**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 hypothesis 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 approach. 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 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 CYP2C219 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.