The Genetics of MDD – A Review of Challenges and Opportunities

Eva C. Verbeek¹, Marianna R. Bevova², Wim J.G. Hoogendijk³, Peter Heutink⁴

¹Department of Clinical Genetics, VUMC Medical Center, Amsterdam, the Netherlands
²European Research Institute for the Biology of Ageing, University Medical Center Groningen, the Netherlands
³Department of Psychiatry, Erasmus Medical Center, Rotterdam, the Netherlands
⁴Genome Biology of Neurodegenerative Diseases, German Center for Neurodegenerative Diseases, Tübingen, Germany

Corresponding author: Eva C. Verbeek, Department of Clinical Genetics, VU University Medical Center, Amsterdam, the Netherlands, Tel: 49 7071/9387-362; E-mail: eef.verbeek@gmail.com

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.

Keywords: Major depressive disorder; Cognitive symptoms; Social phobia

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 [1]. 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 [2]. 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 [3]. 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% [4]. 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 [5]. In 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 [6] and in an elderly cohort, when looking at genetic risk scores of MDD in conjunction with anxiety, 2.1% of variation was explained [7]. 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 [8].

Strikingly, in depressive disorders is the 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 [9]. 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 [10], 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 [11]. In addition, although suicide ideation in most Western countries is more apparent in women, mortality from suicide is typically higher for men [12].

**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 descent 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 [13]. The same region showed evidence of linkage with MDD in a smaller cohort of families of heavy smokers [14]. 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 [15-19].

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 [20]. In addition it was shown that administration of BDNF has antidepressant effects [21]. In several studies of genetic polymorphisms in BDNF, no significant effect was found [22-24]. However, several studies report gene x environment interaction of BDNF and associations with depressive states in bipolar disorder and schizophrenia [25-27]. 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 [28]. 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 [29]. 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 [30]. 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 [31], which was performed on a cohort of 3540 individuals of Western European ancestry. 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 [32-34]. However, several top signals mapped back to a genomic region overlapping the gene PCLG. 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 [35-36].

However, this finding was not replicated by Shyn et al. on a different cohort of European ancestry [37]. 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 CIDI1 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 [38]. 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 low 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 [39]. A common variant in the same gene was previously associated with MDD in a cohort of ±700 individuals and later replicated in six cohorts [40] of similar magnitude. 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 [41]. More specifically in MDD, the estimate lies around 1% [42] and thus the vast majority of genetic risk factors remain to be identified. 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 [43], 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 [44]. 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 [45] and type two diabetes [46]. 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 effect against the 5-HTTLPR length polymorphism in SLC6A4 [47]. 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 [48]. 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 [49]. 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 [50]. These influences on gene expression may provide an additional explanation for the currently missing heritability. 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 [51], a reduced expression of fatty acid biosynthesis genes in the prefrontal cortex of depressed patients [52], decreased expression of thyrotropin-releasing hormone [53] and increased expression of vasopressin [54] 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 clearcut 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 [55]. In addition, dysfunction of the reward system has been suggested to be heritable as well [56].

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 [57]. 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 [58]. 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


