Fine-mapping candidate genes for major depressive disorder: connecting the dots

Eva Caroline Verbeek
The research for this thesis was performed within the framework of TIPharma project T5-203 (www.tipharma.com). The infrastructure for the NESDA study (www.nesda.nl) is funded through the Geestkracht program of the Netherlands Organization for Health Research and Development (Zon-Mw, Grant number 10-000-1002) and is supported by participating universities and mental health care organizations. Further financial support for publication and distribution of this thesis was kindly provided by: VU University Amsterdam, Department of Psychiatry.

Cover design: Benno Hendriks
Lay-out: Persoonlijkproefschrift.nl, Liske van der Vliet
Published by: Ipskamp Printing B.V., Enschede, The Netherlands

Copyright © 2016 Eva Verbeek, The Netherlands
No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval without prior permission from the author. All rights reserved.
Fine-mapping candidate genes for major depressive disorder: connecting the dots

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op donderdag 12 januari 2017 om 9.45 uur
in de aula van de universiteit,
De Boelelaan 1105

doors

Eva Caroline Verbeek
geboren te Hilversum
Promotoren: prof.dr. W.J.G. Hoogendijk
prof.dr. P. Heutink
prof.dr. B.W.J.H. Penninx
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>General introduction and outline of the thesis</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>The Genetics of MDD – A Review of Challenges and Opportunities</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>A fine-mapping study of 7 top scoring genes from a GWAS for major depressive disorder</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>Resequencing three candidate genes for major depressive disorder in a Dutch cohort</td>
<td>67</td>
</tr>
<tr>
<td>5</td>
<td>Pathway-based analyses of candidate genes for major depressive disorder. A pilot study</td>
<td>91</td>
</tr>
<tr>
<td>6</td>
<td>A common polymorphism in the ABCB1 gene is associated with side effects of Pgp-dependent antidepressants in a large naturalistic Dutch cohort</td>
<td>109</td>
</tr>
<tr>
<td>7</td>
<td>General discussion and conclusions</td>
<td>133</td>
</tr>
<tr>
<td>Summary</td>
<td></td>
<td>155</td>
</tr>
<tr>
<td>Nederlandse samenvatting</td>
<td></td>
<td>161</td>
</tr>
<tr>
<td>Dankwoord</td>
<td></td>
<td>169</td>
</tr>
<tr>
<td>Dissertation series</td>
<td></td>
<td>173</td>
</tr>
</tbody>
</table>
CHAPTER 1

Introduction on Major Depressive Disorder and outline of the thesis
Chapter 1
INTRODUCTION ON MAJOR DEPRESSIVE DISORDER AND OUTLINE OF THE THESIS SYMPTOMS AND EPIDEMIOLOGY

Major depressive disorder (MDD) is a psychiatric disorder characterized by a prolonged depressed mood (> 2 weeks) or a loss of interest or pleasure in normally enjoyable activities, accompanied by a depressed mood during the larger part of the day, changes in appetite, changes in sleeping behavior, psychomotor retardation, feelings of guilt, a lack of concentration and suicidal thoughts (DSM-IV, American Psychiatric Association, 1994). It is one of the leading causes of disability in western civilization, and is predicted by the WHO to be the second leading cause of disability worldwide in 2020 (Murray & Lopez, 1996). MDD has a high disease burden for both patients as well as their relatives, with approximately only half of the patients responding to current treatment options (Papakostas & Fava, 2010).

The age of onset of MDD mostly begins at young to mid-adult age groups with a median age of onset in the early to mid-twenties, but the risk of having a depressive episode continues throughout the life course (Andrade et al., 2003). However, prevalence is lower in the population aged 50 years and older (Stewart & Lindesay, 2010). The lifetime prevalence is approximately 16.2% in western countries, with a mean episode duration of 16 weeks (Kessler et al., 2003). MDD is a complex disorder, emerging from an intricate combined effort of both environmental factors and multiple underlying genetic risk factors.

There are several risk factors known to predispose to MDD. Remarkable in depressive disorders is the difference in prevalence between women and men. In both adults and adolescents, MDD is twice as prevalent in women, in both clinical and population-based cohorts (Van de Velde et al., 2010). However, many reports are cross-sectional and do not follow participants over time. In longitudinal cohorts, an early age of onset is significantly correlated with the number of depressive episodes in both genders. Nevertheless, female participants of the cohorts reported a higher number of depressive episodes throughout the course of adolescence and adulthood than male participants (Essau et al., 2010).

Another major factor in the development of depressive symptoms is socioeconomic status (SES). Lower SES is often associated with decreased physical health and an increased level of mortality, high psychiatric morbidity and disability. Results for the association between SES and MDD have been inconsistent, as there are many different methods used to assess SES. When taking into account education level and household income, there is an association with MDD, regardless of gender (Wang et al., 2010). Also, when looking at patients treated for MDD, those with a
lower SES show significantly fewer reductions in depressive symptoms or gains in functioning than patients with a higher SES (Fauconnier, 2009).

Another predictor value for developing depressive symptoms, closely linked to SES, is marital status. Although there has been much methodological variation in the research of the effect of marital status on MDD, in 2010 a cross-sectional survey among households in 15 different countries revealed that marriage versus never married was associated with reduced risk of first onset of various mental disorders in both genders. However, for depression and panic disorder the association was limited to men. Being previously married, but not currently, increased the chance of mental disorders in both genders, but again the association with depression was strongest in men (Scott et al., 2010).

Independent of SES, ethnicity can be a predictor value for MDD. Various studies reported a lower prevalence of 12-month MDD among Blacks, Latinos and Asians compared to non-Hispanic Whites (Gavin et al., 2010). However, the chronicity of MDD is higher in the Latino and Black population compared to non-Hispanic Whites. This might be related to the fact that these groups have less consistent access to primary care than non-Hispanic Whites (González et al. 2010).

Also, 33% of patients suffering from chronic illness report symptoms of MDD. Co-morbid depressive symptoms have been reported –among others- for various forms of cancer, COPD, diabetes mellitus, cardiovascular disorders and autoimmune disorders (Rao, 2009; Aina & Susman, 2006).

Finally, experiences of stressors, particularly negative life events, are thought to play an important role in the development of MDD. Stressful life events (SLEs) that carry a high degree of threat and negativity, such as losing a spouse or becoming unemployed, have been consistently found to precede the onset of MDD (Hammen, 2005). It has been well documented in different studies that one of the most important predictors of emotional and behavioral dysfunction in children is a family history of depression. It is hypothesized that an impaired upbringing due to parental depression is, at least in part, responsible for children’s psychopathology (Silberg et al., 2010). In addition, from epidemiological studies it is known that SLE’s in critical early life stages, like emotional trauma, increase an individual’s risk for MDD and anxiety disorders later in life (Heim et al., 2004).
THE AIM OF THE THESIS

Except for the influence of gender, the aforementioned factors influencing the vulnerability for MDD all suggest environmental causes for the disease. However, being a complex disorder, the development of MDD is not limited to environmental risk factors: individuals without SLEs may be prone to developing (recurrent) MDD, while other individuals, with several SLE’s, may never develop depressive symptoms. Additionally, evidence for gene Environment interactions is mounting and it has even been found that SLEs might have a smaller effect in the absence of susceptibility genes, but a larger effect size in the presence of such genes (Rutter, 2008). From twin studies it is known that the heritability of MDD is approximately 40% (Kendler, 2001). However, depressive disorders show great heterogeneity in symptoms, severity, recurrence and age of onset and different prevalences of comorbidities across different populations (Edwards & Kendler, 2012; Demirkan et al., 2011; Czajkowski et al., 2010). This difference across populations suggests that several genetic variants with different effect sizes may play a role in the development of depressive symptoms and that these variants may differ between populations. Despite the existence of various hypotheses about the etiology of the disease, the exact mechanism behind MDD remains unknown.

Given the heritability of about 40% in MDD, it is important to increase knowledge of the genetic variants underlying the mechanisms of MDD. The aim of this thesis is therefore to contribute to the search for causal variants for MDD, in order to achieve better understanding of the etiology of the disease with the ultimate goal of contributing to more accurate diagnosis and more advanced therapeutic options. Chapter 2 provides a further introduction into the challenges of the genetic field of MDD research.

THE GAIN-MDD GWAS AND FOLLOW-UP RESEARCH

Nowadays, a widely used method for studying the genetics of complex disorders is the genome-wide association study (GWAS). In this type of study hundreds of thousands or even millions of common genetic variants for large numbers of DNA samples in which one compares the frequencies of the variants between cases and controls. A GWAS uses the “common disease, common variant-hypothesis” which assumes that disease arises from the coinheritance of multiple risk variants, each of a relatively small effect and that liability is normally distributed in the population. Since 2005, well over 1000 GWAS have been published, of which 115 with the focus on psychiatric disorders: bipolar disorder, attention deficit hyperactive disorder (ADHD), autism, Alzheimer’s disease, schizophrenia and MDD (Collins &
Sullivan, 2013). Eight GWAS for MDD have been published (Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium, 2013), revealing mostly sub-threshold results, with the exception of one study that found an association between MDD and a common SNP in the neuronal transporter gene \textit{SLC6A15} (Kohli et al., 2011).

The work presented in this thesis is largely based on the Genetic Association Information Network-MDD (GAIN-MDD) cohort. This was one of the first GWAS performed on MDD with the aim of identifying the variants associated with MDD. The GAIN-MDD initiative facilitated genotyping of 600,000 single nucleotide polymorphisms (SNPs) in MDD cases and population-based controls. 1702 cases came from the Netherlands Study of Depression and Anxiety (NESDA). The aim of the NESDA cohort is to study and describe the long-term course and the consequences of both depressive disorders and anxiety disorders. In addition, through this cohort, biological and psychosocial research are integrated. The NESDA cohort is a longitudinal cohort, in which the participants (ages 18-65) were followed for eight years. Individuals were recruited from general practitioners and from mental health organizations. In addition to medical examination and interviews of participants, blood and saliva were collected at baseline. With the method of ascertainment used for the NESDA cohort, extensive information on demography, psychosocial behavior, clinical characteristics, biological measurements and genetic markers was made available (Penninx et al., 2008) (see Table 1). 1700 controls came from the Netherlands Twin Registry (NTR) (Boomsma et al., 2006). Additionally, 160 cases from NTR and 157 controls from NESDA were added, as well as both parents from 33 controls, in order to form 33 trios and 21 duplicate samples (Boomsma et al., 2008). Results in of 1738 cases and 1802 controls of Dutch ancestry used in this study. For the NESDA cohort the Composite International Diagnostic Interview (CIDI) was used to diagnose MDD and to provide information about severity, specific symptoms, number of episodes and co-morbidities (World Health Organization, 1997). For the NTR, depression was diagnosed using Beck's Depression Inventory (BDI) (Beck et al., 1961) and using the Young Adult Self Report (YASR) (Achenbach, 1990). In addition, NTR cases filled out the CIDI.

<table>
<thead>
<tr>
<th>Diagnosis group</th>
<th>Number of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current/recent diagnosis of depression and/or anxiety</td>
<td>1701</td>
</tr>
<tr>
<td>Lifetime diagnoses or at risk</td>
<td>907</td>
</tr>
<tr>
<td>Healthy controls</td>
<td>373</td>
</tr>
</tbody>
</table>

In 2009 Sullivan et al. published a GWAS based on the results of the GAIN-MDD initiative. Many of the top signals of this GWAS mapped to a region spanning the
gene *PCLO* and the non-synonymous coding SNP rs2522833 in the *PCLO* gene became genome-wide significant after post-hoc analysis with a similarly ascertained Australian cohort (Sullivan et al., 2009). In addition, Bochdanovits et al. showed in a joint re-analysis of 29 SNPs in the *PCLO* region that rs2522833 was indeed the most likely causal variant in the GAIN-MDD cohort (Bochdanovits et al., 2009). Since this non-synonymous coding SNP was only marginally significant, it could well be that a previously ungenotyped neighboring SNP that is in high linkage disequilibrium (LD) may in fact be causal for the association found in the GAIN-MDD GWAS. The genotyping chip used for the GAIN-MDD GWAS was not designed in a gene-centric manner, leading to a lack of coverage of several genic regions. Therefore it cannot be excluded that another, less sufficiently covered region may be associated with MDD. In addition, next to the signals spanning *PCLO*, several other genes also showed sub-threshold signals. In order to further explore the possibility of an associated variant missed by lack of coverage, in Chapter 3 we perform a fine-mapping study of *PCLO* and six additional genes that showed top signals in the GAIN-MDD GWAS: *AFAP1L1*, *ANPEP*, *FGF14*, *GZMK*, *PTK2B* and *ST3GAL6*. With this fine-mapping effort, coverage of these genes was increased to contain all SNPs with a minor allele frequency (MAF) of at least 10% and a correlation coefficient ($r^2$), a measurement of LD, of 0.9.
**Box 1:** The function of PCLO in the presynaptic active zone

The results of the GWAS by Sullivan et al. show multiple top signals in the PCLO gene. This gene codes for the protein Piccolo, which is part of a family of presynaptically located cytoskeletal proteins. In the so-called “presynaptic active zone” vesicles containing neurotransmitters dock at and fuse with the cell membrane in a multistep process. Piccolo is part of the protein machinery that facilitates this process.

The non-synonymous coding SNP rs2522833 is located in the C2A calcium-binding domain of the PCLO gene. This SNP changes serine to alanine: a hydrophilic uncharged amino acid is replaced by a charged amino acid. This substitution may affect the stability of the Piccolo protein and hence its function. This is supported by the fact that a rearrangement of the secondary structure is required for binding Ca$^{2+}$ (Garcia et al., 2004) and furthermore, the domain was shown to be a phospholipid binding domain, likely to function in calcium-dependent signaling in synapses (Gerber et al., 2001). A non-synonymous coding SNP in this domain may therefore alter the calcium-dependent signaling, by interfering with the structural rearrangements required for Ca$^{2+}$ binding.

*Source of the image: Ziv & Garner, 2004 Nature reviews neuroscience pages 385-99*
Besides a lack of coverage of genic regions, there is the possibility that an association peak in a GWAS is caused by a yet undiscovered and/or ungenotyped variant. Since rs2522833 is a common variant with its MAF of 0.4, we hypothesized that an unknown causal variant in high LD with it is also likely to be a common variant. As a follow-up study for the GAIN-MDD GWAS, in Chapter 4 we aimed to identify this yet unknown common variant, by sequencing 50 control samples. As we hypothesized that the variant would be common, it should therefore also be present in control samples. Homozygotes for rs2522833 were selected to increase the possibility of finding other variants in high LD with this SNP.

In addition to PCLO, the genes GRM7 and SLC6A4 were also sequenced. Both genes have been studied extensively as candidate genes for MDD. GRM7 codes for a metabotropic glutamate receptor (mGluR) and showed sub-threshold P-values in a meta-analysis of three MDD cohorts (Shyn et al., 2009; Mitsukawa et al., 2006). In mice, having a targeted deletion of mGluR7, an anxiolytic and antidepressant effect was observed (Cryan et al., 2003). In addition, it was found that mGluR7 activation impaired the acquisition, but facilitated the extinction of aversive memories (Fendt et al., 2008). These studies suggest a role for mGluR7 in the mechanisms that regulate response to aversive events. Given the results in meta-analyses and the results in mice models, GRM7 was selected for resequencing in order to find genetic variants that may contribute to the development of MDD.

SLC6A4 codes for the serotonin transporter, which regulates the availability of serotonin in the synaptic cleft. It plays an important role in the monoamine hypothesis of depression, which states that depressive phenotypes are caused by an imbalance of monoamines such as serotonin. Association of the promoter region of this gene with MDD has been inconsistent, in spite of various large scale meta-analyses, and is still topic of discussion (Caspi et al., 2003; Lazary et al., 2008). In addition, variants in SLC6A4 have been found to influence the outcome of antidepressant therapy (Wilkie et al., 2009).

ENDOPHENOTYPES AND CANDIDATE GENE APPROACH

While Chapters 3 and 4 are based on further fine-mapping of the GAIN-MDD GWAS, a different approach was used for Chapter 5.

The candidate gene approach focuses on associations between genes of interest and a disease state. In order to select genes, there must be an a priori knowledge of specific genes and their biological relevance for the disorder that is being investigated. This shows a high contrast with the GWAS: the GWAS does not have an
priori hypothesis about specific target genes. Traditionally, the candidate gene approach is based on the analysis of a small number of variants or single-gene analysis. However, genes generally do not tend to function alone, but in complex networks. Genes are the functional units of the DNA and, although individual genetic variants such as SNPs may show differences in minor allele frequency (MAF) and linkage disequilibrium (LD) structure, genes show a high consistency across populations (Neale et al., 2004). The focus on individual genetic variants or single-gene analysis provides a challenge for elucidating the interactions between genes that underlie the molecular mechanisms of human complex disorders. Traditional genetic analysis through candidate genes and even through GWAS only, identifies only a small portion of the genetic heritability, and thereby only contributes to a limited understanding of the underlying mechanisms. The observation that only a small part of the heritability can be explained by genetic variations, is also known as the “missing heritability problem”. Moreover, replication appears to be difficult. Thus, in order to better interpret the molecular basis for complex disorders, not only the individual genes, but also their functionalities as a network may provide further information. A pathway-based analysis overcomes this problem by testing association between a set of functionally related genes and a disease phenotype. In Chapter 5, with the aim of testing the functionally related genes in the HPA axis, the HPT axis and vitamin D metabolism for association with MDD, we selected genes in these three pathways based on literature and genotyped tagging SNPs with MAF 0.2 and r² 0.8. In addition, several SNPs in related genes were added based on literature. However, psychiatric disorders have the disadvantage that often no clear biomarkers for the disease exist and phenotypes are highly heterogeneous. In order to dissect these heterogeneous phenotypes into more clear-cut phenotypes with a more obvious genetic connection, the “endophenotype concept” was born. Endophenotypes are heritable traits that are associated with illness in the population and that co-segregate with illness within families. In addition, in an ideal situation the endophenotype is primarily state-independent and will manifest in an individual whether or not the illness is active (Gottesman et al., 2003). In Chapter 5 we used the endophenotype approach to look for an association with MDD in three biologically relevant genetic pathways. In the NESDA cohort, biological measurements of participants were taken, including -but not limited to- morning cortisol, thyroid-stimulating hormone (TSH) and vitamin D levels. These measurements offer the opportunity to couple genetic variants to biologically relevant information. Also, the thorough phenotyping procedure of the NESDA cohort supplied additional information on the type of depression: moderate or severe, single episode or recurrent. To further investigate the effect of the endophenotype, ANOVA tests were performed, to see if 1) the biological measurements differed between cases and controls and 2) if this difference was associated with any of the genotyped alleles.
We genotyped variants for genes in the hypothalamus pituitary adrenal gland (HPA) axis, the hypothalamus pituitary thyroid (HPT) axis and genes involved in vitamin D metabolism. Dysregulation of the HPA axis is thought of as an important factor involved in the occurrence of depressive episodes (Palazidou, 2012). During a stressor, either physical or emotional, the HPA axis becomes activated. Upon experiencing the stressor, the starting point of the HPA axis, the hypothalamus, starts secreting corticotropin-releasing hormone (CRH). This hormone, known as corticotropin-releasing factor (CRF) in rodents, binds to receptors on the pituitary gland. The pituitary gland is then stimulated to release adrenocorticotropin hormone (ACTH). ACTH is transported through the bloodstream towards the adrenal glands. The cortex of the adrenal glands contains receptors that interact with ACTH, in order to stimulate production and release of cortisol, the so-called “stress hormone”. The HPA axis is completed by the negative feedback of cortisol on the hypothalamus, the pituitary gland and the hippocampus. When the concentration of cortisol rises, the release of both CRH and ACTH is inhibited (Varghese & Brown, 2001) (Figure 1). Cortisol is bound by two types of receptors: the mineralocorticoid receptor (MR) and the glucocorticoid receptor (GR). The latter has been shown to regulate the expression of neurotrophic factors such as BDNF, to affect neuronal cell death and neurogenesis in the hippocampus (Sousa et al., 2008). In severe MDD, the HPA axis shows signs of hyperactivity: In post-mortem material of MDD patients, the number of CRH expressing neurons in the hypothalamus is increased, supporting the hypothesis of hyperactivation (Raadsheer et al., 1994). In addition, patients were found to have significantly higher levels of cortisol in saliva and blood, while cortisol receptor function was reduced, thus impairing the negative feedback on the hippocampus, hypothalamus and pituitary gland (Pariante & Lightman, 2008). Brouwer et al. found that carriers of the G allele of non-synonymous coding SNP rs6195 in the GR showed higher levels of ACTH after exerting negative feedback on the HPA axis with dexamethasone and CRH. These carriers also showed less reduction in Hamilton Rating scale for depression after 8 weeks of paroxetine treatment (Brouwer et al., 2006). Also, high levels of cortisol inhibit cell proliferation in the hippocampus in both animal models and humans (Anacker et al., 2013). These observations suggest that abnormalities in GR function may contribute to changes in the brain in MDD. In addition to GR function, altered function of other receptors may contribute to the development of MDD. Genetic variants in CRH receptor 1 (CRHR1) have been associated with depression severity, increased risk for a seasonal pattern in depression and an increased risk for an early age of onset (Schatzberg et al., 2014). In addition, variants in CRHR2 show an association with poor response to the antidepressant drug citalopram (Papiol et al., 2007).
Secondly, we genotyped genes in the HPT axis. Upon sensing low levels of thyroid hormone, the hypothalamus releases thyrotropin-releasing hormone (TRH). TRH stimulates the production and release of thyroid-stimulating hormone (TSH) or thyrotropin in the pituitary gland. TSH in turn stimulates the thyroid gland to produce the thyroid hormones thyroxine (T4), and to a lesser extent, triiodothyronine (T3). However, the majority of T3 is produced in peripheral organs by the deiodination of circulating T4. Similar to the HPA axis, the HPT axis is also inhibited by a negative feedback loop: high levels of T3 inhibit the release of TRH from the hypothalamus and the release of TSH from the anterior part of the pituitary gland, by negatively regulating the TSH and TRH genes (Figure 2). Disturbances in the HPT axis, particularly those leading to hypothyroidism, produce symptoms similar to those in MDD: mood disturbances, fatigue, weight changes, irritability and changes in circadian rhythm are common in both groups of patients. In depressed patients levels of the pre-hormone thyroxine are consistently increased, whereas the levels of active thyroid hormone, triiodothyronine, have been shown to be decreased (Kirkegaard & Faber, 1998). In addition, nighttime levels of thyroid stimulating hormone are reduced in MDD patients (Bartalena et al., 1990). Also, a study of the response to the antidepressant drug paroxetine showed that higher serum
levels of TSH were associated with a better response in out-patients with MDD (Brouwer et al., 2005). These findings suggest alterations of the HPT axis in MDD.

Figure 2: The HPT axis

The third pathway that was selected was vitamin D metabolism. Vitamin D, or cholecalciferol, is a lipid soluble pro-hormone. It is taken in from dairy products and fish oils, but mainly synthesized in the skin from 7-dehydrocholesterol under the influence of ultraviolet (UV) light. In order to be biologically active, cholecalciferol is bound to vitamin D binding protein (VDBP or VBP) and transported to the liver. In the liver, cholecalciferol is hydroxylated to create the active form of vitamin D: 25-hydroxycholecalciferol (25(OH)D) (Christakos et al., 2010). The dependence on UV-light suggests a role for vitamin D in seasonal affective disorders, where the majority of patients experience depressive symptoms during winter (Rosenthal et al., 1984). In addition, studies investigating neurological targets of 25(OH)D showed a regulatory role in the production of aminergic messengers (Stumpf & O’Brien, 1987), interactions between the glucocorticoid receptor and 25(OH)D (Obradovic et al., 2006) and 25(OH)D response elements in the promotor regions of serotonin-related genes (Wang et al., 2005). Also, associations have been found between decreased 25(OH)D levels and depression in older adults (Hoogendijk et al., 2008) and between vitamin D receptor polymorphisms and depressive symptoms in adults (Kuningas et al., 2009). Although not all studies could be replicated, a meta-analysis on 31,424 participants found lower vitamin D
levels in patients with depression in comparison to healthy controls (Anglin et al., 2013). In the NESDA cohort, in patients with a current depressive episode, levels of 25(OH)D were significantly lower, showing a dose-response gradient with an increase of severity as 25(OH)D levels decreased (Milaneschi et al., 2014).

**PHARMACOTHERAPEUTIC TARGETS AND CHALLENGES**

Although complete remission of symptoms is the primary goal in the pharmacotherapeutical approach for MDD, it is estimated that eventually approximately 50%-70% of MDD patients benefit from treatment and one third of patients remains treatment resistant (Souery et al., 2006). In addition, patients do not always take their medication according to the prescription. Persistence of medication use after 6 months, the minimum duration of treatment, lies between 42 and 51% (Sawada et al., 2009). Besides the lack of response, side effects are also commonly experienced in antidepressant treatment of MDD. Side effects include, but are not limited to: sexual dysfunction, gastrointestinal disturbances, either somnolence or insomnia, dry mouth, dizziness, headaches and weight gain. The lack of a response and the large number of side effects are an important indicator for discontinuation of antidepressant treatment, as side effects influence quality of life (Fortney et al., 2011; Cohen et al., 2004; Chakrabortry et al., 2009).

In an ideal situation in pharmacological treatment, a balance has to be created where levels of available molecules have to be high enough to create a beneficiary effect, but low enough to keep the occurrence of side effects at a minimum. In antidepressants, availability is regulated by various enzymes located in both the liver and the brain. The drug efflux transporter permeability-glycoprotein (PgP) is an ATP-dependent efflux pump located in the blood brain barrier. It has a broad range of substrates and is therefore pivotal in eliminating drugs from the brain. It has been found to play a role in the pharmacokinetics of antidepressant medication such as venlafaxine, citalopram and paroxetine (Horstmann et al., 2009). Cytochrome P450 2C19 (CYP2C19) is known to particularly influence the pharmacokinetic profile of antidepressants and could therefore also influence the occurrence of side effects (Altar et al., 2013). This gene is known to be highly polymorphic, containing functional variants that have effects on CYP2C19 activity, showing ultrarapid, extensive, intermediate or poor metabolism of pharmaceutical agents (Hicks et al., 2013).

We hypothesized that variants in the genes coding for these enzymes may influence enzyme activity, leading to altered availability and therefore alterations in number of side effects. Therefore, in **Chapter 6** common variants in the genes **ABCB1**,
coding for PgP, and CYP2C19 were genotyped with 100% coverage at MAF=0.2 and r²=0.8 in order to test for an association with the number of side effects patients experienced. A total of 789 antidepressant users were classified as PgP-dependent or non-PgP-dependent antidepressant users. Association was tested for both the total number of side effects as well as categorized side effects (i.e. serotonergic, cholinergic and histaminergic).

Finally, in Chapter 7 the results are summarized and discussed. In addition, implications for future research of the genetic background of MDD are presented, completed with a reflection of the societal relevance of genetic MDD research.

REFERENCES

Achenbach TM (1990) The Young Adult Self Report; Burlington, VT; University of Vermont, Dept of Psychiatry
Anglin RE, Samaan Z, Walter SD, McDonald SD (2013) Vitamin D deficiency and depression in adults: systematic review and meta-analysis; *Br J Psychiatry*; Feb;202:1-10
for behavior genetics and genomics studies. *Twin Res Hum Genet*; Jun;11(3):342-8


Cryan JF, Kelly PH, Neijt HC, Sansig G, Flor PJ, van Der Putten H (2003) Antidepressant and anxiolytic-like effects in mice lacking the group III metabotropic glutamate receptor mGluR7; *Eur J Neurosci*; Jun;17(11):2409-17


Essau CA, Lewinsohn PM, Seeley JR, Sasagawa S (2010) Gender differences in the developmental course of depression; *J Affect Disord*; Dec;127(1-3):185-90


Hoogendijk WJ, Lips P, Dolk MG, Deeg NJ, Beekman AT; Penninx BW (2008) Depression is associated with decreased 25-hydroxyvitamin D and increased parathyroid hormone levels in older adults; *Arch Gen Psychiatry*; May;65(5):508-12

Horstmann S, Binder EB (2009) Pharmacogenomics of antidepressant drugs; *Pharmacol Ther*; Oct;124(1):57-73


Pariante CM, Lightman SL (2008) The HPA axis in major depression: classical theories and new developments; *Trends Neurosci*; Sep;31(9):464-8


Raadsheer FC, Hoogendijk WJ, Stam FC, Tilders FJ, Swaab DF (1994) Increased numbers of corticotropin-releasing hormone expressing neurons in the hypothalamic paraventricular nucleus of depressed patients; *Neuroendocrinology*; Oct;60(4):436-44


Stumpf WE, O’Brien LP (1987) 1,25(OH)₂ vitamin D₃ sites of action in the brain: An autoradiographic study; *Histochemistry*; Sep;87, 5:393-406


Wang JL, Schmitz N, Dewa CS (2010) Socioeconomic status and the risk of major depression: the Canadian National Population Health Survey; *J Epidemiol Community Health*; May;65(5):447-52


Willkie MJ, Smith G, Day RK, Matthews K, Smith D, Blackwood D, Reid IC, Wolf CR (2009) Polymorphisms in the SLC6A4 and HTR2A genes influence treatment outcome following antidepressant therapy; *Pharmacogenomics*; Feb;9(1):61-70

World Health Organization: Composite International Diagnostic Interview, Core Version 2.1: Interviewer’s Manual; Sydney, Australia; World Health Organization; 1997

CHAPTER 2
The Genetics of MDD – A Review of Challenges and Opportunities

Eva C. Verbeek
Marianna R. Bevova
Witte J. Hoogendijk
Peter Heutink

ABSTRACT

Major depressive disorder (MDD) is a psychiatric disorder characterized among others by prolonged depressed mood, a loss of interest in enjoyable activities, psychomotor retardation and various cognitive symptoms. Although exact numbers of prevalence may differ between various western countries, partly due to a social taboo of the illness, the life-time prevalence in the USA and Western Europe lies around 15%. Women are more likely to be struck by MDD than men, often with a first episode between 30-40 years of age, with a smaller second peak around 50-60 years of age. Although MDD may appear as a "stand alone" disease, 33% of patients with a chronic illness report symptoms of major depression. In addition, approximately 72% of patients diagnosed with MDD also have a second mental illness diagnosed, in most cases generalized anxiety disorder or a social phobia. Although some patients may only experience a single episode, MDD often appears in multiple episodes.
INTRODUCTION

Major depressive disorder (MDD) is a psychiatric disorder characterized among others by prolonged depressed mood, a loss of interest in enjoyable activities, psychomotor retardation and various cognitive symptoms (American Psychiatric Association, 1994). Although exact numbers of prevalence may differ between various western countries, partly due to a social taboo of the illness, the life-time prevalence in the USA and Western Europe lies around 15%. Women are more likely to be struck by MDD than men, often with a first episode between 30-40 years of age, with a smaller second peak around 50-60 years of age. Although MDD may appear as a “stand alone” disease, 33% of patients with a chronic illness report symptoms of major depression. In addition, approximately 72% of patients diagnosed with MDD also have a second mental illness diagnosed, in most cases generalized anxiety disorder or a social phobia (Kessler et al., 2005). Although some patients may only experience a single episode, MDD often appears in multiple episodes.

