



CHAPTER 8

SUMMARY

Chapter 8. Summary

In this thesis I applied various computational approaches to SNP and twin data to estimate the relative importance of genetic factors in a range of the complex human phenotypes and explored the shared genetic risk factors between correlated traits.

In chapter 2 the GoNL reference set was used to combine data from two Dutch childhood cohorts, NTR (N = 3,102) and GENR (N = 2,826), through cross-platform imputation. The estimates of SNP-heritability of childhood height were similar across GRMs, built from: 1) pre-combined and cross-platform imputed ($h^2 = 51\%$), 2) cross-platform imputed and post-combined ($h^2 = 52\%$) and slightly lower for 3) just combined datasets ($h^2 = 43\%$). Correction for cohort resulted in $\approx 2\%$ drop in the SNP-heritability estimates for each combination approach. Correction for the Dutch PCs alone resulted in $\approx 11\%$ drop for imputed and combined data, suggesting that imputation against the GoNL reference did not alter similarity between individuals. The SNP-heritability estimates, corrected for both cohort and Dutch population structure in addition to age and sex, were within the range from 32% to 41%. Our results suggest that even a small number of SNPs that overlap between cohorts, allows the estimation of genetic relationships between individuals correctly. We also showed that imputation with a reference set reduces the amount of platform stratification in comparison to imputation without a reference set. Although imputation with a reference set allows for combining the datasets, genotyped on different platforms with little overlap, the cohort should be always included as a covariate.

In chapter 3 the SNP-heritability of a range of childhood behavior problems was estimated based on two Dutch cohorts, NTR and GENR. With increased sample size, we were able to detect the significant SNP-heritability for attention deficit/hyperactivity ($h^2 = 0.37 - 0.71$, SE = 0.14 - 0.22), externalizing problems ($h^2 = 0.44$, SE = 0.22) and total problems ($h^2 = 0.18$, SE = 0.10), rated by mother or teacher. Application of sensitivity analyses involving the exclusion of extreme cases or phenotype quantile normalization, did not affect the statistical significance of the estimates, but resulted in decreased SNP-heritability estimates. The implication of these results would be further continuation of large collaborative GWAS efforts, aiming to detect loci, influencing childhood behavior problems.

Following the results of chapter 3 and for the sake of comparison between heritability estimates, resulting from different raters' perspectives, we explored the rater shared and unique contribution to the variation of the child behavior problems in chapter 4. We estimated the heritability of maternal and paternal ratings of the child behavioral problems, based on CBCL 6-18 empirical scales, in a large Dutch cohort, comprising 12,310 twin pairs at around age 7. On average, mothers rated their children as scoring higher on problem scales compared to fathers. The parental agreement was between 0.62 and 0.74 across all scales. A large part of the heritability was shared between parents, which indicated that to a large extent, parents perceive similar behavioral problems in their

children. A smaller part of the heritability was unique, indicating behavior of the child expressed in presence of one parent exclusively. Since the heritability for the behavior both parents agree upon is large, it suggests pulling paternal ratings together to increase the power in GWAS projects, while correcting for mean differences, is a valid approach. In chapter 5 the genetic correlations between Subjective Well-being (SWB) and two personality traits, Neuroticism (NEU) and Extraversion (EXT), were estimated. I employed the bivariate analysis, implemented in GCTA software, and used both distantly and closely related individuals ($N \approx 9,000$) to estimate the total heritability and genetic correlation and those explained by SNPs, present on current genotyping platforms. The total heritability estimates were 32%, 37% and 42% for SWB, NEU and EXT and genetic correlation estimates were $-.70$ ($SE = .03$) and $.48$ ($SE = .03$) between SWB and NEU and SWB and EXT, respectively. The SNP-heritability for SWB was 7%, 10% for NEU and 16% for EXT. The genetic correlation, based on SNPs was larger between SWB and NEU ($r_g = -.80$), than between SWB and EXT ($r_g = 0.18$), which was in a contrast to the observed correlation ($r = -.43$ and $r = .32$, respectively). A large genetic correlation between SWB and NEU suggests that common loci between these phenotypes are likely to be detected. In contrast, despite the large observed correlation between SWB and EXT, environmental rather than genetic influences could be more pronounced in explaining the role of EXT in SWB variation.

