with anxiety disorders (92). Secondly, genetic studies have so far been more rewarding in BPD than in MDD which is mainly due to the larger heritability rates in BPD but also due to cross-disorder analyses in which the presumably shared genetic backgrounds between SZ and BPD have been an advantage. These observations contribute with valuable knowledge and may guide future studies attempting to reduce heterogeneity and to use forms of common clinical manifestations or common dysfunctional biological systems (gene, gene product, nerve circuit or signaling pathway) as ascertainment criteria.

The effect-size/allele-frequency correlation paradigm (Figure 5) may illustrate the rationale for using certain methods in the search of genetic predictors to common complex disorders such as mood disorders. As illustrated in the figure, no method by itself will be sufficient to cover all aspects of the spectrum of anticipated genetic risk factors. Thus, any chosen approach possesses its strength and weakness and a specific model will not yield a final solution to unravel all risk variables to mood disorder by itself. The heritability rates are higher in BPD but with lesser prevalence with the opposite relationship in MDD, family based study designs are being more suited to identify genetic risk factors for BPD whereas population based study designs are favor unraveling of genetic risk to MDD.

Risk loci at gradually lower allele frequencies and at low effect sizes (approaching towards the lower left quadrant in Figure 5) requires the use of a marker chip for coverage of rare alleles combined with sequencing technologies to be detected. As allele frequencies decline larger study populations are needed. Two general approaches are currently commonly used to address the problem of not robustly reporting genetic risk factors in mood disorder. First, increasing the sample size results in improved statistical

power and is a central solution. However, this approach also increases the chance of an expansion of the heterogeneity. Secondly, an alternative strategy would therefore be to reduce both genetic and environmental diversity with regards to underlying risk and thus search for predisposing factors in smaller groups of individuals with close similarities with regard to background risk. However, the strategy to use an increased sample size when looking for single genetic risk factors, within the paradigm of effect size and allele frequency, will not explain the missing heritability and help to elucidate the genetic architecture for complex mood disorders. The number of risk loci and the mode of interaction with other genetic loci and environmental exposures are also important factors. Incomplete LD between causal and tagging variants leads to reduced effect size estimates, and undetected rare variants may also contribute to the missing heritability (123, 124).

The allele-frequency/genetic-risk paradigm thus does not describe the full picture of the genetic landscape. Other factors that may explain the missing heritability, though not described in this thesis, are epigenetic influences (125) or just due to an overestimation of the heritability estimates (126).

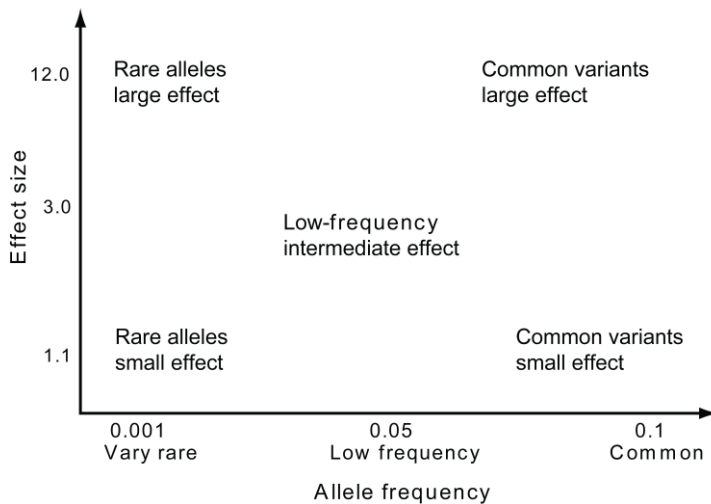


Figure 5. Correlation between effect size and allele frequency for human genetic traits.

The suggested underlying force for these estimates and correlations is selection pressure. The right lower quadrant in the figure represents disorders that have complex genetic background, i.e. many different risk loci, with a high allele frequency and at a low effect size. Disorders of this category are represented by the common disease common variant (CD-CV) hypothesis. Association studies of case-control design are preferentially used to search for genetic risk factors in this category of genetic polymorphisms. Alleles with high penetrance, but which are rare in the population, are depicted in the left-side of the figure. Disorders of this category are represented by the common disease rare variant (CD-RV) hypothesis. Study designs optimal for genetic polymorphisms in this category of disorders are family-based studies using linkage analysis. An increased focus has been directed towards the lower left quadrant to find very rare genetic risk underlying common complex disorders such as mood disorders.

The figure is adapted from ref. 125.

6 HYPOTHESIS

The overall hypothesis of this thesis was:

- Genetic variation from a single locus and interaction from pair-wise sets of loci contribute with risk to mood disorders.

The specific hypotheses in the separate projects include:

- The *FKBP5* gene is a risk gene for MDD and contributes to severity of MDD (paper I).
- The *FKBP5* gene is involved in the treatment response to antidepressants for individuals diagnosed with MDD (paper I).
- The candidate genes for MDD are acting synergistically to contribute with risk to MDD (paper II).
- Rare and highly penetrant large structural genomic variations (CNVs) contribute with risk to BPD (paper III).
- Inherited large structural genomic variations (CNVs), irrespective of the frequency of occurrence, contribute with risk to BPD (paper IV).

7 AIMS

Our overall aim of this thesis was;

- To find genetic predisposing factors to mood disorders.

The specific aims in the separate projects include;

- **Paper I.** In the first project we aimed to evaluate the involvement of the *FKBP5* gene in (i) efficacy to treatment of an antidepressant drug, (ii) to involvement in severity of MDD and (iii) for risk to MDD. These analyses were intended to replicate previous correlation studies of the *FKBP5* gene (127).
- **Paper II.** In the second project we aimed to find interactions between genetic variations that contribute to risk of MDD. We applied three different algorithms to evaluate SNP-SNP interactions and to map the genetic interacting landscape in candidate genes to MDD. We developed a new algorithm to search for genetic interactions on an additive scale.
- **Paper III.** In the third project we aimed to find rare large structural variants (CNV) that have high penetrance to BPD. Individuals diagnosed bipolar disorder, schizophrenia or schizoaffective disorders were included in our analysis.
- **Paper IV.** In the last project we aimed to identify inherited regions containing larger structural variations that irrespective of the frequency of occurrence confer risk to BPD. To achieve this we developed an algorithm that combines genome-wide linkage data with the CNV content within and across families, and searched for regions with CNVs that mediate risk to BPD.

8 MATERIALS

8.1 STUDY POPULATION I; STAR*D (PAPERS I & II)

The Sequenced Treatment Alternative to Relieve Depression (STAR*D) was a prospective treatment-response study designed for the clinical setting of evaluating efficacy of antidepressant treatments and care managing for persons diagnosed with MDD. Of note, no control group was ascertained (128). Participants were recruited during 2001 to 2004 from primary and specialty care practices at 14 regional centers, at 41 clinical sites across USA. The selected patients chosen were seeking their physician for symptoms of depression. Those who met criterion for DSM-IV MDD and who would receive medication or psychotherapy were asked to participate. Inclusion criteria were; age 18-75 years (without regard to race, ethnicity or sex), scoring ≥ 14 on the 17-item Hamilton Rating Scale for Depression (HRSD₁₇) and single or recurrent non-psychotic MDD. To be representative for the clinical situation the exclusion criteria were low (129). A diagnosis of bipolar disorder, schizophrenia, schizoaffective disorder, psychosis or obsessive-compulsive disorder led to exclusion. Otherwise the inclusion criteria were broad to ensure that the study participants (cases) would be representative of the clinical situation to optimize evaluation of the efficacy of the antidepressant treatment. Comorbidity from medical and other psychiatric disorders not described above were thus allowed (129).

For the genetic analyses a control sample obtained from the NIMH Genetic Initiative for Schizophrenia comprising 739 individuals was used. All control samples underwent a psychiatric screening for not having a diagnosis of bipolar disorder, schizophrenia or psychosis (130).

8.2 STUDY POPULATION II; GAIN MDD (PAPER II)

The second study population was from the Genetic Association Information Network major depressive disorder (GAIN MDD) study, a cohort study of major depressive disorder conducted in the Netherlands. Individuals with MDD were recruited from 4 cohort studies in the Netherlands. The main study was a longitudinal study of depression and anxiety, the Netherlands Study of Depression and Anxiety (NESDA), which aimed to examine the prevalence of psychiatric disorders in the Netherlands. The recruitment process took place between 2004 and 2007 and individuals with a diagnosis of MDD ascertained in 1966, 1977 and 1999 were enrolled (83). A smaller sample was also obtained from The Netherlands Mental Health Survey and Incidence (NEMESIS) Study (131), the Adolescents at Risk for Anxiety and Depression (ARIADNE) study (132), and finally from the Netherlands Twin Registry (NTR) (133). Enrolled individuals were mainly recruited from the primary care and mental health care organizations with a small proportion of participants arising from the municipality. All cases underwent a phenotypic characterization, using the Composite International Diagnostic Interview

(CIDI), to meet the stringent DSM-IV criteria for MDD. The study participants were aged 18-65 years (both sexes) and had self-reported western European ancestry. The exclusion criteria were bipolar disorder, schizophrenia, schizoaffective disorder, obsessive-compulsive disorder or substance dependence (83). Control samples were mainly from a longitudinal study, the Netherland Twin Registry (NTR), and were recruited period-wise every second to third year from 1991 until 2005. Only biologically unrelated individuals were included. A smaller subset was also recruited from the NESDA and ARIDANE studies. The control samples had no diagnosis of MDD or anxiety disorders (as assessed by CIDI) (134).

8.3 STUDY POPULATION III; NIMH BP-PEDIGREES (PAPERS III & IV)

The third sample was a family-based study population and was provided from the NIMH Bipolar Disorder Genetic Initiative (135), a nationwide effort in USA to provide resources for identifying the genetic basis of bipolar disorder. Families were collected from multiple sites from 1991 to 2003 through a screening process of clinical and non-clinical treatment facilities. Families were required to have a proband with a diagnosis of bipolar type 1 (BP-I) and a first-degree relative with diagnosis of BP-I or schizoaffective bipolar-type (SABP). The final collection of nuclear and extended BP-pedigrees included a spectrum of bipolar disorders; BP-I, BP-II, SABP or recurrent unipolar disorder (RUDD). The diagnostic instruments were DSM-III-R for BP-I and SABP and research diagnostic criteria (RDC) for BP-II and RUDD (136).

Given the complex background of etiology of BPD, our study design aimed at selecting families presumed to carry a genetic form of BPD. A screening was performed of 637 BP-pedigrees encompassing 3,849 individuals, based on family-wise genome-wide linkage analysis or by analyzing candidate genes for presence of large stretches of deletions in these families.

The specific criteria we used for selecting BP-pedigrees were a parametric family-wise LOD score > 1.1 , or if several families were found to have overlapping family-wise LOD scores of > 1.0 in the same genomic region. We also selected a family if several genomic regions with LOD scores close to but below 1.0 were identified. Linkage analyses were evaluated using several different transmission models and a microsatellite marker map with average marker distance of ~ 10 cM.

Additionally, a sample of 978 individuals was also screened for larger stretches of deletions in 357 candidate genes for BPD using a *runs of homozygosity* (ROH) analysis and a mendelian-error analysis. Finally, we selected 46 BP-pedigrees for our analyses comprising 277 individuals with DNA and 97 individuals for whom DNA was not available.

8.4 REPLICATION SAMPLES (PAPER III)

A collection of unrelated patients with a diagnosis of SZ, BPD or SA from 20 clinical trials conducted at Johnson & Johnson Pharmaceutical Research & Development were used for confirmation of the initial findings in paper III, and were obtained as a part of a collaborative project. The DNA samples were genotyped using the Illumina Human 1M-duo platform. Additional confirmation samples were obtained from a literature search in publicly available databases with information from original articles. In total 3,683 BPD, 7,242 SZ or SA, and 16,747 control samples were included. A complete list of these samples is provided in paper III.

9 METHODS

9.1 CANDIDATE GENES, GENOTYPING AND MARKER SELECTION

9.1.1 STAR*D (papers I & II)

The STAR*D study was a candidate gene study with genes selected on the basis of previous results of association to MDD and, for association to response to antidepressant treatment (Table S1 in paper II display the final collection of 68 the candidate genes) (137). Markers were selected to achieve an optimal coverage of the 68 candidate genes based on the reference sequence from HapMap phase I (CEU Nov 2004) with a minor allele frequency of at least 7.5%, and filtered for linkage disequilibria between markers with an $r^2 \geq 0.8$ (137). DNA samples from 1,953 patients and 739 controls were used for genotyping. Signal intensities from an Illumina bead array chip (130) were decoded using an algorithm that called genotypes with high accuracy (138). A smaller sub-set of markers was genotyped using a TaqMan assay. A complete concordance of genotype calling between the genotyping platforms was ensured using a subset of DNA samples that were genotyped using both platforms (137). A sample consisting of 780 markers from 1,809 cases and 634 control individuals (of different ethnic backgrounds, mainly of White non-Hispanic and Black ancestries) remained after the initial quality control procedures.

9.1.2 GAIN (paper II)

Genotypes for the GAIN MDD sample were derived from a genome-wide SNP array-based platform (600K Perlegen chip) by Perlegen Science (83, 134). The marker selection was based on a tagging approach using the European and Asian reference samples in the HapMap phase II data with an applied stringent threshold for a pair-wise r^2 value > 0.89 and r^2 value > 0.96 for multi-marker analyses. After initial QC analyses genotypes from 599,164 markers in 1,821 cases and 1,822 controls were available for our analyses. We also used data from 35 trio families for the following QC analyses.

9.1.3 NIMH BP-pedigrees (papers III & IV)

The DNA samples were obtained from Rutgers University and Cell Repository (New Jersey, USA) and genotyped using the Illumina Human 610quad chip at Uppsala University, Sweden. To ensure high accuracy of genotypes a Genotype Call (GC) score exceeding 0.8 was required, generating a sample of 598,821 markers in 277 individuals (46 pedigrees).

