Chapter 9: Summary

This thesis focused on the genetics of psychopathology across the lifespan. Genetic contributions to measures of psychopathology were estimated based on twin and family models, and on measured genotypes (i.e., single nucleotide polymorphisms). **Chapter 1** broadly outlined the methods used in behavior genetics, and genetic epidemiology as applied in this thesis, and discussed how measured genetic variants and environmental exposures can be incorporated into genetic studies.

In **chapter 2,** the results were presented of a cohort sequential study of the genetic and environmental influences on Symptoms of Anxiety and Depression (SxAnxDep) between age 3 and 63. Symptoms of anxiety and depression were measured in twins participating in research of the Netherlands twin register (NTR).[1-3] Young twins were rated by their mother at age 3, 7, 10, and 12. Self-report data for these twins was available at age 14, 16 and 18 years, and for adolescent and adult twins of 14 years and older, who participated in up to 8 waves of data collection in the Adult Netherlands Twin Register. SxAnxDep were assessed using an age appropriate version of the Anxious-Depression subscale of the Child Behavior Check List (CBCL; ages 3 through 12) and the Youth or Adult self report (YSR and ASR) inventories (ages 14 through 63).[4] The availability of twin data allowed us to estimate the proportion of variance in SxAnxDep that was explained by genetic and the environment at different ages. The availability of repeated measures (up to a maximum of 8 repeated measures on some participants) allowed us to estimate to what extent the genetic and environmental factors at one age played a role at a later age (transmission), and to what extent novel genetic and environmental factors (innovation) were important. Specifically, after organizing the data into 2 year age bins, we fitted a genetic simplex model to obtain this information concerning stability and innovation.

Results showed a decrease in the heritability of SxAnxDep between childhood and adulthood. The heritability was around .60 in childhood, and decreased to around .40 in adolescents and adults. This decrease was caused by an increase in environmental variance that outpaces a simultaneous increase in genetic variance between the ages of 3 and 18. After age 18 the genetic variance in SxAnxDep remained very stable, genetic influences are highly correlated between ages (around .90), and new genetic variance (innovation) is absent after age 18, except for time point specific sources of genetic variance. The environmental variance in

SxAnxDep was also transmitted from age to age in adulthood. However at each age in adulthood, environmental innovation was present, as new environmental sources of variance (environmental innovation) were present. With increasing age, the transmission of environmental effects rose, as the new environmental influences on SxAnxDep reduced somewhat. This process resulted in increasing environmental correlations between subsequent SxAnxDep scores with increasing age.

Chapter 3 focused on internalizing and externalizing psychopathology. Internalizing psychopathology included measures of depressive and anxiety disorders, and externalizing psychopathology included measures of attention deficit hyperactivity disorder (ADHD), oppositional defiant disorder (ODD) and conduct disorder (CD). The aim of the analyses was to describe the transition between childhood and adolescence and the co-morbidity of internalizing and externalizing psychopathology over development.

Data from the Avon Longitudinal study of parents and children (ALSPAC) were analyzed in Bristol (UK). Psychopathology was measured at age 7, 10, 13, and 15 using the DAWBA diagnostic interview.[5] The DAWBA provide both a diagnosis and an ordered categorical score, with higher scores indicating higher risk to fulfill the criteria for the disorders. The ordered categorical scores, denoted DAWBA bands, are statistically more informative than the binary diagnosis.[5]

Growth mixture models were fitted to identify categorically distinct trajectories for both internalizing and externalizing psychopathologies. For both internalizing and externalizing disorders, we expected an increasing trajectory, a decreasing trajectory, a stable high category and a large and stable low category. Finally, a single model was fitted modeling the co-occurrence between trajectories of internalizing and externalizing psychopathology.

The results of the analyses revealed that internalizing psychopathology was best captured by a model with 5 distinct trajectories. In addition to two trajectories with a consistently low and very low risk for internalizing psychopathology, the other 3 trajectories were a decreasing risk for internalizing psychopathology, an increasing risk of internalizing psychopathology, and an adolescent onset risk of internalizing psychopathology. Externalizing psychopathology was also characterized by 5 distinct trajectories. Four were similar to internalizing psychopathology. However, instead of an adolescent onset risk trajectory, a

trajectory of stable high risk of externalizing disorders through childhood and adolescence was identified.
Combined analysis of internalizing and externalizing categories revealed that increasing internalizing and increasing externalizing trajectories co-occur, as did decreasing internalizing and externalizing trajectories. However, the adolescent onset internalizing trajectory was independent of high externalizing trajectories, and the persistent high externalizing trajectory was mainly associated with the decreasing internalizing trajectories. Sex and early life environmental risk factors predicted externalizing and, to a lesser extent, internalizing trajectories. The analysis reveals the need to screen for co-morbidity in the case of either early onset externalizing or internalizing problems. The only exception seems to be adolescent onset internalizing problems, which are not related to a high risk for externalizing problems.
Thus, Chapters 2 and 3 reveal both genetic and phenotypic continuity between childhood psychiatric problems and adolescent and adult psychiatric outcomes.