As annually 6% of adults have an episode of depression, the burden on primary care is high. MDD is the third most common reason for a consultation in primary care. Besides this impact, there is a burden on society as MDD is predicted to be the second leading cause of disability by 2020. Although the patient and his/her care givers will have to deal with the emotional burden of the disorder, the burden on healthcare and society are also high. The costs of MDD in the US alone were estimated at 83.1 billion in 2000 (Greenberg et al., 2003). MDD can be classified as a complex disorder: it is likely associated with the effects of multiple genes in combination with lifestyles and environmental factors. When looking at demographic data, there is a higher prevalence in individuals with a low socioeconomic status in combination with an urban living area. In addition, stressful life events (SLEs), such as the loss of a spouse or abuse, increase the likelihood for an individual to develop MDD. However, not everyone who has suffered SLEs will develop MDD later in life. In addition, there are patients who have never experienced SLEs, but who do suffer from (recurrent) depressive episodes.

There is strong evidence that genetic factors may predispose individuals to the development of MDD. In a Swedish twin study the heritability of MDD was estimated at approximately 40% (Kendler et al., 2006). In most complex disorders, a large number of genes contribute to the disorder, each gene only responsible for a slight increase in risk. Because of this multigenic aspect, disorders such as MDD show a familial aggregation that does not resemble Mendelian inheritance. In addition, evidence for gene-environment interactions is mounting. It has been shown that even SLEs might have a negligible effect in the absence of relevant susceptibility genes, but a very large effect in the presence of such genes (Rutter et al., 2009).
addition, the heritability of MDD in conjunction with various other environmental traits has been investigated. For nicotine dependence there is a 32% shared liability with MDD (Edwards & Kendler, 2012) and in an elderly cohort, when looking at genetic risk scores of MDD in conjunction with anxiety, 2.1% of variation was explained (Demirkan et al., 2011). However, this percentage shows large differences between different cohorts. For instance, in a sample of Norwegian families, the combined heritability of MDD and generalized anxiety disorder was estimated at 25%. This difference across populations suggests that several variants with different effect sizes may play a role in the development of depressive symptoms, with specific variants playing a role in a specific population (Czaikowski et al., 2010).

In depressive disorders there is a remarkable difference in prevalence between women and men. In western countries, MDD is approximately twice more prevalent in women than in men, in both clinical and population-based cohorts (Van de Velde, 2010). This has been reported in adults as well as adolescents. However, many of these reports are cross-sectional and do not follow participants over time. In longitudinal cohorts, an early age of onset of depression is significantly correlated with the number of depressive episodes in both genders. Nevertheless, female participants reported a higher number of depressive episodes throughout the course of adolescence and adulthood than male participants (Kessler et al., 1993), which may suggest a putative role for sex hormones in the development of MDD. However, the view on prevalence may be skewed due to a higher likelihood of female patients reporting psychological and physical symptoms and to seek medical attention (Spinhoven & Kooiman, 1997). In addition, although suicide ideation in most Western countries is more apparent in women, mortality from suicide is typically higher for men (Canetto & Sakinofsky, 1998).

**Genetic Techniques to Detect Variants**

Genetic research into complex disorders such as MDD has been through an evolution that started with genetic linkage studies. These studies are based on the frequencies of recombination between markers. It is assumed that the greater the recombination frequency between two markers, the greater the distance between them and compares the likelihood of finding the obtained data versus the likelihood of finding the same data by chance. For complex traits, the most commonly used method of linkage studies is to examine marker allele sharing between pairs of affected relatives, for instance pairs of siblings. If sib-pairs share alleles more often than would be expected by chance, then this suggests that a susceptibility locus may be linked to a marker.

Recently, one of the largest linkage studies on depressive disorders, performed by Breen et al., was published. It comprised of 971 affected siblings of European de-
scent with recurrent MDD (RE-MDD) of various severities. Individuals were classified according to severity, after which in the linkage was found on chromosome 3p25-26. Importantly, this was only found for individuals with a moderate phenotype, but not for milder cases or for very severe cases (Breen et al., 2011). The same region showed evidence of linkage with MDD in a smaller cohort of families of heavy smokers (Pergadia et al., 2011). This may suggest different underlying genetic mechanisms for phenotypes with different severities and different comorbidities.

In spite of these findings, most of the results identified by linkage studies were not replicated. For the findings in large multigenerational families this might be explained that the identified risk factors are extremely rare causes of the disease and only very few large families with a Mendelian inheritance pattern have been reported. Sib-pair studies have very often been underpowered, especially to detect common risk factors with small effect size. In the instances where findings were replicated, this was mostly on the same cohort, but with a more stringent phenotype, i.e. in heavy smokers, depression with suicide attempts or early onset recurrent depression. Linkage studies, especially on sib pairs, are a low resolution method and therefore less suited to zoom in on specific genes than newer, higher resolution methods. In extended families however, this is not necessarily the case. If more affected family members are included, this increases the possibility to narrow down the region involved in the disease.

A second widely used method of genetic analysis is the candidate gene association approach. This approach uses genes that have been specified beforehand to look for an association between these genes and a phenotype. Genes are selected based on a priori knowledge of the biological function of the gene, after which a hypothesis is generated on how this biological function is implicated in the development of the phenotype under investigation. This is also the immediate advantage of candidate gene studies: once an association is found, it is usually also known which biological function is involved. However, the information about biological function may not always be complete, leading to incorrect assumptions about these functions. In MDD, candidate gene association studies have implicated various suspected risk genes, but at the same time they are hindered by the focus on single gene.

One of the best known examples of the candidate gene approach in MDD is the serotonin transporter, SLC6A4. This gene regulates the availability of serotonin in the synaptic cleft and it is the target of various antidepressant drugs. The length polymorphism in SLC6A4 has been investigated in numerous studies and associations with both unipolar and bipolar depression have been found, but replication
in different cohorts proved to be a challenge (Collier et al., 1996; Caspi et al., 2003; Mendlewicz et al., 2004; Risch et al., 2009; Munafò et al., 2009).

Another prominent candidate gene is the brain-derived neurotrophic factor gene, BDNF. Nibuya et al. showed that a prolonged exposure to antidepressant medication, including SSRIs, caused an increase of BDNF protein in hippocampal regions (Nibuya et al., 1995). In addition it was shown that administration of BDNF has antidepressant effects (Siuciak et al., 1997). In several studies of genetic polymorphisms in BDNF, no significant effect was found (Musil et al., 2013; Pae et al., 2012; Surtees et al., 2007). However, several studies report gene x environment interaction of BDNF and associations with depressive states in bipolar disorder and schizophrenia (Schumacher et al., 2005; Neves-Pereira et al., 2002; Sun et al., 2013). In addition, postmortem studies have revealed a decrease in brain-derived neurotrophic factor (BDNF) in the hippocampus and an increase of vasopressin- and oxytocin expressing neurons in the hypothalamus of patients suffering from depression (Purba et al., 1996). These two examples of candidate genes illustrate that even though a clear biological function may exist, this is not always reflected in an associated outcome. Here the relatively small sample sizes of many candidate gene studies may interfere with finding an association. In an effort to replicate candidate genes by Bosker et al., sample size was increased by using data of a larger study (Bosker et al., 2011). Candidate genes were gathered from literature and coverage of these genes in existing data was enriched by imputation. Unfortunately, even with this larger sample size, replication was still poor. However, these studies do not take into account indirect associations such as gene x environment interactions.

With the birth of the genome wide genotyping techniques like microarrays, the opportunity arose to perform genome wide association studies (GWAS) without an a priori stringent hypothesis. In contrast to the candidate gene approach, GWAS scan the entire genome for associated variants based on hundreds of thousands or even millions of common genetic variants in which one looks for a difference of frequency of these variants between cases and controls. However, when performing a GWAS, one assumes that common variants are causal for common disorders, which is the so called “common disease, common variant hypothesis”. This hypothesis assumes that disease arises from the coinheritance of multiple risk variants, each of a relatively small effect and that liability is normally distributed in the population. To explain prevalence of a common disorder in a particular population, the variants have to be common and therefore should be observed when performing a GWAS (Busuyi et al., 2012). A major drawback of the GWAS approach is that, in order to obtain sufficient statistical power, a large cohort of comparable cases and
comparable controls is required. Particularly in psychiatric disorders, where the phenotype may be very diffuse, a large sample size is an inevitable necessity.

In 2009, Sullivan et al. published one of the first GWAS for MDD, which was performed on a cohort of 3540 individuals of Western European ancestry (Sullivan et al., 2009). In this GWAS, various sub-threshold signals were detected, but no genome-wide significant results were found. In many GWAS on complex traits such as height and Alzheimer’s disease, the cohort size was considerably larger, so one might argue that the statistical power is too low to detect a common variant with small effect (Hemani et al., 2013; Lambert et al., 2009; Harold et al., 2009). However, several top signals mapped back to a genomic region overlapping the gene PCLO. When replication was performed with the Australian QIMR cohort, which used a similar method of ascertainment as the cohort used by Sullivan et al., the non-synonymous coding SNP rs2522833 became marginally significant (P=6.4E-08). In addition, a fine-mapping study and a joint re-analysis of 29 SNPs surrounding rs2522833 supported the hypothesis of a causal role for this SNP (Verbeek et al., 2012; Bochdanovits et al., 2009).

However, this finding was not replicated by Shyn et al. on a different cohort of European ancestry (Shyn et al., 2011). Differences in inclusion criteria may be crucial in these different findings. Cases were of European ancestry and were determined using DSM-IV criteria, but contrary to the GAIN-MDD GWAS, Shyn et al. used the Hamilton Depression Rating score instead of the CIDI interview. In addition, in the STAR*D cohort used by Shyn et al., the ages were 18-75, whereas in the GAIN-MDD cohort ages were 18-65. These different inclusion criteria may be causal to the lack of replicating PCLO in this cohort. Furthermore, Shyn et al. performed a meta-analysis on three studies: the STAR*D cohort, the GAIN-MDD cohort and the Genetics of Recurrent Early-Onset Depression (GenRED) cohort (Shi et al., 2011). The strongest evidence for association was found for several intronic SNPs in the genes ATP6V1B2, SP4 and GRM7. However, no genome-wide significance was found. Theoretically, the increase of sample size increases the statistical power to detect an associated variant. However, when performing a meta-analysis, it is of the utmost importance that populations are indeed comparable.

In addition to the “common disease, common variant hypothesis”, there is also the possibility of the “common disease, multiple rare variant hypothesis”. This hypothesis suggests that common disorders such as MDD are caused by multiple variants with relatively low minor allele frequencies. Another possibility is a combination of both rare and common variants.
Taking into account the diffuse phenotypes of psychiatric disorders, it may well be that different variants, with different effect sizes, are responsible for different severities and different recurrence patterns found MDD. In general, rare variants do not give a clearly detectable association peak as a GWAS aims for common variants and thus rare variants would only appear as noise in such a study design. With the emergence of next generation sequencing (NGS) techniques, the ability to detect new and especially low frequency variants increased. Over the past years, the capacity of sequencing increased from parts of genes to the systematic sequencing of entire genomes. With the decrease of complexity to detect new variants, the door is now opened to not only find high numbers of previously undetected common variants, but also high numbers of rare variants specific to a certain population or a certain disorder. When applying this to complex disorders, the era of GWAS brought a substantial number of associated common variants, but a considerable void in the heritability remains. Research of complex disorders is currently shifting from common variants towards lower frequency (1-5%) and even rare variants (<1%), but this improved cataloguing of variation in the human genome does not necessarily lead to successful association analyses. In a 2013 sequencing effort by Quast et al., two rare variants were discovered, validated and found to be significantly associated with neutral amino acid transporter SLC6A15 functioning (Quast et al., 2013). A common variant in the same gene was previously associated with MDD in a cohort of ± 700 individuals and later replicated in six cohorts of similar magnitude (Kohli et al., 2011). The fact that these variants are associated in these limited sample sizes implies that they have a larger effect size than most common variants found in a GWAS.

**Considerations in Genetic Analysis of MDD**

In complex disorders, multiple genetic risk factors play a role. However, when searching for associated variants, the individual effect sizes of these variants may be so small, that an association may go overlooked. Even in the previous example of Quast et al., only a minority of the heritability can be explained. In psychiatric disorders in general, it is estimated that the currently associated variants explain roughly 2% of the heritability (Crow 2011). More specifically, in MDD, the estimate lies around 1% and thus the vast majority of genetic risk factors remain to be identified (Lubke et al., 2012). With the evolution of sequencing techniques, the possibility to find rare variants has increased. However, when looking for associations with rare variants, sample sizes would have to increase dramatically. An alternative is to look for pathway or network-based associations. In schizophrenia, the analysis of functional gene groups has identified new variants (Lips et al., 2012), but in MDD this method has not been widely used yet.
In addition, it has been suggested that de novo mutations may contribute to a part of the heritability for complex genetic diseases that is not detectable by genome-wide association studies, because their frequency in the population is too low.

De novo mutations occur in every individual, so as a phenomenon they are not rare. Veltman & Brunner suggested that it is possible that de novo mutations are responsible for an important fraction of more commonly occurring diseases by disrupting any one of a large number of genes (Veltman & Brunner, 2012). This implies that there may be low numbers of mutations that represent a relatively large effect and stands in sharp contrast with the thought that high numbers of common variants cause complex disorders. Of course, a mixed model with common and rare variants and de novo mutations is also one of the possibilities. With an estimation of 74 new SNPs per generation, it is not unthinkable that new mutations are also part of the picture.

Additionally, research into combinations of variants could be an worthwhile investment. This combinatory effect of genes, epistasis, has been investigated in conjunction with various disorders such as Alzheimer’s disease and type two diabetes (Mateo et al., 2009; Wiltshire et al., 2006). Epistasis is more than the sum of single locus effects, and therefore assumes that the phenotypic effect of one variant depends on the genotype of another variant. In MDD, there has been a report of variants in SLC6A4 and BDNF that show interaction, where a certain variant of BDNF shows a protective function against the 5-HTTLPR length polymorphism in SLC6A4 (Pezawas et al., 2008). Nonetheless, in spite of the obvious necessity of research into epistatic effects, there is still much debate on how to model and test for both main effects and interactions when one expect epistasis to be present (Verhoeven et al., 2010). As computational power would have to be tremendous in order to predict all possible gene-gene interactions genome-wide, algorithms to calculate epistasis are in need of improvement. One method of reducing the required computational power is described by Bochdanovits et al., in which the number of pair-wise tests is reduced by enriching for gene pairs predicted to be more likely to jointly affect variation in complex traits (Bochdanovits et al., 2008). However, for such a method a good knowledge of gene function and function of protein domains is required.

Besides studying association of variants in the DNA sequence itself, the search for heritable changes in gene activity, epigenetics, may be a valuable addition. It was shown previously that there is an association between SLEs and epigenetic modification of gene expression (Dalton et al., 2014). These influences on gene expression may provide an additional explanation for the currently missing herit-
ability. In addition to epigenetics, gene expression studies on postmortem material may be a valuable extension of determining differences in the healthy brain versus the brain of an individual that suffered from MDD. Studies have already revealed differences in expression of BDNF and nerve growth factor (NGF) and their receptors in the hippocampus (Banerjee et al., (2013), a reduced expression of fatty acid biosynthesis genes in the prefrontal cortex of depressed patients (McNamara & Liu (2011), decreased expression of thyrotropin-releasing hormone (Meynen et al., 2006) and increased expression of vasopressin (Alkemade et al., 2003) in the hypothalamus. As not only the techniques for DNA-sequencing, but also the techniques for RNA-sequencing have dramatically improved over the last decade, the doors are now open to more insight in differential expression profiles in the brain of MDD patients.

In addition to an increased catalogue of variants in the human genome and the search for less than straightforward associations, the correct assembly of a cohort is also be of vital importance for finding a causal variant. In a common disorder a large detrimental effect of a single variant is not expected. Otherwise the disease would not be common: there would have been selection against this variant in the population. This automatically implies that a cohort of substantial size is mandatory to find an association, or variants will remain sub-threshold.

Also, in complex disorders the phenotype is often fickle, thus making careful and clear phenotyping a requirement. Where some patients will experience a single moderate depressive episode, another patient will suffer from severe recurrent unipolar depression throughout a substantial part of his or her life. Although the type of symptoms will be the same, the intensity and recurrence of symptoms are different, so one could argue that these are different phenotypes, caused by different genetic variants.

In order to create more clear-cut phenotypes with a more obvious genetic connection, endophenotypes may be useful. Endophenotypes are heritable properties that are associated with the disease and show co-segregation with the disease within families. In MDD, anhedonia, the impairment of the reward system, may be a good candidate for an endophenotype. Anhedonia is very specific to depression and enhanced rewarding effects of dextroamphetamine have been found in patients with MDD. This suggests hypofunction of the dopaminergic system associated with anhedonia (Tremblay et al., 2002). In addition, dysfunction of the reward system has been suggested to be heritable as well (Dreher et al., 2009). One of the suggested endophenotypes in psychiatric disorders is brain imaging, although there is a lot of debate about the heritability of brain activity patterns (Blokland et al., 2011). In line with this suggestion for an endophenotype is the imaging study
of Woudstra et al. In this study, the PCLO risk allele that was found by Sullivan et al., is associated with altered emotion processing. In addition, during processing of fearful emotions, the PCLO risk allele was associated with increased activation in the amygdala of MDD patients (Woudstra et al., 2012). This example shows the benefits of clear-cut (endo) phenotypes, when trying to look for a functional connection.

In summary, the possibilities to detect and map genetic variation have taken a giant leap forward and the detection of variants is no longer a rate-limiting step. This provides the research of complex disorders with new tools to find associations. However, effect size still presents a challenge, for which strict phenotyping and substantial cohort size are mandatory. With these new developments in genetics, the view on complex disorders may have to be adjusted. During the era of the GWAS, common disease was mostly hypothesized to be caused by common variants. However, the discovery of rare associated variants and the putative contribution of de novo mutations forces us to reconsider the common disease common variant hypothesis. With current techniques and knowledge, it now seems more likely that the recipe for common disease is a mixture of common, rare and new variants with variations in effect sizes. Additionally, combining the effect of variants by means of epistatic research may prove valuable, as complex disease such as MDD is caused by multiple variants that may interact. Research into the pathophysiology of MDD has the ultimate goal of improving treatment, but the interpretation of genetic findings in this respect is still a challenge. Despite the increase in identified variants, few of the SNPs found in studies have clear functional implications. The process of translating an association to the comprehension of a variant’s functionality is the next barrier in the genetics of MDD and other complex disorders.

Thus, as a future perspective, the rate-limiting step in MDD research may no longer be the detection of variants, but the even more complicated boundary of developing assays in cells and animal models to assess the biological effects of implicated variants.

**ACKNOWLEDGEMENTS**

This review was written within the framework of the Center for Medical Systems Biology (CMSB) and TIPharma project T5-203. We thank the Netherlands Study of Depression and Anxiety (NESDA) and the Dutch Twin Registry (NTR) for the opportunity to work with their cohort data and the Neuroscience Campus Amsterdam for their support.
REFERENCES


Blokland GA, McMahon KL, Thompson PM, Martin NG, de Zubicaray GI, Wright MJ (2011) Heritability of working memory brain activation; J Neurosci; 31: 10882-10890


McNamara RKI, Liu Y (2011) Reduced expression of fatty acid biosynthesis genes in the prefrontal cortex of patients with major depressive disorder; J Affect Disord; 129: 359-363
Purba JS, Hoogendijk WJ, Hofman MA, Swaab DF (1996) Increased number of vasopressin- and oxytocin-expressing neurons in the paraventricular nucleus of the hypothalamus in depression; Arch Gen Psychiatry; 53: 137-143


Siuciak JA, Lewis DR, Wiegand SJ, Lindsay RM (1997) Antidepressant-like effect of brain-derived neurotrophic factor (BDNF); Pharmacol Biochem Behav; 56: 131-137


Tremblay LK, Naranjo CA, Cardenas L, Herrmann N, Busto UE (2002) Probing brain reward system function in major depressive disorder: altered response to dextroamphetamine; Arch Gen Psychiatry; 59:409-416
CHAPTER 3
A Fine-Mapping Study of 7 Top Scoring Genes from a GWAS for Major Depressive Disorder

Eva C. Verbeek
Ingrid M. Bakker
Marianna R. Bevova
Zoltán Bochdanovits
Patrizia Rizzu
David Sondervan
Gonneke Willemsen
Eco J. de Geus
Johannes H. Smit
Brenda W. Penninx
Dorret I. Boomsma
Witte J. Hoogendijk
Peter Heutink

ABSTRACT

Major depressive disorder (MDD) is a psychiatric disorder that is characterized amongst others by persistent depressed mood, loss of interest and pleasure and psychomotor retardation. Environmental circumstances have proven to influence the aetiology of the disease, but MDD also has an estimated 40% heritability, probably with a polygenic background. In 2009, a genome wide association study (GWAS) was performed on the Dutch GAIN-MDD cohort. A non-synonymous coding single nucleotide polymorphism (SNP) rs2522833 in the PCLO gene became only nominally significant after post-hoc analysis with an Australian cohort which used similar ascertainment. The absence of genome-wide significance may be caused by low SNP coverage of genes. To increase SNP coverage to 100% for common variants (m.a.f.>0.1, r²>0.8), we selected seven genes from the GAIN-MDD GWAS: PCLO, GZMK, ANPEP, AFAP1L1, ST3GAL6, FGF14 and PTK2B. We genotyped 349 SNPs and obtained the lowest P-value for rs2715147 in PCLO at P=6.8E−7. We imputed, filling in missing genotypes, after which rs2715147 and rs2715148 showed the lowest P-value at P=1.2E−6. When we created a haplotype of these SNPs together with the non-synonymous coding SNP rs2522833, the P-value decreased to P=9.9E−7 but was not genome wide significant. Although our study did not identify a more strongly associated variant, the results for PCLO suggest that the causal variant is in high LD with rs2715147, rs2715148 and rs2522833.
INTRODUCTION

Major depressive disorder (MDD) is a psychiatric disorder characterized by persisting depressed mood, loss of interest or pleasure in normally enjoyable activities, psychomotor retardation and changes in e.g. sleep and appetite (American Psychiatric Association, 1994). The lifetime prevalence in western civilization is estimated to be approximately 10–15% and the World Health Organization has predicted that by the year 2020, MDD will be the second leading cause of disability worldwide (Murray & Lopez, 1996).

Though the etiology of the disease remains elusive, a genetic component is recognized and, based on twin studies, heritability is estimated to be around 40% (Kendler et al., 2006; Sullivan et al., 2000; Levinson, 2006). However, MDD is a complex disorder and so far causal variants have proven to be difficult to find. For candidate genes, many association studies have been conducted, but this has not resulted in reproducible identification of susceptibility genes, because findings have often been inconsistent. This may be explained by methodological differences (i.e. difference in study design, study population, diagnostic criteria) or small sample sizes (Lopez-Leon et al., 2008).

With the introduction of genome-wide association studies (GWAS), a systematic hypothesis-free search for common susceptibility genes became possible. The Netherlands Study for Depression and Anxiety and the Netherlands Twin Registry both took part in the Genetic Association and Information Network (GAIN) to conduct the first GWAS for MDD.

In this GWAS, 11 single nucleotide polymorphisms (SNPs) of the 200 SNPs with the lowest P-values located to a 167 kb segment overlapping the gene \textit{PCLO}. This gene encodes the presynaptic protein piccolo, which has a possible role in facilitating monoamine transporter internalization (Cen et al., 2008). In addition, it negatively regulates synaptic vesicle exocytosis by decreasing transport of vesicles from reserve pools to readily-releasable pools through an action on synapsin (Léal-Ortiz et al., 2008). This suggests a possible role for \textit{PCLO} in the regulation of mood-related monoaminergic neurotransmission.

Though multiple SNPs reached P-values in the order of 10E–7, genome-wide significance was not reached. 30 SNPs were included in a replication effort using an additional five MDD cohorts. These replication studies only partly confirmed the results. Only after post-hoc analysis with an Australian cohort that used similar
ascertainment, the non-synonymous coding SNP rs2522833 showed nominal genome-wide significance (6.4E–8).

The lack of conclusive evidence for the involvement of any gene suggests that different factors are involved in different types of MDD. MDD is quite a heterogeneous disorder, with diagnosis based on levels of severity, depression subtypes and suggested underlying etiology. In order to obtain a more specific phenotype, one could use so-called endophenotypes: a concept with the purpose to divide for example behavioral symptoms into more stable phenotypes with a clearer genetic connection.

A second cause for sub-threshold P-values may be a lack of statistical power to detect a variant at a genome-wide level, due to the sheer number of variants genotyped. In addition, the effect size of a variant may be small in case of a common complex disorder. Thirdly, in order to accurately distinguish an association, it is imperative to have sufficient SNP-coverage within the regions of interest. Despite the intragenic association in PCLO, the SNP genotyping microarray that was used for the GWAS was not designed in a gene-centered manner. This implies that SNP coverage was generally not optimal for genic regions, including most genes for which small but not genome-wide significant p-values were found. We cannot rule out that these genes contain genetic risk factors, as there is no full coverage of them.

We therefore selected seven genes from the GAIN-MDD GWAS, with low SNP-coverage and multiple SNPs with a P-value ≤0.05, for further fine mapping. We aimed to increase coverage for these genes to capture all common variation in order to find a variant with stronger association with MDD in the GAIN-MDD cohort.

**MATERIALS AND METHODS**

**Samples**

The subjects for this study originated from two longitudinal studies, the Netherlands Study for Depression and Anxiety (http://www.nesda.nl), designed to be representative of individuals with depression and/or anxiety disorders, and the Netherlands Twin Registry (http://www.tweelingenregister.org) for both of which sample collection and DNA isolation has been extensively described previously (Boomsma et al., 2008; Sullivan et al., 2009). Genotyped samples contained 1738 cases and 1802 controls, of which 1216 male and 2324 female. All individuals had an age of 18–65 years and had self-reported western European ancestry.
Ethical Issues
The NESDA and NTR studies were approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Center, Amsterdam, an Institutional Review Board certified by the US Office of Human Research Protections (IRB number IRB-2991 under Federal-wide Assurance-3703; IRB/institute codes, NESDA 03-183; NTR 03-180). All subjects provided written informed consent. As part of the GAIN application process, consent forms were specifically re-reviewed for suitability for the deposit of de-identified phenotype and genotype data into the controlled-access dbGaP repository (Mailman et al., 2007).

Gene and Tag SNP Selection
We made a selection of the 25 genes with the lowest SNP P-values in the GAIN-MDD GWAS and ranked them according to 1) expression in the brain (yes or no), 2) high number of SNPs that reached P≤0.05 per total number of SNPs genotyped for this gene, 3) low SNP coverage of the gene in the GAIN-MDD GWAS, 4) low number of haplotype blocks per kb. Genes were tagged using the online Tagger tool (De Bakker et al., 2005) with r²>0.8 and m.a.f.>0.1 (Table 1). A margin of 5 kb around each gene was included, to tag possible regulatory regions as well. In addition, for each gene we included several SNPs that showed low P-values in the GAIN-MDD GWAS as a quality check.

Table 1: Selected genes, their function and the number of tag SNPs required to reach 100% coverage at m.a.f.>0.1 and r²>0.8.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function/Description</th>
<th>Tag SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFAP1L1</td>
<td>Actin filament associated protein</td>
<td>21</td>
</tr>
<tr>
<td>ANPEP</td>
<td>Alanyl (membrane) aminopeptidase</td>
<td>17</td>
</tr>
<tr>
<td>FGF14</td>
<td>Fibroblast growth factor</td>
<td>167</td>
</tr>
<tr>
<td>GZMK</td>
<td>Granzyme K precursor</td>
<td>7</td>
</tr>
<tr>
<td>PCLO</td>
<td>Presynaptic active zone protein Piccolo</td>
<td>70</td>
</tr>
<tr>
<td>PTK2B</td>
<td>Protein tyrosine kinase</td>
<td>37</td>
</tr>
<tr>
<td>ST3GAL6</td>
<td>Beta-galactoside alpha-2,3-sialyltransferase 6</td>
<td>25</td>
</tr>
</tbody>
</table>

Genotyping
Forty 96-well plates were made, blind to case-control status. Cases and controls were randomly allocated to plates and positions within plates. Each plate contained 93 samples from Dutch subjects, plus 3 QC samples at a concentration of 50 ng/µl of DNA. The three QC samples included two parents of one control sample on that plate, to add up to a total of 40 trios. Half of the plates contained a randomly selected duplicate case sample. Several samples were removed for analysis: offspring from trios, duplicates and various samples based on a principal component
analysis described previously (Sullivan et al., 2009), leading to a total of 3540 samples (1738 cases and 1802 controls).

All genotyping was performed using the OpenArray® Real-Time PCR System (Life Technologies, Carlsbad, USA), in accordance with the protocol of the manufacturer (version: 7/2010). Arrays were designed to have 128 assays for 24 samples per array and were loaded using the OpenArray Accufill robot or using the AutoLoader, manually loaded into a cassette and then PCR was performed in an NT cycler (GeneAmp® PCR System 9700, Life Technologies, Carlsbad, USA). After this, arrays were scanned with the OpenArray NT Imager. 30 assays that were not correctly spotted onto the 128-format arrays were put on a separate 32-format array.

The quality of scanned arrays was checked by visually assessing the location of the array in the scanner (the so-called Spotfind image). The loading of the arrays was checked using the ROX image and the fluorescence signal strength was checked using the VIC and FAM images with the software tool ImageJ (http://rsbweb.nih.gov/ij/). Genotypes for approximately 200 samples were analyzed simultaneously, using Taqman Genotyper Software v 1.0.1. This number of 200 samples was set by optimizing for clear clustering, without getting a bias due to too few data points.

Quality Control and Concordance Rates
For quality control reasons we included duplicated samples in the cohort. After genotyping we have checked the concordance between the two identical samples. To do this we used a home-made Perl script to compare all the genotype data from duplicate samples. Concordance was calculated for every SNP for which the sample and its duplicate both had a genotype. Concordance was 99.0% for duplicate samples. Out of the two duplicates we selected the sample that had the most genotype data for further analysis.

The Y-chromosomal SNP rs2534636 was included for QC. Genotype results for this SNP correspond to female/male distribution on the arrays.