9.1.4 Quality control

Array-based genotyping make use of hybridization intensities that are translated into genotypes using a genotype calling algorithms are not error-free processes (139). Incorrectly called genotypes will mislead the statistical inference and this a potential source of error that may have contributed to the inconsistency of reported associations (140).

In all projects within this thesis we put emphasis on calling genotypes with high accuracy and designed a series of QC filters with the aim of reducing erroneous genotypes, identifying population stratification and identifying cryptic relatedness which otherwise may have misled our statistical analyses. This included a series of QC filters to control for sample mixed-up, poor DNA quality, level of heterozygosity, call rate per marker and per person, cryptic relatedness and population substructure (for details, see *the Methods* sections of the separate papers). In the last project we applied filtering approaches to zero-out genotypes in regions with presence of CNVs. We also applied an approach to avoid rejecting informative markers for the linkage analyses and selected markers for linkage equilibrium (LE) based on a case-control population.

9.2 STATISTICAL MODELS TO MAP RISK LOCI TO MOOD DISORDERS

9.2.1 Background

Mapping genes to human traits has undergone substantial changes over time. In the early 20th century the discovery that Mendel's law of inheritance could explain transmission of genes for traits in model organisms spurred geneticists and statisticians to develop linkage analysis for genetic mapping in human families (141, 142).

Linkage analysis makes use of the pattern of inheritance to an offspring of two loci that are closely linked on the same chromosome. Two loci are co-segregated if no recombination occurs between them during meiosis (143). The initial attempts with linkage studies were not successful for complex diseases, which led to focus on association analyses, a contrasting method that does not make use of inheritance of risk alleles within relatives. In this approach correlation between genotype and phenotype is tested in unrelated individuals making use of the assumption that individuals with the same trait share the same alleles and that shared genetic risk factors are organized in block structures in the human genome (142). Early association studies were limited in that they were restricted to searching for potential risk loci based on biological presumptions of underlying disease pathology. The discovery of naturally occurring

DNA polymorphisms and the creation of the human map offered the advantage of linkage analysis to screen the whole genome without the need to prioritize regions with assumed biological function (141).

Mapping genes using linkage analysis and association analysis have different advantages and limitations, and there are fairly strong arguments for the both approaches given a certain assumption of the genetic architecture underlying the trait of interest, as illustrated in the allele frequency – effect size relation paradigm model of Figure 5 (144). Both association analyses and linkage analyses have been successful for mapping genes for many human traits (141, 142), but the efforts have been less rewarding in analyses of highly complex disorders such as mood disorders (see section 1.5).

Human complex traits, with no exception to mood disorders, are characterized by the genetic influences from multiple variants of incomplete penetrance, locus and allelic heterogeneities and different forms of genetic variants occurring in various frequencies reflecting selection pressure. Environmental exposures also affect the risk of developing the trait by often interacting with genetic risk factors. Accordingly, different theoretical models have been proposed which aim to explain the occurrence of the disease genes in the population. The most prevailing theory is *the common disease common variant (CD-CV) hypothesis* (145). The opposing theory is *the common disease rare variant (CD-RV) hypothesis* (146). These hypotheses have practical implications for how to detect a disease causing variant in the population in terms of choice of statistical models, variants to investigate and populations to analyze (147, 148). The latter theory implies that risk genes are relatively new in the population, and thus due to modest selection pressure existing risk genes are mildly or highly deleterious and thus not so frequent in the population. It also implies that these forms of genetic variations are more likely to be functional as they have not been under selection pressure for a long time period. Conversely, the former theory implies that the human population has expanded rapidly over a long time and has been under selection pressure for a long period. For this reason risk genes are common and they each confer a small risk. The CD-CV hypothesis has inspired and led to the GWAS era with great hope of finding polymorphisms responsible for genetic risk. The CD-RV theory motivated linkage studies since this method is better suited to find genes with a modest or high risk.

Thus far, no consensus model that fully characterizes the complete genetic contribution to mood disorders has been presented. Possible explanations for the missing heritability have been proposed, varying from (i) a multiple explanatory model in which both theories are correct but a variety of genetic risk factors exists, (ii) an alternative model which incorporates both theories (iii) pleiotropy effects or (iv) that gene-gene and gene-environmental mechanisms hold the understanding why people are at risk for mood disorders which then cannot be explained by existing theoretical frameworks (142, 147). For these reasons it has been a great challenge to map risk genes of mood disorders. A variety of approaches are in use and are emerging. New sophisticated statistical models that can take into account gene-gene and gene-environment effects, prediction of markers with high prior probability of impact on disease thus leading to screening of

even larger datasets with a reduction of the necessity to correct for a large number of tests, and better ascertainment procedures with higher accuracy for correct diagnosis are ongoing, and this will undoubtedly improve the success of identifying risk genes.

In this thesis work a variety of currently available statistical methods of linkage-, association- and gene-gene interactions effects were applied to analyze family-based and population-based study populations for genetic risk factors to mood disorders. In a collaborative project we extended an existing software (149) to test for additive interaction between two genetic loci under a logistic regressions model. The software is coded in JAVA and is intended to screen large datasets for interaction effects and to simulate data under the null hypothesis of no association, using a permutation analysis. We also developed an algorithm to search for inherited larger genomic variations (CNVs) that convey risk to BPD.

9.2.2 Linkage analysis

The central theorem of linkage depends on the probability of transmission (segregation) of two alleles located on the same chromosome to be passed on to the same gamete (egg or sperm) and thus to be transmitted together to the offspring. This probability depends on the occurrence of recombination (estimated as a theta value) when homologous chromosomes align during meiosis. More closely located loci have a higher probability to be co-segregated. When using linkage to identify disease loci, linkage between a genetic marker and the disease is tested. A factor that influences the probability of co-segregation is the genetic model which concerns the relative distribution (frequency) of the risk allele in the population and the mode of disease transmission. As the genetic model is in most instances unknown for complex traits these factors have to be estimated. Such estimates are given by the penetrance vectors. Model-based linkage tests for recombination and estimates of linkage are generated using the logarithms of odds (LOD) scores that two loci on the same chromosome are inherited together under the assumption of the penetrance vector used. The parameters of the genetic model have to be specified, which is the frequency as well as the penetrance for the risk allele.

Test for linkage is calculated as a function of the joint probability of trait and marker genotype. In detail, the calculation is conducted to test for the probability of linkage (i.e. occurrence of recombination, thus leading to joint occurrence of trait and marker phenotype, $\theta < 0.5$) divided by the probability of no linkage (free recombination, $\theta \geq 0.5$). Morton suggested a method to test for linkage across families using LOD scores (150). However, if the genetic model is not correctly specified the power of the parametric linkage analysis diminishes. In such circumstances a non-parametric (model-free) linkage analysis could be preferred (151).

The non-parametric linkage relies on evaluating the proportion of shared alleles among affected individuals at a particular locus without needing to specify the genetic model. Non-parametric linkage analysis makes use estimates of allele sharing *identical-by-descent* (IBD) to test for linkage between the marker and disease loci. Estimation of IBD tests if alleles are inherited from the same parent. For full siblings the numbers of shared alleles 0, 1 or 2 at each locus have the probabilities $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{1}{4}$ respectively under the null hypothesis that this locus is not linked to disease. If a locus is linked to disease it is expected that the IBD sharing is increased for affected sib-pairs. Making use of the observed allele sharing calculation for linkage can then be made employing a variety of different algorithms.

In spite of extensive debates and discussions in favor of one or other of the different approaches it is still not clear whether parametric or non-parametric linkage models will be the most powerful statistical model for mapping genes in mood disorders (151). We therefore chose both parametric and non-parametric linkage models under a variety of parameter settings which then were then applied on different liability classes of the disorder, known as *affection status models* (ASMs).

9.2.3 Association analysis

Association studies aim at localizing risk genes through comparisons of allele frequencies between unrelated affected and unaffected individuals. As opposed to linkage analysis, association testing requires linkage disequilibrium (LD) between the marker locus and the causative locus. Population studies reflect a historical perspective of LD pattern, haplotype blocks generated over many generations being much shorter than what is observed in family-based population samples (142). These narrow LD blocks provide one advantage to linkage in that the statistical test captures a causative locus within a short genomic distance. Allele-wise and genotype-wise tests were used to calculate P values, odds ratios (OR) and 95% confidence intervals (C.I.) in unrelated cases versus unrelated controls. The disadvantages of association analyses is the necessity to correct for many independent tests, the sensitivity for population stratification, and the occurrence of genotyping errors leading to spurious associations that reduce the probability of reporting true associations (152).

9.2.4 Gene-gene interaction analysis

An increasing interest has been directed towards studying the combined effects of genes to understand the underlying mechanisms of mood disorders (93, 153). The combined effect, or joint effect, of two genetic variants resulting in an unexpected phenotype that is not explained by the combined effect from the individual genetic variants is known as *genetic interaction* (154). In genetics, the term 'interaction' has been described as

epistasis or synergy (155, 156). In this thesis these terms are used interchangeably and refer to the probability to develop a disease given a certain combination of genotypes at two loci. There has been confusion about what interaction means, and how to detect interaction effects with statistical models and how interpret these findings biologically (155, 157). Numerous methods are available for detecting gene-gene interactions, each one designed with a different assumption of how biological interactions may be detected using statistical methods (158, 159). As there is no consensus for which method is best to detect interaction effects with relevance to disease status (160) we used two logistic regression methods (multiplicative and additive) and one data mining and machine learning (multifactor dimensionality reduction, MDR) approach to examine what extent these different statistical methods report synergy effects with relevance to diagnosis of MDD.

Regression approaches

The two logistic regression methods are based on the multiplicative and additive interaction models. For both methods, interaction is tested under recessive and dominant genetic models. To accomplish this, the binary genetic risk factors, SNP_A and SNP_B , are coded risk (=1) or non-risk (=0) respectively according to each separate genetic model.

Odds ratio (OR) for each combination of risk and no-risk at SNP_A and SNP_B are calculated in a logistic regression model according to:

(Eq 1):

$$\text{Log odds} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

The multiplicative interaction model tests for interaction using the interaction term β_3 . Test for significance is made by testing the null hypothesis $\beta_3 = 0$ against a two-sided alternative hypothesis.

The additive interaction model is based on a theory in which genetic risk factors (or environmental) are organized in sufficient causes that may lead to disease. The model assumes that there is additivity between these sufficient causes but not within them (161).

To address this hypothesis the binary genetic risk factors SNP_A and SNP_B with risk (= 1) or no-risk (= 0) were re-coded into 4 new variables (149) A0B0, A1B0, A0B1, A1B1:

If: $SNP_A = 0$; $SNP_B = 0$ then A0B0 = 1, A1B0 = 0, A0B1 = 0, A1B1 = 0

If: $SNP_A = 1$; $SNP_B = 0$ then A0B0 = 0, A1B0 = 1, A0B1 = 0, A1B1 = 0

If: $SNP_A = 0$; $SNP_B = 1$ then A0B0 = 0, A1B0 = 0, A0B1 = 1, A1B1 = 0

If: $SNP_A = 1$; $SNP_B = 1$ then A0B0 = 0, A1B0 = 0, A0B1 = 0, A1B1 = 1

The disease model is then defined as:

(Eq 2):

If A0B0: odds = e^{β_0}

If A1B0: odds = $e^{\beta_0+\beta_1}$

If A0B1: odds = $e^{\beta_0+\beta_2}$

If A1B1: odds = $e^{\beta_0+\beta_1+\beta_2+\beta_3}$

Test for interaction is defined as departure from additivity by estimating the attributable proportion (AP) due to interaction according to:

(Eq 3):

$$AP = (OR_{A1B1} - OR_{A1B0} - OR_{A0B1} + 1) / OR_{A1B1} \quad (149).$$

The ORs are calculated relative to the reference group A0B0 so that the baseline parameter (β_0) is cancelled out. An AP value > 0 measures the proportion individuals among those who are exposed to risk at both SNPs and who are diseased due to interaction between the risk factors, whereas $AP < 0$ measures the proportion of individuals among those who are exposed to risk at both SNPs and who are protected from developing diseased due to interaction from both genetic factors. Assuming a normal distribution of the estimated AP value under the null hypothesis of no additive interaction (true $AP = 0$), test for significance is made by evaluating how much the risk with double exposure differs from what is expected under the additive model. Deviation from additivity is interpreted as the two risk factors being in the same sufficient cause (161).

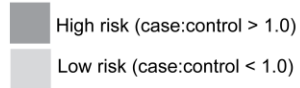
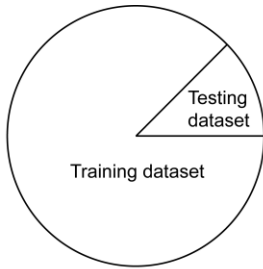
MDR approach

As it is widely recognized that the inheritance patterns deviate from classic Mendelian ratios in mood disorders we also used a genetically model-free MDR approach to evaluate interaction effects.

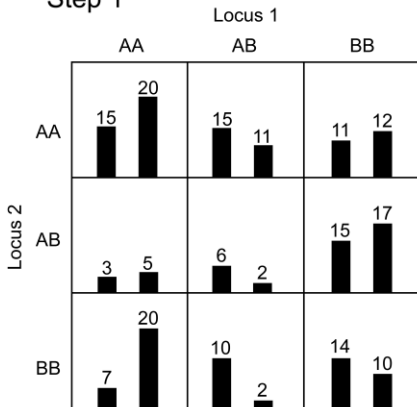
MDR is a data mining and machine learning approach that seeks to identify combinations of multi-locus genotypes that predict high versus low risk to a certain discrete phenotypic outcome. The dataset is divided into training and testing sets (157) (procedure illustrated in Figure 6). First, using the training dataset, for a two-locus model with three genotypes at each locus a matrix (multi-factor classes) is generated containing 9 genotype combinations (i.e., a high dimension of data). The ratio of cases to controls is calculated for each cell in the matrix. If this ratio exceeds a threshold $T =$

1.0 the corresponding combination of two locus genotypes is classified high risk, if the ratio is less than 1.0 it is classified low risk (step 2 in Figure 6). Based on this classification two new groups are generated, high versus low risk groups (step3). Thus, this process has now reduced the high dimension of data, of multifactor's, by pooling the data into one dimension with two classes (high and low risk), i.e. a multifactor reduction. In the fourth step the sum of errors for each genotype combination at all tested loci is ranked. The sum of errors refers to the total deviation from the threshold $T = 1.0$ for the nine different genotype combinations at each pairs of markers. This process is known as *feature selection*. The model (i.e. the combination of two SNPs) with the lowest rank best predicts the phenotype. The consistency for the best model is then calculated. To achieve this, the training dataset is divided into 10 equal parts that are used for a 10-fold cross-validation analysis. The best model is selected. In this process the balanced accuracy value is calculated (162). Finally, the test dataset is used to evaluate the best model from the training dataset.