Chapter 6 describes an application of a recently developed method (GWIS) [222] to analytically derive the results of HOMA-B and HOMA-IR meta-analysis. We evaluated the performance of the method by comparing the summary statistics of current study to the summary statistics of previous meta-analysis of HOMA-B/-IR. Sample size was increased in GWIS in comparison to previous analyses and, thus, there was a gain in power. We replicated seven loci from previous meta-analyses and detected four new loci for HOMA-B. For HOMA-IR, two loci were identified previously and three loci in the current analysis were novel. In addition, we explored the genetic correlation between HOMA-B/-IR and range of glycaemic and metabolic traits, namely FI, FG, high-density lipoprotein (HDL), low-density lipoprotein (LDL), total cholesterol (TC), triglyceride (TG), body mass index (BMI) and type-2 diabetes (T2D). We found significant genetic correlations between FI and HOMA-B/-IR ($r_g = 0.76/0.98$, $SE = 0.05/0.005$); FG and HOMA-B/-IR ($r_g = -0.38/0.49$, $SE = 0.12/0.07$); BMI and HOMA-B/-IR ($r_g = 0.39/0.62$, $SE = 0.05/0.05$) and T2D and HOMA-IR ($r_g = 0.53$, $SE = 0.08$). We did not find significant genetic correlations between HOMA-B/-IR and LDL/TC and between T2D and HOMA-B. Results from analysis of analytically derived HOMA-B/-IR genome-wide summary statistics, demonstrate the advantage of GWIS method over direct HOMA GWAS. GWIS allows for analytical derivation of summary statistics in partly overlapping samples and thus gain in power to detect new genetic loci when studies do not have both phenotypes measured. It also allows for the more powerful LD score regression analysis as derived summary statistics are similarly based on the larger sample size as well.

In chapter 7 the shared genetic aetiology between two comorbid disorders was studied, namely between T2D and MDD through the measure of Fasting Insulin (FI), fasting Glucose (FG), β -cell function (HOMA-B) and insulin resistance (HOMA-IR). A Polygenic Risk Score (PRS) method was used to predict MDD status in a combined sample of NTR and NESDA. We extended the study by inclusion of MDD subtypes, characterized by increased or decreased appetite. PRS profiles based on FI, FG, HOMA-B and HOMA-IR with various cut-offs for significance did not predict MDD or their subtypes. We also selected a set of SNPs, previously reported for association with glycaemic traits, lipids, waist-to-hip ratio, and weighted them by summary statistics from previous MDD mega-analysis. In addition, we used LD score regression to estimate the genetic correlation between glycaemic traits and MDD and its risk factors (Depressive Symptoms and Neuroticism). None of the SNP sets significantly predicted MDD status or its subtypes. In addition, BMI as a covariate did not have a large effect on the estimates. LD score regression showed a small overlap between HOMA-IR/FI and Depressive Symptoms ($r_g = 0.16$, $SE = 0.07$ and $r_g = 0.17$, $SE = 0.07$ respectively), however, this was not statistically significant. These results suggest that FG and FI as well as indices of Insulin resistance (HOMA-IR) and β -cell function (HOMA-B) have distinct genetic aetiology with MDD and its symptoms. Therefore future studies should focus on possible other influences, such as behavioral, demographical and socio-economic factors.

In conclusion, computational approaches along with Genome of the Netherlands (GoNL) reference set formed the basis of this thesis, in which the previously collected data from Netherlands Twin Register (NTR) and data, meta-analyzed by various genetic consortia, were explored. Analytical derivation of summary statistics from partly overlapping samples generated new data for the future use and research in T2D and related glycaemic traits. Insights gained from analysis of co-morbid phenotypes or the same phenotype, rated by different informants, suggested different strategies to analyze such data by including new risk factors or employing new models, depending on shared - or distinct genetic aetiology. In this thesis the genetic analyses, which are usually performed in unrelated subjects, were considered with a focus on twin data, employing the information from relatives to increase power. From a molecular perspective, the high resolution of GoNL reference set, helped to reduce the bias, introduced by different genotyping platforms, preserving as much information as possible of the genetic variation in Dutch population. Overall, a range of different approaches employed in the current thesis, showed that efficient use of existing genotype and phenotype data together with new analytical approaches should be extensively exploited to gain new biological insights.