Using the genome analysis tool PLINK, we performed quality control. As this is a follow up study of the initial GAIN-MDD GWAS, we chose to use the same quality control settings. Samples were excluded if more than 25% of data was missing, according to the standards that were used in the GAIN-MDD GWAS. SNPs were excluded if a) m.a.f. was lower than 1%, b) missing genotype rate was higher than 5%, c) more than one Mendelian error occurred in 38 trios, or d) P<10E–5 for the Hardy-Weinberg Equilibrium exact test in PLINK.
For each gene, several SNPs with a low P-value in the GAIN-MDD GWAS were also genotyped using the OpenArray system. Concordance between genotypes of both platforms was calculated to be 99.5% using PLINK (Purcell et al., 2007).

**Statistical Analysis**

The results of each analysis that was performed with the Taqman Genotyper Software were exported as a text file. Text files for all analyses were combined using a home-made script written in Perl (Wall et al., 2000); With this script sample IDs, rs-numbers and genotypes were extracted and, with an additional script these data were merged into a ped-file.

All statistical analyses were performed using PLINK. We used an allelic chi-square test with one degree of freedom to perform association analysis, to compare the allele frequencies between MDD cases and controls for each SNP. Since this project entails the fine mapping of the results of a GWAS, we corrected for genome-wide significance when performing the association analysis. A P-value of $5 \times 10^{-8}$ or lower was considered to be genome-wide significant.

Haplotype blocks were calculated with PLINK, using the method of Gabriel et al., which defines pairs to be in strong LD if the one-sided upper 95% confidence bound on $D'$ is larger than 0.98 and the lower bound is above 0.7 (Gabriel et al., 2002). The association of haplotypes with MDD was calculated with a chi-square test using one degree of freedom.

**Calculation of Coverage**

In order to calculate coverage, we used the online Tagger tool from De Bakker et al (De Bakker et al., 2005). We force included all the tag SNPs that we selected and force excluded all other SNPs for tagging at m.a.f.>0.1 and $r^2>0.8$. This resulted in a calculation of how many SNPs out of all the present SNPs are covered by the force included tag SNPs.

**Imputation**

MaCH was our imputation method of choice, based on its high imputation accuracy and efficacy, its user-friendly data handling (Nothnagel et al., 2009), and high compatibility with 1000 genomes data. 1000 genomes 2010-06 release CEU data was used as a reference, because of its high number of variants and its novelty (1000 Genomes Project Consortium, 2010).

We did not use imputation data for the entire chromosome, as we were only interested in seven genes and their regulatory regions. However, to leave the underlying LD-structure intact, we used a margin of 100 kb around each gene.
To extract the genes +/-100 kb from the full chromosome data of the 1000 genomics project, we used a home-made script written in Python (Van Rossum & De Boer, 1991). According to MaCH protocol, an estimation of imputation parameters was created with 100 random control samples and 100 random cases, to get information about the length of haplotype stretches shared between our data and the reference panel (Li et al., 2006). After estimating parameters, imputation was performed with 100 Markov chain iterations for the entire cohort, per gene. All imputation was performed on the Lisa system cluster (www.sara.nl/systems/lisa). For each gene, we filled in the missing genotypes by imputation and left genotyped SNPs intact.

**Joint Reanalysis**
We performed a joint reanalysis of 77 PCLO SNPs surrounding rs2522833 and rs2715147. For this analysis, we calculated Z-scores by performing logistic regression and dividing the slope for each data point by its standard error, similar to the method used by Sullivan et al. (Sullivan et al., 2009; Bochdanovits et al., 2009). The absolute values of these Z-scores were then plotted against the root of the r² between one of these 77 SNPs with either rs2522833 or rs2715147.

**Epistasis Analysis**
To perform an analysis of epistasis, we selected 52 genes that, on a protein level, interact with PCLO, using the method of Lips et al. for 47 synaptic genes and using the InWeb database for 5 additional genes (Lips et al., 2012; Lage et al., 2007). Genotypes for the SNPs existing in these genes were extracted from the GAIN-MDD GWAS data, after which epistasis analysis was performed with PLINK (Purcell et al., 2007). We tested a total of 94 PCLO SNPs against the 1579 SNPs in the selected genes.

**RESULTS**

**Genotyping**
The seven selected genes were tagged in order to reach 100% coverage at r²>0.8 and m.a.f.>0.1. A total of 349 tag SNPs were selected for genotyping. After genotyping, five SNPs were removed due to poor clustering. 51 SNPs and 64 samples failed because of high levels of missing data, after which the average call rate per sample was 96.7% and average call rate per SNP was 96.7%.

**Coverage**
In order to compare the coverage of the seven selected genes, coverage was calculated before and after additional genotyping, using the online Tagger tool (De Bakker et al., 2005). SNPs that were genotyped in the original GWAS were merged
with the 298 SNPs that were genotyped and passed the quality control that we performed with the genome analysis tool PLINK (Purcell et al., 2007). After merging, 459 SNPs and 3476 individuals remained (1712 cases and 1764 controls) for the seven genes. Total genotyping rate in remaining individuals was 98.8%.

As not all our tag SNPs passed quality control, we did not reach 100% coverage for all genes, but adding these SNPs to those genotyped in the initial GWAS resulted in significantly higher coverage (Table 2).

<table>
<thead>
<tr>
<th>Table 2: Coverage calculated for each gene at $r^2 &gt; 0.8$ m.a.f. &gt;0.1 before and after fine mapping.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>AFAP1L1</td>
</tr>
<tr>
<td>ANPEP</td>
</tr>
<tr>
<td>FGF14</td>
</tr>
<tr>
<td>GZMK</td>
</tr>
<tr>
<td>PCLO</td>
</tr>
<tr>
<td>PTK2B</td>
</tr>
<tr>
<td>ST3GAL6</td>
</tr>
</tbody>
</table>

**Association Analysis**

After quality control, we performed association analysis for the newly genotyped SNPs using PLINK, for association with MDD. For six genes the result of fine mapping did not improve P-values compared to the P-values that were detected in the original GAIN-MDD GWAS (Table 2). However, for **PCLO** we found that rs2715147 had a P-value of 6.8E−7. This is lower than rs2715148, which showed the lowest P-value ($P = 7.7E−7$) for **PCLO** in the GAIN-MDD GWAS. This finding did not reach genome-wide significance ($P = 5E−8$).

We then compared rs2715147 and rs2715148, while only using the samples that were genotyped for both SNPs to exclude a bias due to unequal numbers of cases and controls. We thus excluded all samples with missing genotypes for either of these SNPs, after which rs2715148 had a P-value of 5.3E−7 and rs2715147 had a P-value of 6.8E−7.

As 51 SNPs were excluded from the analysis after quality control, this prevented reaching full coverage for 6 genes, except for **GZMK**. To increase coverage for these genes after exclusion of these SNPs, we imputed missing genotypes using the 1000 genomes CEU data.
After imputation we again performed an association analysis (Table 3). rs2715147 and rs2715148 showed a similar P-value: 1.223E−6. In addition, the P-values for \textit{FGF14} and \textit{PTK2B} decreased. However, none of the genotyped and imputed SNPs reached genome-wide significance (P=5E−8). After imputation, rs2715147 and rs2715148 show a slightly better P-value (P=1.172E−6) than the non-synonymous coding SNP rs2522833 (P=1.223E−6). When using a logistic model with sex as a covariate, P-values for rs2715147 and rs2715148 increased slightly to P=1.763E−6, showing only a marginal effect of sex when taken along as a covariate.
### Table 3: rs-numbers and P-values for the SNPs with the lowest P-values

<table>
<thead>
<tr>
<th>Gene</th>
<th>GAIN-MDD</th>
<th>P-value</th>
<th>Fine-mapping</th>
<th>P-value</th>
<th>OR; CI</th>
<th>Imputed data</th>
<th>P-value</th>
<th>OR; CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFAP1L1</td>
<td>rs4705335</td>
<td>1.90E-04</td>
<td>rs352355</td>
<td>1.30E-04</td>
<td>0.83; 0.72-0.96</td>
<td>rs4705335</td>
<td>2.70E-04</td>
<td>1.26; 1.11-1.43</td>
</tr>
<tr>
<td>ANPEP</td>
<td>rs6496603</td>
<td>5.60E-05</td>
<td>rs8035089</td>
<td>3.90E-04</td>
<td>0.82; 0.72-0.92</td>
<td>rs6496603</td>
<td>8.20E-05</td>
<td>0.82; 0.75-0.90</td>
</tr>
<tr>
<td>FGF14</td>
<td>rs17688345</td>
<td>1.20E-04</td>
<td>rs9518638</td>
<td>1.60E-03</td>
<td>0.84; 0.75-0.94</td>
<td>rs17688345</td>
<td>8.20E-05</td>
<td>0.75; 0.65-0.87</td>
</tr>
<tr>
<td>GZMK</td>
<td>rs2112938</td>
<td>5.10E-05</td>
<td>rs6875666</td>
<td>4.90E-03</td>
<td>0.86; 0.78-0.96</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PCLD</td>
<td>rs2715148</td>
<td>7.70E-07</td>
<td>rs2715147</td>
<td>6.80E-07</td>
<td>0.79; 0.72-0.87</td>
<td>rs2715147+rs271548</td>
<td>1.20E-06</td>
<td>0.79; 0.72-0.87</td>
</tr>
<tr>
<td>PTK2B</td>
<td>rs7000615</td>
<td>1.50E-04</td>
<td>rs748281</td>
<td>3.70E-04</td>
<td>0.84; 0.76-0.93</td>
<td>rs7000615</td>
<td>5.40E-05</td>
<td>1.30; 1.14-1.47</td>
</tr>
<tr>
<td>ST3GAL6</td>
<td>rs999147</td>
<td>1.60E-04</td>
<td>rs704586</td>
<td>1.00E-03</td>
<td>0.84; 0.76-0.93</td>
<td>rs14310</td>
<td>1.70E-04</td>
<td>1.2; 1.09-1.33</td>
</tr>
</tbody>
</table>
Joint Reanalysis
In addition, we performed a joint reanalysis of 77 SNPs surrounding rs2522833 and rs2715147. The absolute values of Z-scores were plotted against the square root of the $r^2$ between one of these 77 SNPs with either rs2522833 or rs2715147. When assuming the null-hypothesis of no association, one would expect that the slope of the linear fit would approximate 0, since SNPs in high LD with a causal variant will reflect the Z-score of this causal variant. When we assume that rs2522833 is the causal variant, the slope of the linear fit is 4.17, which increases slightly to 4.24 when assuming that rs2715147 is the causal variant (Figure 1), supporting the hypothesis that an unknown variant between rs2715147 and rs2522833 may be causal for MDD in the GAIN-MDD cohort.

![Figure 1: Linear fit for the Z-scores and correlation (\(\sqrt{r^2}\)) between markers and rs2715147](image)

Epistasis Analysis
Since PCLO gave the lowest P-values of the seven genes selected for fine mapping, epistasis analysis was performed for PCLO only. The lowest P-value (1.6E−05; OR 0.5928) was found for PCLO SNP rs6947662 in conjunction with rs16946196, which is located in DLGAP1. Since this epistasis analysis did not lead to a lower P-value than a single SNP analysis, we found no evidence for an epistatic effect of PCLO SNPs with SNPS from interacting proteins.

Using the merged data of the GAIN-MDD GWAS, our fine mapping study, plus the
imputed data, we generated an r²-plot of the region spanning PCLO in the haplotype analysis program Haploview (Barrett et al., 2005), since PCLO provided the lowest P-value. rs2715147 and rs2715148 are in high r² (0.99) with one another. In addition, both SNPs show an r² of 0.77 with the non-synonymous coding SNP rs2522833 (Figure 2).

Figure 2: The LD-structure of PCLO shown in an r²-plot created in Haploview. The plot shows the LD-block in which the SNPs with the lowest P-values were found. Non-synonymous coding SNP rs2522833, rs2715147 and rs2715148 are in high r² with each other.

Based on the haplotype structure as seen in Haploview, we performed a haplotype association test with for rs2715147, rs2715148 and rs2522833, as we find the lowest P-values in this region. For this haplotype we found a P-value of 9.9E−7, meaning that the combination of these three SNPs as a haplotype will give a slightly better association than any of them as a single SNP.
DISCUSSION

In 2009 a GWAS for MDD was performed (Sullivan et al., 2009). Unfortunately, the proprietary microarrays used for this GWAS (Perlegen Sciences Inc., Mountain View, CA, USA) were not designed in a gene-centered manner resulting in incomplete coverage of genic regions. From the 25 genes that harbored the SNPs with lowest P-values, we selected seven genes for fine mapping. We used the Hapmap Tagger tool to tag these genes with $r^2 > 0.8$ and m.a.f. > 0.10, in order to capture all common variation.

After genotyping, several SNPs were excluded through quality control. Even though we did improve coverage significantly, due to this exclusion we did not acquire full coverage for all genes. Full coverage was reached only for *GZMK*. We performed an association test with all SNPs and samples that made it through cut-off values. For the SNP rs2715147 in *PCLO* we found a P-value of $P = 6.8E^{-7}$, which is lower than the lowest P-value for *PCLO*-SNPs in the original GAIN-MDD GWAS ($P = 7.7E^{-7}$ for rs2715148). This small decrease in P-value could also be due to technical variability, however, in both the GAIN-MDD GWAS and our fine mapping project, the lowest P-values are found in this area of the *PCLO* gene. For the other six genes we did not find a variant with better association than in the GAIN-MDD GWAS.

Since we reached 100% coverage only for the *GZMK* gene, we filled up missing genotypes by performing imputation with MaCH for the remaining six genes. Previously, for the GAIN-MDD GWAS, two imputation approaches have been used: MaCH was used for imputing 2037829 autosomal SNPs with $r^2 \geq 0.5$ (which removes approximately 90% of SNPs with unreliable imputation results, while dropping only 2–3% of reliably imputed SNPs) and using the SNPMStat method (Lin et al., 2008), 246 SNPs in the *PCLO* area were imputed. The HapMap2 CEU panel was used as a reference (International HapMap Consortium, 2003).

In this study, imputation was only performed for missing genotypes, rather than for all new tag SNPs. The rationale behind this is that there are local differences in LD structure between the GAIN-MDD cohort and the HapMap CEU population. This might decrease the validity of the genotypes estimated by imputation (Pardo et al., 2009). We used MaCH to impute for six genes. Imputation decreased P-values for *FGF14* and *PTK2B*, albeit for the SNP that showed the lowest P-value for those genes in the original GAIN-MDD GWAS. For *PCLO*, rs2715147 and rs2715148, which are in strong LD, both showed the same P-value at $P = 1.2E^{-6}$, which was also the lowest P-value for this gene. For *ANPEP*, *AFAP1L1* and *ST3GAL6*, P-values were not improved by means of imputation. None of the genes showed a genome-wide significant association with MDD after imputation.
In addition, we wanted to investigate whether SNPs in *PCLO* are interacting with SNPs in synaptic genes. To determine this, we performed an epistasis analysis using PLINK (Purcell et al., 2007). As the lowest P-value was in the range of 10E−5, we cannot conclude whether there is epistasis between these SNPs or not.

In a joint reanalysis of 77 *PCLO* SNPs we show a graphical representation of the Z-scores for each SNP versus the correlation of this SNP with rs2715147. In comparison with rs2522833, the slope for rs2715147 is slightly steeper. This supports the hypothesis that the low P-values in this area may be caused by an unknown variant located between rs2715147 and rs2522833, or an unknown variant that is in strong LD with these SNPs.

We can conclude that fine mapping of these seven genes did not provide a variant with a stronger association than reported in the original GAIN-MDD GWAS, where the lowest P-value was obtained for rs2715148 and rs2522833 showing nominal significance after post-hoc analysis with an Australian cohort. However, there could be a number of reasons for this apparent lack of association. First of all, diagnosis of MDD is based on relatively subjective assessments of symptoms. By specifying endophenotypes within an MDD cohort, for instance brain activity, cortisol levels and pharmacological response, one might find variants that are exclusive to that particular endophenotype, with a higher effect size.

Another possibility would be to expand the cohort in order to increase the power for detecting an associated variant. Park et al. show that for a number of complex traits, the sample size has to be at least around 10,000 in order to reliably detect new variants (Park et al., 2010).

One way to create such an expansion would be to perform a meta-analysis of several cohorts. Nevertheless, despite the increase in sample size, one has to take into account that a meta-analysis, in case of MDD, may also increase any heterogeneity caused by inconsistencies in ascertainment.

Other GWAS for depression are troubled by equal predicaments. So far, marginally significant associations have been found for -among others- *FKBP5, SP4, GRM7, CSORF20* and *NPY*. However, many of these results cannot be replicated in another cohort (Velders et al., 2011; Shi et al., 2011; Shyn et al., 2011; Bosker et al., 2011). Here again, sample size may be crucial to acquire the statistical power necessary to find an associated variant. In addition, not all studies use the same method of ascertainment. Even though cases are mostly obtained for research through a DSM-IV diagnosis of MDD, more specific secondary interviews may deviate in de-
Chapter 3

...terminating depression subtypes, severity, age of onset, recurrence and comorbidity (Shi et al., 2011).

Although we did not select our seven genes based on their function, several of them are linked to the central nervous system and brain physiology. First of all, the product of ANPEP, aminopeptidase N, metabolizes angiotensin III (AngIII), which is one of the main effector peptides of the brain renin-angiotensin system. This system controls vasopressin release in the brain. When aminopeptidase N is inhibited, both AngIII and vasopressin increase, which in turn causes an increase of ACTH (Reaux et al., 1991). An increase in ACTH ultimately stimulates the release of cortisol, which is a major stress hormone. This connects aminopeptidase N to the HPA-axis, which is linked to MDD as it elicits the stress-response in the brain (Holsboer, 2000).

Both PTK2B and GZMK have been linked to brain physiology and depression through animal models. Following acute stress, PTK2B (also known as pyk2) expression is increased, whereas increasing PTK2B activity in lateral septum neurons reverses the behavioral deficits of acute, inescapable stress. These findings establish a role for PTK2B in the behavioral response to stress and may suggest a possible role in the pathophysiology of depression (Sheehan et al., 2003).

GZMK is part of a network of genes that are co-expressed higher in mice that have a high predisposition to freezing behavior or catalepsy (Kondaurova et al., 2011). This reaction is a natural passive defensive strategy, but in chronically stressed animals, for instance in models for post-traumatic stress disorder or MDD, animals show enhanced catalepsy (Tomida et al., 2009).

The protein product of PCLO, Piccolo, can be found in the presynaptic active zone (Fenster et al., 2002). If Piccolo is knocked out, synapse formation or morphology is not affected, suggesting that piccolo is not necessary for formation of synapses. However, synapses lacking Piccolo exhibit faster rates of synaptic vesicle exocytosis, indicating that Piccolo is a negative regulator of the exocytotic process (Leal-Ortiz et al., 2008). This may suggest a role for Piccolo in the monoamine hypothesis of depression, which states that depression is caused by an imbalance of monoamine availability (Schildkraut, 1965). In addition, the non-synonymous coding SNP that was found to be significant in the GWAS by Sullivan et al., changes a serine to an alanine in a calcium-binding C2A-domain. Overexpression of this C2A-domain causes a depression-like phenotype in mice (Sullivan et al., 2009; Furukawa-Hibi et al., 2010).
These genes may still be interesting candidate genes, when looking at monoamine availability (PCLO), or more specific (endo-)phenotypes like cortisol levels (ANPEP) and co-morbid anxiety (PTK2B and GZMK). Despite the fact that the other selected genes that are also expressed in the brain, based on exploring literature, they do not show an obvious link with MDD. In combination with their apparent lack of genome-wide associated variants, this makes them less likely to be successful candidate genes.

None of the SNPs for any of the seven genes showed a P-value in the magnitude of P=5E−8, which leads to the conclusion that in the scenario of common variation and corrected for genome-wide testing, these genes show no genome-wide significant association with MDD for this cohort. However, considering the fact that in PCLO there are several signals in the magnitude of P=1.0E−6 and P=1.0E−7 and the “Fundamental Theorem of the HapMap”, which states that all tested SNPs are expected to reflect the true association of the unknown causal variant proportional to their LD with it, one cannot disregard the possibility that a rare variant may still be associated.

Previously, we showed that most of the association between genotype data and MDD is statistically explained by the association of the non-synonymous coding SNP rs2522833 with MDD. The data from the GWAS are consistent with the hypothesis that either rs2522833 or a variant in high LD with it is a causal risk factor for MDD. However, our data do not favor rs2522833 as the causal variant, as it does not show the lowest P-value in our data set. We do see a very high LD (r²=0.99) between rs2715147 and rs2715148 and a high LD between these two SNPs and rs2522833 (r² =0.77). In addition, the haplotype which includes SNPs rs2715147, rs2715148 and rs2522833 shows a lower P-value than the P-values calculated for these SNPs individually. This implies that between rs2715148 and rs2522833 there may be an unknown variant that has an r² of at least 0.77 with both variants and has a slightly better association with MDD (9.9E−7). Nevertheless, this observation could also be caused by missing data. Ideally, the study should be replicated in a larger cohort or in a meta-analysis in order to confirm or decline the improved P-value in case of this haplotype.

In addition, instead of looking at SNPs as individual units of association studies, one might jointly analyze all variants within a putative gene to obtain a single P-value for the association of the entire gene, as it is the functional unit of the genome. A pitfall for joint analysis is that one would have to assign weights to the individual SNPs, as not every SNP will have the same impact on a putative association. In tools for gene-based P-values, this matter is still an open question, as
we do not yet have a full understanding of the relationship between sequence and function (Li et al., 2011).

In conclusion, the current study suggests that using common variation to fine map the GAIN-MDD GWAS results, does not lead to lower P-values or the identification of a stronger associated variant. The genomic region in *PCLO* between rs2715147 and rs2522833 covers approximately 5 kb. It is estimated that SNPs occur every 100–300 bp in the human genome. That would imply that between rs2715147 and rs2522833 approximately 16–50 variants could occur. With new, powerful approaches for DNA analysis such as next generation or massive parallel sequencing (MPS), these variants could be identified and subsequently genotyped in the whole cohort. This could lead to the discovery of a causal variant that is in high LD with rs2715147, rs2715148 and/or rs2522833. Accordingly, we should perform MPS for *PCLO*, in order to confirm the existence of such a variant and find its association with MDD.

**ACKNOWLEDGMENTS**

We thank SARA Computing and Networking Services for their support in using the Lisa Compute Cluster.

We thank Nanne Aben for writing the Python script that was used for extracting haplotype and SNP data from 1000 genomes data.

We thank Esther Lips for allowing us to use the JAG tool to annotate SNPs to genes for the epistasis analysis [Lips et al., unpublished].

**AUTHOR CONTRIBUTIONS**

Conceived and designed the experiments: ECV MRB PH WJGH ZB BWP. Performed the experiments: ECV IMCB. Analyzed the data: ECV. Contributed reagents/materials/analysis tools: ZB PR DS GW EJG JHS BWP DIB. Wrote the paper: ECV IMCB.
REFERENCES

1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing; Nature; 467: 1061–1073

American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders: DSM-IV; Washington, DC


Chapter 3

Levinson DF (2006) The genetics of depression: a review; Biol Psychiatry; 60: 84–92
Li MX GH, Kwan JSH, Sham PC (2011) GATES: A Rapid and Powerful Gene-Based Association Test Using Extended Simes Procedure; The American Journal of Human Genetics; 283–293
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses; Am J Hum Genet; 81: 559–575
association study of Sequenced Treatment Alternatives to Relieve Depression and meta-analysis of three studies; Mol Psychiatry; 16: 202-215
Van Rossum G, De Boer J (1991) Linking a Stub Generator (AIL) to a Prototyping Language (Python); EurOpen Conference Proceedings
CHAPTER 4
Resequencing Three Candidate Genes for Major Depressive Disorder in a Dutch Cohort

Eva C. Verbeek
Marianna R. Bevova
Zoltán Bochdanovits
Patrizia Rizzu
Ingrid M. Bakker
Tiny Uithuisje
Eco J. de Geus
Johannes H. Smit
Brenda W. Penninx
Dorret I. Boomsma
Witte J. Hoogendijk
Peter Heutink

ABSTRACT

Major depressive disorder (MDD) is a psychiatric disorder, characterized by periods of low mood of more than two weeks, loss of interest in normally enjoyable activities and behavioral changes. MDD is a complex disorder and does not have a single genetic cause. In 2009 a genome wide association study (GWAS) was performed on the Dutch GAIN-MDD cohort. Many of the top signals of this GWAS mapped to a region spanning the gene PCLO, and the non-synonymous coding single nucleotide polymorphism (SNP) rs2522833 in the PCLO gene became genome wide significant after post-hoc analysis. We performed resequencing of PCLO, GRM7, and SLC6A4 in 50 control samples from the GAIN-MDD cohort, to detect new genomic variants. Subsequently, we genotyped these variants in the entire GAIN-MDD cohort and performed association analysis to investigate if rs2522833 is the causal variant or simply in linkage disequilibrium with a more associated variant. GRM7 and SLC6A4 are both candidate genes for MDD from literature. We aimed to gather more evidence that rs2522833 is indeed the causal variant in the GAIN-MDD cohort or to find a previously undetected common variant in either PCLO, GRM7, or SLC6A4 with a higher association in this cohort. After next generation sequencing and association analysis we excluded the possibility of an undetected common variant to be more associated. For neither PCLO nor GRM7 we found a more associated variant. For SLC6A4, we found a new SNP that showed a lower P-value (P=0.07) than in the GAIN-MDD GWAS (P=0.09). However, no evidence for genome-wide significance was found. Although we did not take into account rare variants, we conclude that our results provide further support for the hypothesis that the non-synonymous coding SNP rs2522833 in the PCLO gene is indeed likely to be the causal variant in the GAIN-MDD cohort.
INTRODUCTION

Major depressive disorder (MDD) is a psychiatric disorder that is characterized by persistent dysphoria, loss of interest and pleasure, changes in appetite and sleep, psychomotor retardation, feelings of guilt or worthlessness, inability to concentrate and recurrent thoughts of death or suicide (American Psychiatric Association, 1994). Environmental circumstances have proven to influence the aetiology of the disease. It is more prevalent in women than in men and though MDD may develop at any age, the mean age of onset is 32 years of age, with a lifetime prevalence of 16.5%. Worldwide, MDD is one of the leading causes of disability (Kessler et al., 2005). The etiology of MDD is still largely an enigma, but stressful life events (SLEs) are a predictor for developing a depressive episode (Kendler et al., 1999). However, from twin studies it is known that heritability of MDD is approximately 40% (Kendler et al., 2006).

In 2009, Sullivan et al. performed a GWAS for MDD on the Dutch GAIN-MDD cohort. Genome-wide significant association with MDD was not reached, but after post-hoc analysis including an Australian cohort the non-synonymous coding SNP rs2522833 in the gene PCLO showed nominal significance (P=6.4E-8)(Sullivan et al., 2009). The Perlegen chip used for this GWAS did not have full genome tagging capacity nor a gene-centered design, which is why we previously performed fine-mapping for seven genes that showed low P-values in the GAIN-MDD GWAS (Verbeek et al., 2012). The increase of SNP coverage did not lead to the discovery of a more strongly associated variant. However, when combining the SNPs with the lowest P-value in PCLO with non-synonymous coding SNP rs2522833 in one haplotype, the P-value decreased, suggesting a possible undetected variant that is more strongly associated with MDD in the GAIN-MDD cohort (Verbeek et al., 2012). In addition, in 2009 Bochdanovits et al. showed that either rs2522833 or an unknown variant that is in high LD with it, is most likely the causal variant in the GAIN-MDD cohort (Bochdanovits et al., 2009).

The non-synonymous coding SNP rs2522833 is a common variant with a minor allele frequency (m.a.f.) of 0.4. Since it is a common variant, we hypothesize that if this SNP is not the causal variant, the unknown variant that may be causal for the GAIN-MDD cohort will also be a common variant, as we expect this variant to be in high LD with rs2522833.

Besides the study of Sullivan et al., in literature there are other case-control studies replicating the role of PCLO in MDD (Aragam et al., 2011; Hek et al., 2010). Moreover, Minelli et al found that the PCLO gene was involved in personality traits that
predispose to depression, showing a role of \textit{PCLO} in MDD using endophenotypes (Minelli et al., 2011).

As a follow-up study for the GAIN-MDD GWAS, the aim of this study is therefore to identify this common causal variant, by increasing the resolution of genotyping with next generation sequencing (NGS) followed by association analysis between the newly identified variants and MDD in the GAIN-MDD cohort.

To accomplish this, we sequenced 50 control samples from the GAIN-MDD cohort. Controls were used since we expect the undetected variant to be common and therefore also present in control samples. In addition, this will allow us to witness this variant against the background of the normal LD-structure of the Dutch population. Although we selected controls for sequencing, it was our aim to find the most associated variant within the cohort. Bochdanovits et al. in 2009 stated that either rs2522833 would be causal, or a variant in high LD with it. If homozygotes are selected for this variant rather than heterozygotes, it increases the possibility to detect other variants in high LD with the risk allele.

In addition to \textit{PCLO}, which was selected based on our previous results, we also sequenced the genes \textit{GRM7} and \textit{SLC6A4}, which have been studied extensively as functional candidate genes for MDD in the literature. \textit{GRM7} codes for the metabotropic glutamate receptor 7 and an intronic SNP in this gene showed a \textit{P}-value that approximated genome-wide significance in a meta-analysis of three depression cohorts (Cryan et al., 2003; Mitsukawa et al., 2006; Caspi et al., 2003). The \textit{SLC6A4} gene codes for the serotonin transporter gene and plays an important role in the monoamine hypothesis of depression, according to which depressive phenotypes are caused by an imbalance in monoamines like serotonin. This gene has long been the topic of discussion, as there have been inconsistent results for the association with MDD of the promoter polymorphism in \textit{SLC6A4} combined with SLEs. However, it has been shown that the polymorphism in the promoter in conjunction with SNPs within the gene itself and SLEs is more associated with depressive phenotype than just the promoter and SLEs (Lazary et al., 2008).

\section*{METHODS AND MATERIALS}

\textbf{Samples}

The subjects for this study originated from two longitudinal studies, the Netherlands Study for Depression and Anxiety (http://www.nesda.nl) (Penninx et al., 2008), designed to be representative of individuals with depression and/or anxiety disorders, and the Netherlands Twin Registry (http://www.tweelingenregister.
org) for both of which sample collection and DNA isolation have been extensively described previously (Sullivan et al., 2009; Boomsma et al., 2008).

50 control samples from the GAIN-MDD cohort were used for variant detection. Samples were selected based on their genotype for rs2522833 in the PCLO gene, with C being the risk allele, and consisted of 22 males and 28 females. Of the 22 males, 10 were CC and 12 AA. Of the females, 14 were CC and 14 AA.