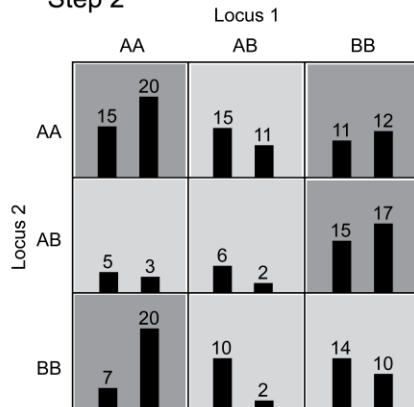
Decision to determine whether a pair-wise set of markers predicts the phenotype is made through observation from (i) the metrics for how accurate data is classified into high versus low risk groups, denoted by a balanced accuracy value, and (ii) the cross-validation consistency (CVC) value. Effect size (odds ratio) can readily be generated through calculation of number of cases and controls in high versus low risk groups. Test for significance is made through permutation analysis. A combination of markers that receive a high balanced accuracy and high CVC values are selected for further interpretation for relevance to disease. Based on the matrix of risk/no-risk classifications, a genetic model is inferred (Figure 6).



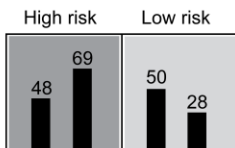
Step 1



Step 2



Step 3



Step 4

Factors	\sum Error
Locus 1: Locus 2	12.43
Locus 3: Locus 4	13.78
Locus 5: Locus 6	15.21
etc...etc	

Figure 6. Principles of the MDR approach.

Workflow of the MDR-analysis is illustrated by means of a hypothetical number of cases and controls. The figure is adapted from ref. 159.

9.2.5 Approach to find inherited CNVs with risk to BPD

To define genomic regions with CNVs that confer risk for BPD we developed an algorithm that calculates for linkage only in the presence of CNVs that are shared between individuals of BPD-ascertained families. The sum of average family-wise parametric LOD scores, or non-parametric Z scores, are calculated over regions for which families share overlapping CNVs (illustrated in Figure 7). For families with at least two members with overlapping CNVs the average linkage score in the region is calculated and added to those observed in the same region in other families. The algorithm thus generates a CNV-weighted linkage scores for genomic segments representing regions with CNVs that are shared within and across families. This approach thus identifies inherited regions harboring CNVs that could convey risk to BPD. The CNV-weighted linkage scores are then ranked with higher score values considered to have greater impact on the phenotypic trait. Test for significance is made through a permutation analysis (label switch of case-control status) to generate null expectations of linkage data. Empirical level of significance is defined based on family-wise error rate (FWER) analysis.

Figure 7.

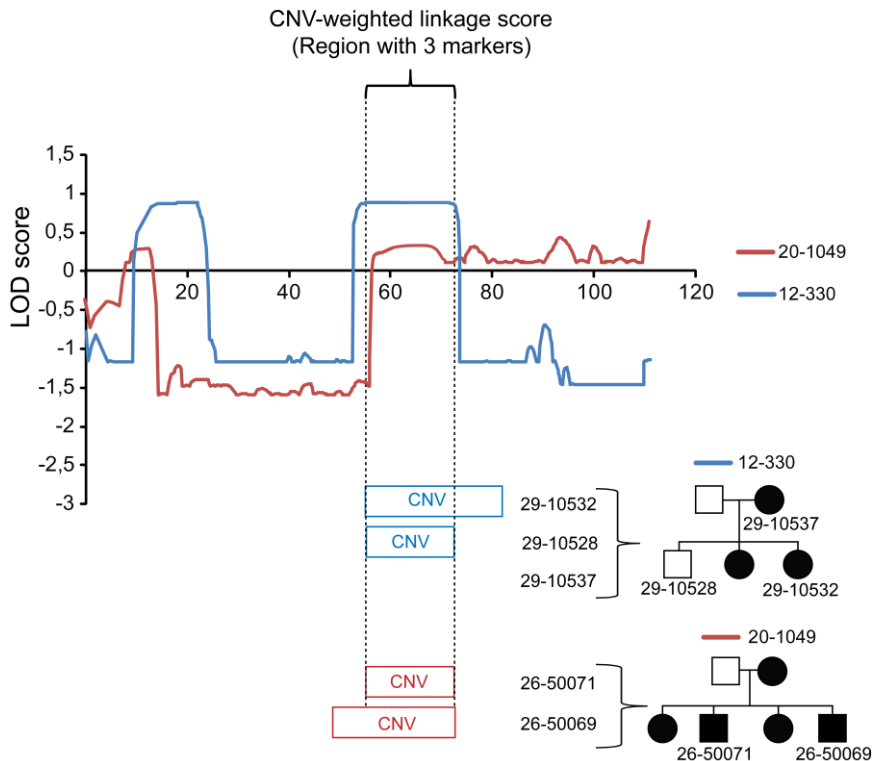


Figure 7. Principles for region for which one CNV-weighted linkage score is calculated.

Lod scores from two families (family ID: 20-1049 and 12-330) for a certain chromosome are depicted. The occurrence of CNVs for individuals within these families is also illustrated. The average linkage score (parametric lod score or non-parametric Z-score) is calculated over the region with overlapping CNV from at least two individuals in the same family. The average linkage score from all families with overlapping CNV in at least two individuals is added. A CNV-weighted linkage score is thus generated for a defined region that share overlapping CNVs for more than 1 individual per family. To illustrate the calculation; an example is given with three markers that are located in a region with four overlapping CNVs, at the position 60- 75cM.

For these 3 markers, the lod scores for the two families are:

Fam-ID 12-330: 0.9, 0.9 and 0.9

Fam-ID: 20-1049: 0.2, 0.4 and 0.1

The CNV-weighted linkage score (the sum of average linkage score) is then calculated as:

Fam-ID 12-330: $(0.9 + 0.9 + 0.9)/3 = 0.9$

Fam-ID: 20-1049: $(0.2 + 0.4 + 0.1)/3 = 0.23$

Total CNV-weighted linkage score = $0.9 + 0.23 = 1.13$

10 RESULTS

10.1 PAPER I

In a previous study, Binder *et al.* (127) determined that hospitalized MDD diagnosed individuals of Caucasian ancestry who were homozygous for the T-allele at the SNP rs1360780 in the *FKBP5* gene were associated with a better response when treated with antidepressants and the number of depressive episodes. This marker was not found to be associated to disease status. Two additional markers, rs4713916 and rs3800373, in the *FKBP5* gene were also tested but did not display association to either disease, number of depressive episodes or outcome in treatment response. The antidepressant medication varied among the individuals enrolled for the analysis and could be classified into three groups: selective serotonin reuptake inhibitors, tricyclic antidepressants and drugs targeting the serotonin 2C and α 2A-adrenergic receptors.

Our study aim was to replicate the association in the Binder study using a better powered study population, provided from the STAR*D study. The same three markers in the *FKBP5* gene were tested for correlation to risk of MDD, to number of depressive episodes of MDD and to treatment-response. The study sample used in this study consisted of non-hospitalized MDD diagnosed individuals of a more ethnically diverse, White non-Hispanic and Black, origin.

Case-control analysis

We determined significant association (corrected for multiple comparisons) for marker rs1360780 to disease status in the White non-Hispanic group (Table 1). A genotype-wise based analysis revealed that the TC-genotype was more frequent in cases (46%) than in controls (38%), whereas the CC-genotype was more frequent in controls (50%) than in cases (44%). None of the other markers were significantly associated to MDD after correction for multiple testing. An allele-wise test did not reveal significant results. No significant signal was observed within the ethnic Black group.

Table 1. Results of case-control and treatment-response analyses.

A. Case-control analysis, genotype-wise test

SNP id	Min/ Maj	Cases			Controls			OR (95% C.I.)	OR (95% C.I.)	Over all		
		Minor Hom	Het	Major Hom	Minor Hom	Het	Major Hom			Nominal P value	Corrected P value	
White non-Hispanics	rs3800373	C/A	0.09	0.43	0.48	0.09	0.39	0.52	CC/AA 1.05 (0.78-1.42)	AC/AA 1.22 (0.99-1.51)	0.18	ns
	rs1360780	T/C	0.10	0.46	0.44	0.11	0.38	0.50	TT/CC 1.01 (0.74-1.40)	TC/CC 1.39 (1.14-1.70)	0.0038	0.046
	rs4713916	A/G	0.092	0.44	0.47	0.095	0.39	0.52	AA/GG 1.08 (0.77-1.52)	AG/GG 1.29 (1.05-1.58)	0.046	ns
Blacks	rs3800373	C/A	0.17	0.51	0.31	0.15	0.44	0.41	CC/AA 1.51 (0.76-2.99)	AC/AA 1.52 (0.91-2.53)	0.23	ns
	rs1360780	T/C	0.17	0.49	0.34	0.16	0.45	0.39	TT/CC 1.19 (0.62-2.29)	TC/CC 1.26 (0.76-2.07)	0.66	ns
	rs4713916	A/G	0.004	0.18	0.81	0.01	0.19	0.80	AA/GG 0.34 (0.02-7.38)	AG/GG 0.95 (0.53-1.70)	0.8	ns

B. Treatment-response analysis, allele-wise test

SNP id	Remitters		Non-Remitters		OR (95% C.I.)	Nominal	Corrected	
	Minor	Major	Minor	Major		P value	P value	
All	rs3800373	C (0.33)	A (0.67)	C (0.31)	A (0.69)	1.05 (0.88-1.25)	0.59	ns
	rs1360780	T (0.34)	C (0.66)	T (0.33)	C (0.67)	1.01 (0.89-1.27)	0.48	ns
	rs4713916	A (0.30)	G (0.70)	A (0.24)	G (0.76)	1.37 (1.14-1.65)	0.00074	0.013
White non-Hispanics	rs3800373	C (0.32)	A (0.68)	C (0.28)	A (0.72)	1.19 (0.95-1.49)	0.12	ns
	rs1360780	T (0.34)	C (0.66)	T (0.30)	C (0.70)	1.21 (0.97-1.50)	0.088	ns
	rs4713916	A (0.34)	G (0.66)	A (0.28)	G (0.72)	1.35 (1.09-1.69)	0.0064	ns
Blacks	rs3800373	C (0.42)	A (0.58)	C (0.46)	A (0.54)	0.84 (0.54-1.28)	0.42	ns
	rs1360780	T (0.40)	C (0.60)	T (0.45)	C (0.55)	0.84 (0.54-1.29)	0.42	ns
	rs4713916	A (0.10)	G (0.90)	A (0.08)	G (0.92)	1.17 (0.56-2.26)	0.67	ns

Treatment response analysis

In the treatment response analysis two measures of outcome were evaluated, remission (defined as symptom-free) and response (defined as having residual symptoms). The 16-item clinician-rated measurement tool QIDS-C₁₆ (Quick Inventory of Depressive Symptomatology) was used for evaluation of treatment response. Patients with remissions were defined as having a QIDS-C₁₆ score ≤ 5 at an end point of the treatment period whereas non-remitters were defined as having a QIDS-C₁₆ score ≥ 10 . Responders were defined to have a 50% reduction of QIDS-C₁₆ scores at the last treatment visit in comparison to initiation of the treatment. Non-responders were defined to have a 40% reduction of the QIDS-C₁₆ scores.

In a genotype-wise analysis, of the two ethnic groups together, the marker rs4713916 survived correction for multiple comparisons in a test for remission. Test for outcome to response did not yield significance levels (Table 2, paper I). In a follow-up analysis, for the comparison of remitters and non-remitters, an allele-wise association test was also significant after correction for multiple comparisons, and which showed that the A-allele was over-represented (30%) within the remitters versus non-remitters (24%) (Table 1).

Testing the three markers for association in the two ethnic groups separately did not reach significant levels. However, we observed that for the two intronic markers, rs1360780 and rs4713916, the OR estimates of the allele-wise test went in opposite directions between two ethnic groups (White non-Hispanics and Blacks) (Table 1). This led us to test the LD structure between the three markers in the *FKBP5* gene within these two groups. In the population of the White non-Hispanics the LD structure between the three markers were high with an r^2 of 0.67 between rs1360780 and rs4713916, and an r^2 of 0.86 between rs1360780 and rs4713916, and finally an r^2 of 0.54 between rs3800373 and rs1360780. This indicates that these markers reside in one haplotype block for individuals of this ethnical origin. For the Black population this pattern looked somewhat different. Two of the markers, rs3800373 and rs1360780, resides in one block (r^2 of 0.86), whereas marker rs4713916 and rs1360780 belongs to a different block ($r^2 = 0.09$). The LD was also weak ($r^2 = 0.08$) between rs3800373 and rs4713916.

Outcome in treatment response over the 14-week treatment period was evaluated using the mean QIDS-C₁₆ estimates. Analyses were performed for genotypes of the three markers at; entry of treatment, after 2 weeks, 4 weeks, 6 weeks, 9 weeks, 12 weeks and finally at 14 weeks. Although a reduction of symptoms was observed this was not correlated to genotypes of any one of the three selected markers, rs3800373, rs1360780 and rs4713916 (Figure 1, paper I).

Number of previous episodes of depression

Finally, a genotype-wise analysis was conducted to test for correlation to the number of self reported episodes of previous depression (indication for severity of MDD). None of the markers reached nominal statistical significance thresholds (Figure 2, paper I).

