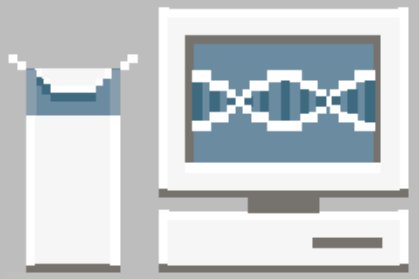
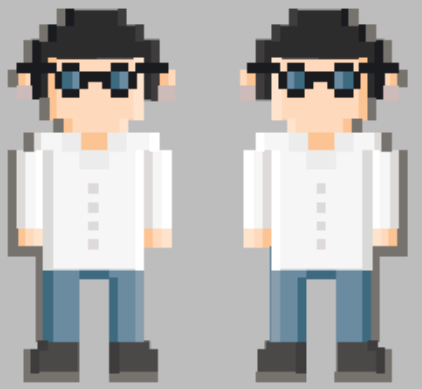


IRYNA FEDKO

COMPUTATIONAL APPROACHES IN GENETICS WITH A FOCUS ON GoNL AND TWIN STUDIES



COMPUTATIONAL APPROACHES IN GENETICS WITH A FOCUS ON GoNL AND TWIN STUDIES

IRYNA FEDKO

Computational approaches in genetics with a focus on GoNL and twin studies

Iryna Fedko

Reading committee

Prof. dr. Gareth Davies
Dr. Inga Prokopenko
Dr. Morris Swertz
Prof. dr. Brenda Penninx
Dr. Sophie van der Sluis
Dr. Gonneke Willemsen

Paranymphs
Jorien Treur
Laura Wesseldijk

Printed by:	Ipskamp Drukkers (https://www.ipskampprinting.nl/)
Artwork and inner layout by:	Aleksandra Masiukevych (a. masiukevych@gmail.com)
ISBN:	

VRIJE UNIVERSITEIT

Computational approaches in genetics with a focus on GoNL and twin studies

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Gedrags- en Bewegingswetenschappen
op donderdag 15 december 2016 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Iryna Fedko

geboren te Kiev, Oekraïne

promotor: prof.dr. D.I. Boomsma
copromotor: dr. J.J. Hottenga

Table of contents

CHAPTER 1

INTRODUCTION

Imputation in genetics studies: concept and application

During the past decade genetic studies, in particular Genome Wide Association Studies (GWAS)[1], grew from small studies to large international collaborations, aiming to detect new genetic loci associated with phenotype of interest. Such collaborations require pulling data together in case of mega-analysis or puling summary statistics of independently performed analysis together in case of meta-analysis. Data usually comes from different cohorts and are often genotyped on different platforms, which may have overlapping Single Nucleotide Polymorphisms (SNPs) between each other to a different extent. In genetics, imputation is a predictive technique which allows to assign (impute) unobserved or missing genotypes based on an individuals' haplotype and on a reference set, representative of the population a person comes from. Imputation is often used to overcome the issue of missing genotypic data [2-4]. Imputation increases power to detect new genetic loci, and allows for cohorts combination in GWAS [5]. To achieve high quality imputation results, however, several steps should be performed, including quality control (QC) before imputation, ancestry-differences inference, appropriate reference set selection and stringent post-imputation QC. Several imputation software packages exist such as IMPUTE [6] and MACH [7]. Each package provides an imputation quality metric, which aims to quantify the performance of imputation for particular SNP. Poorly performed imputation may result in a larger number of SNPs, which will fail the post imputation QC and reduce the SNP coverage in GWAS. Imputed genotypes are usually expressed as probabilities of observing one of the three genotypes or as dosages, which reflects the expected allele counts and therefore introduce uncertainty. Imputation is largely dependent on various parameters, which may affect imputation accuracy, and have to be taken into consideration prior to imputation or even prior to genotyping of the samples [5, 8, 9].