For genotyping the detected variants and tag SNPs, we used the entire GAIN-MDD cohort, consisting of 1738 cases and 1802 controls, of which 1216 were male and 2324 female. All individuals had an age of 18–65 years and had self-reported western European ancestry. Ascertainment of cases was from outpatient specialist mental health facilities and by primary care screening. Inclusion criteria were a lifetime diagnosis of DSM-IV MDD as diagnosed by the Composite International Diagnostic Interview psychiatric interview, age 18–65 years, and self-reported western European ancestry.

Controls mainly came from the longitudinal cohort of the NTR. Longitudinal phenotyping includes assessment of depressive symptoms (via multiple instruments), anxiety, neuroticism and other personality measures. Inclusion required availability of both survey data and biological samples, no report of MDD at any measurement occasion, and low genetic liability for MDD. No report of MDD was determined by specific queries about medication use or whether the subject had ever sought treatment for depression symptoms and/or through the CIDI interview. Low genetic liability for MDD was determined by the use of a factor score derived from longitudinal measures of neuroticism, anxiety and depressive symptoms (mean 0, s.d. 0.7); controls were required never to have scored highly (≥0.65) on this factor score. Finally, controls and their parents were required to have been born in the Netherlands or western Europe. Only one control per family was selected.

**Ethical Issues**

The NESDA and NTR studies were approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Center, Amsterdam, an Institutional Review Board certified by the US Office of Human Research Protections (IRB number IRB-2991 under Federal-wide Assurance-3703; IRB/institute codes, NESDA 03-183; NTR 03-180). All subjects provided written informed consent. As part of the GAIN application process, consent forms were specifically re-reviewed for suitability for the deposit of deidentified phenotype and genotype data into the controlled-access dbGaP repository (Mailman et al., 2007). NESDA and NTR subjects were informed of participation in GAIN by means of newsletters.
Gene Selection

The genes that were selected for targeted resequencing were \textit{GRM7}, \textit{PCLO}, \textit{SLC6A4}. \textit{PCLO} was selected based on the results from the GAIN-MDD GWAS and our previous fine mapping efforts, which suggested that either rs2522833 or an undetected variant in high LD with it would be the causal variant in the GAIN-MDD cohort. \textit{GRM7} and \textit{SLC6A4} were selected based on literature. \textit{GRM7} codes for the metabotropic glutamate receptor 7 and was one of the top genes from a meta-analysis for MDD. \textit{SLC6A4} encodes the serotonin transporter, which regulates serotonin availability in the synaptic cleft. The promoter contains a length polymorphism that is thought to modulate MDD in conjunction with SLEs (Aragam et al., 2011; Hek et al., 2010; Minelli et al., 2012; Cryan et al., 2003; Mitsukawa et al., 2006; Caspi et al., 2003).

Library Construction

385 K NimbleGen Sequence Capture (Roche NimbleGen, Inc., Madison, WI, USA) arrays were custom-designed to capture the complete genomic locus of the \textit{PCLO}, \textit{GRM7} and \textit{SLC6A4} genes, plus a 10 kb region upstream and downstream of the gene as defined by the UCSC Genome Browser B37 (http://genome.ucsc.edu) to capture possible regulatory regions as well. The total area consisted of 1,388,868 bp of which 1,217,056 bp was captured on the array (Table 1). A repeat mask was applied to reduce interference of genomic regions with a similar sequence.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Length(+/-10 kb)</th>
<th>Covered on array</th>
<th>% of region covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{GRM7}</td>
<td>900.4 kb</td>
<td>793.3 kb</td>
<td>88.1</td>
</tr>
<tr>
<td>\textit{PCLO}</td>
<td>478.9 kb</td>
<td>376.4 kb</td>
<td>87.8</td>
</tr>
<tr>
<td>\textit{SLC6A4}</td>
<td>59.6 kb</td>
<td>47.4 kb</td>
<td>82.0</td>
</tr>
</tbody>
</table>

Sequence capture was performed according to the manufacturer’s protocol version 3.0 from December 2008, except for elution of DNA from the arrays, which was performed according to the manufacturer’s protocol for elution using sodium hydroxide version 1.0. After elution, arrays were cleaned using the NimbleGen Array Reuse Kit according to protocol. As a means of quality control, qPCR was performed for four control loci and enrichment was calculated.

At the time that this research was performed, the Roche Nimblegen and Illumina Solexa (Illumina, Inc., San Diego, CA, USA) platforms worked with different fragment sizes. In order to overcome this situation, an intermediate protocol was performed. Nimblegen fragments were ligated and then fragmented to the size corresponding to the Illumina protocols. For multiplexing purposes, index tags were ligated to the samples in order to identify them after sequencing.
Resequencing Three Candidate Genes for Major Depressive Disorder in a Dutch Cohort

**Sequencing**

Sequencing was performed using the Illumina Solexa Genome Analyzer IIx (Illumina Inc., San Diego, CA, USA), by BaseClear (BaseClear BV, Leiden, The Netherlands). 50 base pair paired-end reads were generated in a multiplex fashion. Per lane 7 samples were loaded onto the flow cell, except for one lane that had 8 samples. As a control for the sequencing process, a PhiX DNA sample was also sequenced in a separate lane.

**Assembly and Variant Detection**

The assembly of reads as well as variant calling was performed using CLC Bio Genomics Workbench (CLC Bio, Aarhus, Denmark). Full chromosome data (Build 37.1, hg19) was downloaded from NCBI and known variants were annotated using dbSNP131 from the UCSC Genome Browser. Reads were mapped back to the entire chromosome, allowing up to one mismatch and two unaligned nucleotides at the end of the reads or no mismatches and five unaligned nucleotides at the end of the reads.

Assembly was performed per sample.

SNP detection was performed for each sample individually. Minimum coverage was set at 20x and maximum coverage was set as the theoretical highest average coverage for that particular sample, by taking the number of reads × 50 (bp) and dividing that number by the length of the sequenced region. 35% of reads had to have an alternative allele in order to be called a heterozygous variant.

**Genotyping Procedure**

All genotyping was performed using the Taqman OpenArray system (Life Technologies, Carlsbad, CA, USA), in accordance with the protocol of the manufacturer (version: 7/2010). 71 of the newly identified SNPs with unique sequences were spotted onto the arrays. Of the newly detected SNPs with a m.a.f. ≥10%, 50 bp flanking sequences upstream and downstream were checked for similarities with other genomic regions using the BLAST tool (Altschul et al., 1990).

In addition, 185 tag SNPs that were not previously genotyped in this cohort, were spotted on the arrays with m.a.f. 10% and r² 0.9, tagging the genes +/−10 kb to include possible regulatory regions. For **GRM7** 47 new SNPs and 157 tag SNPs were genotyped, for **PCLO** 22 new SNPs and 27 tag SNPs, for **SLC6A4** 2 new SNPs and 1 tag SNP. The tag SNPs were selected by using Tagger software (De Bakker et al., 2005). Arrays were designed to have 256 assays for 12 samples per array and were loaded using the OpenArray Accufill robot, manually loaded into a cassette and...
then PCR was performed in an NT cycler. After this, arrays were scanned with the OpenArray NT Imager. These SNPs were deposited at dbSNP.

The quality of scanned arrays was checked by visually assessing the location of the array in the scanner (the so-called Spotfind image) and ROX, VIC and FAM signals using ImageJ, (http://rsbweb.nih.gov/ij/). Genotypes for approximately 200 samples were analyzed simultaneously, using Taqman Genotyper Software v 1.0.1. This number of 200 samples was set by optimizing for clear clustering, without getting a bias due to too few data points. A home-made Perl script (Wall et al., 2000) was then used to combine all data and to create a pedigree file.

**Association Analysis**

After the genotyping procedure, data was merged with genotyping data from the GAIN-MDD GWAS (Sullivan et al., 2009) and for PCLO with fine-mapping data (Verbeek et al., 2012). We used the genome analysis tool PLINK to perform an association analysis (Purcell et al., 2007). We excluded samples with missing data >25%, SNPs with missing genotypes >10%, SNPs with m.a.f. <1% and HWE P-value <1E-05. A chi-squared test with one degree of freedom was used to perform the actual association analysis. A P-value of P=5E-08 was considered to be genome-wide significant.

**Imputation**

For imputation we used Beagle software (Browning & Browning, 2009). 1000 genomes 2010-06 release CEU data was used as a reference (Nothnagel et al., 2009; The 1000 Genomes Project Consortium, 2010).

We did not use imputation data for the entire chromosome, as we were only interested in three genes and their regulatory regions. However, to leave the underlying LD-structure intact, we used a margin of 100 kb around each gene.

To extract the genes +/-100 kb from the full chromosome data of the 1000 genomes project, we used a home-made script written in Python (Sanner, 1999). Imputation was performed per gene with 100 Markov chain iterations, for all samples. All imputation was performed on the Lisa system cluster (www.sara.nl/systems/lisa).

**Gene-based Association**

Since PCLO shows several sub-threshold association peaks in an 10E-06 magnitude, we also performed a gene-based association test and generated a single P-value for this gene rather than P-values for each SNP, by means of the VEGAS-tool (Liu et al., 2010). VEGAS tests the evidence for association on a per-gene basis by summarizing the full set of markers and takes LD between markers into account.
by using simulation based on the LD structure of a set of reference individuals. We used our individual genotype data as a reference set, so that LD would be estimated specifically for the Dutch population. For each gene one million simulations were run.

**Epistasis Analysis**
For each gene we performed an analysis of epistasis. Genes were classified into gene groups based on cellular function as determined by previous protein identification and data mining for synaptic genes and gene function, according to the method of Ruano et al. (Ruano et al., 2010), in which synaptic genes are subdivided into 17 functional groups of genes on the basis of shared function into a biological process. We selected genes known to interact with, or be in the same functional gene group as either GRM7 (G protein-coupled receptors), or PCLO (proteins involved in regulated secretion) or SLC6A4 (ion and solute carriers and exchangers). Using the method of Lips et al. (Lips et al., 2012), genotypes for the SNPs existing in these genes were extracted from the GAIN-MDD GWAS data, after which epistasis analysis was performed with PLINK (Purcell et al., 2007). Table 2 depicts how many SNPs were tested for each gene.

**Table 2: The number of SNPs used for the epistasis analysis**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Number of SNPs</th>
<th>Number of genes tested against</th>
<th>Number of SNPs tested against</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRM7</td>
<td>416</td>
<td>41</td>
<td>1220</td>
</tr>
<tr>
<td>PCLO</td>
<td>113</td>
<td>52</td>
<td>1579</td>
</tr>
<tr>
<td>SLC6A4</td>
<td>8</td>
<td>29</td>
<td>419</td>
</tr>
</tbody>
</table>

**Joint Reanalysis**
We performed a joint reanalysis of 92 PCLO SNPs surrounding rs2522833 and rs2715148. For this analysis, we calculated Z-scores by performing logistic regression and dividing the slope for each data point by its standard error, similar to the method used by Sullivan et al. (Sullivan et al., 2009; Bochdanovits et al., 2009). The absolute values of these Z-scores were then plotted against the square root of the $r^2$ between one of these 92 SNPs with either rs2522833 or rs2715147.

**RESULTS**

**Sequencing**
A total of 219 million reads were generated for all samples with 59 million reads mapping back to the region of interest (27%). All three genes reached an average coverage of at least 25 times and both GRM7 and SLC6A4 showed coverage of 20 times or higher for more than 50% of their base pairs. Using only basepairs with
a minimum coverage of 20 times, we detected 4026 SNPs in total, of which 2658 were known previously in dbSNP131, 406 were found in the 1000 genomes 2010-06 release CEU data and 961 were newly discovered (Table 3).

Table 3: Coverage data and newly detected variants over all samples

<table>
<thead>
<tr>
<th>Gene</th>
<th>GRM7</th>
<th>PCLO</th>
<th>SLC6A4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Coverage</td>
<td>32.56</td>
<td>26.61</td>
<td>39.26</td>
</tr>
<tr>
<td>% bp covered ≥10x</td>
<td>75.47</td>
<td>61.63</td>
<td>78.98</td>
</tr>
<tr>
<td>% bp covered ≥20x</td>
<td>56.65</td>
<td>39.84</td>
<td>60.85</td>
</tr>
<tr>
<td>SNPs detected</td>
<td>2953</td>
<td>954</td>
<td>119</td>
</tr>
<tr>
<td>Exonic</td>
<td>8 (1 non-synonymous)</td>
<td>15 (4 non-synonymous)</td>
<td>0</td>
</tr>
<tr>
<td>In dbSNP</td>
<td>2893</td>
<td>885</td>
<td>78</td>
</tr>
<tr>
<td>In 1000 genomes project</td>
<td>374</td>
<td>29</td>
<td>3</td>
</tr>
<tr>
<td>Newly discovered</td>
<td>655</td>
<td>266</td>
<td>40</td>
</tr>
<tr>
<td>% SNPs with m.a.f. &lt;5%</td>
<td>32.6</td>
<td>28.4</td>
<td>52.1</td>
</tr>
<tr>
<td>% SNPs with m.a.f. 5–10%</td>
<td>16.2</td>
<td>14.0</td>
<td>15.1</td>
</tr>
<tr>
<td>% SNPs with m.a.f. &gt;10%</td>
<td>51.2</td>
<td>57.6</td>
<td>32.8</td>
</tr>
</tbody>
</table>

**Association Analysis**

After variant detection, the GAIN-MDD cohort was genotyped for high resolution fine mapping using a tagging approach that included 71 newly discovered SNPs and 185 reported tag SNPs, as mentioned in methods and materials. The tag SNPs were selected so that all genes were covered 100% with m.a.f. >10% and r²=0.9, since the newly identified SNPs alone did not provide 100% coverage and we also aimed to recover the underlying LD-structure of the genes. For **GRM7** 47 new SNPs and 157 tag SNPs were genotyped, for **PCLO** 22 new SNPs and 27 tag SNPs and, for **SLC6A4** 2 new SNPs and 1 tag SNP. Several SNPs failed genotyping as the assays did not cluster very well, as they were either monomorphic or clusters were too close together to distinguish between genotypes. This lead to a total genotyping rate of 96.5% for SNPs. 293 samples (of which 60 cases and 233 controls) failed because of high levels of missing data, leaving a total of genotyping rate of 97.2% for samples. After quality control, genotyping data was merged with SNPs from the GAIN-MDD GWAS and for **PCLO** also with SNPs from the fine mapping study that we performed previously (Verbeek et al., 2012), to add up to a total of 479 SNPs in three genes. After performing an association test with depression status as the dependent variable, the lowest P-value was found for in **PCLO** for rs2715147 at P = 1.5E-06 (OR = 0.79). For **GRM7** and **SLC6A4** the lowest P-values were P = 6.6E-05 (rs17664833, OR = 0.73) and P = 0.07 (SSNP38, OR = 1.18), respectively. For **GRM7**, the P-value was not lower
than the lowest P-value in the GAIN-MDD GWAS. For \textit{SLC6A4}, SSNP38 showed a lower P-value than the lowest in the GAIN-MDD GWAS (P=0.09).

\textbf{Imputation}

Since several SNPs were excluded from the analysis after quality control and several samples had missing genotypes, we imputed these missing genotypes using Beagle with the 1000 genomes CEU data as a reference panel for all missing genotypes.

We then again performed an association analysis. The lowest P-value was found for rs2715147 and rs2715148 at 2.3E-06 (OR=0.80), located in the \textit{PCLO} gene. These two SNPs are in strong LD with each other ($r^2=0.99$) and with rs2522833 ($r^2=0.77$); (Figure 1), in our data as well as in the 1000 genomes data and show a similar m.a.f. in the Dutch population when compared to the 1000 genomes CEU data. For \textit{GRM7} and \textit{SLC6A4} the lowest P-values were 2.61E-05 (rs17664833, OR=0.71) and 0.08 (SSNP38, OR=1.17), respectively.

\textbf{Figure 1:} LD-plot of the region of interest in \textit{PCLO}
These P-values did not provide a better association than the initial GAIN-MDD GWAS, and are therefore consistent with the hypothesis that rs2522833 may indeed be the causal variant in this cohort.

Haplotypes
Using PLINK, we calculated the architecture of haplotype blocks for each gene, for the genotype data completed with imputed data. The lowest P-value was found for PCLO at P=1.19E-05 (Table 4), showing no genome-wide significance. However, this block did not contain the SNPs that showed the lowest single SNP association (rs2715147 and rs2715148). When assessing the haplotype blocks in Haploview (Barrett et al., 2005), we found that several SNPs surrounding rs2715147 and rs2715148 had an $r^2$ lower than 0.2 and a single SNP P-value in the range of 0.5-0.1. Because of this lack of $r^2$ and their high P-values, we created new haplotype blocks in which these SNPs were not included. Haplotype-based association analysis was performed again, which revealed the same block to have the lowest P-value. Nonetheless, this haplotype block does not show a better association with MDD than our single SNP association data. Moreover, the haplotype-based association test does not yield a P-value lower than rs2522833 in the GAIN-MDD GWAS.

Table 4: Haplotypes constructed using PLINK and their respective P-values

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNPs in haplotype with lowest P-value</th>
<th>Lowest P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRM7</td>
<td>rs3804925</td>
<td>rs17664792</td>
</tr>
<tr>
<td>PCLO</td>
<td>rs2371364</td>
<td>rs13237603</td>
</tr>
<tr>
<td>SLC6A4</td>
<td>SSNP38</td>
<td>rs1042173</td>
</tr>
</tbody>
</table>

Gene-based Association
Using the VEGAS tool, we generated P-values for all three genes by performing one million simulations. Since the human genome contains approximately 20,000 genes (International Human Genome Sequencing Consortium, 2004), we corrected for this number and considered 2.5E-06 (0.05/20,000) to be significant. Generating a gene-based P-value did not lead to a lower P-value result, since the lowest P-value was found for PCLO at P=1.8E-05.

Epistasis Analysis
For all three genes we performed an epistasis analysis in PLINK. First we tested all the SNPs that had P-values lower than 10E-05 in the single SNP association analysis. These SNPs yielded no P-values under 10E-04 in the epistasis analysis. Subsequently we tested all genotyped SNPs present in GRM7, PCLO and SLC6A4. The lowest P-value was found for GRM7 rs1516569 in conjunction with rs9479791.
(P=3.8E-06), located in the intronic region of OPRM1, which codes for the Opioid Receptor Mu 1. For PCLO the lowest P-value was found at P=9.4E-06 for rs17157173 together with rs16946196 in DLGAPI, which codes for Guanylate Kinase-Associated Protein (GKAP). The lowest P-value for SLC6A4 (P=2.42E-03) was found for rs4251417 with rs233112 in DDAH1, which regulates nitric oxide production. Since after correction for multiple testing, this epistasis analysis did not lead to a lower P-value than our single SNP analysis, we conclude that there is no evidence for an epistatic effect for SNPs of any of these genes with SNPs from interacting proteins. In addition, in literature, no effects of interaction between these genes have been described as yet.

**Joint Reanalysis**

We then performed a joint reanalysis of 92 SNPs surrounding rs2522833 and rs2715147. The absolute values of Z-scores were plotted against the square root of the $r^2$ between one of these 92 SNPs with either rs2522833 or rs2715147. When assuming the null-hypothesis of no association, one would expect that the slope of the linear fit would approximate 0, since SNPs in high LD with a causal variant will reflect the Z-score of this causal variant. When we assume that rs2522833 is the causal variant, the slope of the linear fit is 4.00, which increases slightly to 4.15 when assuming that rs2715147 is the causal variant (Figure 2).
Chapter 4

Figure 2: A joint re-analysis of 92 SNPs, in which Z-scores for each SNP are tested against the relative correlation of each SNP with rs2715147

DISCUSSION

For this study our aim was to detect all common variants in the genes PCLO, GRM7 and SLC6A4 in 50 control samples of the Dutch GAIN-MDD cohort and then genotype these variants for the full cohort, in order to test if we could identify a more likely causal variant than rs2522833 for MDD in this Dutch cohort.

Rs2522833 was the variant with the lowest P-value in the GAIN-MDD GWAS and the variant with the lowest P-value in our fine-mapping study (rs2715147) are both common variants in the Dutch population. Since we expect a causal variant to be in high LD with these SNPs and these SNPs are common, we would expect an undetected causal variant also to be common in our population, allowing control samples to be used. In addition, when using control samples, one can detect the underlying LD-structure of the common Dutch population, rather than a putatively skewed LD-structure in cases.

After genotyping newly identified SNPs and tag SNPs, several SNPs were excluded by our quality control. In order to acquire genotypes for all genotyped SNPs, we imputed using Beagle. Both before and after imputation, we did not find a stronger associated variant suggesting that rs2522833 may indeed be the causal variant in the GAIN-MDD GWAS. However, there may be several other reasons why we did not find a variant with a lower P-value than rs2522833.

First of all, in the Sullivan GWAS, rs2522833 only became nominally significant after post-hoc analysis with a cohort that used a similar method of ascertainment. This could imply that the sample size is too limited to detect variants with a small effect size. When looking at GWAS for other complex traits, successes mostly occur with a substantially larger sample size. This has already led to the discovery of new loci for example for Parkinson’s disease (International Parkinson Disease Genomics Consortium, 2011), multiple sclerosis (International Multiple Sclerosis Genetics Consortium, 2011) and breast cancer (Ahmed et al., 2009).

Previously, we investigated another cause for the apparent lack of associated variants: poor SNP coverage (Verbeek et al., 2012). The array that was used for the GAIN-MDD GWAS, was a relatively early design and did not fully tag a substantial amount of the genome and was not designed in a gene-centered manner. This could lead to poor SNP coverage of certain genes, giving information only about the variants that have been genotyped for that gene in the GWAS and those variants that are in strong linkage disequilibrium (LD) with them. Therefore an associated
variant that is not in LD with the genotyped variant, may go undetected. We tested whether an increase of SNP coverage may lead to a more associated variant, and though P-values slightly decreased, no genome-wide significance was found.

Thirdly, the phenotype ‘MDD’ may yet be too diffuse to find an associated variant. One way to solve this predicament is to create more specialized phenotypes, the so-called ‘endophenotypes’ that link together genetic factors and biological markers. There are many physiological steps to go from genetic variants to a psychiatric disorder, which is why psychiatry hopes to use the endophenotypes to move closer to the DNA level. A distinct endophenotype may increase effect size and therefore yield more significant results. In particular psychiatric disorders may benefit from endophenotypical descriptions, as their etiology is often complex and is thought to be a mixture of environmental and genetic causes (Kendler et al., 2010). However, this may be a laborious task, since the complexity of the disorder would lead to many different endophenotypes to investigate.

Additionally, if a single common variant only has a small effect size, one would expect epistasis to occur; several variants, which together cause an increase in risk. However, with the methodology of a GWAS or a case-control genotyping study, one will not easily detect all the variants involved in epistasis, exactly because of the small effect size. An alternative approach to this problem is to perform gene-based association tests, as genes are the functional units of the genome. For this, we used the VEGAS method, which tests the evidence for association on a per-gene basis by summarizing the full set of markers. It also takes LD between markers into account by using simulation based on the LD structure of a set of reference individuals. However, when taking all SNPs from a certain gene, a weight has to be assigned to each SNP, for which methods are still under debate. In addition, only part of the gene –i.e. a single domain- may be involved in the etiology of the disease. In this case taking the whole gene as a functional unit may cause a weaker association than when looking at the association with a specific domain (Li et al., 2011), but the means to perform such tests are still limited. It may also be required for these tests to expand knowledge about the functions of protein domains in order to make a logical cut off which SNPs are to be included in a test. When more is known about the biological functions of various parts of the protein, one could for instance perform a joint re-analysis of SNPs located in specific domains, to increase the likelihood to find an association that has biological implications as well.

Also, we selected the newly detected variants that we genotyped based on a m.a.f. of more than 10% rather than on physical position, to increase the probability that the SNPs that we detected were actual variants instead of artifacts due to sequencing errors or contamination. It could well be that the variant(s) responsible for
the pathology of MDD have a m.a.f. of less than 10% in our 50 control samples and therefore were not genotyped on the full GAIN-MDD cohort.

By sequencing 50 control samples we aimed to find a previously undetected common variant. The region between rs27175147 and rs2522833 has an average coverage of 25x. However, in the same region, on average 10% of base pairs had not been covered. This may explain why an additional variant was not found in this region. The lack of an associated common variant may also suggest that the “common disease, common variant” hypothesis may not hold true for either MDD or for this particular cohort. Since the beginning of the GWAS era, over 500 associated common variants have been found for a range of disorders. However, they usually only explain a small portion of the heritability and only account for a small increase in risk. An alternative scenario would encompass multiple rare variants with a m.a.f. of less than 5% to cause an increase in risk. To detect variants with an m.a.f. of 1–5%, at least 100 cases would have to be sequenced. With the per base costs of NGS lowering, it becomes more feasible to sequence larger groups, enabling the detection of multiple rare variants which may contribute to complex disorders (Cirulli et al., 2010; Gibson, 2011). In the GAIN-MDD GWAS however, had rare variants been causal, there would not have been a marginally significant signal, unless if these rare variants would all have been recent and in the same haplotype. If these rare variants would cluster together in the same haplotype, then the variance explained by them should be so high, that they would have been expected to appear in linkage studies, which for MDD is not the case. Mixed models of both rare and common variants are currently under discussion, as it is indeed likely that complex disorders are under the influence of variants with various frequencies (Wray et al., 2011).

When taking all these factors into account, the fact remains that in this study as well as in three additional publications an identical area located in PCLO appears to contain the causal variant (Sullivan et al., 2009; Verbeek et al., 2012; Bochdanovits et al., 2009). The area in which rs2715147, rs2715148 and rs2522833 are situated shows high r² values, suggesting that the non-synonymous coding SNP rs2522833 or a SNP in high LD with it should be causal for the GAIN-MDD cohort. This SNP was found to be significant in the GWAS after post-hoc analysis with an Australian cohort, that used a similar method of ascertainment. The SNP changes a serine to an alanine in Piccolo’s calcium-binding C2A-domain. Overexpression of this C2A-domain causes a depression-like phenotype in mice (Furukawa-Hibi et al., 2010), which makes the PCLO gene still an interesting candidate gene for MDD.

The selection of the three genes was based on previous results (PCLO) and on literature (GRM7, SLC6A4). The first gene we selected, PCLO, is situated on chromo-
some 7q11.23–q11.30. It encodes the protein Piccolo, which is located in the presynaptic active zone. These specialized areas of the presynaptic terminal have specific cytoskeletal properties to facilitate the preparation and release of vesicles into the synaptic cleft. In 2008, Leal-Ortiz et al. showed that Piccolo is not essential for excitatory synapse formation, but it is a negative regulator of exocytosis, through modulation of Synapsin dynamics (Leal-Ortiz et al., 2008). This was later supported by Mukherjee et al., who suggested that Piccolo and its highly homologous brother Bassoon function as tethering proteins that mediate efficient synaptic vesicle clustering (Mukherjee et al., 2010). These observations make PCLO an interesting functional candidate for modulating the pathophysiology of MDD, as MDD is suggested to be caused by an imbalance in monoaminergic neurotransmission (Schildkraut 1975). Besides the GWAS from Sullivan in 2009, a meta-analysis of three population-based studies also showed a genome-wide significant P-value for rs2522833, which further underscores a possible role for PCLO in MDD (Hek et al., 2010).

The second gene, GRM7, encodes the protein mGluR7. This is a metabotropic glutamate receptor, which mediates slowly modulating actions of glutamate on the release of neurotransmitters and the excitability of cells (Shyn et al., 2011). It is abundant in brain regions which are known to be critical in anxiolysis and antidepressant action, such as the amygdala and hippocampus. This suggests that mGluR7 is involved in the regulatory circuits that influence anxious and/or depressed behavior. In 2003, Cryan et al showed that GRM7−/− mice displayed less immobility following various stress paradigms. However, these anxiolytic/antidepressant results were still less pronounced than when animals were treated with anxiolytic drugs like benzodiazepines (Cryan et al., 2003). Furthermore, Mitsukawa et al., found an increase in glucocorticoid receptors in the hippocampus of GRM7−/− mice after stress paradigms. This connects mGluR7 to the hypothalamus-pituitary-adrenal axis (HPA-axis) which in turn is thought to be a key regulator in the stress response. In addition, GRM7−/− mice showed lower levels of the stress hormone corticosterone after stress paradigms than their GRM7+/+ litter mates (Mitsukawa et al., 2006). Moreover, in a meta-analysis of three studies for MDD, one of the strongest association peaks was observed for GRM7 (Shyn et al., 2011). In summary, though GRM7 may be an eligible candidate gene in animal models, in this study we did not find evidence for a variant that showed a stronger association than in the GAIN-MDD GWAS.

And finally, SLC6A4 encodes the serotonin transporter, which plays a pivotal role in the monoamine hypothesis of depression. The monoamine hypothesis states that depression is caused by the underactivity/imbalance of monoamines in the brain. The serotonin transporter regulates the availability of serotonin in the synaptic

83
cleft by terminating the action of serotonin and recycling it in a sodium-dependent manner. Consequently, the serotonin transporter is a target for antidepressant drugs like selective serotonin reuptake inhibitors (SSRIs) which block the transporter and thereby increase available serotonin. SLC6A4 has a length polymorphism in the promoter region, of which the short allele leads to less transcription of the gene. In 2003, Caspi et al. found a gene-environment interaction between the short allele and stressful life events as a predictor for MDD. However, replication efforts have been inconclusive and a meta-analysis in 2009 did not show this interaction (Caspi et al., 2003; Risch et al., 2009). Although this length polymorphism may be associated with MDD in interaction with the environment in the cohort used by Caspi et al. and we found a slightly lower P-value for this gene, we do not find evidence for genome-wide association.

In conclusion, while in the 5 kb area between rs2715147 and rs2522833 in PCLO an average coverage of 25x was reached, we did not detect an additional common variant. Both in GRM7 and SLC6A4 previously undetected variants were found as well, but in neither genes we detected a variant that was more associated with MDD than rs2522833 in PCLO.

Although we cannot exclude the presence of multiple rare variants, our results suggest that, in accordance with the findings of Sullivan et al., non-synonymous coding SNP rs2522833 (or a variant in high LD with it) in PCLO gene is the causal variant responsible for the association peak in the GAIN-MDD cohort.

ACKNOWLEDGMENTS

Statistical analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org) which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003) along with a supplement from the Dutch Brain Foundation and the VU University Amsterdam.

We thank Nanne Aben for writing the Python script that was used for extracting haplotype and SNP data from 1000 genomes data.