Chapter 4 reports on a genome-wide association study (GWAS) on preschool internalizing problems. Results from 3 cohorts, including NTR, were meta-analyzed, and the variance explained by all measured genetic variants (SNPs) was estimated. Three cohorts participated: the NTR, The Western Australian Pregnancy Cohort Study (Raine), and the generation R from Rotterdam. Internalizing scores in 2-3 year olds, based on the Child Behavior Check List[6], were harmonized. Genotypes were imputed against HAPMAP 2.[7] After post imputation quality control, 2.4 million SNPs were available for analysis. A total of 2037 children had genotype and internalizing scores available in generation R; 1475 children from 1031 families in the NTR; and 1084 children in the Raine cohort. In each cohort, the association between Internalizing and genotypes was tested, with the inclusion of principle components to correct for population stratification and sex. This was followed by meta-analysis of 4566 children. The variance explained by all SNPs for internalizing behavior was estimated using 2 methods, Genomic Relationship Matrix Restricted Maximum Likelihood (GREML) as implemented in Genomic Complex Trait Analysis (GCTA) and density estimation (DE).[8; 9]
The SNPs were found to explain between 13 and 43% of the total variance in internalizing problems. As the heritability in twin studies was estimated at 59%, this implies that the genetic variants analyzed in this study captured between 22

and 72% of the genetic variance. The meta-analysis revealed no SNPs associated with Internalizing problems at a genome-wide significant p-value < $1x10^{-8}$. In two regions, there were SNPs, which reached a p-value below $1x10^{-5}$ in the meta-analysis. One SNP was located in an intergenic region on chromosome 9. The other region was on chromosome 20, and included SNPs of the PCSK2 gene. PCSK2 is an important protein in the processing of pro-insulin to insulin, and PCSK2 variants are correlated with insulin resistance,[10; 11] myocardial infarction[12] and age at menarche[13]. The link between depression and cardiovascular disease has long been recognized. Post hoc analysis of SNPs that were previously associated with adult internalizing psychopathology, psychopathology that usually presents in childhood (ADHD, conduct disorder), or psychotic disorders, and of SNPs in candidate genes[14] did not show a significant association of any of these SNPs with internalizing problems in preschool children. Collectively, the SNPs previously associated with adult internalizing disorders did not show lower p-values than expected by chance. However, the SNPs previously associated with adult internalizing disorder, adult or childhood psychiatric disorders usually diagnosed in childhood or psychotic disorders did collectively show lower p-values than expected by chance in the GWAS of preschool internalizing problems. Adding SNPs associated with treatment response diminished this signal, while subsequently adding SNPs in candidate genes slightly strengthened the signal. The analyses performed in chapter 4 show that childhood preschool internalizing problems are heritable, and that a substantial part of this heritability can be explained by common genetic variation. The results further show that childhood internalizing problems are a complex trait, and no single genetic variant explains a substantial part of the phenotypic variation. The significant signal of SNPs previously associated with adult and other psychiatric disorders was suggestive of common genetic causes.

Chapter 5 looked at polygenic score prediction of childhood psychopathology. The most recent schizophrenia GWAS meta-analysis included 36,989 cases and 113,075 controls and revealed 108 loci significantly associated with schizophrenia.[15] This study provided the starting point to test for associations between genetic risk for schizophrenia and childhood psychopathology directly at the molecular genetics level. Polygenic risk scores were calculated based on the schizophrenia GWAS to predict childhood psychopathology scores at ages 7, 10, 12, and 15 years. The analysis was performed in samples from the Netherlands Twin

Registrer (NTR) and the Avon longitudinal study of parents and children (ALSPAC). In both cohorts, DSM based measures of anxiety, depression, attention deficit hyperactivity disorder (ADHD), and oppositional deviant disorder, and conduct disorder (ODD & CD) were available. The NTR scores were based on the DSM oriented CBCL or YSR scales[16] and the scores in ALSPAC on the DAWBA bands.[5] The regression of the psychopathology phenotype on the polygenic risk score included as covariates principle components to control for population stratification and sex. Meta analysis of the results of both studies revealed an false discovery rate (FDR) corrected significant association between schizophrenia risk and anxiety at age 10. This result seemed mainly to be driven by results in the NTR. The analysis further revealed associations at uncorrected $p < 0.05$ between schizophrenia polygenic risk scores and anxiety at age 7 and depression at age 7, age 10, and age 12 to 13. Based on these results the initial hypothesis of a broad positive association with childhood psychopathology was not confirmed. Post hoc test revealed a stronger effect on internalizing psychopathology than on externalizing psychopathologiy. Note that the results were consistent with the PGC cross disorder study[17]. The PGC cross disorder group found a genetic correlation between adult MDD and schizophrenia, but not between schizophrenia and ADHD[17].