Conclusions

This study provided a better power than the initial study by Binder and collaborators, and thus was able to report significant association to disease status for the rs1360780 marker in the White non-Hispanic population. Inclusion of White non-Hispanics and Black individuals and analyzing them as one group for outcome to treatment-response allowed for identification of association of rs4713916 to remission, but not to response. Inclusion of the Black population provided an indication for the A-allele as the functional allele in the treatment-response. We suggest that the *FKBP5* gene is an interesting candidate for further studies to better understand part of the complex etiology in MDD and for prediction in outcome of antidepressant treatment.

10.2 PAPER II

Test for heterogeneity

Preceding the interaction analyses in paper II we evaluated if the two study populations, STAR*D and GAIN had a shared genetic risk to MDD. Assessment of heterogeneity was based on a nominal P value ($P < 0.05$) generated from single marker association analysis. None of the nominally associated markers in STAR*D were found to be nominally associated in either genotyped or and imputed markers in the GAIN sample. Moreover, none of the nominally associated markers in the two study populations were in high LD. In a final meta-analysis using the identical marker with the same risk allele (10 markers) only one nominally associated SNP was identified ($P = 0.039$). A shared genetic risk between the two study-populations could not be confirmed. Therefore only the GAIN sample, which was the sample with more complete genotyping in the candidate genes, was used for the subsequent gene-gene interaction analyses.

Interaction analyses

Overall results

Genotyped and imputed genotypes in 63 candidate genes (chromosome X genes excluded and two autosomal genes not represented with markers in GAIN) were tested for a 2-way SNP-SNP interaction analysis. For each of the regression models 92 markers were tested against 3,704 markers (entire dataset) for evaluation of interaction effects, which constituted 340,768 pair-wise tests. The 92 markers were selected based on two sources of information, main effect (i.e. single marker association analysis at $P < 0.05$, $n = 43$) or using an algorithm (163) that predict interaction effects ($n = 49$). In the MDR method 3,704 markers were tested in an exhaustive interaction analysis (representing ~6.1 million interactions).

The large number of tests performed penalized our approach with regard to report a significant interaction, and none of the tests generated a P value that survived correction for multiple testing (Figure 2 A-C, paper II). Without rejecting that all results are due to random variation we hypothesized, based on the observation that our study sample was lacking power, that the 0.5% most significant interactions would harbor interactions involved in MDD-susceptibility. Adopting a conservative approach, we therefore next considered the top 10 ranked interactions (LD filtered, $r^2 \leq 0.2$) in each of the different interaction methods as a source of information to provide a better understanding of the genetic interactions in these candidate genes that may explain part of the risk to MDD (Table 1, paper II).

For the additive model our first observation concerns the unrealistic large fraction of negative AP values that exceeded the theoretical limit of -1.0 (i.e. $AP < -1.0$). We therefore aimed to investigate if the negative AP values were valid estimates for reduction of risk (Supplementary Figure 2A-D, paper II). We observed a skewing of AP values towards the negative side of the distribution and a correlation between large negative AP values and few individuals in one of the exposure groups. There did not appear to be any correlation between the lowest observed OR and negative AP, since $AP < -1.0$ was observed for both OR close to 1.0 and large OR's for single exposure. However, there was a negative correlation between the largest observed single exposure OR and AP when $AP < 0$. No deflecting pattern was observed for the positive AP values. These observations led us to simulate AP values (data not shown) with the aim to find out situations that give rise to negative AP estimates, especially those below the theoretical limit of -1. Our result identifies an inborn error in the definition of AP, which results in occasional estimates below -1.0 that often have a large magnitude. We concluded that although these negative AP values may indicate negative interactions, their magnitudes are far too large to warrant statistical analysis. Consequently, we decided not to draw any statistical conclusions from these estimates for susceptibility of MDD.

Top 10 interactions

Of the ten strongest interactions generated from the three different algorithms, none represents pairs of markers that have previously been reported to interact in the susceptibility to MDD. We also noted that within this group of interactions the markers did not represent those selected based on predictive interaction effects using main effect or an algorithm that predicts SNP-SNP interactions. Although the three different methods generated metrics of interaction effects with different assumptions and were thus not readily comparable, we further noted that the estimates of interaction effects were surprisingly strong for the regression models. In the multiplicative method (recessive interaction model), the strongest estimate was for two glutamate genes (*GRIN2B* and *GRIN2A*) with an OR (generated from the interaction term β_3) of 4.99 (95% C.I., 2.26-11.03). All of the ten strongest interactions of the multiplicative method had OR (from the β_3 term) values higher than 1.84.

For the additive method (recessive interaction model) markers in the genes *ARHGAP10* and *GRIK4* had an AP value of 0.72 (95% C.I., 0.42-1.02). All of the ten strongest interactions of the additive method had AP values higher than 0.51.

The effect measures of the MDR approach were modest. None of the top 10 interactions had OR exceeding 1.51. The strongest interaction was for markers in the *GRIN2B* and *HTR3C* genes with a balanced accuracy value of 0.55 and a cross validation consistency value of 4/10.

The different algorithms used detected different pairs of interacting markers with relevance for MDD-susceptibility. As a consequence we further noted that the three algorithms reported on a divergent set of genes. We did not, however, determine whether these findings are simply due to chance. Moreover, as the number of markers per gene varied substantially (Supplementary Table 1, paper II) we did not make inference with the observation that genes from e.g. the glutamate system are frequently reported. We did note, however, that genes with only a few markers were also reported among the top 10 interactions.

Test for correlation between the different interaction methods

There was little overlap of identified interactions between the 10 most significant observations for each model. To find out to what extent the different algorithms identified pairs of markers of relevance to disease we performed a test for correlation between the three interaction methods. We computed the Pearson's correlation coefficient (r) to test the covariance of interactions among the 0.5% region in one method against the entire dataset of another method. This test was used for combinations

of all methods (Figure 3 A-L, paper II). Overall, the measures of correlations were weak, with r values < 0.33 .

Model fitting and quantifying the genotype-total MDD relation

To assess the importance of these interactions for susceptibility to MDD we estimated the population attributable fraction (PAF) for the top 10 interactions in each model. We also calculated the number of cases predicted as cases (true positive rate) against the number of controls predicted as cases (false positive rate) at each level of combination of risk and no-risk of two tested loci (Table 1 A-E, paper II).

Although the estimates of the proportion of MDD in the population that is attributable to the exposure of the risk factors were low (i.e. the PAF values), these estimates were larger for the additive model in comparison to the multiplicative and MDR approaches. However, considering the PAF values for the interaction analysis versus the estimates from the single-marker analysis, the proportion population risk attributable risk due to interaction was only marginally larger, and in some instance these estimates were even lower (Supplemental Table 4 A-E, paper II).

Results of 3-way interaction analyses

We also conducted a 3-way interaction analysis based on the observation that several markers of the top 10 interactions were involved in more than one interaction (Table 1 A-E, paper II). This analysis was performed in the model-free (MDR) approach for 371 markers which were included based on results from the 2-way analysis with a balanced accuracy value > 0.54 . The results of the 3-way analysis were weak. The classification (accuracy value) for the best model was 0.57, with a cross-validation consistency value of 3/10. None of the top 3-way models confirmed the assumed higher order of interactions observed from the separate 2-way models (data not shown).

Conclusions

None of the tested interactions were significant after correction for multiple comparisons. We assumed that the top 10 ranked interactions would provide information to better understand the genetic interacting landscape in these candidate genes, and can conclude that none of the top ranked interactions have previously been reported for synergy effects to MDD-susceptibility. We further note that the three different statistical approaches are not likely to identify the same pairs of interacting markers involved in risk for MDD. We also observe that the two logistic regression models identify

interactions with a large effect sizes, whereas the MDR approach reports weak interactions effects. None of the top interactions explain a large proportion of MDD in the general population.

10.3 PAPER III

Recent genetic studies have revealed that part of the genetic risk for BPD is shared with SZ (92, 164). An accumulation of data further indicates that high penetrant rare CNVs may constitute a part of the genetic risk to SZ and BPD (93). To increase the chance to find high penetrant rare CNVs we applied a multistage analysis and searched for CNVs in individuals diagnosed with BPD, SZ or SA. Of note, due to some differences (see section 1.1), in classification of mood disorders in paper III, the diagnosis classification bipolar affective disorder (BPAD) is used but for the phenotypic classification this is identical to BPD.

In the initial phase, a sample consisting of 46 BP-pedigrees (275 individuals) were selected presumed to carry a genetic form of BPD through screening process consisting of family-wise parametric linkage analysis and analysis of large regions of deletion (runs of homozygosity). This sample was analyzed for the presence of CNVs larger than 10 kb and not overlapping with CNVs reported in the Database of Genomic Variants (DGV). This process ensures that CNVs are robustly called and that they are rare in the general population. A ranking system was designed to select presumptive risk families. CNVs were ranked based on the number of affected individuals per family in which they were found and selected for further analysis.

This raking approach identified one family (family 11-158), consisting of seven individuals of which six were diagnosed as bipolar type 1 and of which all had a CNV in the same region (Figure 1, paper III). The CNV is a 200 kb deletion and maps to intron 1 of the *MAGII* gene. Due to the selection process, the identified CNV in the *MAGII* gene in the six bipolar affected individuals (family 11-158) was considered as a hypothesis-generating step and excluded from further analyses. Another BP-family (11-130) with a shared CNV in the *MAGII* genes was found.

To investigate if the CNV finding could be observed in an independent sample, we first analyzed unrelated individuals (sampled for the purpose of clinical trials) with a diagnosis of BPD, SZ and SA disorder were provided by Johnson & Johnson Pharmaceutical Research & Development. In this sample three individuals with a diagnosis of SZ had a CNV (one deletion and two duplications) which also maps to the first intron in the *MAGII* gene and that partially overlapped with the CNV identified in family 11-158. In the final stage we searched publicly available databases of case-control samples and unaffected control samples. Three additional cases (BP and SZ) and 2 healthy controls, with a CNV overlapping with intron 1 of the *MAGII* gene were identified (Figure 1, paper III). A statistical test was performed to assess if the frequency

of *MAGII* CNV was higher in cases versus controls. Excluding the initial BP-family (11-158) (considered as a hypothesis-generating step), a test for association of CNVs in cases with BPD, SZ or SA (n=7) versus controls (n=2) was significant at $P = 0.023$.

Conclusions

Our multistep approach to search for rare highly penetrant CNVs involved in BPD by searching across diagnostic boundaries including individuals with BPD, SZ and SA allowed for identification of a CNV in the *MAGII* gene at 3p41.1. We identified 7 cases and 2 control samples with CNV in the *MAGII* gene. A statistical test reveals that there is a significant difference between cases and controls and we therefore suggest that this gene is an interesting candidate for future studies in the etiology to BPD and SZ.

10.4 PAPER IV

The same sample as in paper III consisting of 46 BP-pedigrees was selected and presumed to carry a genetic form of BPD. Two forms of analyses were performed (i) linkage analysis and (ii) a combined CNV and linkage analysis. To account for both clinical and genetic heterogeneity we analyzed the dataset using three affection status models (ASM1-3), and calculated linkage using multipoint parametric heterogeneity (HLOD) scores and non-parametric linkage analysis.

Linkage analysis

The non-parametric (NPL_{ALL}) linkage analysis reached a suggestive level on 3p14.1 for all affection status models. The broad affection status model (ASM3) exhibited the strongest peak, NPL $Z = 3.56$. For the parametric linkage analyses, four regions reached a level of suggestive linkage (HLOD), all in the ASM3. The strongest signal of linkage occurred at 6p12.3, HLOD = 2.64 (Table 2, paper IV).

CNV-weighted linkage analysis

We identified 2,806 CNVs in our dataset that were used to generate CNV-weighted linkage scores in nine different linkage and affection status models (genome-wide plots in Figure 1, paper IV). The parametric dominant model (ASM1), exhibited a significant CNV-weighted linkage score on 19q13, after a 1,000-fold simulation including FWER

correction. A detailed analysis revealed that the significant CNV-weighted linkage score was generated from 12 individuals (5 families) who shared a CNV in this region. Figure 2, paper IV illustrates the genomic location of the CNV in relation to the linkage scores from these 5 families and the complete set of genes that resides in the linkage region.

In a subsequent analysis we aimed to find out if the CNVs are inherited or arise *de novo* by performing a haplotype analysis in the 5 families that contributed with the significant CNV-weighted linkage score in 19q13 (Figure 3A, paper IV). The phased haplotype analysis revealed that there are frequently occurring recombinations in this region, or alternatively that we have missed the detection of CNVs in some individuals. We also observed that within the 5 BP-families contributing to the CNV-weighted linkage score, not all affected under the ASM1 model were CNV carriers, and moreover, that some unaffected individuals were also CNV carriers. These results suggest that if the CNV is functional in causing a risk for BPD, and if we detected all CNVs in these families, the CNV has an incomplete penetrance and that some individuals in these families developed BPD due to other reasons. The CNV stretches over a region that harbors a family of genes encoding the pregnancy-specific glycoproteins (*PSGs*). The *PSG* genes have not previously been reported to be involved in BPD. A bioinformatic analysis revealed that two markers within the *PSG* gene are likely eQTLs for *FBXO30* protein with regulatory properties of *NFκB* (165), a transcription factor for neurogenesis and inflammatory response (166, 167). Of note, within the linkage peak-region there were several other candidate genes (*GRIK5*, *GSK3A* and *CEACAM21*) previously implicated in BPD and SZ (75, 168-170).

Conclusions

For the linkage analysis, we identified four loci that reached suggestive levels of linkage to BPD of which three confirmed previously published results (1q23, 3p14 and 10q26). The CNV-weighted linkage analysis identified one significant region with a CNV. This region mapped to 19q13 and stretched over a gene family, the *PSG* genes. Of interest, the *PSGs* have recently been shown to have a regulatory role in secretion of inflammatory cytokines and to activate transforming growth factor TGF-β (171). TGF-β proteins are involved in regulation of cell growth and immune cell functions, two suggested mechanisms underlying the pathophysiology of mood disorders.