We thank Ruben van ’t Slot for assisting us with the adjusted elution protocol for the sequence capture arrays.
FUNDING STATEMENT

Funding support for NESDA and NTR was provided by the Netherlands Scientific Organization (904-61-090, 904-61-193, 480-04-004, 400-05-717, 912-100-20) Centre for Medical Systems Biology (NWO Genomics), the Neuroscience Campus Amsterdam (NCA) and the EMGO+ institute; ZonMW (Geestkracht program, 10-000-1002), National Institute of Mental Health (NIMH, R01 MH059160) and matching funds from participating institutes in NESDA (Netherlands Study of Depression and Anxiety) and NTR. Genotyping was funded by the Genetic Association Information Network (GAIN) of the Foundation for the U.S. National Institutes of Health, and analysis was supported by grants from GAIN and the NIMH (MH081802). Genotype data were obtained from dbGaP (http://www.ncbi.nlm.nih.gov/dbgap, accession number phs000020.v1.p1). The additional genotyping project was performed within the framework of the Center for Medical Systems Biology (CMSB) and TIPharma project T5-203. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing; Nature; 467: 1061–1073
American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders: DSM-IV; Washington, DC
Resequencing Three Candidate Genes for Major Depressive Disorder in a Dutch Cohort


Li MX, Gui HS, Kwan JS, Sham PC (2011) GATES: a rapid and powerful gene-based association test using extended Simes procedure; Am J Hum Genet; 88: 283–293

Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM; AMFS Investigators, Hayward NK, Montgomery GW, Visscher PM, Martin NG, Macgregor S (2010) A versatile gene-based test for genome-wide association studies; Am J Hum Genet; 87: 139–145


Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses; Am J Hum Genet; 81: 559–575


Schildkraut JJ (1975) Catecholamines in affective disorders; Psychopharmacol Bull; 11: 35


Resequencing Three Candidate Genes for Major Depressive Disorder in a Dutch Cohort

Martin NG, Wray NR, Boomsma DI, Penninx BW (2009) Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo; Mol Psychiatry; 14: 359–375


Wray NR, Purcell SM, Visscher PM (2011) Synthetic associations created by rare variants do not explain most GWAS results; PLoS Biol; 9: e1000579
CHAPTER 5
Pathway-based analyses of candidate genes for major depressive disorder.
A pilot study.
ABSTRACT

Patients with Major Depressive Disorder (MDD) show changes in activity of the Hypothalamus Pituitary Adrenal (HPA) axis, Hypothalamus Pituitary Thyroid (HPT) axis and Vitamin D metabolism. This study aimed to find common genetic variants that may underlie alterations in these pathways in MDD in general and, more specific, in severe recurrent cases. Individuals from the Netherlands Study for Depression and Anxiety (NESDA) with a lifetime diagnosis of MDD as diagnosed by the Composite International Diagnostic Interview (CIDI) were classified as moderate or severe with or without recurrence. Tag SNPs were then selected within key genes of each pathway, to reach 100% coverage. In addition, candidate SNPs from literature were genotyped. We identified a borderline significant association between severe recurrent MDD and rs6198, a SNP located in the 3'UTR region of NR3C1 gene (P = 3.54*10^{-3}, OR 1.48). We detected a difference in the biological measurements between cases and controls in both the HPA axis and vitamin D metabolism, but the genotyped SNPs did not show an additional effect on these measurements. NR3C1 encodes the glucocorticoid receptor, which is the location where the negative feedback on the HPA axis takes place. The gene has been previously implicated in MDD. The fact that only borderline significance was reached, may be explained by the small number of severe recurrent MDD cases. Although differences in the levels of biological measurements were detected, the genotyped SNPs did not show an additional effect.
INTRODUCTION

Major depressive disorder (MDD) has a lifetime prevalence of 15%, and is characterized by prolonged low mood, loss of interest/pleasure in normally enjoyable activities, decrease in cognitive functioning (memory and concentration) and changes in behavior, such as sleep and appetite (American Psychiatric Association, 1994). MDD is considered to be a complex disorder, since both stress-related and other environmental factors as well as genetic mechanisms may play a role in its pathogenesis. Probably, a large number of genes contributes to the disorder, each gene only responsible for a slight increase in risk. Although it is known from twin studies that MDD has an approximate heritability of 40%, discovering the genetic variants responsible is still a challenge (Kendler et al., 2006).

During the last decade, for most complex disorders, genome-wide association studies (GWAS) have been performed. However, the variants found in these studies have relatively small effect sizes and therefore only explain a small part of the heritability of the disorder. In addition, the variants that are detected with this method do not always show a clear functional connection.

In a second approach, the candidate gene approach, genes are selected based on a hypothesis concerning their function within a trait. Although small successes are found, in general these studies are difficult to replicate (Bosker et al., 2011). However, candidate genes are seldomly viewed in a pathway specific manner. Lips et al., studied how genes are organized in functional groups to test their accumulated variants for association with the psychiatric disorder schizophrenia (Lips et al., 2012). However, in this study genes were classified according to molecular location rather than their function(s) in a specific pathway. One study combined genotype data with underlying biological pathways in five major psychiatric disorders, such as MDD (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013). This study showed genome-wide significance for SNPs located in genes coding for calcium channel subunits after pathway analysis. This finding adds to several earlier studies suggesting a role for SNPs in calcium-associated genes.

In this study we aim to analyze pathways involved in the etiology of MDD in order to find variants associated with the disorder and to find whether these variants together with disease status may have a combined effect on the levels of the biological end products of these pathways. The first pathway we investigate is the hypothalamic-pituitary-adrenal (HPA) axis. The HPA axis plays a prominent role in the neuroendocrine system and controls the physiological reaction to stress as well the immune system, mood and emotion. Corticotropin releasing hormone
(CRH) from the hypothalamus stimulates the release of adrenocorticotropic hormone (ACTH) from the pituitary gland, which in turn stimulates the release of cortisol from the adrenal glands. Cortisol itself has a negative feedback effect on the hypothalamus; when cortisol levels are high, CRH release decreases (Pariante, 2003). In both humans and animals, high levels of stress are associated with elevated plasma concentrations of cortisol (Burke et al., 2005; Dallman et al., 2004). Rats with prolonged exposure to high levels of cortisol or chronic stress show decreased neurogenesis in the dentate gyrus of the hippocampus, a part of the brain associated with memory and cognition. This effect can be counteracted by the use of antidepressant medication. In healthy subjects, high levels of cortisol inhibit the HPA-axis, but in MDD patients this feedback loop is frequently hampered. In addition, healthy controls with a parental history of depression and/or anxiety show an increased cortisol awakening response similar to depressed patients (Vreelburg et al., 2010). In this pathway-based study, we therefore genotyped SNPs in the genes of the receptors and other proteins, involved in the HPA axis pathway to find whether the effect of cortisol levels can be mediated through genetic variants.

The second hypothalamic pathway we studied is the hypothalamus-pituitary-thyroid (HPT) axis. In addition to its many functions in basic metabolic rate, thyroid hormone has been suggested to act as a neurotransmitter, as it can regulate the amount and activity of serotonin, norepinephrine and GABA in the brain (Dratman & Gordon, 1996). This would imply that a dysfunction in the thyroid system may contribute to an imbalance in serotonin and norepinephrine. This imbalance was originally hypothesized to be the cause of MDD, in the monoamine hypothesis of depression (Schildkraut 1975). For these neurotransmitters, a connection with depression has been established. Secondly, serum concentrations of thyroid stimulating hormone were found to be slightly higher in depressed patients than in controls in a study by Brouwer et al. (Brouwer et al., 2005). This is supported by the finding that hypothyroidism was associated with depression (Kirkegaard & Faber, 1998). In this study, we aimed, therefore, to genotype tag SNPs in genes involved in thyroid hormone metabolism. In addition, we aim to find whether the effect of low thyroid hormone can be mediated through genetic variants.

The third pathway that we analyzed in conjunction with MDD was vitamin D metabolism. Although vitamin D is present in some animal source foods, production in the skin, under the influence of UV-B light from the sun, is the main source of vitamin D. Low levels of vitamin D are hypothesized to be linked to MDD (Hoogendijk et al., 2008; Annweiler et al., 2013). Milaneschi et al. found that the presence and severity of MDD were linked to 25-hydroxy-vitamin D (25(OH)D), the vitamin D metabolite that can be measured in blood (Milaneschi et al., 2013). In this study we investigated whether variants in genes, that code for receptors or other proteins
involved in vitamin D metabolism, are associated to MDD through altered vitamin D levels.

This study shows a borderline significant association between the glucocorticoid receptor variant rs6198 and severe recurrent MDD cases, confirming a possible role for this variant in the etiology of MDD. In addition, differences were found between AUC in cases and controls. However, there was no significant difference between carriers and non-carriers of the minor allele of rs6198.

**METHODS AND MATERIALS**

**Sample description**
The subjects for this study originated from a longitudinal study, the Netherlands Study for Depression and Anxiety (http://www.nesda.nl), designed to be representative of individuals with depression and/or anxiety disorders, including 1766 cases and 900 controls. All individuals had an age of 18-65 years and had self-reported western European ancestry. Ascertainment of cases was from outpatient specialist mental health facilities and by primary care screening. Inclusion criteria were a lifetime diagnosis of DSM-IV MDD as diagnosed by the Composite International Diagnostic Interview (CIDI). Individuals were classified as moderate or severe with or without recurrence, according to CIDI guidelines (Kessler et al., 2003). Collection of samples and isolation of DNA has been thoroughly described elsewhere (Penninx et al., 2008).

**Ethical Issues**
The NESDA study was approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Center, Amsterdam, an Institutional Review Board certified by the US Office of Human Research Protections (IRB number IRB-2991 under Federal-wide Assurance-3703; IRB/institute code, NESDA 03-183). All subjects provided written informed consent.

**Gene and SNP selection**
Genes were selected based on literature for each pathway. Key genes in the activation of HPA axis, the HPT axis and in Vitamin D metabolism were then tagged at $r^2=0.8$ and m.a.f.=0.2, using tagger software (De Bakker et al., 2005) to select SNPs to reach 100% coverage with these parameters. In the tagging procedure, SNPs with a previous association with MDD were preferred over SNPs with no known associations (Table 1). Genes *NR3C1* and *NR3C2* are of considerable size, leading to numbers of tag SNPs that would fall beyond the scope of this project. Therefore,
for these genes, SNPs were selected from literature, that were not previously genotyped in the NESDA cohort.

Table 1: Genes selected to be genotyped in each pathway

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Genes</th>
<th>Number of genotyped SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cortisol</td>
<td>NR3C1, NR3C2, AVPR1A, AVPR1B, FKB5, CRHR1, CRHR2</td>
<td>43</td>
</tr>
<tr>
<td>Vitamin D</td>
<td>DHCR7, VDR, GC, CYP27A1</td>
<td>41</td>
</tr>
<tr>
<td>Thyroid hormone</td>
<td>THR, DIO1, DIO2, SLC01C1</td>
<td>7</td>
</tr>
</tbody>
</table>

**Genotyping procedure**

All genotyping was performed using the OpenArray® Real-Time PCR System (Life Technologies, Carlsbad, USA), in accordance with the protocol of the manufacturer (version: 7/2010). Arrays were designed to have 128 assays for 24 samples per array and were loaded using the OpenArray Accufill robot or using the AutoLoader, manually loaded into a cassette and then PCR was performed in an NT cycler (GeneAmp® PCR System 9700, Life Technologies, Carlsbad, USA). After this, arrays were scanned with the OpenArray NT Imager. The quality of scanned arrays was checked by visually assessing the location of the array in the scanner (the so-called Spotfind image). The loading of the arrays was checked using the ROX image and the fluorescence signal strength was checked using the VIC and FAM images with the software tool ImageJ (http://rsbweb.nih.gov/ij/). Genotypes for approximately 200 samples were analyzed simultaneously, using Taqman Genotyper Software v 1.0.1. This number of 200 samples was set by optimizing for clear clustering, without getting a bias due to too few data points. A home-made Perl script was then used to merge all data and create a pedigree file. SNPs were required to have a m.a.f.>1%, missing genotypes <5% and a HWE P-value >0.0005 in order to be analyzed.

**Imputation**

For the imputation of genotypes for missing individuals, we used Beagle software (Browning & Browning, 2007). 1000 genomes 2010-06 release CEU data was used as a reference. To leave the underlying LD-structure intact, we used a margin of 100kb around each gene. To extract the genes +/- 100kb from the full chromosome data of the 1000 genomes project, we used PLINK-Seq (http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell et al., 2007). Imputation was performed per gene with 100 Markov chain iterations, for all samples. All imputation was performed on the Lisa system cluster (https://www.surfsara.nl/systems/lisa).
Pathway-based analyses of candidate genes for major depressive disorder: A pilot study.

**Association analysis**

Single SNP association analysis was performed using SPSS (IBM® SPSS® Statistics version 22, IBM Corp, Armonk, NY, USA) by means of a chi square test for association. In addition, logistic regression was performed to include biological measurements as continuous covariates. Gene-based P-values were calculated using the VEGAS tool [Liu et al., 2010]. VEGAS tests the evidence for association on a per-gene basis by summarizing the full set of markers and takes LD between markers into account by using simulation based on the LD structure of a set of reference individuals. We used our individual genotype data as a reference set, so that LD would be estimated specifically for the NESDA cohort. For each gene one million simulations were run.

In addition, all single SNP analyses were repeated using only the severe recurrent cases (N=354) of the NESDA cohort, versus healthy controls (N=900).

We used the Bonferroni method of correcting for multiple testing in all our association analyses, which entails correcting for 43 SNPs in the HPA-axis (P=1.2E-3), for 41 SNPs for vitamin D metabolism (P=1.2E-3) and 7 SNPs for thyroid hormone metabolism (P=7.1E-3).

To correct for biological measurements, binary logistic regression was performed with affection status as the dependent variable and SNPs and biological measurements as independent variables. Finally, when using biological measurements as an outcome variable, an ANOVA test was performed to assess the difference between groups.

**Pathway analysis**

For the analysis of the pathways for HPA-axis, we used the InRich tool, designed for detecting enriched association signals of LD-independent genomic regions within biologically relevant gene sets (Lee et al., 2012). This pathway tool was selected as it can conduct pathway analysis on any type of genomic variation data, including but not limited to SNPs and genes, as well as their combination. The underlying principle of the InRich program assumes association in at least one calculated interval, which is why this analysis was performed on the HPA-axis only, as the other pathways did not show any (borderline) significant SNPs.

Association tests for biological measurements

In addition to mapping genetic variants in the three selected pathways, we tested for an association between genotypes in cases and levels of the biological markers of these pathways. Four different groups were assessed, assuming a dominant model: depression + minor allele, depression no minor allele present, no depression + minor allele present and no depression no minor allele present.

We performed ANOVA tests followed by post-hoc Scheffé tests, to test for an association of the biological measurements of these pathways for the biological meas-
measurements for these pathways. For the HPA axis, we assessed total cortisol output (AUCo), cortisol increase (AUCi), the slope of the cortisol increase, evening cortisol levels and cortisol levels after the dexamethasone suppression test (DST). For the HPT axis, levels of TSH in blood were measured. For vitamin D, vitamin D levels in blood were measured.

RESULTS

Genotyping rates
13 SNPs failed genotyping as the assays did not cluster properly, as they were either monomorphic or clusters were too close together to distinguish between genotypes. With these SNPs excluded, this lead to a total genotyping rate of 96.4% for the HPA axis, 96.1% for vitamin D metabolism and 96.6% for the HPT axis across all genes. Although our aim was to reach a minimum genotyping rate of 98%, concordant with the generally accepted rates in literature, several SNPs were not previously validated by the manufacturer, leading to a lower genotyping rate.

Individual SNP association analysis
In our analysis of individual SNPs, 43 SNPs were tested for the HPA axis, 41 for the vitamin D pathway and 7 SNPs, for the HPT axis. For the single SNP association analysis in the full NESDA cohort, we did not acquire any significant findings. However, when performing single SNP analysis, using the severe recurrent cases only, SNP rs6198 (A/G) in the glucocorticoid receptor gene NR3C1 reached significance (P=3.54E-3, OR= 1.48). When applying Bonferroni correction however, this P-value is not significant, as there would have to be corrected for 43 tests (P=1.16E-3).

<table>
<thead>
<tr>
<th>Pathway in which SNP is located</th>
<th>SNP with lowest P-value</th>
<th>Gene</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPA-axis</td>
<td>rs6198</td>
<td>NR3C1</td>
<td>0.01</td>
</tr>
<tr>
<td>Vitamin D metabolism</td>
<td>rs12797951</td>
<td>DHCR7</td>
<td>0.06</td>
</tr>
<tr>
<td>Thyroid hormone metabolism</td>
<td>rs13063628</td>
<td>THRB</td>
<td>0.23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pathway in which SNP is located</th>
<th>SNP with lowest P-value</th>
<th>Gene</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPA-axis</td>
<td>rs6198</td>
<td>NR3C1</td>
<td>3.54E-3</td>
</tr>
<tr>
<td>Vitamin D metabolism</td>
<td>rs7299460</td>
<td>VDR</td>
<td>0.03</td>
</tr>
<tr>
<td>Thyroid hormone metabolism</td>
<td>rs10444412</td>
<td>SLCO1C1</td>
<td>0.16</td>
</tr>
</tbody>
</table>
Gene-based association analysis
As genes create the functional product of the genome, we combined the genotyped SNPs into a gene-based association analysis, using the VEGAS-tool [20]. To correct for multiple testing, we performed a Bonferroni correction for the number of genes tested in each pathway. This did not lead to any significant findings (supplementary table 1).

Pathway analysis
The Inrich software performed a total of 6000 permutations on 43 SNPs and their corresponding genes in the HPA-axis. The lowest P-value in this analysis was found not to be significant at P=0.2054, which suggests that the borderline significance of rs6198 in severe recurrent MDD cases relies on the SNP itself, rather than the NR3C1 gene or the HPA-axis pathway. Correcting for the various measurements for cortisol did not lower P-values.

ANOVA tests for biological measurements
Based on previous associations between cortisol levels and depression and the borderline significance between rs6198 and depression, we hypothesized that the minor allele (G) of rs6198 in combination with depression status may play a role in the levels of cortisol in blood.
To assess the effect of depression status and the rs6198 risk allele on cortisol levels, ANOVA analyses were performed on total cortisol output (AUCG), cortisol increase (AUCI), the slope of the cortisol increase, evening cortisol levels and cortisol levels after the dexamethasone suppression test. Four different groups were assessed, assuming a dominant model: depression + minor allele, depression no minor allele present, no depression + minor allele present and no depression no minor allele present. The oneway ANOVA for AUCG showed P = 2.83E-4, indicating a significant difference between the four groups. In order to analyze between which groups the largest difference occurs, a post-hoc Scheffé test was performed. This test takes into account the fact that unequal group sizes were used. The Scheffé test demonstrated P=6.04E-3 (95% C.I. = 0.62 – 5.3) for the difference in means between C- and D+. In addition, P=9.00E-3 (95% C.I. = 0.57 – 5.79) for C+ and D+. These results suggest that the difference in AUCG is not due to the effect of rs6189, but due to depression status itself.
For Vitamin D, SNP rs12721364 gave the lowest P-value (P=8.6E-5 for the ANOVA, indicating a difference in Vitamin D levels between groups. The Scheffé test showed P=8.62E-3 (95% C.I. = 1.22 – 12.19) for the difference in mean vitamin D level between C- and D- and P=1.48E-3 (95% C.I. = 2.46 – 14.56), suggesting that also in vitamin D levels there is a difference between cases and controls, but this effect is not increased by the minor allele of the SNP.
For the HPT axis, no difference between groups was found with the ANOVA, indicating no difference in TSH levels for genotyped individuals.

**DISCUSSION**

In this study we genotyped variants in three biological pathways linked to MDD: the HPA-axis, the HPT axis and vitamin D metabolism. Our aim was to test the hypothesis that genetic variants and biological measurements of these pathways are associated with MDD. In order to obtain maximal information about genetic variants in these pathways, we selected genes from literature, for which we selected and genotyped tag SNPs with a high $r^2$ (0.8). No significant association was found for vitamin D and thyroid pathways, but borderline significance was reached for rs6198 in 3’UTR region of the *NR3C1* gene. In a subsequent step to test for association between this SNP in cases and cortisol levels, a difference was found in the AUC_{6} measurement between cases and controls, but this was not associated with the minor allele of rs6198.

In addition, a difference was found in vitamin D levels between cases and controls, as is known from literature (Hoogendijk et al., 2008; Annweiler et al., 2013). However, there was no additional effect of the minor allele of the genotyped SNPs.

*NR3C1* encodes the glucocorticoid receptor (GR), is a crucial part of the HPA-axis, as the end product of the HPA-axis, cortisol, binds to this receptor. A consistent finding in MDD patients is the high level of cortisol that can be measured in blood (Maletic et al., 2007). High cortisol levels suggest a dysfunction of the HPA-axis, in particular the GR-regulated negative feedback loop via the hippocampus and the hypothalamus. Rs6198 was previously shown to influence GR mRNA stability. Moreover, Rs6198 together with another SNP, rs10482605, were shown to form a haplotype that can influence both mRNA stability and transcription (DeRijk et al., 2001; Kumsta et al., 2009). A change in mRNA stability may reduce available GRs, reducing the inhibitory signal that the hypothalamus will receive in the negative feedback loop of the HPA-axis. Although most of the research concerning this topic is of course performed on animals, in human subjects with MDD, the number of GRs has been shown to be reduced in peripheral tissues (Gormly et al., 2005). Interestingly, the GR-antagonist mifepristone ameliorates symptoms of psychotic depression and in rats it can rapidly reverse the decrease in neurogenesis in the hippocampus, caused by chronic administration of corticosteroids (Mayer et al., 2006). With a low clinical efficacy of most antidepressant medication and a large number of side effects, it is desirable to explore towards alternatives for selective serotonin reuptake inhibitors or norepinephrin reuptake inhibitors. The HPA-ax-
is, specifically the GR, may provide a target for the pharmacological treatment of MDD.

In addition to our analysis of the HPA-axis, we performed genotyping for the HPT axis as well. Kirkegaard and Faber showed that lack of thyroid hormone, hypothyroidism, may cause clinical depression (Kirkegaard & Faber, 1998). In our research, variants in the genes of the thyroid metabolism pathway did not show an association with MDD or with thyroid levels.

Finally, we performed an analysis of the metabolism of Vitamin D. Vitamin D levels have been previously associated with MDD and with seasonal affective disorder (Hoogendijk et al., 2008; Milaneschi et al., 2013). However, the genes that we selected in the pathway of vitamin D metabolism show no association with MDD. We found a difference in vitamin D levels between cases and controls, but there was no additional effect of genotype. Although this cohort consisted of MDD cases, a larger effect may be witnessed when looking at seasonal affective disorder, as vitamin D levels fluctuate during seasonal changes.

Methods of analysis for detecting association may be limited, as in our gene-based tests the individual results of the variants are all taken along when looking for a gene-based association. Therefore, the lack of an association in our gene-based test may be explained by the sheer length of the genes: even if a variant close to the 3'UTR region shows borderline significance, a variant that lies several exons away may show no inclination towards significance at all. This may be caused by the underlying LD structure: the variant that does show borderline significance may be in low $r^2$ with other variants in the gene, so that their combined P-values do not reflect the significance anymore.

As genes are the functional units of the genome, it is highly desirable that more advanced gene-based tests will be developed, since the current tests depend too much on the average of P-values. In an ideal situation, gene-based tests would incorporate information about exons and regulatory regions, as a variant in these regions may have a larger effect than a variant in an intronic region.

In this study we have shown a borderline significant result with rs6198 in the GR receptor. However in literature, there are more suggestions of SNPs in this gene to be associated with MDD (Szczepankiewicz et al., 2011). A recurring challenge in these studies is solid replication of the results, which is hampered by the poor specificity of the phenotyping. Although diagnostic interviews for MDD have greatly improved, there are large differences between cases from a general practitioner and cases selected from psychiatric clinics. Our results suggest that there is indeed a difference between a full cohort of various MDD cases and the more se-
vere recurrent cases, as our most associated variant was only found using severe recurrent cases only.
Besides more phenotypes, possible endophenotypes for the disorder may be an alternative to find a valid association. In psychiatric disorders, diagnosis is often based on a set of behaviors, rather than the identification of anatomical or physiological dysfunction. Endophenotypes allow researchers to dissect a complex disorder into its various components, thus not only increasing comprehension of the pathophysiology, but also increasing knowledge of the genetic background. The endophenotype has to be associated with the trait in the population that is being studied, has to be heritable and will cosegregate with the disease within families. A controversial part of working with physiological pathways is that there are lacunas in the knowledge of which genes/proteins are involved in which pathway and whether there are interactions between different pathways. In the case of working with pathways, an a priori hypothesis is required, which in turn requires understanding of the biological effect of said pathway. This is in sharp contrast with the large numbers of GWAS that have been performed. Although GWAS have dramatically increased the number of variants associated with complex disorders, much of the variability is still unexplained. The endophenotype approach could be an alternative to overcome the shortcomings of GWAS and create biologically relevant phenotypes.
In addition, variants associated with complex disorders have relatively low effect sizes. Variants that have a larger deteriorating effect might be selected against, as they are not beneficial for the individual. However, this is dependent of the age of onset of the disorder (Park et al., 2011).

In conclusion, we have found a borderline significant association between rs6198 situated in NR3C1 and the severe recurrent MDD cases of the Dutch NESDA cohort. No significance was found for gene-based and pathway-based test, before and after correcting for various biological measurements of cortisol. NR3C1 and its product the GR are an important part of the negative feedback loop in the HPA-axis. Although only borderline significance was reached in this cohort, rs6198 has been associated with depression in other cohorts both as a single SNP and in a haplotype, indicating a putative role for this variants in the etiology of the disorder (Szczepankiewicz et al., 2011; Lahti et al., 2011). Furthermore, the rs6198 SNP is associated with glucocorticoid resistance, which is also commonly established in MDD patients (Pariante et al., 2001). Further research into the HPA-axis in conjunction with MDD may be required to further explore the etiology of the disorder and eventually to find novel targets for antidepressant medication.
Pathway-based analyses of candidate genes for major depressive disorder. A pilot study.

ACKNOWLEDGEMENTS

Funding support for NESDA was provided by the Netherlands Scientific Organization (904-61-090, 904-61-193, 480-04-004, 400-05-717, 912-100-20) Centre for Medical Systems Biology (NWO Genomics), the Neuroscience Campus Amsterdam (NCA) and the EMGO+ institute; ZonMW (Geestkracht program, 10-000-1002), NIMH (RO1 MH059160) and matching funds from participating institutes in NESDA. Genotyping was funded by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health, and analysis was supported by grants from GAIN and the NIMH (MH081802). Genotype data were obtained from dbGaP (http://www.ncbi.nlm.nih.gov/dbgap, accession number phs000020.v1.p1). The additional genotyping project was performed within the framework of the Center for Medical Systems Biology (CMSB) and TIPharma project T5-203.

Statistical analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org) which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003) along with a supplement from the Dutch Brain Foundation and the VU University Amsterdam.

REFERENCES

1000 Genomes Consortium (2010) A map of human genome variation from population-scale sequencing
Cross-Disorder Group of the Psychiatric Genomics Consortium (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs; Lancet; Apr 20;381(9875):1371-9

Dratman MB, Gordon JT (1996) Thyroid hormones as neurotransmitters; Thyroid; Dec;6(6):639-47


Hoogendijk WJ, Lips P, Dik MG, Deeg DJ, Beekman AT, Penninx BW (2008) Depression is associated with decreased 25-hydroxyvitamin D and increased parathyroid hormone levels in older adults; Arch Gen Psychiatry; May;65(5):508-12


Lee PH, O’Dushlaine C, Thomas B, Purcell SM (2012) INRICH: interval-based enrichment analysis for genome-wide association studies; Bioinformatics; Jul 1;28(13):1797-9


Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM; AMFS Investigators, Hayward NK, Montgomery GW, Visscher PM, Martin NG, Macgregor S (2010) A versatile gene-based test for genome-wide association studies; Am J Hum Genet; Jul 9;87(1):139-45


Pathway-based analyses of candidate genes for major depressive disorder. A pilot study.

Pariante CM, Miller AH (2001) Glucocorticoid receptors in major depression: relevance to pathophysiology and treatment; Biol Psychiatry; Mar 1;49(5):391-404
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses; Am J Hum Genet; Sep;81(3):559-75
S R Browning and B L Browning (2007) Rapid and accurate haplotype phasing and missing data inference for whole genome association studies by use of localized haplotype clustering; Am J Hum Genet; 81:1084-1097
Schildkraut JJ (1975) Catecholamines in affective disorders; Psychopharmacol Bull; Jul;11(3):35
SUPPLEMENTARY DATA

The assay for rs6198 showed three clearly separated clusters with sufficient amplification to distinguish from negative control samples (Figure 1), ensuring that any results distilled from this assay were in fact reliable.

Supplementary figure 1: The clustering for rs6198 in NR3C1

Supplementary table 1: Results of gene-based analysis

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene P-value</th>
<th>Best SNP</th>
<th>SNP P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR3C1</td>
<td>0.0566</td>
<td>rs6198</td>
<td>0.01</td>
</tr>
<tr>
<td>CYP27A1</td>
<td>0.133</td>
<td>rs645163</td>
<td>0.08</td>
</tr>
<tr>
<td>SLCO1C1</td>
<td>0.695</td>
<td>rs10444412</td>
<td>0.51</td>
</tr>
</tbody>
</table>
Pathway-based analyses of candidate genes for major depressive disorder. A pilot study.
A Common polymorphism in the ABCB1 gene is associated with side effects of PGP-dependent antidepressants in a large naturalistic Dutch cohort.

Pierre M. Bet*
Eva C. Verbeek*
Yuri Milaneschi
Daniëlle B. Straver
Tiny Uithuisje
Marianna R. Bevova
Jacqueline G. Hugtenburg
Peter Heutink
Brenda W. Penninx
Witte J. Hoogendijk

Pharmacogenomics J. 2016 Apr;16(2):202-8. doi: 10.1038/tpj.2015.38
*Both authors contributed equally
Chapter 6

ABSTRACT

The drug efflux transporter permeability glycoprotein (PGP) and cytochrome P450 (CYP) 2C19 are important for eliminating antidepressants from the brain and body. The ABCB1 gene, encoding for PGP, and CYP2C19 gene have several variants that could influence enzyme function and thereby the effect of PGP- and 2C19-dependent antidepressants. We investigated the association of antidepressant side effect and common genetic variation in 789 antidepressant users. In PGP-dependent antidepressant users, the A-allele of the rs2032588 single-nucleotide polymorphism (SNP) was associated with a lower number of side effects after adjusting for gender, age, dosage and duration of use, \( B = -0.44, q = 4.6 \times 10^{-3} \). This association was different from and absent in non-PGP-dependent antidepressant users. Other SNP associations as well as an interaction analysis between the rs2032588 SNP and the CYP2C19 SNPs were not statistically significant after adjusting for covariates and multiple comparisons. The association of rs2032588 with antidepressant side effects suggests the involvement of the ABCB1 genotype in the clinical pharmacology of PGP-dependent antidepressants.
INTRODUCTION

Major depressive disorder and anxiety disorders are disabling psychiatric disorders with a lifetime prevalence in western society of 14.7% and 14.5%, respectively (Alonso et al., 2004). Currently, treatment of major depressive disorder and anxiety disorders consists of psychotherapy, antidepressant medication or a combination of both (Baldwin et al., 2005; Weihs et al., 2011; Davidson et al., 2010). The use of antidepressants is often complicated by the occurrence of side effects, which affect persistence, drug compliance and quality of life (Kelly et al., 2008; Kikuchi et al., 2013; Hu et al., 2004; Demyttenaere et al., 2001; Goethe et al., 2007).