In **chapter 6** the aim was to replicate a finding reported by the rat genome sequencing and mapping consortium. This consortium obtained evidence for an association between the CTNND2 gene and anxiety in rats.[18] Replication was sought in a sample of adult participants from NTR and from the Netherlands Study of Depression and Anxiety (NESDA). The phenotype was based on the CIDI anxiety Diagnosis. All individual SNPs in the CTNND2 gene were tested for association with anxiety in the NTR/NESDA sample. To test for an association between CTNND2 and Major Depressive Disorder (MDD), Bipolar Disorder, and Schizophrenia lookups in the results of the PGC mega and meta-analyses of MDD, bipolar disorder and schizophrenia were performed. No SNPs reached significance for any disorder corrected for the number of SNPs tested. A gene-based test for enrichment of all P-values in the gene was performed, and revealed tentative evidence for enrichment of the CTNND2 gene in anxiety, MDD, and schizophrenia, but not bipolar disorder. This chapter shows that follow up of findings in animal studies can reveal potential associations in human data, and may provide a useful addition tool to explore genetic associations.

The current REML model as implemented in GCTA allows for genotype by environment (GxE) moderation in the case that the environmental exposure is dichotomous.[8] In **chapter 7** the model specified in GCTA was extended to include continuous moderation of genetic and environmental effects, given a sample of closely related and nominally unrelated individuals. This involved a reparameterization of the model proposed by Zaitlen et al.[19]. This resulted in a model in which the concurrent moderation of the variance specifically attributable to SNPS, and the total additive genetic variance can be tested. We applied this model to symptoms of anxiety and depression (AnxDep), attention problems (AP), height, and body mass index (BMI). The analysis revealed that the (genetic) variance components for the different phenotypes were differently moderated by age (or birth year in the case of height).

Fitting the Zaitlen model to the four phenotypes revealed moderated additive genetic effects (~40%) for AnxDep and A) and strong additive genetic effects for Height (90%) and BMI (75%). The portion of variance explained by measured SNPs was moderate for AnxDep (~10%) and AP (~11%), but larger for height (~55%) and BMI (~40%).

We proceeded to fit moderation models. The variance explained by SNPs, additive variance and residual variance for AnxDep were not moderated. For BMI the additive genetic variance and the residual variance were moderated, but the variance explained by SNPs was not. In the analysis of height and AP, the residual variance was moderated, but the additive variance or variance explained by SNPs was not. The analysis revealed differences in the way age and or birth year moderated these 4 phenotypes.

chapter 8 presented a model that allows for the estimation of genetic (co)variance between traits based on measured genotypes. This in itself is not new. In bivariate GCTA[20] and GEMMA[21], it is also possible to estimate the genetic covariance between traits given all SNPs. However, increasing the number of traits or the number of separate genetic variance components will increase computational burden. In this chapter, a method was developed that breaks the multivariate analysis up in to a series of univariate analysis. This method relies on the fact that the variance of the sum of two variables equal to the variance of each individual variable and twice the covariance between two variables. Simulations showed that this method yields unbiased estimates of genetic (co)variance. Moreover, approximate standard errors were obtained using a Taylor approximation, first used by Visscher et al.[22] The model was extended to

allow for multiple genetic effects. We simulated data to show that the model produced unbiased estimates of separate genetic covariance matrices for each genetic relatedness matrix, and that their standard errors are correctly.
The method was applied to 24 items derived from the NEO PI personality inventory.[23; 24] Twelve items indicate the construct of neuroticism, 12 items indicate the construct of extraversion. We estimated the genetic correlation between the items based on SNPs in a sample containing related individuals. Based on previous results[25] neuroticism items were expected to correlate positively, extraversion items were expected to correlate positively. Negative correlations were expected between the neuroticism and extraversion items. Only a modest proportion of variance in the individual items was attributable to SNPs (0 to 14.8% for extraversion, 5 to 16.7% for neuroticism). Variance explained by SNPs in the total scale was 6.3% for extraversion and 22.6% for neuroticism. Despite these moderate SNP heritability's we were able to retrieve the expected covariance structure. The first principle component of the genetic covariance matrix of the items separated the neuroticism and extraversion items. A second feature of the model is the possibility to estimate a separate genetic (co)variance for multiple sets of SNPs. This would allow for estimation of separate co-variances between traits for multiple distinct sets of SNP. Each of these sets of SNPs could be selected to reflect a set of genes in a biological pathway, a specific chromosome or any other biologically interesting subset of all measured SNPs.

In **chapter 10** I discuss the developments in the field of behavior genetic in the periode that I was writing this dissertation. I discuss how current (methodological) developments will allow a deeper understanding of the genetics of complex traits in the coming few years.