In addition, two SNP-markers are likely eQTL's for *FBXO30* (located on chromosome 6). The *FBXO30* protein has a regulatory function of the transcription factor *NFκB*, making this pathway an interesting candidate to partly understand the complex etiology to BPD (166, 167). The role of *NFκB* in the pathophysiology of mood disorders is suggested to be mediated through increased levels of the cytokine TNF-alpha. During periods of depressed mood increased levels of TNF-alpha leads to lower levels of BDNF which in turn reduce *NFκB* translocation. This process induces apoptosis and thus serves as proposed model for the regional reduced brain volume that is observed in mood

disorders (166). Finally, within the linkage peak region there are several candidate genes with putative roles in BPD, e.g. *GRIK5* and *GSK3A*.

Our results indicated that the CNV is inherited. However, as not all the affected were CNV carriers and since recombinations frequently occur we conclude that the CNV has a low penetrance to mood disorder. A prioritized follow-up analysis would be to molecularly (PCR technology) confirm the signal in 19q13 before a more functional link to BPD could be made by the candidates in this region.

11 DISCUSSION

11.1 PAPER I

The A-allele of the rs4713916 marker was more frequent among remitters (30%) than non-remitters (24%) when both White non-Hispanics and Blacks were included in the analysis. This test yielded a $P = 0.013$ after correction for multiple comparisons. However, we observed that the allele frequencies for the three markers were not identical between the two ethnic groups (Table 1. For marker rs4713916 the allele frequencies for the minor and major alleles in the White non-Hispanic population were 34% and 28% respectively. These estimates were 10% respectively 8% in the Black population. This observation prompted us to clarify whether there is a stratification issue when both groups are analyzed together. We therefore performed a meta-analysis using ethnicity as a covariate and found that there is a genetic heterogeneity between the two ethnic groups (Figure 8). The test for heterogeneity was significant ($P = 0.046$). This implies that there is stratification with a different allele frequency between the two study populations for rs4713916. These results indicate that our approach to combine the two ethnic groups into one sample when testing association to treatment response was not correct and no significant association of *FKBP5* genotypes to treatment-response remained after correction for multiple comparisons.

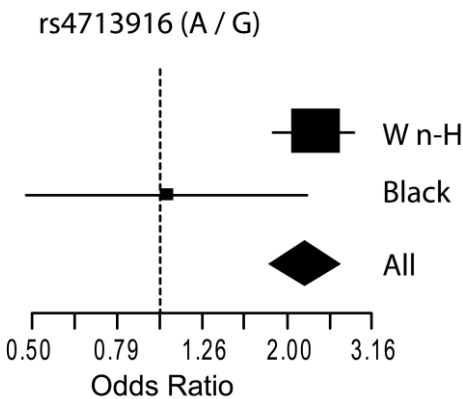


Figure 8. Test for heterogeneity.

Marker rs4713916 was tested in a meta-analysis (Mantel-Haenzel) with White non-Hispanics and Blacks, for A/G genotype to remission. Test for heterogeneity, chi-square test (1 d.f.), reached significance level with $P = 0.046$.

The HPA-axis is an established neuroendocrine stress-response pathway, which has been proven to be important in both in humans and in animal models (41). In these feed-forward and feed-backward pathways, glucocorticoids are suggested to be a critical component for activation and termination of the stress-response. The weak result regarding the correlation of SNP-marker to risk of MDD, and the fact that the association to treatment-response was not significant when combining data from the two ethnic groups in a correct way, urged us to address the questions of whether the *FKBP5* gene is a key regulatory factor for conferring risk to MDD and whether the *FKBP5* gene has a regulatory role in treatment-response with antidepressants.

First of all, in the study by Binder *et al.* 2004 the marker rs1360780 was reported to be significantly associated with response to antidepressants. However, the reported 95% C.I. was 0.60-48.3 in the replication sample. No significant association to disease status was reported by Binder *et al.* while it is associated to disease status among White non-Hispanics in our study. Secondly, a limitation which makes it difficult to distinguish true drug effects from placebo is the absence of a control group in both our study and the Binder study. Thirdly, the therapeutic effects of antidepressant drugs are slow, for most patients effects appearing after 2-4 weeks of treatment (172). In light of this observation the results from our study and the Binder study should be taken with caution as both studies indicate a rather rapid decline in symptoms after the first weeks of treatment. Fourthly, a treatment-response analysis (Citalopram treatment) of the STAR*D sample using genome-wide marker coverage did not report markers of the *FKBP5* region among the strongest candidates (173). Lastly, results from genome-wide association studies of the STAR*D-sample did not report *FKBP5* markers among the highest ranked candidates to risk of MDD (89).

Taken together, these notions suggest that the *FKBP5* gene may still be one important, although, not likely the only system mediating influences to MDD-pathophysiology.

11.2 PAPER II

At the time of the planning of this study there was no clarity about the magnitude of the interaction effect we were expected to discover. We therefore assumed that the effect would be more marked than our present results. As there is no evidence for which statistical model that has practical consequences, i.e. that may detect interacting loci involved in disease susceptibility (160), we decided to use different algorithms. As there is also no robust method for predicting markers with interaction effects, we decided to screen as many markers as was computationally practical.

None of the identified interactions survived correction for multiple comparisons, and as a consequence we cannot exclude that the results are just caused by chance and have no

biological meaning. However, we speculate that our results may hold some important information that leads to a better understanding of the genetic landscape in MDD within these candidate genes. We argue that our results performed indicate a direction that supports previously made assumptions about the contribution of genetic background for mood disorders and therefore make sense and merits discussion.

First, our approach to rank interaction based on test for significance may be questioned as our aim was to search for strong interaction effects. A sufficient large sample size will generate small P values, although, the most relevant question concerns the risk estimates. When looking into these estimates we can see that the regression models generated marked effect size estimates for the top 10 strongest interactions. We note however, that small P values do not necessarily correlate to largest effect size estimates. We next examined the top interactions for determining disease status (calculated from sensitivity and specificity scores). We noted that none of the interactions were able to correctly classify diseased individuals and non-diseased individuals to a large extent (Table 1, paper II). Finally, the PAF estimates for the top interactions were only marginally larger than for the single-marker analysis (Supplementary Table 4, paper II). Taken together, these observations illustrate that none of these interactions could explain their contribution to risk of MDD to a large extent. The relevance of these findings can be explained. These findings concord with emerging data from genetic studies in psychiatric disorders (see section 1.3), evolutionary theories of mood disorder (174, 175) as well as from interaction studies in model organisms (154). Traits that are evolutionary old and that are crucial for the organism for survival are thought to have an extensive network of interactions (156, 159) which supports the assumption that no gene or signaling pathway stands out being significantly more important than any other. The last argument is further supported by interaction studies in model systems in which the majority of the genes have been shown to be dispensable (176, 177).

These results provide an interesting lead to future interaction studies. It implies that new approaches will be needed not only to predict specific disease-associated genes but also to predict systems of genes. A possible solution to reduce the number of independent tests would thus be accomplished by testing systems of genes or network of genes rather than testing separate genes (154). Another important note that requires explanation concerns the interpretation of the additive model. AP values do not indicate how many sufficient causes there are, but only refer to a measure of the currently tested double exposure. In respect to the negative AP estimates, we have identified an inherent problem with the algorithm to quantify these estimates and we cannot interpret these negative AP measures other than to say that they indicate an interaction reducing the risk to developing MDD.

Our approach to test for interactions, which was not restricted among predicted markers, yielded important information since none of the currently prediction methods seems to be optimal. Previously reported gene interactions with suggested relevance for involvement in MDD-etiology were not evident among the strongest reported interactions in this study. The high LD-structure between markers is likely not a

disadvantage for the interpretation of the results since we have simulated data to generate exact P values by randomizing case-control status and not genotypic correlates. Genetic interaction analysis is an important instrument aiming to find genes with relevance in disease pathology, but our knowledge of how to bridge the gap between statistical signals to function is still limited. Interpretation of genetic interaction into etiology of mood disorders is therefore a challenging task as the identification of genetic interactions using statistical models does not imply that the two gene products physically interact, or that identified genes are in the same pathway or even that they are temporally co-expressed (178).

Our approach, using three different algorithms, did not deliver any conclusive result for which approach is preferable. The three approaches define interaction with different assumptions and therefore do not identify the same pairs of interacting markers. This is further supported from interaction analysis in model systems which reveal that all genetic interactions are not simplistic additive or multiplicative, and that different definitions of interaction results in dramatically divergent identification of genetic interactions (160). This line of evidence motivates the application of different statistical methods for the detection of all forms of interactions. A standard procedure in genetic studies is replication of results in an independent sample set. In order for such a process to be valid the two sample populations are required to have been exposed to similar genetic and environmental risk factors. We demonstrate that the STAR*D and GAIN samples do not appear to share similar genetic background, maybe because of exposure to different environmental risk factors. Thus, the STAR*D sample is not optimal for replicating interaction findings from the GAIN study. As no optimal replication sample was available in our study and no functional studies have been performed our results can only be referred to as being exploratory (179). The replication issue will be of significant challenge for future interaction studies in mood disorders as there is a pronounced environmental influence. In addition, even small differences in allele frequency in replication sample between tagging marker and risk marker has a large impact on the power to replicate initial results (180).

11.3 PAPERS III & IV

In studies III & IV we addressed several proposed reasons as to why genetic risk factors for BPD have not been robustly reported. As an initial step we selected a study sample likely to have genetic risk for bipolar disorder by screening a large sample of BP-pedigrees with family-wise parametric genome-wide linkage analyses and for the presence of large stretches of deletions in candidate genes for BPD. In this process 46 BP-pedigrees were selected. According to our pre-agreed analysis strategy we would not report any locus that constituted a criterion for selection of these families. The notion that we in fact did not identify any of the final results in any of the loci that constituted the selection criteria confirms the pronounced heterogeneity in BPD, since a slightly different composition of individuals in the BP-families points to another risk locus.

In study III our hypothesis was that rare high penetrant CNVs contribute with risk to BPD. To increase power to detect such genomic variants we took advantage of that genetic risk has been shown to be shared across disorder boundaries. To ensure that the CNV should be inherited our first criterion was that the CNV should be identified in BP-pedigrees. We also aimed to ensure that the CNV should be rare and therefore screened the results from our CNV-calls for overlapping variations catalogued in a database containing CNVs from healthy control samples. To identify inherited highly penetrant CNVs involved in BPD our method was to rank the number of CNV based on the number of affected individuals per family in which they were found. To investigate if this CNV could also be found in the general population we searched a large case-control sample consisting of cases diagnosed as BPD, SZ and SA and a sample consisting of unaffected individuals. One region was reported, located in the intronic region of the *MAGII* gene.

A limitation of the study design in paper III is due to the lack of evidence for CNV being correlated to BPD in the pedigree sample in any other respect than that the CNV is observed among BPD individuals. Other regions with possible stronger correlations to disorder may exist. Nevertheless, the confirmation analysis correlates the CNV to the disorder and we presented a possible functional role for the *MAGII* protein in the etiology to BPD. But since the frequency difference of the CNV between affected and non-affected sporadic cases was small, although significant, it could be questioned whether this CNV has a high relevance for BPD.

In study IV we tested the hypothesis that inherited large structural genomic variations (CNVs), irrespective of the frequency of occurrence, contribute with risk to BPD. We developed an algorithm that combines genome-wide linkage data with the CNV content within and across families and searched for regions enriched for CNVs that mediated risk for BPD. This method has the advantage that an identified region with a CNV is correlated to BPD using a linkage analysis. Thus this approach ensures that a potential region with CNV has implications to disease etiology. To reduce the problem with incorrect called genotypes in CNV regions we designed a QC filtering with the aim to zero-out markers in CNV regions. As accumulating data reveals that there are both genetic and clinical heterogeneity for those affected by BPD, we sub-categorized the sample in three affection status models (ASM1-ASM3) that are commonly used (136). This means that in this study we applied an opposite approach to what was performed in study III, and aimed to take advantage of information of heterogeneity to search for genetic risk in a sample assumed to be more genetically convergent in its risk to BPD. All of our CNV calls were tested for overlap (performed 2013) to 150,000 catalogued CNVs from unaffected individuals. This should be compared to the previous analysis in paper III (performed 2009), with only 30,000 catalogued CNVs from unaffected individuals. Our approach has potential limitations and caveats that merits discussion. The subset of pedigrees used in our analyses was selected from a large sample of possible pedigrees using data from linkage analysis as a proof that they segregated genetic variants affecting BPD. This has a potential bias ascertainment as it will have the effect of increasing the effect size of any genetic effect that is detected. A possible

alternative approach would therefore to account for this by selecting pedigrees based on non-genomic criteria for linkage to BPD.

The CNV-weighted linkage analysis identified 19q13 as a potential risk locus for BPD. However, a bioinformatic search revealed an eQTL in the CNV region with functional role in BPD etiology, but also that several previously found candidate genes are located in the linkage region. Our results support a recent notion in that CNVs are more common than previously expected and confer risk to disorder with less penetrance. We also see that the CNV occur in a region reported with frequent recombinations. We checked our results to find out how the identified CNV in the *MAGII* gene correlated to BPD (Table S5, paper IV), and observed that this CNV is ranked in 23rd place and confirms that it correlates with BPD, although with small effect on the linkage score.

Taken together, the results from studies III & IV indicate that the potential risk loci for BPD can be identified through two complementary approaches to handle genetic and clinical heterogeneity. Although no candidate gene could conclusively be pinpointed we suggest that the CNV-weighted linkage approach could be a useful instrument for future studies in complex disorders.

12 FUTURE PERSPECTIVES

In this thesis, I tested four hypotheses, all concerning identification of susceptibility genes for mood disorders. I addressed several known and debated issues regarding difficulties in successfully identifying genetic risk factors. In the analysis of major depressive disorder we identified the *FKBP5* gene contributing with risk to MDD in the STAR*D cohort, no effect was observed to treatment response. Testing for interaction effects in the GAIN sample we did not detect any pairs of markers contributing to risk of MDD. Nonetheless, we contributed valuable information for the analysis of genetic interaction effects and provided a speculative insight into how the genetic interaction pattern looks in candidate genes for MDD. In the analysis of BPD we presented two potential risk loci for future studies to understand part of the complex etiology for BPD.