The drug efflux transporter permeability-glycoprotein (PGP) located on the blood brain barrier is important for eliminating drugs from the brain. PGP has been found to have a role in the pharmacokinetics of antidepressants such as paroxetine, citalopram and venlafaxine (Horstmann et al., 2009). The \textit{ABCB1} gene, formerly known as \textit{MDR1}, encodes for PGP. Three \textit{ABCB1} single-nucleotide polymorphisms (SNPs), rs1128503 (1236C>T) rs2032582 (2677G>T/A) and rs1045642 (3435C>T), have been associated with both higher and lower PGP expression levels, PGP functionality and specificity for PGP substrates (Hoffmeyer et al., 2000; Gow et al., 2008; Kimchi-Sarfaty et al., 2007; Salama et al., 2006; Leschziner et al., 2007). However, these findings have not been confirmed in replication studies.

These three SNPs inconsistently have been associated with antidepressant efficacy (Singh et al., 2012; Gex-Fabry, et al., 2008; Dong et al., 2009; Kato et al., 2008; Lin et al., 2011; Mihaljevic et al., 2008; Nikisch et al., 2008; Peters et al., 2008; Sarginson et al., 2010; Uhr et al., 2008) dose variability (Singh et al., 2012; Mas et al., 2012; Noordam et al., 2013) antidepressant blood levels (Gex-Fabry et al., 2008; Nikisch et al., 2008; Uhr et al., 2008; Fukui et al., 2007; Yoo et al., 2012; Xiang et al., 2010) and occurrence of side effects (Roberts et al., 2002; Jensen et al., 2012; Bly et al., 2013; Zourkova et al., 2013; Menu et al., 2010; Perroud et al., 2011; Laika et al., 2006). In the Netherlands Study of Depression and Anxiety (NEPSDA), De Klerk et al. studied the association between PGP SNPs and SSRI side effects (De Klerk et al., 2013). Two imputed SNPs, rs2235040 and rs2302583, were associated with the occurrence of side effects but there was no association for the three most common PGP SNPs mentioned above. Other PGP SNPs have been associated with the antidepressant side effects such as orthostatic hypotension, sexual dysfunction (Bly et al., 2013; Zourkova et al., 2013), weight gain (Menu et al., 2010), suicidal ideation (Perroud et al., 2011), and amitriptyline side effects assessed with the Dosage Record and Treatment Emergent Symptoms scale (Laika et al., 2006). However, studies on antidepressant side effects covering the whole \textit{ABCB1} gene have not been performed.
Cytochrome P450 (CYP) enzymes, particularly CYP2C19 and CYP2D6, are also known to influence the pharmacokinetic profile of antidepressants and therefore resulting in the possible occurrence of side effects (Altar et al., 2013; Kirchheiner et al., 2004). The CYP2D6 gene contains not only SNPs but also a multitude of length polymorphisms that influence enzyme activity. This makes it a challenge to capture the full gene using only tag SNPs. In contrast, the CYP2C19 gene can be captured using seven tag SNPs. These SNPs influence CYP2C19 enzyme activity and have been associated with pharmacological effects of CYP2C19-dependent antidepressants, like citalopram (Mrazek et al., 2011).

We hypothesized that one or more variants in the ABCB1 and CYP2C19 genes may modulate the number of side effects while on antidepressants that are substrates for the enzymes for which these genes code (see Figure 1). Our aim was to cover both genes with tag SNPs and look for an association between the number of side effects and the selected tag SNPs in a Dutch cohort of antidepressant users. For this purpose we selected 31 tag SNPs in ABCB1 and 7 SNPs in CYP2C19, consisting of 3 SNPs from literature and 4 additional tag SNPs. As complex traits, such as side effects resulting from antidepressant use, are caused by multiple genes, where interactions between genes are likely to contribute. Therefore, an additional interaction analysis of the genotyped ABCB1 and CYP2C19 tag SNPs was also included.

Vice versa, mutations in ABCB1 and CYP2C19 genes, that decrease PGP- and 2C19-activity, could increase AD concentrations and therefore increase the number of side effects.
A common polymorphism in ABCB1 is associated with side effects of PgP-dependent antidepressants

Figure 1: Hypothesizing that low antidepressant (AD) concentrations in the brain and in the body cause less side effects, mutations in ABCB1 and CYP2C19 genes could increase PGP- and 2C19-activity and decrease AD concentrations and therefore decrease the number of side effects.

MATERIALS AND METHODS

Sample
Data used in the present study originate from the NESDA, an ongoing cohort study conducted among 2981 adults, aged between 18 and 65 years at the baseline assessment. The NESDA sample consists of 652 persons without depression or anxiety disorders and 2329 with a (remitted or current) diagnosis of a depressive and/or anxiety disorder. To represent various stages of psychopathology, depressed or anxious subjects were recruited at three different regions in the Netherlands and at different settings, that is, the general population (n=564), general practices (n=1610) and mental health care organizations (n=807). Community-based subjects were previously identified in two earlier population-based studies. Primary care subjects were identified through a two-stage screening procedure, conducted among patients of 65 general practitioners. Mental health care patients were
recruited when newly admitted to 1 of the 17 participating mental health care organizations. Subjects with clinically overt psychotic disorder, obsessive compulsive disorder, bipolar disorder or severe addiction disorder were excluded. Further details of the NESDA study have been described elsewhere (Penninx et al., 2008).

The NESDA study was approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Center, Amsterdam, an Institutional Review Board certified by the US Office of Human Research Protections (IRB number IRB-2991 under Federal-wide Assurance-3703; IRB/institute codes, NESDA 03-183). All subjects provided written informed consent.

All NESDA subjects who reported antidepressant use at the first (1 year) and/or second (2 year) follow-up, were included in the study sample \( n=907 \). Subjects using two antidepressants at the same time \( n=61 \) were excluded in order to avoid attribution of side effects to the wrong antidepressant. This sample of 846 antidepressant users has been described elsewhere (Bet et al., 2013). Each subject’s first assessment of side effects of single antidepressant use was included in this analysis. Because of incomplete fine mapping of the genotype 38 subjects were excluded. Subjects concurrently using PGP-inhibitors (ciclosporine, erythromycin, itraconazole, ketoconazole, quinidine, ritonavir and verapamil) or PGP-inducers (rifampicine and St. John’s worth) (FDA, 2013) were excluded from the analysis, rendering a final sample of 789 antidepressant users.

**Medication use**

Medication use including name, dose and daily amount of both psychotropic and non-psychotropic medication, was assessed using each subjects’ medication boxes and bottles. Medication was classified using the World Health Organization Anatomical Therapeutic Chemical system. For all antidepressants a derived daily dose was calculated by dividing the subject’s daily dose used by the defined daily dose recommended by the World Health Organization (WHO, 2013).

Subjects were categorized into two groups based on the use of antidepressants known to be PGP substrates (PGP-dependent: citalopram, fluvoxamine, paroxetine, sertraline and venlafaxine) or antidepressants either without a PGP-interaction or without sufficient evidence supporting a PGP interaction (non-PGP dependent: amitriptyline, clomipramine, duloxetine, escitalopram, fluoxetine, imipramine, maprotiline, mianserin, mirtazapine, nortriptyline, St. John’s Worth, tranylcypromine and trazodone) (De Klerk et al., 2013; O’Leary et al., 2014). To examine the associations with CYP2C19 variants, the sample was categorized in users of CYP2C19-dependent (sertraline, citalopram, escitalopram, amitriptyline and clomipramine) (Kirchheiner et al., 2004) and non-CYP2C19-dependent antidepressants (all other antidepressants).
A common polymorphism in ABCB1 is associated with side effects of P-gp-dependent antidepressants

Side effects
Side effects of psychotropic drugs that were used on a daily basis were collected with the 12-question self-report Antidepressant Side Effect Checklist. The 12-question self-report Antidepressant Side Effect Checklist specifically aims to identify patient-perceived side effects related to antidepressant therapy: insomnia, sleepiness during the day, restlessness, muscle spasms/twitching, dry mouth, profuse sweating, sexual disorders, nausea, constipation, diarrhea, weight gain and dizziness. Although no specific time period in which side effects should have occurred was defined, we expected to identify current side effects or side effects that occurred recently. Detailed data on these side effects have been published elsewhere. Side effects were categorized in three groups to explore potential differences in SNP associations: serotonergic (insomnia, restlessness, muscle spasms/twitching, profuse sweating, sexual disorders, nausea and diarrhea), cholinergic (dry mouth and constipation) and histaminergic (sleepiness during the day and weight gain). Side-effect data were obtained by a written questionnaire at the 1-year follow-up measurement of NESDA (n=579) and by an interview at the 2-year follow-up (n=210) between 2005 and 2009. Overall, 76% of the subjects reported side effects on current use, whereas others were not currently taking this antidepressant anymore.

Genotyping procedure
ABCB1 tag SNPs were selected using Tagger software, tagging the gene at minor allele frequency of 0.20 and $r^2=0.8$ (De Bakker et al., 2005). A 5-kb margin around the gene was included to capture possible regulatory regions, leading to a total of 31 tag SNPs. For CYP2C19, seven tag SNPs were selected, that is, the *2, *3 and *17 variants and four additional SNPs. These variants are known to alter CYP2C19 enzyme activity (http://www.cypalleles.ki.se/cyp2c19.htm). DNA sample collection and DNA isolation have been extensively described elsewhere (Penninx et al., 2008). All genotyping was performed using the Taqman Open Array system (Life Technologies, Carlsbad, CA, USA) in accordance with the protocol of the manufacturer (version: 7/2010). Genotypes for ~200 samples were analyzed simultaneously using Taqman Genotyper Software v 1.0.1 (Life Technologies, Carlsbad, CA, USA). The number of 200 samples was set by optimizing for clear clustering, without getting a bias due to too few data points.

The results of each individual analysis were exported as a text file. Text files for all analyses were combined using a home-made script written in Perl. With this script sample IDs, rs-numbers and genotypes were extracted and using an additional script these data were merged into a pedigree file. In addition to the quality control of the arrays and assay clustering, SNPs were limited to a 5% of missing data and a Hardy–Weinberg equilibrium $P$-value of 0.0005.
Imputation
Beagle Software was used for imputation of genotypes that could not be determined using the Taqman Open Array system (±5%), owing to poor clustering or poor amplification (Browning & Browning, 2009). 1000 genomes 2010-06 release Central EUrope data was used as a reference. As the study only concerned the \textit{ABCB1} gene and its regulatory regions, reference data for the entire chromosome were not used. However, to leave the underlying linkage disequilibrium-structure intact, a margin of 100 kb around the gene was applied. P-SEQ software was used to extract the gene ±100 kb from the full chromosome data of the 1000 genomes project on the Lisa system cluster (www.sara.nl/systems/lisa). For all samples imputation was performed with 100 Markov chain iterations (Purcell et al., 2007).

Statistical analysis
Statistical analyses were performed using IBM SPSS Statistics 21 (IBM Corporation, Armonk, NY, USA). Poisson regression on allele dosage was performed per SNP with number of side effects (overall, serotonergic, cholinergic and histaminergic) as the dependent variable and including the covariates gender, age, treatment dosage and treatment duration. Results were corrected for multiple testing using the false-discovery rate (FDR) and were deemed significant for $q$-values<0.05 (Chen et al., 2010). For \textit{ABCB1} SNPs that showed statistically significant associations with side effects, interaction analyses with all \textit{CYP2C19} SNPs were performed using the same method, with the addition of an interaction term in the model.

RESULTS

The demographic characteristics of all 789 antidepressant users are listed in Table 1. About two-thirds of the study population was female. PGP-dependent antidepressant users were younger (median 42 vs 46 years, $P<0.001$), used a higher dosage ($P<0.001$), had a longer duration of antidepressant use ($P=0.025$) and experienced more serotonergic side effects ($P=0.009$) as compared with non-PGP-dependent antidepressant users. The number of side effects did not differ between the two groups. Overviews of the numbers of subjects with overall side effects (Figure 2) and specific serotonergic, cholinergic and histaminergic side effects are provided (Supplementary Figures 1).
A common polymorphism in ABCB1 is associated with side effects of Pgp-dependent antidepressants

Figure 2: Overview of antidepressant users (n=789) and the number of side effects
### Table 1: Demographic characteristics of antidepressant users

Abbreviations: DDD, defined daily dose; IQR, inter quartile range. *Dichotomous variables were tested with the Chi-square test and non-normally distributed variables were tested with the Mann–Whitney U-test between the groups of PGP-dependent and the non-PGP-dependent antidepressant users.*

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All antidepressant users (n = 789)</th>
<th>PGP-dependent antidepressant users (n = 557)</th>
<th>Non-PGP dependent antidepressant users (n = 232)</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (female (%))</td>
<td>582 (66.9)</td>
<td>371 (66.6)</td>
<td>157 (67.7)</td>
<td>0.084</td>
</tr>
<tr>
<td>Age (mean, (s.d.) in years)</td>
<td>43 (12)</td>
<td>42 (12)</td>
<td>46 (11)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Antidepressant medication (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Citalopram</td>
<td>127 (16.1)</td>
<td>127 (22.8)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Fluvoxamine</td>
<td>45 (5.7)</td>
<td>45 (8.1)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Paroxetine</td>
<td>211 (26.7)</td>
<td>211 (37.9)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Sertraline</td>
<td>55 (7.0)</td>
<td>55 (9.9)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Venlafaxine</td>
<td>119 (15.1)</td>
<td>119 (21.4)</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Amitriptyline</td>
<td>37 (4.7)</td>
<td>-</td>
<td>37 (15.9)</td>
<td></td>
</tr>
<tr>
<td>Clomipramine</td>
<td>28 (3.5)</td>
<td>-</td>
<td>28 (12.1)</td>
<td></td>
</tr>
<tr>
<td>Duloxetine</td>
<td>6 (0.8)</td>
<td>-</td>
<td>6 (2.6)</td>
<td></td>
</tr>
<tr>
<td>Escitalopram</td>
<td>16 (2.0)</td>
<td>-</td>
<td>16 (6.9)</td>
<td></td>
</tr>
<tr>
<td>Fluoxetine</td>
<td>62 (7.9)</td>
<td>-</td>
<td>62 (26.7)</td>
<td></td>
</tr>
<tr>
<td>Imipramine</td>
<td>4 (0.5)</td>
<td>-</td>
<td>4 (1.7)</td>
<td></td>
</tr>
<tr>
<td>Maprotiline</td>
<td>1 (0.1)</td>
<td>-</td>
<td>1 (0.4)</td>
<td></td>
</tr>
<tr>
<td>Mianserin</td>
<td>2 (0.3)</td>
<td>-</td>
<td>2 (0.9)</td>
<td></td>
</tr>
<tr>
<td>Mirtazapine</td>
<td>42 (5.3)</td>
<td>-</td>
<td>42 (18.1)</td>
<td></td>
</tr>
<tr>
<td>Nortriptyline</td>
<td>11 (1.4)</td>
<td>-</td>
<td>11 (4.7)</td>
<td></td>
</tr>
<tr>
<td>St. John’s worth</td>
<td>16 (2.0)</td>
<td>-</td>
<td>16 (6.9)</td>
<td></td>
</tr>
</tbody>
</table>
A common polymorphism in ABCB1 is associated with side effects of PgP-dependent antidepressants

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All antidepressant users (n = 789)</th>
<th>PGP-dependent antidepressant users (n = 557)</th>
<th>Non-PGP dependent antidepressant users (n = 232)</th>
<th>P-valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tranylcypromine</td>
<td>2 (0.3)</td>
<td>-</td>
<td>2 (0.9)</td>
<td></td>
</tr>
<tr>
<td>Trazodone</td>
<td>5 (0.6)</td>
<td>-</td>
<td>5 (2.2)</td>
<td></td>
</tr>
<tr>
<td>Dosage of antidepressants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(median, (IQR 25–75%) in DDD)</td>
<td>1.0 (1.0–1.5)</td>
<td>1.0 (1.0–1.5)</td>
<td>1.0 (0.7–1.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Duration of antidepressant use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(median, (IQR 25–75%) in years)</td>
<td>1.0 (0.5–2.1)</td>
<td>1.0 (0.5–2.5)</td>
<td>1.0 (0.3–2.0)</td>
<td>0.025</td>
</tr>
</tbody>
</table>

Number of side effects:

| Overall                          | 1 (0-3)                          | 1 (0-3)                                   | 1 (0-3)                                      | 0.125    |
| Serotonergic                     | 0 (0-1)                          | 0 (0-2)                                   | 0 (0-1)                                      | 0.009    |
| Cholinergic                      | 0 (0-1)                          | 0 (0-1)                                   | 0 (0-1)                                      | 0.155    |
| Histaminergic                    | 0 (0-1)                          | 0 (0-1)                                   | 0 (0-1)                                      | 0.289    |
Although various ABCB1 SNPs (rs10808072, rs10276036, rs2188526, rs13233308 and rs2032588) were associated with the occurrence of side effects in the total sample (Supplementary Table 1), none of the associations remained significant after FDR correction. However, in the group of PGP-dependent antidepressant users, rs2032588 SNP A-alleles were nominally associated with a lower number of side effects ($B=-0.37$, $P=6 \times 10^{-5}$) (Figure 3). This association remained statistically significant after adjusting for gender, age, dosage and duration of use, ($B=-0.44$ with Wald 95% confidence interval $-0.54$ to $-0.18$, $P=1.22 \times 10^{-4}$, after FDR: $q=4.6 \times 10^{-3}$). The rs2032588 SNP showed a trend towards association with the number of serotonergic side effects (after FDR: $q=0.057$). No associations of the rs2032588 SNP with histaminergic and cholinergic side effects were detected in the group of PGP-dependent antidepressant users. In the group of non-PGP-dependent antidepressant users, no rs2032588 SNP associations were found (Table 2). The interaction effect of the rs2032588 SNP and use of a PGP-dependent antidepressant in predicting the number of side effects was examined by introducing an interaction term in a Poisson regression model already including the individual terms; the rs2032588 SNP-by-PGP-antidepressant use was significant ($P=0.008$), indicating that the relationship between genotype and the number of side effects differed based on the use of PGP-dependent antidepressants.

* For easy comparison of the number of subjects without (n=465) and with A-allele(s) (n=92), the ratio of the scales left vs right is 5 vs 1.

**Figure 3:** Overview of PGP-dependent antidepressant users (n=557) and the number of side effects according to rs2032588 allele status
A common polymorphism in ABCB1 is associated with side effects of Pgp-dependent antidepressants

Table 2: P-values and q-values of associations between the rs2032588 SNP and antidepressant side effects

<table>
<thead>
<tr>
<th></th>
<th>Total sample (n=789)</th>
<th>PGP-dependent AD users (n=557)</th>
<th>Non-PGP dependent AD users (n=232)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unadjusted</td>
<td>Adjusted</td>
<td>Unadjusted</td>
</tr>
<tr>
<td><strong>All side effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>2.0 × 10⁻³</td>
<td>4.6 × 10⁻³</td>
<td>6.0 × 10⁻⁵</td>
</tr>
<tr>
<td>Beta-estimates</td>
<td>0.23</td>
<td>0.59</td>
<td>-0.37</td>
</tr>
<tr>
<td>q-value after FDR</td>
<td>0.08</td>
<td>0.17</td>
<td>2.3 × 10⁻³b</td>
</tr>
<tr>
<td><strong>Serotonergic side effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>4.8 × 10⁻³</td>
<td>8.6 × 10⁻³</td>
<td>1.1 × 10⁻³</td>
</tr>
<tr>
<td>Beta-estimates</td>
<td>-0.22</td>
<td>-0.23</td>
<td>-0.46</td>
</tr>
<tr>
<td>q-value after FDR</td>
<td>0.18</td>
<td>0.33</td>
<td>0.04b</td>
</tr>
<tr>
<td><strong>Cholinergic side effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>0.38</td>
<td>0.42</td>
<td>0.05</td>
</tr>
<tr>
<td>Beta-estimates</td>
<td>0.16</td>
<td>0.14</td>
<td>0.47</td>
</tr>
<tr>
<td><strong>Histaminergic side effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>0.06</td>
<td>0.07</td>
<td>0.04</td>
</tr>
<tr>
<td>Beta-estimates</td>
<td>0.32</td>
<td>0.30</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Abbreviations: AD, antidepressant; FDR, false-discovery rate; PGP, permeability-glycoprotein; SNP, single-nucleotide polymorphism. *Adjusted for gender, age, dosage and duration of use. **Statistically significant after correcting for multiple testing by means of false-discovery rate q<0.05.

All other ABCB1 and CYP2C19 SNP associations with the number of side effects (overall, serotonergic, cholinergic and histaminergic) were not statistically significant after adjusting for gender, age, dosage and duration of use, and after FDR correction.

Testing the interaction between ABCB1 and CYP2C19 for the rs2032588 SNP, a P-value of 0.042 was found in the overall sample. This relationship, however, was not significant after correcting for multiple testing, nor did it exist in the PGP-dependent antidepressant users group nor in association with the number of serotonergic side effects.

**DISCUSSION**

In the present study the A-allele of the rs2032588 SNP in the ABCB1 gene was associated with a lower number of side effects in a large group of PGP-dependent antidepressant users. As the association was not found in the group of non-PGP-de-
A common polymorphism in ABCB1 is associated with side effects of PGP-dependent antidepressants. Other SNP associations as well as an interaction analysis between the rs2032588 SNP and the CYP2C19 SNPs were not statistically significant after adjusting for covariates and FDR correction.

The present study is the first pharmacogenetic study on antidepressant side effects that covers the whole ABCB1 gene. Although several ABCB1 SNPs showed associations before adjusting for co-variates and multiple testing, only the association of the rs2302588 SNP with a lower number of side effects remained statistically significant. As yet little is known of the rs2302588 SNP in relation to PGP function and alterations in the pharmacological activity of PGP-dependent drugs. To date, with respect to antidepressant activity only two candidate gene studies examined the whole ABCB1 gene with more than 30 SNPs including the rs2032588 SNP. However, exploring the association between ABCB1 SNPs and treatment response in Mexican Americans or studying remission, plasma and brain concentrations of antidepressants in Caucasians, a significant association between effects and the rs2032588 SNP was not found (Dong et al., 2009; Uhr et al., 2008).

Until now the present study is the largest ABCB1 gene association study on antidepressant side effects. Previously, the NESDA cohort has also been the subject of a candidate gene study which included the ABCB1 gene (De Klerk et al., 2013). This study led to associations of antidepressant side effects with two SNPs, rs2032583 and rs2235040, located within 19 kb from the rs2032588 SNP. However, although a new association was found for the rs2032588 SNP, in the present study with its larger sample size, the results of De Klerk et al. could not be confirmed. There are several possibilities that may explain these different results. First, about 80% of our side effect phenotypes and about 30% of our tag SNP genotypes were not included in the analysis of De Klerk et al. Second, in the present study the ABCB1 gene was tagged to reach 100% coverage at $r^2=0.8$ and minor allele frequency of 0.2. In contrast, De Klerk et al. selected only six SNPs that did not cover the entire gene. Third, only 5% of the genotype data of the present study was imputed because of missing data, whereas de Klerk et al. imputed all genotype data. Although the quality of imputation algorithms has greatly improved over the years, small differences in underlying LD between the reference set and the imputed set already may cause a difference between genotyping and imputation (Pardo et al., 2009).

Although de Klerk et al. obtained significant associations for the rs2032583 and rs2235040 SNPs, they did not include the rs2032588 SNP in their study. It is possible that none of these SNPs is a true causal variant. All three SNPs are located in the intronic region of ABCB1 gene and therefore do not exert a clear effect on the
A common polymorphism in \textit{ABCB1} is associated with side effects of PgP-dependent antidepressants

PGP-transporter configuration, as a non-synonymous coding SNP would. However, it is stated in the ‘Fundamental Theorem of the HapMap’, that all tested SNPs are expected to reflect the true association of the unknown causal variant proportional to their LD with it (Terwilliger et al., 2006). This might suggest that there is a variant located in the 19-kb area between the rs2032583 and rs2032588 SNPs that is causal and that in these two studies the significant SNPs only reflect the association of a variant, proportional to their LD with this unknown variant.

When looking at the gene structure and the functional domains of the \textit{ABCB1} gene, the majority of SNPs that have been associated with pharmacological changes such as efficacy and side effects, are located in the intronic regions of the gene. This is in concordance with a large number of genetic association studies, in which intronic SNPs give the best association results. Although these SNPs may reflect an association with a functional SNP in one of the neighboring exons, there are numerous other mechanisms that could underlie these associations. Intronic SNPs may have an effect on gene splicing. To date, 10 different splice variants of the \textit{ABCB1} gene have been validated, each with their own pattern of expression (www.ensembl.org). However, more insight into the effect of associated intronic SNPs can only be obtained by functional testing.

Although there have been experiments exploring the complete \textit{ABCB1} gene in knockout mice, the effects of antidepressants have not been studied extensively, nor has there been focus on the region of interest indicated by the results of the present study (Uhr et al., 2008; Karlsson et al., 2013; Tang et al., 2002). By means of functional experiments, the importance of this region may be corroborated on a molecular and/or cellular level, thereby leading to a better understanding of PGP-transporter function in individuals with a high number of side effects. Ideally, functional experiments using the 19-kb area between the rs2032583 and rs2032588 SNPs should be performed.

As three \textit{ABCB1} SNPs (rs1128503 (1236C>T), rs2032582 (2677G>T/A) and rs1045642 (3435C>T)) have previously been associated with an altered PGP expression level and are therefore likely to have an altered PGP-transporter function, these SNPs have been examined extensively in earlier studies in relation with the occurrence of antidepressant side effects. However, these findings have not been confirmed by the results of later studies. In a sample of 160 patients with depression Roberts \textit{et al.} found a relationship between the rs1045642 SNP (3435C>T) and orthostatic hypotension (Roberts et al., 2002). Jenssen \textit{et al.}, however, failed to replicate his finding (Jensen et al., 2012). Recently, two small studies on sexual dysfunction reported associations between the rs1128503 SNP (1236C>T) and sexual dysfunction in a sample of 57 depressed women and between the rs2032582 SNP
(2677G>T/A) and difficulties with orgasm and lubrication in 55 women with an anxiety disorder or bulimia (Bly et al., 2013; Zourkova et al., 2013). Weight gain in 117 depressed patients, suicidal ideation in 131 depressed patients and amitriptyline side effects in 50 patients with a mild depression were only associated once with one of these three SNPs (Menu et al., 2010; Perroud et al., 2011; Laika et al., 2006). In the present and the largest study to date as well as in the study by de Klerk et al., associations between these three SNPs and the occurrence of antidepressant side effects have not been found. This might be due to the use of the 12-question self-report Antidepressant Side Effect Checklist questionnaire that asks for 12 of the most commonly reported side effects by patients, whereas other studies focused on side effects such as suicidal ideation and postural hypotension that are not covered by this questionnaire (Perroud et al., 2011; Roberts et al., 2002). Another possible explanation is that the present study only concerned the occurrence of long-term side effects and therefore failed to detect associations related to short-term side effects.

Major strengths of the present study are the large sample size of antidepressant users in a naturalistic setting, full coverage of the \textit{ABCB1} gene and the \textit{CYP2C19} gene with only 5% missing data that needed imputation. The variety of antidepressants used in this study and the classification based on PGP-interaction are considered as a limitation, because studies on PGP-interaction of antidepressants show mixed results (O’Leary et al., 2014). Moreover, endophenotype data on PGP-transporter activity or antidepressant concentrations were not available. Further research on the nature of the impact of the rs2032588 SNP on PGP-function and clinical outcome, as well as confirmation in another sample is needed.

In the sample of the present study, the A-allele of the rs2032588 SNP in \textit{ABCB1} was associated with a lower number of side effects in PGP-dependent antidepressants users. Although the nature of the impact of the rs2032588 SNP on PGP-transporter function is unknown, these results provide further evidence for an involvement of the \textit{ABCB1} genotype in the clinical pharmacology of PGP-dependent antidepressants. Elucidating the mechanism of this relationship is important for the successful use of personalized genotype based treatments in clinical practice and the development of antidepressant drugs.

\textbf{CONFLICT OF INTEREST}

The authors declare no conflict of interest.
A common polymorphism in ABCB1 is associated with side effects of Pgp-dependent antidepressants

REFERENCES


Dong C, Wong ML, Licinio J (2009) Sequence variations of ABCB1, SLC6A2, SLC6A3, SLC6A4, CREB1, CRHR1 and NTRK2: association with major depression and antidepressant response in Mexican-Americans; Mol Psychiatry; 14: 1105–1118

FDA (2013) US Food and Drug Administration, Drug development and Drug interactions: Table of Substrates, Inhibitors and Inducers

A common polymorphism in ABCB1 is associated with side effects of PgP-dependent antidepressants.

