SUMMARY OF THESIS FINDINGS

This thesis aimed to study the complex link between genetic effects and measured environmental risk factors in major depressive disorder (MDD) in empirical data, and to explore boundaries of the consequences of two Genome-Wide Association study (GWAS) designs and assortative mating from a theoretical perspective. Both genes and environment affect MDD risk, but it remains unclear whether both act independently, whether they interact, or whether environmental risk might actually reflect shared genetic effects between MDD risk and e.g. behavioral traits that might increase environmental stress.

The most studied candidate gene in MDD is the serotonin transporter gene. The length polymorphism in the promoter region of this gene (5-HTTLPR) has been hypothesized to influence MDD risk, because an important group of antidepressant drugs acts on the serotonin transporter. A Science paper by Caspi et al suggested that this gene mainly exerts its influence when persons had been exposed to childhood trauma.[1] In **Chapter 2**, the 5-HTTLPR polymorphism was analyzed in 1593 cases and 1411 controls from the Netherlands Study of Depression and Anxiety (NESDA) and the Netherlands Twin Register (NTR). In these individuals, four different outcome measures were defined: lifetime MDD, suicidal MDD, chronic MDD, and course of MDD (chronic versus non-chronic). No evidence was found for either direct effects of 5-HTTLPR on these outcome measures or interaction effects between 5-HTTLPR and five environmental risk factors for MDD: lifetime stressful life-events, recent stressful life-events, sexual abuse, childhood trauma, and educational attainment (as proxy for social economic status associated with increased stress).

In **Chapter 3**, the relation between MDD and educational attainment was investigated in approximately 25,000 individuals from the Psychiatric Genomics Consortium (PGC wave 1) with additional Dutch and Estonian data. An increased risk for MDD was confirmed in individuals with lower educational attainment. Subsequently, the possible contribution of shared genetic effects to this link was assessed with three different methods applying data of 884,105 autosomal common single-nucleotide polymorphisms (SNPs). Firstly, polygenic risk scores (PRS) based on GWAS results on education attainment in ~120,000 individuals (EA-PRS) did not affect MDD risk, and PRS based on MDD GWAS results in ~20,000 individuals (MDD-PRS) did not affect EA. Secondly, a non-consistent weak significant negative genetic correlation was found with bivariate genomic-relationship-matrix restricted maximum likelihood (GREML). Thirdly, no concordance was found in either significance or direction of SNP effects across MDD GWAS and EA GWAS results based on SNP effect concordance analysis

(SECA). To conclude, these findings indicate that it is unlikely that shared genetic effects explain a large proportion of the link between MDD risk and lower education attainment, but a small genetic contribution to this deleterious link could not be excluded.

Up to 2014, research on gene-by-environment (GxE) interaction in MDD had mainly focused on candidate genes, such as 5-HTTLPR. However, with the emergence of MDD cohorts with genome wide SNP data, novel methods were developed that allowed to tag genome-wide genetic MDD risk with polygenic risk scores (PRS). In **Chapter 4**, PRS were constructed in 1645 cases and 340 controls from NESDA based on discovery results from the large Psychiatric Genomics Consortium (PGC wave 1). These PRS were found to have an increased impact on MDD risk in individuals exposed to childhood trauma (CT), which suggested gene-by-environment interaction on a genome-wide scale. This interaction-effect was found both as departure from multiplicativity (combined impact of PRS and CT larger than the *product* of the individual effects) and as departure from additivity (combined impact larger than the *sum* of the individual effects), the latter of which has been hypothesized to be more plausible from a biological perspective.

In **Chapter 5**, the interaction between polygenic risk scores (PRS) and childhood trauma (CT) was further tested in seven cohorts from PGC (wave 2) with CT information available in 3,024 cases and 2,741 controls. CT had consistent impact across cohorts, with similar impact in males and females. However, the interaction effects were heterogeneous with a positive interaction effect in NESDA (as in Chapter 4), negative interaction effect in the Radiant-UK study, and no interaction in the other contributing five cohorts, resulting in no overall evidence for interaction between PRS and CT in MDD. The results from Chapter 5 illustrate the heterogeneity of MDD, and suggest that the results of Chapter 4 can best be interpreted as a single cohort phenomenon.

The focus was switched from analyzing empirical data to theoretic work in **Chapter 6.** Two GWAS study designs applied in the Psychiatric Genomics Consortium (PGC) were considered with respect to their power of SNP association analysis and SNP-heritability estimates (proportion of population-variance in disease-risk attributable to genome-wide common SNPs). First, parent-affected-offspring trio data are regularly applied in the subgroups of the PGC analyzing autism spectrum disorder (ASD) and attention deficit/hyperactivity disorder (ADHD). Trio data are essential to detect *de novo* mutations, but its use may results in reduced power in association analysis and underestimation of the SNP-heritability compared to analyses in case screened-control data. This difference is attributable to (i) potential oversampling of multiplex families (with more than

one affected offspring) and to (ii) assortative mating, which describes the correlation between mating partners in vulnerability for ADHD or ASD. Second, the use of poorly or unscreened controls for common disorders will result in decreased power in association analysis and decreased SNP-heritability estimates. In particular, for MDD with a lifetime prevalence of around 15%, the anticipated underestimation of the SNP-heritability was analytically derived at 28% when none of the controls would be screened. An updated equation was provided to properly scale the SNP-heritability when including unscreened controls (Equation 3 in **Chapter 6**). When aiming to analyze the polygenic effects in psychiatric disorders, it is advisable not to use trio data and to properly scale the SNP-heritability when applying data with unscreened controls.

Research has found significant partner-resemblances for psychiatric disorders, that is, mating partners are more often concordant in psychiatric disorder-status than expected by chance. This phenomenon is often referred to as assortative mating. However, the potential consequences of these partner-resemblances have not been quantified and have been left implicit despite available theory in the quantitative genetics literature. Therefore, in **Chapter 7**, boundaries were quantified for the anticipated consequences for disorder prevalence and heritability under various inevitable assumptions. The consequences are most pronounced when partner-resemblance is attributable to phenotypic assortment (partner-resemblance driven by the psychiatric trait), and are reflected in increased population prevalence and heritability in the offspring generation. From the first generation in which assortative mating takes place, the consequences add generation after generation to reach equilibrium asymptotically over generations. Because of this equilibrium, assortative mating is unlikely to balance the impact of reduced fecundity of psychiatric patients in the long term, as analytically derived in **Chapter 7**. Modeling suggests that the heritability of none of the psychiatric disorders considered is likely to increase with more than 5% from one generation of assortative mating (or 13% over several generations). The population prevalence will increase most for rare disorders with high heritability, such as the prevalence of ASD that might maximally increase 1.5-fold after one generation of assortative mating (or 2.4-fold over several generations). While emphasizing the limitations inherent to the inevitable model assumptions, genetic theory suggests that the consequences of assortative mating are, at most, modest for the heritability, but may be considerable for the population prevalence.