The path towards identification of biological processes leading to mood disorders through the identification of susceptibility genes will be a challenging task. Studies designed to test correlation of genomic variation to human trait and disease is only a first initial step in a long series of subsequent analyses until conclusions about function can be made. The raising prevalence of the disorders is likely to be influenced by environmental risk exposures and interactions from gene-environmental risk factors (181). Identification of such risk effects would be of a great value. Although the success will likely depend on collaborative projects designed for translational research work across domains of genetic analyses, statistical analyses and animal models, much of the progress will lie in the genetic epidemiology. Prospective studies and better diagnostic system that has a high accuracy to discern the different subcategories of mood disorders will bring valuable improvements for finding genetic risk factors. However, it seems that there is a 'Catch-22' situation. A validated genetic risk factor would be an initial step to understand molecular mechanisms underlying risk for the disorder, and could thus lead to an improvement of the diagnostic system. Conversely, a more accurate diagnose classification would improve the ability to find a genetic predisposing factor. A central element for the relation of the cause-effect-loop concerns non-genetic risk factors. Non-genetic factors refer to a wide range of social adversities and its contribution to vulnerability to mood disorders is firmly established (4, 5). One of the most well known risk factors is early stressful life events (childhood adversities, such as parental loss, sexual abuse or parental mental disorders). A variety of exposure factors (socio-economical related or comorbidity from other psychiatric or medical disorders) may also trigger debut of the disorders in late adolescence or even later in life (13). However, there are many controversies. In spite of the growing awareness that such risk factors have substantial influence on vulnerability there is little evidence that a specific event can be attributed to a manifest disorder. Far from all that are exposed to social adversities develop the disorders. This scenario has consequences for the success of genetic studies and could explain why any initial genetic observation is as yet not robustly replicated in an independent study sample. This could be explained by differences in environmental exposures in the different study populations. Thus, different social-economical exposures that interact with the genetic background are

likely to be part of the reason as to why predisposing genetic factors have still not been identified.

Animal models have added a significant insight into the etiology and new strategies for finding better treatments in mood disorders, and will indisputably be a central instrument in the future gene validation process. However, there are certain points worth mentioning. The observation that mapping genes using linkage analysis in animal models has been more successful than in human populations can be explained by reason of power. A first possible argument is that an inbred strain might have a fixation of a rare allele or have a limited number of other involved loci (182). A second argument is that animal models have a higher probability for being more informative for linkage, which will then result in a higher power for finding segregating loci than in a human population (183). Conversely, it is unclear if these depression-like phenotype models are satisfactory proxies for the human diseases (184).

To conclude, identification of genetic determinants would be a considerable step towards unraveling the complex background of mood disorders. Besides providing a substrate for functional analyses and molecularly evaluating the effect in affected versus unaffected individuals it would serve as a valuable platform for studying the effect of social stressors. It would also provide an instrument to identify neuronal circuits in an animal model related to mechanisms underlying human depression and bipolar disorder pathophysiology. It appears that different methods generate results suggesting different genetic risk factors to mood disorders. One of the questions concerns whether larger study populations should be used to search for risk genes including broader phenotypic boundaries, or if a more narrow disease classification approach should be used. The former approach will find susceptibility genes (a large enough sample will find a risk locus), but it is not given that a certain risk gene pertains all sub-phenotypes of the disorders and these may be missed. Conversely, searching for genes restricted to a certain sub-phenotype may be inapplicable for other sub-phenotypes. As yet, there are several uncertainties as to which way that will be the most successful. Continued research will answer these questions. It is unlikely that a single approach will unravel all genetic risk to mood disorder (illustrated in Figure 6). This notion implies that there is no room for a reductionistic thinking, various methods and approaches will contribute with valuable information. Indeed, currently there are arguments that both concede or refute the different proposed theories for how a potential genetic risk profile might appear (182). Thus, the proposed hypotheses aiming to describe the mutation-selection balance and the presence of susceptibility genes may serve as a platform for testing our hypothesis and form a valuable instrument for deciphering the genetic influences to mood disorders. However, it should not restrict our mind to think in other dimensions.

Success will lie in translational research, sharing data and open discussions.

13 ACKNOWLEDGEMENTS

Many people have contributed to the completion to this thesis. I hope I have not forgotten too many of you who have helped me along this journey. I would like to express my sincere gratitude to:

Ingrid Kockum – My main supervisor. No words can express how grateful I am to have been your student. I have learned more than I could ever expect. You have always been standing by my side even though there have been hard times. You are both an exceptional strong person and supervisor. With your great scientific knowledge, personality and intellectuality I have developed a lot throughout these years. Your analytical skills in many different areas of science have brought absolutely critical and new methods to my project that finally has come to a completion. You have always been encouraging, positive, caring and most of all open for my never ending desire for having discussions. Our discussions have been stimulating, open, challenging, yet warm, in subjects that have been more than human genetics. Owing to your scientific expertise I have learned science and developed as a person. No one else had managed what you have done. Of most importance, I have found a new friend!

Ola Hössjer – My co-supervisor. Without your help, my PhD project would not have been possible to complete. Your analytical skills, intellectual mind and excellent scientific advices have brought my project into new perspectives. I have always found the discussion with you tremendous stimulating, it has truly inspired me. You've had the ability to bring complicated statistical theories down to my level. With great patients have you explained statistical theories and applied these into human genetics. You have truly brought me into a world I never thought exist. Your enthusiasm and encouragement has meant a lot for me!

Maria Anvret – My co-supervisor. Owing to your expertise, wise advice and personal contributions my PhD project have come to a fine finish.

Silvia Paddock – My former supervisor. Thanks for accepting me as a PhD student. I have learned a lot in human genetics which I can bring along in science.

Tomas Olsson – My mentor and Professor. Thank you for accepting me as a PhD student in your group. It has meant a lot for me to be in your group.

Markus Dagnell – My best friend! You have been present 24 hour per day, always open to help and to discuss all kind of questions. With your empathy and friendship I have made this thesis. No words can describe how grateful I am.

Magdalena Lindén - Thanks for all interesting discussions and meetings I have had with you. In your company I have learned genetics and developed as a person. Your presence and friendship has meant a lot for me. Especially, it has been important that you have listen, being encouraging and that you have an open heart. I have found a new friend which is very important for me!

Margarita Diez – You have been an enormous support for me. With your dedication, professional skills and great personality my project has come to a completion. I am deeply touched and very grateful for all your help.

Karin Wallin Blomberg – I know you have found (and taken) all possibilities to help me in my project. Without your personal dedication and professional skills my thesis would not have come to a completion. I am very grateful!

Robert Harris – Your support and scientific advices have been profoundly important. Special thanks for manuscript and thesis revisions.

Daniel Uvehag – My private teacher in computer science. You have been invaluable support for all computer work I had to struggle with. Without your help, my project would never have come to a completion.

Pernilla Stridh – My deepest appreciation for your all personal support. I am also very grateful for all scientific discussions. I am impressed by your commitment to support me to finish my thesis. Special thanks for all help in the revision of the thesis.

André Ortlieb – My deepest gratitude for all your help during these years. Special thanks for all advices in thesis revision.

Nada Abdelmagid – You have been an important support. You have always been kind, positive and encouraging.

Mohsen Khademi – Thanks for all your help and interest in my project. You always put a lot of effort to make the best for members for the Olson group and I have felt that you have cared a lot for me. Deepest appreciation for the memorable celebration!

Venus Azhary - You have been a great support for me. You are always listened and being present for meetings and support. You have a good sense of humor and a great personality. Thanks!

Maja Jagodic – Thanks for all occasions with scientific discussions. Special thanks for all input to my project.

Emilie Sundqvist – Thanks for all scientific discussions and help. It has meant a lot to me!

Henrik Källberg – Thanks for all scientific discussions, it has been very educating. I am very grateful for your patients to describe epidemiology and additive interaction analysis.

Mattias Frånberg – Through all discussions with you have learned and understood model-free interaction analysis. I appreciate your scientific attitude and open mind during our discussions.

Robert Karlsson – My former group colleague. Thanks for all scientific discussions and all help in my projects. I am very grateful for all your help!

Lisette Graae – My former group colleague. Thanks for all scientific discussions and help throughout these years. I have so many memories from our time together at The Department of Neuroscience.

Alexandra Gyllenberg, Izaura Lima, Cecilia Domingues, Tojo James, Jenny Link and Samina Asad – I want to thank You all for all scientific discussions, help with my thesis and friendship during these years. I am looking forward on future friendship with all of You!

All members of Tomas Olsson's group, **Lara, Roham, Karl, Petra, Marie, Sabrina, Sohel, Hannah, Faiez, Richard, Cyntia, Xingmei, Sreeni, Sevi, Shahin, Hannes, Rasmus, Harald, Andreas, Rux, Anatoly and Maria** - Thank you all for all interesting discussions and a nice time together.

Jason Moore and Peter Andrews – With your help I have learned interaction analysis using the MDR approach. My deepest appreciation for this! Your scientific input for the manuscript has been enormous valuable. For some years ago I knocked at your door,

this led to a collaborative project and many educating scientific discussions. Thanks for your broad-mindedness and scientific attitude!

Chee-Seng Ku – My friend! Thanks for all conversations, input and suggestions for improvements in my project. I am very grateful for your invitation for being support in my project. I have enjoyed your company while we were travelling together. Many memorable moments together that means a lot for me.

John Rush – My deepest appreciation for your commitment to help me with the manuscript. I am very grateful for your excellent scientific advices. Your teaching and never ending efforts to help me to better understand psychiatric genetics has been of great value for me. You have always answered and made the best to help me. I will never forget this!

Francis McMahon – Thanks for all your help with the manuscripts. Your input has been tremendous valuable. Your scientific input has made me to understand human genetics.

Henrik Zazzi – Thanks for helping me with the software. Owing to your work I was able to complete the project and finally to write the manuscript. I am very grateful for this!

Helga Westerlind – With your expertise and kind help I have learned a lot about principle component analysis.

Izar ul Hassan – My personal support at the computational center at KTH. Without your help the calculations would never have been done. I am tremendous thankful for your help!

My family! I am sincerely and deeply grateful and touched for all your help and support! My father and mother Sven and Märta, my brother Fredrik with my sister-in-law Anne together with their lovely children Hugo and Blanca have all been a great support whenever I needed.

Thanks to anyone I have I accidentally forget to mention. If there is anyone who is not mentioned it doesn't mean I don't have you in my heart!

14 REFERENCES

1. Sadock BJ, Sadock VA, Ruiz P (2005): *Kaplan & Sadock's Comprehensive Textbook of Psychiatry* 8th ed.: Lippincott Williams & Wilkins
2. Association AP (2003): *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR)*. American Psychiatric Publishing, 1000 Wilson Boulevard. Arlington, VA 22209-3901.
3. <http://www.nimh.nih.gov/about/strategic-planning-reports/index.shtml>.
4. Joska JA, Stein DJ (2008): *The American Psychiatric Publishing Textbook of Psychiatry, 5th Edition*. American Psychiatric Publishing, Inc.
5. Hirschfeld RMA, Weissman MM (2002): *Neuropsychopharmacology. The 5th generation of progress*. Lippincott, Williams & Wilkins.
6. Kessler RC, Birnbaum HG, Shahly V, Bromet E, Hwang I, McLaughlin KA, et al. (2010): Age differences in the prevalence and co-morbidity of DSM-IV major depressive episodes: results from the WHO World Mental Health Survey Initiative. *Depression and anxiety*. 27:351-364.
7. Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu HG, et al. (1996): Cross-national epidemiology of major depression and bipolar disorder. *JAMA : the journal of the American Medical Association*. 276:293-299.
8. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. (2013): Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet*. 382:1575-1586.
9. Murray CJL, Lopez AD (1996): *Global Burden of Disease A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. Cambridge, MA: Harvard Univ. Press.
10. Baghai TC, Blier P, Baldwin DS, Bauer M, Goodwin GM, Fountoulakis KN, et al. (2011): General and comparative efficacy and effectiveness of antidepressants in the acute treatment of depressive disorders: a report by the WPA section of pharmacopsychiatry. *European archives of psychiatry and clinical neuroscience*. 261 Suppl 3:207-245.
11. Baghai TC, Blier P, Baldwin DS, Bauer M, Goodwin GM, Fountoulakis KN, et al. (2012): Executive summary of the report by the WPA section on pharmacopsychiatry on general and comparative efficacy and effectiveness of antidepressants in the acute treatment of depressive disorders. *European archives of psychiatry and clinical neuroscience*. 262:13-22.
12. Connolly KR, Thase ME (2012): Emerging drugs for major depressive disorder. *Expert opinion on emerging drugs*. 17:105-126.
13. Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, et al. (2011): Cross-national epidemiology of DSM-IV major depressive episode. *BMC medicine*. 9:90.