Chapter 6


Horstmann S, Binder EB (2009) Pharmacogenomics of antidepressant drugs; Pharmacol Ther; 124: 57–73


A common polymorphism in ABCB1 is associated with side effects of PgP-dependent antidepressants


Mas S, Gasso P, Alvarez S, Parellada E, Bernardo M, Lafuente A (2012) Intuitive pharmacogenetics: spontaneous risperidone dosage is related to CYP2D6, CYP3A5 and ABCB1 genotypes; Pharmacogenomics J; 12: 255–259


Mrazek DA, Biernacka JM, O’Kane DJ, Black JL, Cunningham JM, Drews MS, Snyder KA, Stevens SR, Rush AJ, Weinshilboum RM (2011) CYP2C19 variation and citalopram response; Pharmacogenet Genomics; 21: 1–9


O’Leary OF, O’Brien FE, O’Connor RM, Cryan JF (2014) Drugs genes and the blues: pharmacogenetics of the antidepressant response from mouse to man; Pharmacol Biochem Behav; 123: 55–76


Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis; Am J Hum Genet; 81: 559–575


Sarginson JE, Lazzeroni LC, Ryan HS, Ershoff BD, Schatzberg AF, Murphy GM Jr. (2010) ABCB1 (MDR1) polymorphisms and antidepressant response in geriatric depression; Pharmacogenet Genomics; 20: 467–475
Singh AB, Bousman CA, Ng CH, Byron K, Berk M (2012) ABCB1 polymorphism predicts escitalopram dose needed for remission in major depression; Transl Psychiatry; 2: e198
WHO (2013) WHO Collaborating Centre for Drug Statistics Methodology; Anatomical Therapeutic Chemical (ATC) Classification; http:/ /www.whocc.no/atcddd/

ACKNOWLEDGEMENTS

The infrastructure for the NESDA study (www.nesda.nl) is funded through the Geestkracht program of the Netherlands Organisation for Health Research and Development (Zon-Mw, grant number 10-000-1002) and is supported by participating universities and mental health care organizations VU University Medical Center, GGZ inGeest, Arkin, Leiden University Medical Center, GGZ Rivierduinen, University Medical Center Groningen, Lentis, GGZ Friesland, GGZ Drenthe, Scientific Institute for Quality of Healthcare (IQ healthcare), the Netherlands Institute for Health Services Research (NIVEL) and the Netherlands Institute of Mental Health and Addiction (Trimbos). The funding agents were not involved in the design or conduct of the study or in the collection, management, analysis, or interpretation of the data, or preparation, review, or approval of the manuscript.
A common polymorphism in ABCB1 is associated with side effects of PgP-dependent antidepressants.

**Supplementary figure 1:** Overview of antidepressant users and the number of serotonergic side effects (n=789)

**Supplementary figure 2:** Overview of antidepressant users and the number of cholinergic side effects (n=789)
A common polymorphism in ABCB1 is associated with side effects of P-gp-dependent antidepressants.

Supplementary figure 3: Overview of antidepressant users and the number of histaminergic side effects (n=789)
A common polymorphism in ABCB1 is associated with side effects of Pgp-dependent antidepressants.
CHAPTER 7
General Discussion and Conclusions
GENERAL DISCUSSION

In order to improve diagnosis and treatment for MDD, it is important to create a better understanding of the underlying mechanisms of the disease. Although genetic research has shown that MDD has an estimated heritability of 40%, its molecular etiology is still largely unknown. In this thesis the overall aim was to contribute to finding causal genetic variants that may offer clarification of the etiology of MDD. For this search two different approaches were used: first the results of the GAIN-MDD GWAS were fine-mapped in order to investigate whether the non-synonymous coding SNP rs2522833 is indeed causal. Secondly, a candidate gene and pathway approach was used to find genetic variants associated with biological measures for MDD and variants associated with antidepressant side effects.

SUMMARY

MDD is one of the leading causes of disability world-wide. As a common complex disorder, it has both environmental and multiple genetic causes. Chapter 2 provides a review of genetic approaches to find associated variants for MDD, including linkage, candidate genes and GWAS. The GWAS that was used for the basis of the fine-mapping studies was the GAIN-MDD GWAS by Sullivan et al. In this GWAS, the SNP rs2522833 in the PCLO gene became nominally significant after post-hoc analysis with a similarly ascertained Australian cohort. In addition, many other top signals also mapped back to the region spanning PCLO (Sullivan et al., 2009). This gene codes for the protein Piccolo, that resides in the cell's presynaptic active zone. The protein is suspected to have a function in the vesicle transportation machinery. Taking into account the hypothesis that MDD may be caused by an imbalance in neurotransmitters, PCLO with its role in vesicle transport seems a plausible candidate gene. However, the chip that was used for genotyping in this GWAS, was not designed in a gene-centric manner, causing less than optimal coverage for many genes. Therefore, in Chapter 3 a fine-mapping study was performed. Firstly, because we wanted to find additional evidence that rs2522833 was truly the causal SNP in the GAIN-MDD cohort. Due to the lack of coverage in the original GWAS, the signal could also be caused by a neighboring SNP in high LD with rs2522833. Secondly, there were various sub-threshold signals in other genes that were not covered optimally. To maximize the genotypic information from these genes, tag SNPs were selected and coverage was increased to 100% with an MAF of 10% and an \( r^2 \) of 0.9. Again, the lowest P-values were found in the immediate vicinity of rs2522833, at rs2715147 and rs2715148 (P=1.2E-6). However, in single SNP association analysis, no genome-wide significance was reached. Additionally a joint re-analysis was performed of all genotyped SNPs in PCLO. When assuming that
rs2522833 was the causal variant, the slope of the regression line in this analysis was the steepest. This supports the hypothesis that the causal variant is either rs2522833 or a variant in high LD with it. In a subsequent step, a haplotype analysis was performed. One haplotype yielded a lower P-value than the single SNP analysis (P=9.9E-7). Even though this result did not reach genome-wide significance, it suggests that a yet unknown variant in this area may be causal for the top signal in the GAIN-MDD GWAS.

To find further evidence that rs2522833 or a yet unknown variant is causal in the GAIN-MDD cohort, a second strategy was used in Chapter 4. The results from the haplotype analysis in chapter 3 suggested that an unknown variant within the PCLO gene may be responsible for the top signal in the GAIN-MDD GWAS. With the intention of finding this unknown variant, the PCLO gene was sequenced, using next generation sequencing methods. In addition, two candidate genes for MDD from literature were sequenced: GRM7 and SLC6A4. GRM7 codes for a metabotropic glutamate receptor. Functional knock-out of this receptor in a mouse model shows an antidepressant effect (Cryan et al., 2003). In addition, one of the top signals of a meta-analysis of three GWAS studies for MDD came from GRM7 (Shyn et al., 2011). The SLC6A4 gene codes for the serotonin transporter. This transporter plays a crucial role in the availability of serotonin in the synaptic cleft: it is located on the presynaptic membrane and transports serotonin back into the cell, recycling serotonin to regulate the availability for the postsynaptic membrane. With this sequencing study, 961 new SNPs were discovered. 71 of these newly identified SNPs in addition to 185 tag SNPs were used for further fine-mapping, as the new SNPs alone did not cover the genes 100%. This approach of next generation sequencing and fine-mapping did not lead to a lower P-value than the original GAIN-MDD GWAS, with rs2715147 showing the lowest P-value at P=1.5E-6. These results show that the previously unknown variants detected in this study were not more associated with MDD than rs2522833. Although rare variants were not taken into account in this study, chapter 4 provides additional evidence for the hypothesis that rs2522833 is the causal variant in the GAIN-MDD cohort.

In chapters 3 and 4 a fine-mapping studies were performed based on the GAIN-MDD GWAS. A GWAS does not use an a priori hypothesis when searching for an associated variant. In chapters 5 and 6 a different approach is used: the candidate gene approach. In this approach, genes are selected based on previously known functional or biological connections to a trait. In Chapter 5, genes from three pathways were selected for genotyping in the NESDA cohort: the HPA axis, the HPT axis and vitamin D metabolism. In MDD, the HPA axis is dysregulated, leading to higher levels of its end product cortisol. In the HPT axis deviations are found as well in MDD patients, while patients with hypothyroidism often exhibit characteristics
of depression. Lower vitamin D levels in blood are associated with depression in several patient groups, including the NESDA cohort. Of these three pathways, biological measurements are available for the NESDA cohort. Genes were tagged and association analyses were performed on SNP level, gene level and pathway level. No significant associations were found in these analyses, neither in the uncorrected data, nor in the data corrected for the biological measurements. However, one SNP in the 3’ UTR of \textit{NR3C1}, rs6198, showed a borderline significant association (P=3.54E-3 OR 1.48) with severe recurrent MDD. Caution is advised however, as the severe recurrent MDD group consisted of only 354 cases, which limits the statistical power of detecting a truly associated variant. Also, even though rs6198 was not genotyped in the GAIN-MDD GWAS, several neighboring SNPs were, and none of them displayed an association with MDD. In addition, in a literature-based analysis of candidate genes for MDD, a haplotype containing rs6198 did not reveal an association (Bosker et al., 2010).

Whereas the previous chapters focused on finding variants associated with the etiology of MDD, \textbf{Chapter 6} addresses the treatment of MDD with antidepressant medication. The study in this chapter was not based on a GWAS, but on the candidate gene approach. With this approach, genes are selected from literature, based on knowledge of the function of the gene. In the treatment of MDD, the occurrence of side effects is an obstacle. In the NESDA cohort, 927 patients that used a single antidepressant, 64% of cases reported on average 2.9 side effects, in which TCA’s were associated with more side effects than SSRI’s. Interestingly, the number of side effects was associated with severity of depression, higher dosage, and the occurrence of more psychiatric diagnoses (Bet et al., 2013). In a study of more than 400 patients treated with SSRI’s, over 55% of patients experienced side effects in the first two weeks of treatment (Hu et al., 2004). Another hindrance in treating MDD is that a substantial part of the patients does not show remission of symptoms with antidepressant use. In a study of approximately 2900 patients, the response rate to the commonly used SSRI citalopram was 47%, with remission in 28-33% of patients (Trivedi et al., 2006). When treating patients with antidepressant medication, a balance has to be found to reach therapeutic levels with a minimum number of side effects. It was hypothesized that variants in the liver enzyme gene \textit{CYP2C19} and in the drug transporter gene \textit{ABCB1} may impact the number of side effects by increasing or decreasing the availability of antidepressants. An association was found between a common variant in \textit{ABCB1} and the number of side-effects in PgP-dependent medication users. The A-allele of rs2032588 was associated with a lower number of side effects after correcting for age, gender, duration of therapy and dosage, (B=-0.44, P=1.22*10^{-4}) and remained significant after control for false discovery rate (B=-0.44, q=4.6E-3). This association was not found in non-PgP-dependent antidepressant users. This study is one of the first to demonstrate an as-
sociation within a naturalistic setting with a considerable sample size, but would ideally have to be replicated in a comparable cohort.

**REFLECTION ON APPROACH AND METHODOLOGY**

**The common disease common variant hypothesis**

Work in this thesis has been performed by the genotyping of common variants. It is hypothesized that if the disease under investigation is common, the variants responsible for it are common as well. Consequently, all these variants have a mild effect, as a larger effect size would increase the disease burden for the individual. Variants with a large deleterious effect should therefore, based on selection, diminish within a population and not be common. However, the common disease common variant hypothesis did not lead to clear results in most of the GWAS that were performed. Although numerous new loci have been found for complex traits and common disease, such as for body height (Wood et al., 2014), schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) and Parkinson’s disease (Nalls et al., 2014), the total effect of these variants only explains a small part of the heritability. In MDD, various GWAS have been performed, but each with very moderate results (Table 1). Regarding these GWAS, two things are remarkable: first, the top signals from each study are all in different genes. Second, these signals by far do not explain the so-called missing heritability of MDD.

### Table 1: An overview of the GWAS for MDD

<table>
<thead>
<tr>
<th>GWAS</th>
<th>Year of publication</th>
<th>Top signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sullivan et al</td>
<td>2009</td>
<td>PCL0 (P=6.4E-8)</td>
</tr>
<tr>
<td>Rietschel et al</td>
<td>2010</td>
<td>CPM (P=3.24E-6), HOMER1 (P=1.48E-6)</td>
</tr>
<tr>
<td>Lewis et al</td>
<td>2010</td>
<td>BICC1 P=1.3E-7</td>
</tr>
<tr>
<td>Shyn et al</td>
<td>2011</td>
<td>ATP6V1B2 (P=6.78E-7), SP4 (P=7.68E-7), GRM7 (P=1.11E-7)</td>
</tr>
<tr>
<td>Shi et al</td>
<td>2011</td>
<td>rs17077540 (P=1.83E-7)</td>
</tr>
<tr>
<td>Kohli et al</td>
<td>2011</td>
<td>SLC6A15 (P=5.53E-8)</td>
</tr>
<tr>
<td>Wray et al</td>
<td>2012</td>
<td>CACNA1C (P=0.020)</td>
</tr>
<tr>
<td>Green et al</td>
<td>2013</td>
<td>SYNE1 (P=2.9E-8)</td>
</tr>
</tbody>
</table>

There are several explanations for the gaps in heritability that the GWAS approach and subsequent fine-mapping in chapters three and four did not resolve by their use of common variation. First of all, one plausible explanation is that the common disease common variant hypothesis holds true for only a subset of the variants. Opposed to the common disease common variant hypothesis stands the common
disease rare variant hypothesis. It hypothesizes that uncommon variants (MAF < 5%) and rare variants (MAF < 1%) may be causal to common disorders, with multiple risk alleles that are each individually rare within the population. By means of next generation sequencing, these risk alleles have become detectable. However, in order to detect these variants, the sequenced cohort size should be considerable to guarantee the presence of the variant. By definition, a single rare variant cannot be causal for a common disorder. If the variant is truly rare, then this will also be reflected within the studied population. Assuming that approximately 10 to 15% of the western population will experience depression during their lifetime, this means that a variant would have to be present in 10 to 15% in the population and therefore cannot be rare. Rather, multiple rare variants may contribute to a combined risk for MDD. Even though detection has become easier and more affordable with next generation sequencing, considering the expected small effect sizes, finding an association may still be a challenge. In psychiatry, associations have been found between rare variants and schizophrenia, bipolar disorder and depression (Sebat et al., 2009; Grozeva et al., 2010; Knight et al., 2009; Curceanu et al., 2013). In addition, in schizophrenia there is evidence that areas that have common variants with small effects, also have enrichment for rare mutations of larger effect after sequencing studies (Fromer et al., 2014). In this thesis, sequencing was performed on common variants in control samples, as MDD is a common disorder. Given the before mentioned evidence, it might prove worthwhile to sequence with the aim of detecting rare variants in MDD patients, opposed to the strategy in chapter 4, where the aim was to find common variants and for which controls were used. However, given the fact that some GWAS did yield modest results, common variants cannot be disregarded completely. Individual common variants may have small effect sizes, but their cumulative effects could be larger. In addition, the results of several successful GWAS point to known drug targets and relevant biological pathways (Wray et al., 2014). Results of both approaches, rare and common variants, suggest that a mixed model of the two may lead to further success in unraveling the etiology of common disease.

A second explanation for the missing heritability may be sheer cohort size. The experiments featured in this thesis have been performed on the GAIN-MDD cohort and the NESDA cohort. The GAIN-MDD cohort consists of 3540 individuals and the NESDA cohort consists of 2840 individuals, of which subsets were used in the different chapters in this thesis. When looking at GWAS performed for other common complex traits and disorders, it is remarkable that often cohorts are used consisting of many more individuals. In psychiatric disorders, cohorts have grown substantially over the past couple of years. In 2014, the Schizophrenia Working Group of the Psychiatric GWAS Consortium published a combined effort of 49 GWAS, leading to the discovery of 108 associated genetic loci (Schizophre-
nia Working Group of the Psychiatric GWAS Consortium, 2014), of which 83 were not previously reported in the individual cohorts. Studies of this size indicate that the success of finding common variants for common disorders may lie in the size of the cohort. Since the effect size of common variants is small, a large cohort is required to have sufficient statistical power to detect them. This is also reflected in the first mega-analysis for MDD, in which 18759 individuals were genotyped (Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium, 2013). This effort yielded 15 genome-wide significant signals in a cross-disorder analysis of MDD and bipolar disorder. However, the results for MDD alone did not reach genome-wide significance. As this mega-analysis is the largest one to date, the lack of results in genetic studies in MDD in general, may suggest that statistical power and effect sizes are overestimated in the situation of the GWAS. However, an exome sequencing study on more than 5000 Chinese women with recurrent MDD and more than 5000 controls, identified two variants, one near the SIRT1 gene and one in an intron of LHHP with a genome-wide association, that was subsequently replicated in an independent sample (CONVERGE Consortium, 2015). This study was performed using whole exome sequencing rather than a 'traditional' GWAS, but it increases the cohort to 10 000 individuals of the same gender, descent and same recurrent phenotype, suggesting that a larger cohort with more specific phenotyping may indeed be the key to successfully identifying variants associated with MDD.

In addition, the study of the CONVERGE Consortium addresses another issue common in psychiatric disorders: phenotypic heterogeneity. In psychiatric disorders, phenotypical heterogeneity causes a large obstacle when investigating the etiology of the disease, for instance: MDD can occur in single or multiple episodes, can be moderate to severe, may show comorbid anxiety disorders and may be prone to seasonal fluctuations. In addition, inclusion of individuals greatly relies on self-report and interviews. These methods are highly sensitive to bias introduced by the method of testing, as they report of an individual’s experience of the disorder rather than using objective measurements. To reduce the effect of phenotypic heterogeneity on the probability of finding associated variants, there are several available strategies. As it has been found that recurrent major depression is the subtype that most consistently identifies an increased familial risk (Sullivan et al., 2000), performing further association studies on this subtype may increase association results. The study of Chinese women of the CONVERGE Consortium addresses this issue by selecting only recurrent MDD, rather than a mixture of patients with single episodes and patients with recurrence. Levinson et al, suggest that, in addition to increasing the number of samples for an MDD GWAS, it is valuable to collect information on variables such as gender, age of onset, comorbidities, and course
of the disorder (such as the recurrence of the disorder in the CONVERGE study) in order to later on model heterogeneity (Levinson et al., 2014).

**FUTURE ENDEAVORS IN THE GENETICS OF MDD**

**Aiming for PCLO**

In this thesis supportive evidence is gathered that indeed rs2522833 within the *PCLO* gene is the causal variant in the GAIN-MDD cohort. The *PCLO* gene is interesting as a candidate gene for MDD, because of the location of the PCLO protein at the presynaptic active zone. Several efforts have been made to find a functional connection between the *PCLO* gene and the etiology of MDD. Leal-Ortiz et al. found enhanced synaptic vesicle release in cultured neurons in which Piccolo was acutely knocked down (Leal-Ortiz et al., 2008). Rs2522833 is located in the calcium binding C2A-domain of the gene, which functions as a calcium sensor. Mice overexpressing this domain displayed a depression-like phenotype, with more immobility than their wildtype counterparts in forced swim tests and tail suspension tests (Furukawa-Hibi et al., 2010). However, this research did not particularly focus on the non-synonymous coding SNP. In a knock-in model of the rs2522833 serine to alanine change in mice, excitatory synaptic transmission was increased as well as levels of synaptic Piccolo protein. However, calcium-dependent phospholipid binding, synapse formation in vitro and synaptic accumulation of synaptic vesicles remained unaltered (Giniatullina et al., 2015). In contrast to the model of Furukawa-Hibi et al., in these mice no significant alterations in behavior were found. In addition, neither studies used a sucrose preference test to ascertain anhedonia, which is a key characteristic of MDD.

In the human situation, several studies have focused on connecting *PCLO* variants to endophenotypes, to overcome the heterogeneity in the MDD phenotype and to acquire a more straightforward genetic connection. It was shown that rs2522833 has a functional effect on emotional processing in MDD patients, with increased amygdala activation in risk allele carriers (Woudstra et al., 2012). In addition, *PCLO* risk allele carriers showed lower memory performance and reduced encoding-related hippocampal activation in a functional MRI study (Schott et al., 2014). In addition, dysregulation of the HPA axis is less pronounced in individuals with the AA genotype, with only a marginal change in HPA activity in 4 weeks of antidepressant treatment compared to individuals carrying one or two C alleles (Schumacher et al., 2011).

In spite of these efforts to show a functional connection between the Piccolo protein and the etiology of MDD, replication in other cohorts is still an obstacle. This
was also witnessed in the original GWAS by Sullivan et al., where six replication cohorts were used. An association was only found after post-hoc analysis with the Australian QIMR cohort, but not after analyses with the remaining five cohorts. This is also the case in another replication effort by Hek et al.: the original study in elderly patients did not lead to genome-wide significance, but after a meta-analysis with the GAIN-MDD cohort, the QIMR cohort and a subset of their own cohort with depressive disorders, a P-value of $P=1.93\times10^{-9}$ was found (Hek et al., 2009). Both the studies of Sullivan and Hek attribute this to differences in ascertainment: all three cohorts are population-based. However, as shown in table 1, different GWAS have supportive evidence for different causal variants and causal genes.

In summary, although changes are witnessed on the molecular level when looking at the effect of rs2522833, animal models prove difficult to replicate. In an ideal situation, interactions of PCLO with the many proteins in the presynaptic active zone would have to be mapped further. Such an effort would retrieve further information on the molecular mechanisms that lie behind the altered synaptic vesicle release and, on a larger scale, the effect this has on memory, emotional processing and HPA axis functioning in MDD patients. However, caution is advised, since SNP rs2522833 in PCLO was only found after post-hoc analysis with the QIMR cohort, and replication has been troublesome. In addition, in the mega-analysis of the Psychiatric Genomics Consortium, the largest cohort on MDD thus far, PCLO did not emerge as an associated variant.

MOvING BEyOND SNPS

The work in this thesis has been solely based on SNP genotyping. However, other types of genetic variation may also increase knowledge of MDD. Copy number variations (CNVs) have shown associations with psychiatric disorders. Although CNVs exist in many lengths and frequencies, the more rare CNVs have yielded more results. Both large duplications and microdeletions have shown associations with autism spectrum disorders (ASD) and schizophrenia (Gillberg, 1998; Murphy et al., 1999). In addition, several studies revealed a 5-10% rate of \textit{de novo} CNVs in ASD (Sanders et al., 2011; Pinto et al., 2010). Also in bipolar disorder, results suggest that \textit{de novo} CNVs play a role (Malhotra et al., 2011). In MDD, CNVs have been studied in patients that attempted suicide versus patients that did not attempt suicide, but no association was found (Perlis et al., 2012). In addition, rare CNVs were identified in treatment resistant MDD, but no genome-wide significance was found (O'Dushlaine et al., 2014). However, cohort sizes were limited in these studies. In a larger cohort of recurrent depression, an association was found with CNV deletions, but not with duplications (Rucker et al., 2013). However, in a re-anal-
ysis of this set, it was found that the deletions were no more frequent in cases than in controls, but the deletions in cases did contain more genes (Rucker et al., 2016). These deletions were found in chromosomal regions 15q13.3 and 22q11.2. Both structural variants are also associated with a wide variety of other disorders such as mental retardation, attention deficit disorder and seizures (Sharp et al., 2008; Williams et al., 2012; Armando et al., 2012). Yet, when looking at its relation to MDD, the 22q11.2 deletion has been associated with catechol-O-methyltransferase (COMT). In addition to this deletion, this gene also harbors a SNP changing a valine to a methionine residue which has been implicated in various psychiatric disorders such as schizophrenia and bipolar disorder (Silberschmidt & Sponheim, 2008). However, the step from associated CNV to functional candidate for MDD has yet to be established.

In addition to looking for associations with genetic variants at the DNA level, gene expression is a logical next step for finding a functional connection between the genome and the etiology of MDD. In the NESDA cohort, differences in expression in 129 genes were found between current MDD patients and controls. The genes that were enriched in current MDD patients were in the interleukin-6 and natural killer cell pathways (Jansen et al., 2015), suggesting a possible role for the immune system in MDD. In addition, in lipopolysaccharide-stimulated (LPS) blood cells seven genes were identified to be differentially expressed between MDD cases and controls (Spijker et al., 2010), serving as a possible biomarker for MDD. However, this research was performed on peripheral blood samples, rather than brain samples. A global brain gene expression analysis found changes in expression of glutamatergic and GABAergic related genes in suicide victims, with the highest number of alterations in the hippocampus and prefrontal cortex (Sequeira et al., 2009). Both areas of the brain are associated with cognitive function, which is impaired in MDD. The connection to glutamatergic signaling was also supported in a study of the brain region locus coeruleus. It is activated in stress partly through glutamatergic input, as glutamate forms the most prominent excitatory input of this region. In locus coeruleus neurons from MDD subjects, several glutamate receptor subunit genes showed increased expression, implicating a role for glutamatergic neurotransmission in the etiology of MDD (Chandley et al., 2014).

Although gene expression was traditionally always thought to be regulated by transcription factors and alternative splicing, evidence suggests that there is also a role for micro (mi)RNAs: small non-coding RNA transcripts that target mRNA for degradation or steric hindrance (Bartel, 2004). Interestingly, e.g. rat strains sensitive to stress have increased levels of the miRNA’s miR-18 and miR-12a in the paraventricular nucleus of the hypothalamus. These miRNAs target the glucocorticoid receptor and are, therefore, a likely contributing factor to stress susceptibil-
ity (Vreugdenhil et al., 2009). In humans, decreased levels of miR-1202 were found in blood samples of individuals with MDD and patients who responded to antidepressant medication had lower levels of miR-1202 than non-responders at baseline (Lopez et al., 2014). However, in this miRNA study, also peripheral blood samples were used rather than brain tissue, which complicates the search for a functional connection with changes in the brain of the MDD patient.

An additional perspective for understanding the etiology of MDD may lie in epigenetics. This branch of genetics focuses on changes in the chromatin structure of a DNA molecule, rather than the information on the DNA itself. These changes give rise to changes in gene expression. Environmental stressors are one of the biggest risk for the development of MDD and data of various studies suggests that environmental stressors are associated with epigenetic changes (Dalton et al., 2014). In a study on the epigenome of the hippocampus and prefrontal cortex of 6 MDD cases and 6 control subjects, differences in methylation profiles were detected (Kaut et al., 2015). In addition, in a twin study there were differences in methylation profile associated with current depressive symptoms in twin pairs (Córdova-Palomera et al., 2015). Also, clinical levels of depression were associated with overall higher methylation levels of two promoters for brain-derived neurotrophic factor (BDNF) in late-life depression (Januar et al., 2015). These results show a promising role for MDD research, but currently replication is still a challenge.

Clinical implications
In chapter 6 an association was shown between the number of side effects in PgP-dependent antidepressant users and a common SNP in ABCB1. The occurrence of side effects is a major obstacle in successful treatment: even if a patient experiences an alleviation of symptoms, if side effects outweigh this benefit, compliance will be low. In data collected over the past ten years, it is shown that only 25 to 50 percent of MDD patients will adhere to their medication (Sansone & Sansone, 2012). In order to increase medication compliance, it may be beneficial to find more genetic variants associated with the number of adverse effects. In addition to the identification of rs2032588 as a possible modulator of side effects in PgP-dependent antidepressant users, several other SNPs in the ABCB1 gene were found to be associated with treatment outcome, the number and type of side effect and remission (Breitenstein et al., 2015; Schatzberg et al., 2014; Ray et al., 2014). This suggests that genetic variants in the ABCB1 gene may provide a starting point to evaluate antidepressant effects in MDD patients with a specific set of alleles.

In addition to the large number of side effects caused by antidepressants, up to 50% of the patients do not experience remission after antidepressant treatment (Papakostas et al., 2010). Similar to the occurrence of side effects, also in anti-
depressant response, various candidate genes have been investigated. The most extensively studied gene in this context is \textit{CYP2D6}, which encodes a member of the cytochrome P-450 enzyme family, that is involved in drug metabolism. Variants in the \textit{CYP2D6} gene influence the metabolic activity of the enzyme it encodes and thus influence medication availability (Malhotra et al., 2004). Also, it has been shown that concurrently used medication that inhibits the CYP2D6 enzyme can influence response to antidepressant medication (Gressier et al., 2014).

Also, several GWAS have been performed on pharmaco-genes (Uher et al., 2010; Ising et al., 2009), but no genome-wide significance was found. In an effort to increase sample size, a meta-analysis of three genome-wide cohorts was performed, but no individual variant met criteria for genome-wide significance. Nevertheless, when only two of the cohorts were used for analysis, an intergenic region on chromosome 5 was found to be associated with an early improvement of symptoms after two weeks (Uher et al., 2013). The finding of significance only when combining two out of three cohorts may indicate that the ascertainment of samples was different across cohorts. The limited sample size of pharmacogenetic cohorts almost forcibly steers towards meta-analysis of several cohorts, making similar ascertainment a necessity.

Although there are clear limitations to present antidepressant drugs, there have been only little advances in the development of antidepressant medication that is beneficial for a substantial part of the affected population. Therefore, in spite of increased knowledge of genetic variants that influence drug efficacy and side effects, there is still a demand for novel therapeutic targets. With the knowledge that an individual’s genetic make-up may predispose to outcome and the number of side effects, individualized treatment programs (precision medicine) is an interesting topic for future research.

**GENERAL CONCLUSION**

The aim of this thesis was to contribute to finding causal variants for MDD. To reach this goal, two strategies were used. Firstly, fine-mapping was performed on the GAIN-MDD GWAS by means of genotyping tag SNPs and thus increasing coverage, and sequencing top genes. In addition, a candidate gene approach was used to find genetic variants that contribute to the number of side effects during antidepressant use.

Although the results from Chapters 3 and 4 support the hypothesis that SNP rs2522833 was causal in the GAIN-MDD cohort, the question remains whether \textit{PCLO} is still a solid candidate gene for MDD. Other GWAS have found top signals in
other genes, and the largest GWAS to date of the Psychiatric Genomics Consortium did not report genome-wide significance. Does this mean that the GWAS is obsolete as a study design for complex traits? Given the many results in other GWAS for complex traits—even in psychiatry, this is not the case. However, in order to reach genome-wide significance for MDD, the number of samples will have to increase dramatically or study groups have to become more homogeneous.

This thesis predominantly used a fine-mapping approach to further investigate the results of a GWAS for MDD. However, given the meager results of fine-mapping studies in general and the hypothesis that cohorts require a substantial increase in size, it may be too early on in the genetic research of MDD to utilize fine-mapping as a tool to find causal variants. It is more likely that causal variants, or variants in high LD with them, may be found when cohort sizes increase drastically. One may even argue that fine-mapping becomes obsolete if an increase of cohort size indeed leads to significant results. The GWAS results on which the research of this thesis was based, brushed the border of significance, so further fine-mapping was required to find a causal SNP. In this case: future research could aim to find further evidence that rs2522833 was indeed causal and not one of its neighbors. However, when a significant hit is found, the search for a causal neighbor becomes less vital. In addition to the expected significance of this finding in a larger cohort, imputation quality has also increased over the past years. This means that the genotype of neighboring SNPs can be predicted in high density and with great precision, making fine-mapping a less pressing matter. Although cohorts have not yet climbed to a size that yields a significant association, for future efforts it may be beneficial to move from fine-mapping to functional work as larger cohorts may increase the certainty of where a functional variant resides.

Chapter 5 explored the possibility of a pathway-based analysis. Although in the single SNP analysis one SNP in NR3C1 displayed borderline significance in an association analysis with severe recurrent MDD, neither gene nor pathway-based analyses showed promising P-values. Also in this chapter, sample size is low, leading to limited statistical power to detect an associated variant. In addition, the genes that were genotyped in this analysis were mainly genes involved in the metabolism of the biological measurements, when in reality many more genes contribute to a network. This gives room for further exploration of the pathway-based analysis with a more comprehensive set of genes, for example by imputing GWAS data for specific pathways.