14. Tsuang M.T., M. T, P J (2011): *Textbook of Psychiatric Epidemiology. 3rd ed.* Wiley-Blackwell.
15. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE (2005): Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of general psychiatry.* 62:617-627.
16. Kessler RC, Merikangas KR, Wang PS (2007): Prevalence, comorbidity, and service utilization for mood disorders in the United States at the beginning of the twenty-first century. *AnnuRevClinPsychol.* 3:137-158.
17. Bostwick JM, Pankratz VS (2000): Affective disorders and suicide risk: a reexamination. *The American journal of psychiatry.* 157:1925-1932.
18. Kessler RC (2012): The costs of depression. *PsychiatrClinNorth Am.* 35:1-14.
19. Kleinman L, Lowin A, Flood E, Gandhi G, Edgell E, Revicki D (2003): Costs of bipolar disorder. *PharmacoEconomics.* 21:601-622.
20. <https://pgc.unc.edu/index.php>.
21. Alexander FG, Selsnick ST (1966): *The History of Psychiatry. An evolution of psychiatry thought and practice from prehistoric times to the present.* First edition ed.: Harper & Row Publishers, New York.
22. Marneros A, Angst J (2002): *In Bipolar Disorders, 100 years after manic-depressive insanity.* Kluwer Academic Publishers New York, Boston, Dordrecht, London, Moscow.
23. Fischer BA (2012): A review of American psychiatry through its diagnoses: the history and development of the Diagnostic and Statistical Manual of Mental Disorders. *The Journal of nervous and mental disease.* 200:1022-1030.
24. Kawa S, Giordano J (2012): A brief historicity of the Diagnostic and Statistical Manual of Mental Disorders: issues and implications for the future of psychiatric canon and practice. *Philosophy, ethics, and humanities in medicine : PEHM.* 7:2.
25. Charney DS, Manji HK (2004): Life stress, genes, and depression: multiple pathways lead to increased risk and new opportunities for intervention. *Science's STKE : signal transduction knowledge environment.* 2004:re5.
26. Timmermans W, Xiong H, Hoogenraad CC, Krugers HJ (2013): Stress and excitatory synapses: from health to disease. *Neuroscience.* 248:626-636.
27. Palazidou E (2012): The neurobiology of depression. *British medical bulletin.* 101:127-145.
28. aan het Rot M, Mathew SJ, Charney DS (2009): Neurobiological mechanisms in major depressive disorder. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne.* 180:305-313.
29. Roxo MR, Franceschini PR, Zubaran C, Kleber FD, Sander JW (2011): The limbic system conception and its historical evolution. *TheScientificWorldJournal.* 11:2428-2441.

30. Damasio AR, Tranel D, Damasio H (1990): Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioural brain research*. 41:81-94.
31. Ongur D, Price JL (2000): The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral cortex*. 10:206-219.
32. Price JL (1999): Prefrontal cortical networks related to visceral function and mood. *Annals of the New York Academy of Sciences*. 877:383-396.
33. Koolschijn PC, van Haren NE, Lensvelt-Mulders GJ, Hulshoff Pol HE, Kahn RS (2009): Brain volume abnormalities in major depressive disorder: a meta-analysis of magnetic resonance imaging studies. *Human brain mapping*. 30:3719-3735.
34. Brambilla P, Glahn DC, Balestrieri M, Soares JC (2005): Magnetic resonance findings in bipolar disorder. *The Psychiatric clinics of North America*. 28:443-467.
35. Drevets WC (2000): Functional anatomical abnormalities in limbic and prefrontal cortical structures in major depression. *Progress in brain research*. 126:413-431.
36. Hamilton JP, Siemer M, Gotlib IH (2008): Amygdala volume in major depressive disorder: a meta-analysis of magnetic resonance imaging studies. *Molecular psychiatry*. 13:993-1000.
37. Usher J, Leucht S, Falkai P, Scherk H (2010): Correlation between amygdala volume and age in bipolar disorder - a systematic review and meta-analysis of structural MRI studies. *Psychiatry research*. 182:1-8.
38. Frodl T, Meisenzahl EM, Zetzsche T, Hohne T, Banac S, Schorr C, et al. (2004): Hippocampal and amygdala changes in patients with major depressive disorder and healthy controls during a 1-year follow-up. *The Journal of clinical psychiatry*. 65:492-499.
39. Neumeister A, Wood S, Bonne O, Nugent AC, Luckenbaugh DA, Young T, et al. (2005): Reduced hippocampal volume in unmedicated, remitted patients with major depression versus control subjects. *Biological psychiatry*. 57:935-937.
40. Charlton BG, Ferrier IN (1989): Hypothalamo-pituitary-adrenal axis abnormalities in depression: a review and a model. *Psychological medicine*. 19:331-336.
41. Binder EB (2009): The role of FKBP5, a co-chaperone of the glucocorticoid receptor in the pathogenesis and therapy of affective and anxiety disorders. *Psychoneuroendocrinology*. 34 Suppl 1:S186-195.
42. Gadek-Michalska A, Spyрка J, Rachwalska P, Tadeusz J, Bugajski J (2013): Influence of chronic stress on brain corticosteroid receptors and HPA axis activity. *Pharmacological reports : PR*. 65:1163-1175.
43. Wang Q, Verweij EW, Krugers HJ, Joels M, Swaab DF, Lucassen PJ (2013): Distribution of the glucocorticoid receptor in the human amygdala; changes in mood disorder patients. *Brain structure & function*.

44. Duman RS, Monteggia LM (2006): A neurotrophic model for stress-related mood disorders. *Biological psychiatry*. 59:1116-1127.
45. Valentino RJ, Foote SL, Page ME (1993): The locus coeruleus as a site for integrating corticotropin-releasing factor and noradrenergic mediation of stress responses. *Annals of the New York Academy of Sciences*. 697:173-188.
46. Guidotti G, Calabrese F, Anacker C, Racagni G, Pariante CM, Riva MA (2013): Glucocorticoid receptor and FKBP5 expression is altered following exposure to chronic stress: modulation by antidepressant treatment. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*. 38:616-627.
47. Mitchell ND, Baker GB (2010): An update on the role of glutamate in the pathophysiology of depression. *Acta psychiatrica Scandinavica*. 122:192-210.
48. Sanacora G, Zarate CA, Krystal JH, Manji HK (2008): Targeting the glutamatergic system to develop novel, improved therapeutics for mood disorders. *Nature reviews Drug discovery*. 7:426-437.
49. Sapolsky RM (2000): Glucocorticoids and hippocampal atrophy in neuropsychiatric disorders. *Archives of general psychiatry*. 57:925-935.
50. McEwen BS (1999): Stress and hippocampal plasticity. *Annual review of neuroscience*. 22:105-122.
51. McEwen BS (2001): Plasticity of the hippocampus: adaptation to chronic stress and allostatic load. *Annals of the New York Academy of Sciences*. 933:265-277.
52. Krishnan V, Nestler EJ (2008): The molecular neurobiology of depression. *Nature*. 455:894-902.
53. Kimpton J (2012): The brain derived neurotrophic factor and influences of stress in depression. *Psychiatria Danubina*. 24 Suppl 1:S169-171.
54. Hirschfeld RM (2000): History and evolution of the monoamine hypothesis of depression. *The Journal of clinical psychiatry*. 61 Suppl 6:4-6.
55. Berton O, Nestler EJ (2006): New approaches to antidepressant drug discovery: beyond monoamines. *Nature reviews Neuroscience*. 7:137-151.
56. Leistedt SJ, Linkowski P (2013): Brain, networks, depression, and more. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology*. 23:55-62.
57. Blier P (2013): Neurotransmitter targeting in the treatment of depression. *The Journal of clinical psychiatry*. 74 Suppl 2:19-24.
58. Jones KA, Thomsen C (2013): The role of the innate immune system in psychiatric disorders. *Molecular and cellular neurosciences*. 53:52-62.
59. Weissman MM, Kidd KK, Prusoff BA (1982): Variability in rates of affective disorders in relatives of depressed and normal probands. *Archives of general psychiatry*. 39:1397-1403.
60. Waraich P, Goldner EM, Somers JM, Hsu L (2004): Prevalence and incidence studies of mood disorders: a systematic review of the literature.

Canadian journal of psychiatry Revue canadienne de psychiatrie. 49:124-138.

61. Sullivan PF, Neale MC, Kendler KS (2000): Genetic epidemiology of major depression: review and meta-analysis. *The American journal of psychiatry.* 157:1552-1562.
62. Peterson BS, Wang Z, Horga G, Warner V, Rutherford B, Klahr KW, et al. (2014): Discriminating risk and resilience endophenotypes from lifetime illness effects in familial major depressive disorder. *JAMA psychiatry.* 71:136-148.
63. Goodwin FK, Jamison KR (2007): *Manic-Depressive Illness Bipolar Disorders and Recurrent Depression.* Oxford University Press; 2nd edition (2007).
64. Craddock N, Jones I (1999): Genetics of bipolar disorder. *Journal of medical genetics.* 36:585-594.
65. Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, et al. (2009): Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet.* 373:234-239.
66. McGuffin P, Rijdsdijk F, Andrew M, Sham P, Katz R, Cardno A (2003): The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Archives of general psychiatry.* 60:497-502.
67. Kieseppa T, Partonen T, Haukka J, Kaprio J, Lonnqvist J (2004): High concordance of bipolar I disorder in a nationwide sample of twins. *The American journal of psychiatry.* 161:1814-1821.
68. Edvardsen J, Torgersen S, Roysamb E, Lygren S, Skre I, Onstad S, et al. (2008): Heritability of bipolar spectrum disorders. Unity or heterogeneity? *Journal of affective disorders.* 106:229-240.
69. Mortensen PB, Pedersen CB, Melbye M, Mors O, Ewald H (2003): Individual and familial risk factors for bipolar affective disorders in Denmark. *Archives of general psychiatry.* 60:1209-1215.
70. McGuffin P, Katz R (1989): The genetics of depression and manic-depressive disorder. *The British journal of psychiatry : the journal of mental science.* 155:294-304.
71. Smoller JW, Finn CT (2003): Family, twin, and adoption studies of bipolar disorder. *Am J Med Genet C Semin Med Genet.* 123C:48-58.
72. Laursen TM, Labouriau R, Licht RW, Bertelsen A, Munk-Olsen T, Mortensen PB (2005): Family history of psychiatric illness as a risk factor for schizoaffective disorder: a Danish register-based cohort study. *Archives of general psychiatry.* 62:841-848.
73. Cohen-Woods S, Craig IW, McGuffin P (2013): The current state of play on the molecular genetics of depression. *Psychological medicine.* 43:673-687.
74. Lohoff FW (2010): Overview of the genetics of major depressive disorder. *CurrPsychiatry Rep.* 12:539-546.

75. Serretti A, Mandelli L (2008): The genetics of bipolar disorder: genome 'hot regions,' genes, new potential candidates and future directions. *Molecular psychiatry*. 13:742-771.
76. Hayden EP, Nurnberger JI, Jr. (2006): Molecular genetics of bipolar disorder. *Genes Brain Behav*. 5:85-95.
77. Craddock N, Sklar P (2013): Genetics of bipolar disorder. *Lancet*. 381:1654-1662.
78. Green EK, Hamshere M, Forty L, Gordon-Smith K, Fraser C, Russell E, et al. (2013): Replication of bipolar disorder susceptibility alleles and identification of two novel genome-wide significant associations in a new bipolar disorder case-control sample. *Molecular psychiatry*. 18:1302-1307.
79. Camp NJ, Cannon-Albright LA (2005): Dissecting the genetic etiology of major depressive disorder using linkage analysis. *Trends in molecular medicine*. 11:138-144.
80. Levinson DF (2006): The genetics of depression: a review. *BiolPsychiatry*. 60:84-92.
81. Bosker FJ, Hartman CA, Nolte IM, Prins BP, Terpstra P, Posthuma D, et al. (2011): Poor replication of candidate genes for major depressive disorder using genome-wide association data. *Molecular psychiatry*. 16:516-532.
82. Rietschel M, Mattheisen M, Frank J, Treutlein J, Degenhardt F, Breuer R, et al. (2010): Genome-wide association-, replication-, and neuroimaging study implicates HOMER1 in the etiology of major depression. *Biological psychiatry*. 68:578-585.
83. Sullivan PF, De Geus EJ, Willemsen G, James MR, Smit JH, Zandbelt T, et al. (2009): Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *MolPsychiatry*. 14:359-375.
84. Shi J, Potash JB, Knowles JA, Weissman MM, Coryell W, Scheftner WA, et al. (2011): Genome-wide association study of recurrent early-onset major depressive disorder. *Molecular psychiatry*. 16:193-201.
85. Muglia P, Tozzi F, Galwey NW, Francks C, Upmanyu R, Kong XQ, et al. (2010): Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. *Molecular psychiatry*. 15:589-601.
86. Wray NR, Pergadia ML, Blackwood DH, Penninx BW, Gordon SD, Nyholt DR, et al. (2012): Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *MolPsychiatry*. 17:36-48.
87. Kohli MA, Lucae S, Saemann PG, Schmidt MV, Demirkan A, Hek K, et al. (2011): The neuronal transporter gene SLC6A15 confers risk to major depression. *Neuron*. 70:252-265.
88. Lewis CM, Ng MY, Butler AW, Cohen-Woods S, Uher R, Pirlo K, et al. (2010): Genome-wide association study of major recurrent depression in the U.K. population. *The American journal of psychiatry*. 167:949-957.

89. Shyn SI, Shi J, Kraft JB, Potash JB, Knowles JA, Weissman MM, et al. (2011): Novel loci for major depression identified by genome-wide association study of Sequenced Treatment Alternatives to Relieve Depression and meta-analysis of three studies. *Molecular psychiatry*. 16:202-215.
90. Major Depressive Disorder Working Group of the Psychiatric GC, Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, et al. (2013): A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular psychiatry*. 18:497-511.
91. Barnett JH, Smoller JW (2009): The genetics of bipolar disorder. *Neuroscience*. 164:331-343.
92. Schulze TG, Akula N, Breuer R, Steele J, Nalls MA, Singleton AB, et al. (2012): Molecular genetic overlap in bipolar disorder, schizophrenia, and major depressive disorder. *The world journal of biological psychiatry : the official journal of the World Federation of Societies of Biological Psychiatry*. 15:200-208.
93. Sullivan PF, Daly MJ, O'Donovan M (2012): Genetic architectures of psychiatric disorders: the emerging picture and its implications. *NatRevGenet*. 13:537-551.
94. Lee JA, Lupski JR (2006): Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron*. 52:103-121.
95. Lachman HM, Pedrosa E, Petruolo OA, Cockerham M, Papolos A, Novak T, et al. (2007): Increase in GSK3beta gene copy number variation in bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet*. 144B:259-265.
96. Cook EH, Jr., Scherer SW (2008): Copy-number variations associated with neuropsychiatric conditions. *Nature*. 455:919-923.
97. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, et al. (2007): The diploid genome sequence of an individual human. *PLoS biology*. 5:e254.
98. Haraksingh RR, Snyder MP (2013): Impacts of variation in the human genome on gene regulation. *Journal of molecular biology*. 425:3970-3977.
99. Feuk L, Carson AR, Scherer SW (2006): Structural variation in the human genome. *Nature reviews Genetics*. 7:85-97.
100. Stankiewicz P, Lupski JR (2002): Genome architecture, rearrangements and genomic disorders. *Trends in genetics : TIG*. 18:74-82.
101. Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, et al. (2006): Copy number variation: new insights in genome diversity. *Genome research*. 16:949-961.
102. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. (2010): Origins and functional impact of copy number variation in the human genome. *Nature*. 464:704-712.
103. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. (2006): Global variation in copy number in the human genome. *Nature*. 444:444-454.