In chapter 6 a variant was found to be associated with the number of side effects during antidepressant use. Currently, many more similar studies appear. This may contribute to a better understanding of the molecular mechanisms underlying the method of action of antidepressants. This increased knowledge of antidepressant action may in the future contribute to more personalized treatment for patients, and alleviate symptoms for this prevalent disorder.
In summary, at the current time, it seems too early to perform thorough fine-mapping studies, as it is now estimated that cohort sizes would need to be increased to approximately 100,000 samples in order to find variant that is significantly associated with MDD. On the other hand, endophenotypes such as type and number of side effects may contribute to a more comprehensive understanding of current pharmaceutical treatment options for MDD, ultimately leading to a more effective treatment of this incapacitating disorder.

REFERENCES

Armando M, Papaleo F, Vicari S (2012) COMT implication in cognitive and psychiatric symptoms in chromosome 22q11 microdeletion syndrome: a selective review; CNS Neurol Disord Drug Targets; May;11(3):273-81


CONVERGE consortium (2015) Sparse whole-genome sequencing identifies two loci for major depressive disorder; Nature; Jul 30;523(7562):588-91


Cryan JF, Kelly PH, Neijt HC, Sansig G, Flor PJ, van Der Putten H (2003) Antidepressant and anxiolytic-like effects in mice lacking the group III metabotropic glutamate receptor mGluR7; Eur J Neurosci; Jun;17(11):2409-17

Dalton VS, Kolshus E, McLaughlin DM (2014) Epigenetics and depression: return of the repressed; J Affect Disord; Feb;155:1-12

notype for psychosis: evidence from healthy relatives and a common risk variant in CACNA1C; Biol Psychiatry; Sep 15;76(6):466-75


Gressier F, Verstuyft C, Hardy P, Becquemont L, Corruble E (2015) Response to CYP2D6 substrate antidepressants is predicted by a CYP2D6 composite phenotype based on genotype and comediations with CYP2D6 inhibitors; J Neural Transm; Jan;122(1):35-42


Januar V, Ancelin ML, Ritchie K, Saffery R, Ryan J (2015) BDNF promoter methylation and genetic variation in late-life depression and genetic variation in major depression; Transl Psychiatry; Aug 18;5:e619

Chapter 7


Levinson DF, Mostafavi S, Milaneschi Y, Rivera M, Ripke S, Wray NR, Sullivan PF (2014) Genetic studies of major depressive disorder: why are there no genome-wide association study findings and what can we do about it?; Biol Psychiatry; Oct 1;76(7):510-2


Murphy KC, Jones LA, Owen MJ (1999) High rates of schizophrenia in adults with velo-cardio-facial syndrome; Arch Gen Psychiatry; Oct;56(10):940-5


O’Dushlaine C, Ripke S, Ruderfer DM, Hamilton SP, Fava M, Iosifescu DV, Kohane IS, Churchill SE, Castro VM, Clements CC, Blumenthal SR, Murphy SN, Smoller JW, Perlis RH (2014) Rare copy number variation in treatment-resistant major depressive disorder; Biol Psychiatry; Oct 1;76(7):536-41


Sansone RA, Sansone LA (2012) Antidepressant adherence: are patients taking their medications?; Innov Clin Neurosci; May;9(5-6):41-6


Sebat J, Levy DL, McCarthy SE (2009) Rare structural variants in schizophrenia: one disorder, multiple mutations; one mutation, multiple disorders; Trends Genet; Dec;25(12):528-35


Vreugdenhil E, Berezikov E (2010) Fine-tuning the brain: MicroRNAs; Front Neuroendocrinol; Apr;31(2):128-33


SUMMARY
SUMMARY

Depression, or Major Depressive Disorder (MDD), is one of the leading causes of disability worldwide with both psychological and physical consequences, having an impact on absenteeism and healthcare. MDD has a high disease burden for both patients as well as their relatives, with approximately only half of the patients responding to current treatment options. In order to improve diagnosis and treatment options, it is of importance to unravel the underlying mechanisms of this disease. Although twin studies have shown a heritability of approximately 40%, the molecular etiology of the disorder is still largely unknown. The aim of this thesis was to find causal genetic variants for MDD, in order to shed light on its etiology. For this aim to methods were used: in chapters 3 and 4 a fine-mapping approach was used to investigate whether rs2522833 is indeed the causal variant in the Dutch GAIN-MDD cohort. In chapters 5 and 6 a pathway and a candidate gene approach were used, respectively, to find genetic variants associated with possible biological markers for MDD and variants associated with the side effects of antidepressant medication.

MDD is a common complex disorder, with both environmental and genetic causes. Chapter 2 is a review of genetic approaches to find associated genetic variants, in which linkage studies, candidate genes and genome wide association studies (GWAS) are discussed. The foundation for the fine-mapping studies in this thesis was the GWAS performed by Sullivan et al., that was performed on the GAIN-MDD cohort. In this GWAS, the single nucleotide polymorphism (SNP) rs2522833 in the PCLO gene showed an association peak, that became nominally significant after post-hoc analysis with an Australian cohort. Many other top signals from this GWAS also mapped to the region spanning PCLO. This gene codes for the protein Piccolo, which is located in the presynaptic active zone of neurons. Piccolo is thought to play a role in transport of vesicles that contain neurotransmitters. One of the hypotheses on the etiology of MDD poses that the disease is caused by imbalance in neurotransmitters. Therefore, PCLO with its role in vesicle transport, seems a plausible candidate gene. However, the chip that was used to perform the GWAS, was not designed in a gene-centric manner, leading to a less than optimal coverage for many genes. In Chapter 3 we performed a fine-mapping study that takes this design into account. First of all, in this study we wanted to find additional evidence that rs2522833 is truly the causal variant in the GAIN-MDD cohort. Due to lack of coverage, the association peak may be caused by a neighbouring SNP that is in high linkage disequilibrium (LD) with rs2522833. In addition, the GWAS showed multiple sub-threshold signals, which were also located in genes with low coverage. To increase genotypic information on these genes, we selected so called “tag SNPs”, to increase the coverage of the genes to 100% at a minor
allele frequency (MAF) of 10% and a correlation coefficient $r^2$ of 0.9. Again, the lowest P-values were found in the region surrounding rs2522833, at SNPs SNPs rs2715147 and rs2715148 ($P=1.2E-6$). However, in the single SNP association analysis, no genome-wide significance was found. We then performed a joint re-analysis of all genotyped SNPs in the PCLO gene. When we assumed that rs2522833 was indeed causal, the regression line in this analysis became the steepest, supporting the hypothesis that either rs2522833 or a neighboring SNP in high LD is causal in the GAIN-MDD cohort. In a subsequent haplotype analysis, we found a lower P-value than in the single SNP analyses ($P=9.9E-7$). Although this is not genome-wide significant, it suggests that a yet unknown variant in this area may be causal for the top signal in the GAIN-MDD GWAS.

In order to find more evidence to support the hypothesis from Chapter 3 that either rs2522833 or a yet unknown variant in high LD with it is the causal variant in the GAIN-MDD cohort, a second strategy for fine-mapping was performed in Chapter 4. The results from the haplotype analysis in Chapter 3 suggest that a yet unknown variant in the PCLO gene may be responsible for the top signal in the GAIN-MDD GWAS. In order to find this unknown variant, the PCLO gene was sequenced using next-generation sequencing methods. Additionally, two other candidate genes from literature were sequenced: GRM7 and SLC6A4. GRM7 codes for a metabotropic glutamate receptor. Functional knock-out models of this receptor show an antidepressant effect (Cryan et al., 2003). In addition, one of the top signals in a meta-analysis of three GWAS studies for MDD came from GRM7 (Shyn et al., 2011). SLC6A4 codes for the serotonin transporter. This transporter plays a pivotal role in the availability of serotonin in the synaptic cleft: the transporter is localized on the presynaptic membrane and transports serotonin back into the presynaptic cell, effectively recycling the neurotransmitter to be used by receptors on the postsynaptic membrane. In this sequencing study 961 new SNPs were discovered. 71 of these newly identified SNPs were genotyped together with 185 tag SNPs for further fine-mapping. The tag SNPs were a necessity, as the newly identified SNPs alone could not cover the genes 100%. This approach of next-generation sequencing, followed by fine-mapping, did not lead to a lower P-value than the original GAIN-MDD GWAS, with a lowest P-value at SNP rs2715147 ($P=1.5E-6$). These results show that the unknown variants that were identified in this research do not show a better association with MDD than rs2522833. Although rare variants were not taken into account in this research, Chapter 4 provides additional evidence that rs2522833 is the causal variant in the GAIN-MDD cohort.

In Chapters 3 and 4, the fine-mapping approach was used, based on the GAIN-MDD GWAS. In a GWAS, there is no a priori hypothesis when looking for an associated variant. In Chapters 5 and 6 a second approach is used: the candidate gene ap-
proach. With this method genes are selected based on prior knowledge of biological functions or connections with a certain trait. In Chapter 5 genes from three biological pathways were genotyped in the NESDA cohort: the HPA axis, the HPT axis and vitamin D metabolism. In MDD, the HPA axis may be dysfunctional, leading to higher levels of its end product cortisol. The HPT axis also shows changes in MDD patients, while patients suffering from hypothyroidism show characteristics of MDD. Vitamin D measurements in blood are associated with MDD in various patient groups, including the NESDA cohort and a cohort of elderly patients. In the NESDA cohort, biological measurements are available for these three pathways: cortisol levels, thyroid levels and vitamin D levels. The key genes from these pathways were fine-mapped using tag SNPs, after which associations analyses were performed on single SNPs, whole genes and on a pathway level. Corrections were made for biological measurements, to exclude other variation than genetic variation. These analyses were repeated for the most severe phenotype within the NESDA cohort: severe recurrent MDD. This approach did not lead to significant results. However, in the analysis of severe recurrent MDD, one SNP approached significance, which was not witnessed in the full cohort of MDD patients. In spite of this observation, it needs to be taken into account that group size in the analysis of the severe recurrent MDD group was reduced to 1200, leading to low statistical power. Also, the gene and pathway analyses showed no significant results, showing that a pathway-based analysis does not lead to improved P-values in this cohort. In addition, analyses were performed to see if the genotyped variants together with MDD-status had a combined effect on the biological measurements of these pathways. Although the HPA axis and vitamin D showed a difference in levels, as known from prior studies, this effect was not enhanced by the presence of a certain genotype.

In contrast to the other chapters, where the focus was mostly on finding variants associated with MDD, in Chapter 6 the focus lies on the treatment of MDD with antidepressant medication. Like Chapter 5, this study was also based on a candidate gene approach. Again in this study, we selected genes based on their function as known from prior studies. When treating MDD with antidepressants, side effects are an obstacle. In the NESDA cohort, in 64% of the 927 patients that used one antidepressant, on average 2.9 side effects were reported. Tricyclic antidepressants (TCA’s) were associated with more side effects than selective serotonin reuptake inhibitors (SSRI’s). Interestingly enough, the number of side effects was associated with the severity of MDD, higher dosage and having multiple psychiatric diagnoses. (Bet et al., 2013). In a study of over 400 patients that were treated with SSRIs, 55% experienced side effects in the first two weeks of treatment (Hu et al., 2004). Another limitation on treating MDD with antidepressants is that a substantial number of patients does not show remission of symptoms. In a study of 2900 pa-
patients, response rate to the SSRI citalopram was (47%), with remission in 28-33% of patients (Trivedi et al., 2006). When treating patients with antidepressant medication, a balance must be found between finding therapeutic dosage and preventing a dosage that causes side effects. In Chapter 6 we hypothesized that common variants in the CYP2C21 gene and in the ABCB1 gene may have an effect on the number of side effects by increasing or decreasing the blood levels of antidepressant medication. A significant association was found between a common variant in ABCB1 and the number of side effects in patients using PgP-transporter-dependent medication. The A-allele of rs2032588 was associated with a lower number of side effects, when we corrected for age, gender, duration of therapy and dosage (B=-0.44, P=1.22E-4). This remained significant after control for false discovery rate (FDR) (B=-0.44, q=4.6E-3). This association was not found in patients that did not use PgP-dependent medication. The study in Chapter 6 is one of the first to show an association between genetic variants and number of side effects in a naturalistic cohort of substantial size. However, in an ideal scenario, results would have to be replicated in a similar cohort.

In summary, although the studies in this thesis support the hypothesis that rs2522833 is the causal variant in the GAIN-MDD cohort, no significant associations were found using the fine-mapping approach. This may be caused by sample size. It is now estimated that that a true association may be found with a sample size of approximately 100,000 patients. In addition, for the pathway approach no significant results were found either. Also in this study, it would be ideal to use a many times larger cohort, in order to increase statistical power. Finally, an association was found between a common variant in ABCB1 and number of side effects in patients using PgP-dependent medication. These results would have to be replicated in a similar cohort, before it even becomes imaginable that the results of this study might contribute to personalized medication.
NEDERLANDSE SAMENVATTING
Nederlandse samenvatting

NEDERLANDSE SAMEN VATTING

Depressie, of Major Depressive Disorder (MDD), is wereldwijd een van de meest voorkomende oorzaken van psychisch en lichamelijk lijden, ziekteverzuim en directe en indirecte zorgkosten. Om de diagnose en behandeling van MDD te kunnen verbeteren, is het belangrijk om meer te weten te komen van de achterliggende mechanismes van de ziekte. Hoewel genetisch onderzoek heeft laten zien dat de erfelijkheid van MDD ongeveer 40% is, is de moleculaire ontstaanswijze nog grotendeels onbekend. In dit proefschrift was het overkoepelende doel om causale genetische varianten te vinden die mogelijk helderheid kunnen verschaffen over deze ontstaanswijze. Voor dit doel werden twee benaderingen gekozen: in Hoofdstuk 3 en 4 werd fine-mapping gebruikt om te onderzoeken of de genetische variant rs2522833 daadwerkelijk een causale variant is; in de hoofdstukken 5 en 6 werd een kandidaatgen- en pathway-benadering gekozen om genetische varianten te vinden die geassocieerd zijn met biologische waardes voor MDD en varianten die geassocieerd zijn met de bijwerkingen van antidepressiva.

MDD is een veel voorkomende complexe ziekte met zowel oorzaken uit de omgeving van het individu als genetische oorzaken. In Hoofdstuk 2 wordt een overzicht gegeven van genetische benaderingen om geassocieerde genetische varianten te vinden. Onder andere linkage studies, kandidaatgenen en de GWAS-methode passeren hier de revue. De GWAS die gebruikt werd als basis voor de fine-mapping studies was de GWAS van Sullivan et al., die werd uitgevoerd op het GAIN-MDD cohort. Met deze GWAS werd in het PCLO-gen het single nucleotide polymorfisme (SNP) rs2522833 gevonden, welke na post-hoc analyse met een Australisch cohort significantie benaderde. Vele andere topsignalen uit deze GWAS bleken ook voort te komen uit het gebied waar PCLO zich bevindt. Dit gen codeert voor het eiwit Piccolo, wat zich bevindt in de presynaptische actieve zone van zenuwwellen. Van dit eiwit wordt gedacht dat het een rol speelt bij het transport van zogenaamde ‘vesicles’ die neurotransmitters bevatten. Een van de hypotheses betreffende de oorzaken van MDD stelt dat de ziekte veroorzaakt wordt door een onbalans van neurotransmitters. Daarom lijkt PCLO met zijn rol in het transporteren van vesicles een plausibel kandidaatgen. Echter, de chip die gebruikt werd voor de GWAS was niet ontworpen op een manier waarbij de genen centraal stonden, waardoor de dekking van deze chip voor vele genen niet optimaal was. In Hoofdstuk 3 wordt een fine-mapping studie beschreven die rekening houdt met dit ontwerp. Ten eerste om additioneel bewijs te vinden dat rs2522833 daadwerkelijk de causale variant is in het GAIN-MDD cohort. Vanwege het gebrek aan volledige dekking in de originele GWAS, zou het ook kunnen zijn dat het signaal veroorzaakt wordt door een nabijgelegen SNP die overerft met rs2522833. Ten tweede waren er meerdere signalen die juist onder de drempelwaarde voor significantie bleven, welke ook voortkwamen.
uit genen die niet optimaal gedekt waren. Om de genotypische informatie van deze genen te maximaliseren, werden zogenaamde tag SNPs uitgezocht, waardoor de dekking van de genen opgehoogd werd naar 100% met een minor allele frequency (MAF) van 10% en een correlatiecoëfficiënt $r^2$ van 0,9. Wederom werden de laagste P-waardes gevonden in de onmiddellijke omgeving van rs2522833, bij de SNPs rs2715147 en rs2715148 ($P=1,2E-6$). In de individuele analyses van deze SNPs werd echter geen genoom brede significantie bereikt. Daarna werd een zogenaamde ‘joint re-analysis’ uitgevoerd van alle gegenotypeerde SNPs in PCLO. Wanneer er werd aangenomen dat rs2522833 de causale variant was, bleek de regressielijn in deze analyse het steil. Dit ondersteunt de hypothese dat de causale variant óf rs2522833 is, óf een variant die naar alle waarschijnlijkheid tezamen met deze SNP overerft. In een volgende stap, werd een haplotype-analyse uitgevoerd van blokken met SNPs die naar alle waarschijnlijkheid samen overerven. Een haplotype leverde een lagere P-waarde dan de individuele SNP-analyses ($P=9,9E-7$). Hoewel dit resultaat geen genoom brede significantie opleverde, suggereert het wel dat een nog onbekende variant in dit gebied causaal zou kunnen zijn voor het topsignaal uit de GAIN-MDD GWAS.

Om meer bewijs te verzamelen voor de hypothese uit Hoofdstuk 3 dat óf rs2522833 óf een nog onbekende variant uit hetzelfde gebied de causale variant is in het GAIN-MDD cohort, werd een tweede strategie van de fine-mapping benadering gebruikt in Hoofdstuk 4. De resultaten van de haplotype-analyse in Hoofdstuk 3 suggereerden dat een nog onbekende variant in het PCLO-gen wellicht verantwoordelijk kan zijn voor het topsignaal in de GAIN-MDD GWAS. Met de insteek om deze variant te vinden, werd het PCLO-gen gesequenced met behulp van next-generation sequencing methods. Daarnaast werden ook twee andere kandidaatgenen voor MDD uit de literatuur gesequenced: GRM7 en SLC6A4. GRM7 codeert voor een metabotrope glutamaatreceptor. Functionele knock-outmodellen van deze receptor laten een antidepressief effect zien (Cryan et al., 2003). Ook gaf GRM7 een van de topsignalen in een meta-analyse van drie GWAS naar MDD (Shyn et al., 2011). Het SLC6A4-gen codeert voor de serotoninetransporter. Deze transporter speelt een cruciale rol bij de beschikbaarheid van serotonine in de synaptische spleet: de transporter is gelokaliseerd op het presynaptische membraan en transporteert serotonine terug de presynaptische cel in, waarbij de neurotransmitter effectief gerecycled wordt voor gebruik door receptoren op het postsynaptische membraan. Met deze sequencing studie werden 961 nieuwe SNPs ontdekt. 71 van deze nieuw gevonden SNPs werden tezamen met 185 tag SNPs gebruikt voor verdere fine-mapping. De tag SNPs waren noodzakelijk omdat een opzet met louter de nieuw ontdekte SNPs de genen niet voor 100% kon dekken. Deze benadering van next-generation sequencen, gevolgd door fine-mapping, leidde niet tot een lagere P-waarde dan de originele GAIN-MDD GWAS, met de laagste P-waarde bij SNP.
Nederlandse samenvatting

rs2715147 van $P=1.5\times10^{-6}$. Deze resultaten laten zien dat de onbekende varianten die gedetecteerd werden in dit onderzoek niet beter geassocieerd zijn met MDD dan rs2522833. Hoewel zeldzame varianten niet meegenomen zijn in deze studie, geeft Hoofdstuk 4 aanvullend bewijs dat rs2522833 de causale variant is in het GAIN-MDD cohort.

In de hoofdstukken 3 en 4 werd de fine-mapping benadering gebruikt, gebaseerd op de GAIN-MDD GWAS. Een GWAS gebruikt geen a priori hypothese in de zoektocht naar een geassocieerde variant. In de hoofdstukken 5 en 6 wordt een andere benadering gebruikt: de kandidaatgenbenadering. Bij deze methode worden genen geselecteerd op basis van eerder bekende functies of biologische connecties met een eigenschap. In Hoofdstuk 5 werden genen uit drie verschillende biologische pathways gegenotypeerd in het NESDA cohort: de HPA-as, de HPT-as en vitamine D-metabolisme. In MDD is de HPA-as ontregeld, hetgeen leidt tot hogere waarden van het eindproduct cortisol. Ook bij de HPT-as worden afwijkingen gevonden bij MDD-patiënten, terwijl patiënten met hypothyroïdie juist vaak kenmerken van depressie laten zien. Vitamine D-waarden in het bloed zijn geassocieerd met depressie in diverse patiënten groepen, o.a. het NESDA-cohort en een cohort van ouderen. Binnen het NESDA-cohort zijn voor deze drie pathways biologische meetingen beschikbaar, in de vorm van cortisolwaarden, schildklierhormoonwaarden en vitamine D-waarden. De meest prominente genen binnen deze pathways werden in kaart gebracht met behulp van tag SNPs, waarna associatieanalyses werden uitgevoerd op individueel SNP-niveau, op genniveau en op pathway-niveau. Er werd gecorrigeerd voor de biologische waarden, om andere oorzaken van variatie buiten dan genetische uit te sluiten. Ook werden deze analyses herhaald voor het meest extreme fenotype: terugkerende ernstige MDD. Deze methode leverde na het corrigeren voor meerdere tests geen significante resultaten op. Echter, voor de terugkerende ernstige groep ontstond benaderde één SNP significantie, hetgeen niet te zien was binnen de gehele groep MDD-patiënten. Ondanks deze observatie moet men in ogenschouw nemen dat de groepsgrootte hierbij gereduceerd was tot 1200 patiënten, waardoor de statistische power zwak is. Ook op genniveau en op pathway-niveau werden geen significante resultaten gevonden, hetgeen laat zien dat een pathway-benadering in dit cohort geen toegevoegde waarde heeft voor het vinden van een associatie met MDD. Verder werd er gekeken of de gegenotypeerde varianten tezamen met depressie een gecombineerd effect hadden op de biologische uitkomstmaten van deze pathways. Hoewel er bij de HPA-as en bij vitamine D een verschil in uitkomstwaardes was, zoals bekend uit eerdere literatuur, werd dit effect niet versterkt door het aanwezig zijn van een bepaald genotype.

Waar bij de eerdere hoofdstukken met name gefocust werd op het vinden van varianten die geassocieerd zijn met MDD, wordt in Hoofdstuk 6 de behandeling van
MDD met antidepressiva onder de loep genomen. Deze studie was, evenals Hoofdstuk 5, gebaseerd op de kandidaatgenbenadering. Ook hier werden genen geselecteerd waarvan de functie reeds bekend was uit de literatuur. Bij de behandeling van MDD met antidepressiva zijn bijwerkingen een obstakel. In het NESDA-cohort werd bij de 64% van de 927 patiënten die één antidepressivum gebruikten gemiddeld 2,9 bijwerkingen gerapporteerd, waarbij tricyclische antidepressiva (TCA) geassocieerd waren met meer bijwerkingen dan selectieve serotonine heropnameremmers (SSRI). Interessant genoeg was het aantal bijwerkingen geassocieerd met de ernst van de depressie, hogere doseringen en het vóórkomen van meerdere psychiatrische diagnoses (Bet et al., 2013). In een studie van meer dan 400 patiënten die met SSRI’s behandeld werden, ervoer 55% van de patiënten bijwerkingen in de eerste twee weken van de behandeling (Hu et al., 2004). Een andere beperking bij het behandelen van MDD is dat een substantieel deel van de patiënten geen remissie van symptomen laat zien bij het gebruik van antidepressiva. In een onderzoek met 2900 patiënten was de respons op de SSRI citalopram 47%, met daadwerkelijke remissie in 28-33% van de patiënten (Trivedi et al., 2006). Bij de behandeling met antidepressiva moet een balans gevonden worden tussen het bereiken van een therapeutische dosering en het voorkomen van een dosis waarbij bijwerkingen optreden. In Hoofdstuk 6 wordt gehypothetiseerd dat varianten in het leverenzym CYP2C9 en in het medicatietransportergen ABCB1 effect kunnen hebben op het aantal bijwerkingen door het verhogen of verlagen van de concentratie van een antidepressivum in het bloed. Een significante associatie werd gevonden tussen een veel voorkomende variant in ABCB1 en het aantal bijwerkingen bij patiënten die medicatie gebruiken die afhankelijk is van het de PgP-transporter. Het A-allel van rs2032588 was geassocieerd met een lager aantal bijwerkingen wanneer gecorrigeerd werd voor leeftijd, geslacht, duur van de therapie en dosering (B=-0,44, P=1,22E-4). Deze significante bleef ook behouden na controle voor vals positieve (false discovery rate, FDR) (B=-0,44, q=4,6E-3). Deze associatie werd niet gevonden in patiënten die géén antidepressivum gebruikten dat afhankelijk is van de PgP-transporter. De studie in Hoofdstuk 6 is een van de eerste die een associatie tussen genetische varianten en aantal bijwerkingen laat zien in een naturalistische setting met een substantiële grootte van het cohort, maar idealiter zouden de resultaten nog gerepliceerd moeten worden in een vergelijkbaar cohort.

Samenvattend, hoewel het werk in dit proefschrift de hypothese lijkt te ondersteunen dat rs2522833 de causale variant is in het GAIN-MDD-cohort, werden geen significante resultaten behaald met deze fine-mapping benadering. Dit valt waarschijnlijk te wijten aan de grootte van het cohort. Inmiddels wordt geschat dat een echte associatie zeer waarschijnlijk pas haalbaar is bij een cohort bestaande uit ongeveer 100.000 patiënten. Ook voor de pathway-benadering werden geen significante resultaten gevonden. Ook hier zou men idealiter een velen malen groter co-
hort willen gebruiken om de statistische power voor een valide ontdekking te kunnen verhogen. Tevens werd een associatie gevonden met een veel voorkomende variant in het \textit{ABCB1}-gen en het aantal bijwerkingen bij patiënten met een PgP-afhankelijk antidepressivum. Deze resultaten dienen echter nog wel gerepliceerd te worden in een vergelijkbaar cohort, voordat men mogelijk zou kunnen denken aan een praktische toepassing van deze resultaten om medicatie verder toe te spitsen op de individuele patiënt.
DANKWOORD
Dankwoord

“Dankbaarheid is de moeder van alle deugden.”
Marcus Tullius Cicero

Gedurende mijn jaren aan de VU heb ik met veel interessante, kundige en hartverwarmende mensen kennis mogen maken, hetgeen mijn leven enorm heeft verrijkt. Hoewel een promotietraject heuvels en dalen kent, waren deze fijne mensen vaak om mij heen om de wetenschap en het leven zonnig in te blijven zien en ik ben ze daarvoor dan ook zeer erkentelijk.

Witte Hoogendijk. Beste Witte, als hoogleraar psychiatrie was jij mijn promotor. Ik wil je bedanken voor je optimisme, enthousiasme en inzet. Door jouw begeleiding heb ik dit proefschrift kunnen afronden. Je bleef positief, ook al zag ik zelf door de bomen het bos niet meer en wilde ik de handdoek in de ring gooien. Bedankt voor de fijne samenwerking.

Peter Heutink. Gekscherd “papa Peter” genoemd, als onze vader van het genetica-lab en mijn promotor qua genetica. Je kennis met betrekking tot genetica was voor mij van onschatbare waarde. Ik wil je enorm bedanken dat je mij uit het lab hebt gehaald en me richting data-mining gestuurd hebt. Tot op de dag van vandaag ben ik hier erg blij mee, het heeft me een hoop opgeleverd. Ook veel dank dat je me er van tijd tot tijd aan herinnerde: “geen resultaat is óók een resultaat”.


Marianna Bevova. Lieve Marianna, dank voor je dagelijkse supervisie, je kennis en je inzet. Je hebt me op vele vlakken geholpen, zowel in het lab als met schrijven als met analyseren.

Zoltán Bochdanovits. Mijn mentor in het programmeren. Ontelbare keren ben ik bij je geweest met syntax-errors die jij binnen een minuut oploste. Ook je kennis van statistiek was onmisbaar bij het schrijven van mijn artikelen.

Graag bedank ik ook de leden van de leescommissie, die tijd hebben vrijgemaakt om dit manuscript te lezen.
Ook veel dank aan alle collega’s van het medical genomics lab: **Patrizia, Deborah, Hans, Luba, Shoey, Javi, Francesca, Hilal, Ashutosh, Margherita, Sasja, Thais, Laura, Ingrid, Sietske, Petra, Fara, Orr, Tiny, Margreet, Esther, Hannemieke, Emiel, Nienke, Valerie, Dennis, Truus, Mohit, Claudia, David**: dank voor al jullie kennis, jullie warme persoonlijkheden en alle gezellige uitstapjes.

Daarnaast ook veel dank aan **TIPharma**, zonder wie dit PhD-project niet bestaan had.

Ook buiten de VU waren er veel mensen die mij ondersteund hebben:

**Papa & Mama.** Dankjewel dat jullie mij altijd in mijn ondernemingen gesteund hebben. Dankzij jullie heb ik doorzettingsvermogen en de wil om te presteren, maar weet ik ook dat er meer is op de wereld dan alleen leren.

**Eline & Martine.** Lieve zusters, dank voor jullie steun. Jullie hebben mij menigmaal horen foeteren als dingen niet zo liepen als ik wilde en hebben me op die momenten laten inzien dat het wel meeviel ("Hoe erg is dit op een schaal van nul tot ebola?"). Bedankt dat jullie mijn zusjes zijn.

**Edith Smit.** Bedankt dat je als alpha-professor naar mijn bèta-proefschrift wilde kijken en me hielp om structuur aan te brengen.

Ook mijn familie op mijn "home away from home", Manege en Dressuurstal Hallinckveld ben ik erg dankbaar en in het bijzonder **Thea en Hellen**: dank voor jullie immer luisterende oor.

Veel dank ook aan al mijn lieve collega’s van **OLVG** die mij gesteund hebben om dit neuro-proefschrift af te ronden. Inmiddels ben ik in hun gezelschap weer terug bij mijn “oude liefde” cardio.

**Last but not least:** lieve **Arne**, dank voor al je positiviteit en je geduld. Je hebt een hoop gemopper van me moeten aanhoren over het afronden van dit proefschrift en je was altijd overtuigd van een goede afloop.
DISSERTATION SERIES
Department of Psychiatry, VU University Medical Center

Dissertation series