104. Iskow RC, Gokcumen O, Lee C (2012): Exploring the role of copy number variants in human adaptation. *Trends in genetics : TIG*. 28:245-257.
105. Henrichsen CN, Chaignat E, Reymond A (2009): Copy number variants, diseases and gene expression. *Human molecular genetics*. 18:R1-8.
106. Stankiewicz P, Lupski JR (2010): Structural variation in the human genome and its role in disease. *Annu Rev Med*. 61:437-455.
107. Inoue K, Lupski JR (2003): Genetics and genomics of behavioral and psychiatric disorders. *Current opinion in genetics & development*. 13:303-309.
108. Hastings PJ, Lupski JR, Rosenberg SM, Ira G (2009): Mechanisms of change in gene copy number. *Nature reviews Genetics*. 10:551-564.
109. Lupski JR (2007): Genomic rearrangements and sporadic disease. *Nature genetics*. 39:S43-47.
110. Weterings E, van Gent DC (2004): The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA repair*. 3:1425-1435.
111. Zhang F, Carvalho CM, Lupski JR (2009): Complex human chromosomal and genomic rearrangements. *Trends in genetics : TIG*. 25:298-307.
112. Lee JA, Carvalho CM, Lupski JR (2007): A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*. 131:1235-1247.
113. Bruder CE, Piotrowski A, Gijsbers AA, Andersson R, Erickson S, Diaz de Stahl T, et al. (2008): Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *American journal of human genetics*. 82:763-771.
114. Piotrowski A, Bruder CE, Andersson R, Diaz de Stahl T, Menzel U, Sandgren J, et al. (2008): Somatic mosaicism for copy number variation in differentiated human tissues. *Human mutation*. 29:1118-1124.
115. Almal SH, Padh H (2012): Implications of gene copy-number variation in health and diseases. *Journal of human genetics*. 57:6-13.
116. McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, et al. (2006): Common deletion polymorphisms in the human genome. *Nature genetics*. 38:86-92.
117. Buckland PR (2003): Polymorphically duplicated genes: their relevance to phenotypic variation in humans. *Annals of medicine*. 35:308-315.
118. Kleinjan DA, van Heyningen V (2005): Long-range control of gene expression: emerging mechanisms and disruption in disease. *American journal of human genetics*. 76:8-32.
119. Kleinjan DJ, Coutinho P (2009): Cis-ruption mechanisms: disruption of cis-regulatory control as a cause of human genetic disease. *Briefings in functional genomics & proteomics*. 8:317-332.
120. Pique-Regi R, Degner JF, Pai AA, Gaffney DJ, Gilad Y, Pritchard JK (2011): Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome research*. 21:447-455.

121. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, et al. (2010): Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS genetics*. 6:e1000952.
122. Dellinger AE, Saw SM, Goh LK, Seielstad M, Young TL, Li YJ (2010): Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. *Nucleic acids research*. 38:e105.
123. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, et al. (2009): Finding the missing heritability of complex diseases. *Nature*. 461:747-753.
124. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. (2010): Missing heritability and strategies for finding the underlying causes of complex disease. *Nature reviews Genetics*. 11:446-450.
125. Marian AJ (2012): Elements of 'missing heritability'. *Curr Opin Cardiol*. 27:197-201.
126. Slatkin M (2009): Epigenetic inheritance and the missing heritability problem. *Genetics*. 182:845-850.
127. Binder EB, Salyakina D, Lichtner P, Wochnik GM, Ising M, Putz B, et al. (2004): Polymorphisms in FKBP5 are associated with increased recurrence of depressive episodes and rapid response to antidepressant treatment. *Nature genetics*. 36:1319-1325.
128. Fava M, Rush AJ, Trivedi MH, Nierenberg AA, Thase ME, Sackeim HA, et al. (2003): Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study. *Psychiatr Clin North Am*. 26:457-494, x.
129. Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA, et al. (2004): Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Control Clin Trials*. 25:119-142.
130. Paddock S, Laje G, Charney D, Rush AJ, Wilson AF, Sorant AJ, et al. (2007): Association of GRIK4 with outcome of antidepressant treatment in the STAR*D cohort. *Am J Psychiatry*. 164:1181-1188.
131. Bijl RV, van Zessen G, Ravelli A, de Rijk C, Langendoen Y (1998): The Netherlands Mental Health Survey and Incidence Study (NEMESIS): objectives and design. *Soc Psychiatry Psychiatr Epidemiol*. 33:581-586.
132. Landman-Peeters KM, Hartman CA, van der Pompe G, den Boer JA, Minderaa RB, Ormel J (2005): Gender differences in the relation between social support, problems in parent-offspring communication, and depression and anxiety. *Soc Sci Med*. 60:2549-2559.
133. Boomsma DI, Beem AL, van den Berg M, Dolan CV, Koopmans JR, Vink JM, et al. (2000): Netherlands twin family study of anxious depression (NETSAD). *Twin Res*. 3:323-334.
134. Boomsma DI, Willemsen G, Sullivan PF, Heutink P, Meijer P, Sondervan D, et al. (2008): Genome-wide association of major depression: description of

- samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *EurJHumGenet.* 16:335-342.
135. <https://www.nimhgenetics.org/>.
 136. Nurnberger JI, DePaulo JR, Gershon ES, Reich T, Blehar MC, Edenberg HJ, et al. (1997): Genomic survey of bipolar illness in the NIMH genetics initiative pedigrees: a preliminary report. *Am J Med Genet.* 74:227-237.
 137. McMahon FJ, Buervenich S, Charney D, Lipsky R, Rush AJ, Wilson AF, et al. (2006): Variation in the gene encoding the serotonin 2A receptor is associated with outcome of antidepressant treatment. *American journal of human genetics.* 78:804-814.
 138. Gunderson KL, Kruglyak S, Graige MS, Garcia F, Kermani BG, Zhao C, et al. (2004): Decoding randomly ordered DNA arrays. *Genome research.* 14:870-877.
 139. Teo YY (2008): Common statistical issues in genome-wide association studies: a review on power, data quality control, genotype calling and population structure. *Current opinion in lipidology.* 19:133-143.
 140. Abecasis GR, Cherny SS, Cardon LR (2001): The impact of genotyping error on family-based analysis of quantitative traits. *European journal of human genetics : EJHG.* 9:130-134.
 141. Altshuler D, Daly MJ, Lander ES (2008): Genetic mapping in human disease. *Science.* 322:881-888.
 142. Stranger BE, Stahl EA, Raj T (2011): Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics.* 187:367-383.
 143. Olson JM, Witte JS, Elston RC (1999): Genetic mapping of complex traits. *Statistics in medicine.* 18:2961-2981.
 144. Risch N, Merikangas K (1996): The future of genetic studies of complex human diseases. *Science.* 273:1516-1517.
 145. Reich DE, Lander ES (2001): On the allelic spectrum of human disease. *Trends in genetics : TIG.* 17:502-510.
 146. Pritchard JK (2001): Are rare variants responsible for susceptibility to complex diseases? *American journal of human genetics.* 69:124-137.
 147. Schork NJ, Murray SS, Frazer KA, Topol EJ (2009): Common vs. rare allele hypotheses for complex diseases. *Current opinion in genetics & development.* 19:212-219.
 148. Pritchard JK, Cox NJ (2002): The allelic architecture of human disease genes: common disease-common variant...or not? *Human molecular genetics.* 11:2417-2423.
 149. Kallberg H, Ahlbom A, Alfredsson L (2006): Calculating measures of biological interaction using R. *EurJEpidemiol.* 21:571-573.
 150. Rice JP, Saccone NL, Corbett J (2001): The lod score method. *Advances in genetics.* 42:99-113.
 151. Strauch K, Fimmers R, Baur MP, Wienker TF (2003): How to model a complex trait. 1. General considerations and suggestions. *Human heredity.* 55:202-210.

152. Brennan P (1999): Design and analysis issues in case-control studies addressing genetic susceptibility. *IARC scientific publications*.123-132.
153. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. (2010): Missing heritability and strategies for finding the underlying causes of complex disease. *NatRevGenet*. 11:446-450.
154. Costanzo M, Baryshnikova A, Myers CL, Andrews B, Boone C (2011): Charting the genetic interaction map of a cell. *Current opinion in biotechnology*. 22:66-74.
155. Cordell HJ (2002): Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Human molecular genetics*. 11:2463-2468.
156. Phillips PC (2008): Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *NatRevGenet*. 9:855-867.
157. Moore JH (2003): The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *HumHered*. 56:73-82.
158. Cordell HJ (2009): Detecting gene-gene interactions that underlie human diseases. *NatRevGenet*. 10:392-404.
159. Steen KV (2012): Travelling the world of gene-gene interactions. *BriefBioinform*. 13:1-19.
160. Mani R, St Onge RP, Hartman JLt, Giaever G, Roth FP (2008): Defining genetic interaction. *Proceedings of the National Academy of Sciences of the United States of America*. 105:3461-3466.
161. Rothman, Greenland, Lash (2008): Concepts of interaction. In: Rothman KJ, Greenland S, Lash TL, editors. *Modern Epidemiology*, 3 ed. Philadelphia: Lippincott, Williams & Wilkins, pp 71-83.
162. Moore JH (2004): Computational analysis of gene-gene interactions using multifactor dimensionality reduction. *Expert review of molecular diagnostics*. 4:795-803.
163. Greene CS, Penrod NM, Kiralis J, Moore JH (2009): Spatially uniform relief (SURF) for computationally-efficient filtering of gene-gene interactions. *BioData mining*. 2:5.
164. Psychiatric GCBWDG (2011): Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature genetics*. 43:977-983.
165. Arabi A, Ullah K, Branca RM, Johansson J, Bandarra D, Haneklaus M, et al. (2012): Proteomic screen reveals Fbw7 as a modulator of the NF-kappaB pathway. *Nat Commun*. 3:976.
166. Brietzke E, Kapczinski F (2008): TNF-alpha as a molecular target in bipolar disorder. *Progress in neuro-psychopharmacology & biological psychiatry*. 32:1355-1361.
167. Barbosa IG, Nogueira CR, Rocha NP, Queiroz AL, Vago JP, Tavares LP, et al. (2013): Altered intracellular signaling cascades in peripheral blood mononuclear cells from BD patients. *J Psychiatr Res*. 47:1949-1954.

168. O'Brien WT, Klein PS (2009): Validating GSK3 as an in vivo target of lithium action. *Biochem Soc Trans.* 37:1133-1138.
169. Li X, Jope RS (2010): Is glycogen synthase kinase-3 a central modulator in mood regulation? *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology.* 35:2143-2154.
170. Alkelai A, Lupoli S, Greenbaum L, Kohn Y, Kanyas-Sarner K, Ben-Asher E, et al. (2012): DOCK4 and CEACAM21 as novel schizophrenia candidate genes in the Jewish population. *Int J Neuropsychopharmacol.* 15:459-469.
171. Blois SM, Sulkowski G, Tirado-Gonzalez I, Warren J, Freitag N, Klapp BF, et al. (2014): Pregnancy-specific glycoprotein 1 (PSG1) activates TGF-beta and prevents dextran sodium sulfate (DSS)-induced colitis in mice. *Mucosal immunology.* 7:348-358.
172. Kasper S, Spadone C, Verpillat P, Angst J (2006): Onset of action of escitalopram compared with other antidepressants: results of a pooled analysis. *International clinical psychopharmacology.* 21:105-110.
173. Garriock HA, Kraft JB, Shyn SI, Peters EJ, Yokoyama JS, Jenkins GD, et al. (2010): A genomewide association study of citalopram response in major depressive disorder. *Biological psychiatry.* 67:133-138.
174. Allen NB, Badcock PB (2006): Darwinian models of depression: a review of evolutionary accounts of mood and mood disorders. *Progress in neuropsychopharmacology & biological psychiatry.* 30:815-826.
175. Stein DJ, Kapfur D, Schatzberg A (2006): *American Psychiatric Publishing Textbook of Mood Disorders.* Washington D.C: American Psychiatric Publishing.
176. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, et al. (2010): The genetic landscape of a cell. *Science.* 327:425-431.
177. Stern DL, Orgogozo V (2009): Is genetic evolution predictable? *Science.* 323:746-751.
178. Bellay J, Atluri G, Sing TL, Toufighi K, Costanzo M, Ribeiro PS, et al. (2011): Putting genetic interactions in context through a global modular decomposition. *Genome research.* 21:1375-1387.
179. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, et al. (2007): Replicating genotype-phenotype associations. *Nature.* 447:655-660.
180. He H, Oetting WS, Brott MJ, Basu S (2009): Power of multifactor dimensionality reduction and penalized logistic regression for detecting gene-gene interaction in a case-control study. *BMC Med Genet.* 10:127.
181. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, et al. (2003): The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA : the journal of the American Medical Association.* 289:3095-3105.
182. Gibson G (2011): Rare and common variants: twenty arguments. *Nature reviews Genetics.* 13:135-145.

183. Risch NJ (2000): Searching for genetic determinants in the new millennium. *Nature*. 405:847-856.
184. Nestler EJ, Gould E, Manji H, Buncan M, Duman RS, Greshenfeld HK, et al. (2002): Preclinical models: status of basic research in depression. *Biological psychiatry*. 52:503-528.