Choosing a reference set for imputation. Genome of the Netherlands Project (GoNL)

The quality of imputation procedures largely relies on the reference set that is representative of a population, from which a person is drawn [10]. This is due to the biological mechanism upon which imputation algorithms are based. All SNPs are not independent from each other and often correlated, which means they are in Linkage Disequilibrium (LD) and are inherited together. The inference of missing genotypes is possible if a block of correlated SNPs (haplotype), from which the SNPs are missing, is known. Initially, imputation was performed in samples consisting of related individuals, where grandparents and parents were genotyped with higher resolution then children [11]. Based on identity by descent (IBD) information shared between relatives missing genotype data in children were inferred from their closest ancestors. In this way the grandparents and parents formed a reference set for children. The idea was that relatives shared long stretches of their genome with each other, which are inherited together (haplotypes). When imputation is carried out for unrelated individuals, genotypes from their relatives are not available and therefore the reference set with data of other

individuals, that are the closest to the sample with respect to their ethnic background, is required. Thus, the European population is usually imputed against HapMap [12, 13] or 1000G [14, 15] CEU reference panels. HapMap is the first project aimed to characterize the variation in the human genome. The 1,184 individuals from 11 populations were genotyped for 1.6 million SNPs. 1000G is the next project, where the genome of 1,092 individuals from 14 populations were sequenced and 38 millions SNPs were provided. Using a reference set from a population closest to the dataset, with respect to their genetic background, ensures that haplotypes are representative of the individuals in the study and occur at the same frequency.

In the Genome of the Netherlands project (GoNL) a group of 250 trio's of two parents and their offspring were whole genome sequenced with an average 12x depth to study genetic variations within the Dutch population [16]. The trios were selected to represent the genetic variation across all provinces in the Netherlands and several large biobanks from the Netherlands, including the Netherlands Twin Register, which collaborates in Biobanking and BioMolecular resources Research Infrastructure (BBMRI) contributed DNA samples for this purpose. The GoNL project allows constructing a Dutch population reference set, which can be used to impute Dutch GWAS samples as well as explore imputation quality, particularly for Dutch specific variants. If a reference set is closer to the dataset with respect to the most common recent ancestor, unrelated individuals will share larger parts of the genome between each other and longer reference haplotypes will be available if individuals belong to the same ethnic group. Moreover, some alleles may be population specific and represented at different frequencies in different populations. The European population is diverse and represented by changes in alleles frequencies from North to South and from East to West [17]. If data are imputed based on European haplotypes and some SNPs appear at low frequency in a CEU panel, they will be poorly imputed due to weak LD with neighbouring SNPs and subsequently filtered out in GWAS. In contrast, imputation based on a population-specific dataset may result in better quality of imputation of such SNPs [18].

Imputation: beyond GWAS

In genetic studies, imputation of genotypes allows for the pooling results of different studies together for meta-analysis, ensuring that all studies are imputed to the same SNPs. In GWAS each SNP is examined individually, for example by regression analyses, and the variation in all significant SNPs together usually explains a small to medium proportion of variation in the phenotype. To quantify the contribution of all SNPs, that do not pass the stringent threshold for significance in GWAS studies, to the phenotype of interest, another approach was developed. Genomic-relatedness-matrix restricted maximum likelihood (GREML) [19] methods, implemented in software such as GCTA [20, 21], allows estimation of the SNP-heritability, that is the variation in the phenotype accounted for by all SNPs, genotyped and imputed. GREML uses a Genomic Relatedness Matrix (GRM), where the relationships between individuals are inferred based on SNPs

rather than on known pedigree structures. The raw genotype data are required to calculate the GRM between individuals. The raw data from multiple cohorts may be required in order to gain power to estimate the SNP-heritability. If data were genotyped on different platforms, what SNP overlap is required to ensure that relationships between individuals are estimated correctly? Will the combination of data result in a bias of SNP-heritability estimates? These questions are explored in this thesis.

Population structure and imputation

A population is not homogeneous across continents or even across a single country. Population subgroups exist within a population and differences in their allele frequencies correlates with the change in geographical coordinates (population stratification) due to ancestry differences [22, 23]. Importantly, such population stratification can confound results of both GWA and GREML studies. Principal Component Analysis (PCA) is typically used to calculate Principal Components (PCs) and correct for the systematic difference in allele frequencies due to ancestry [24]. Usually PCA is performed before the imputation and PCs are further used in analysis of imputed data as covariates.

Heritability: concept and methods

Heritability is the ratio of genetic variation over the total phenotypic variation and reflects the degree of genetic determination [25]. Broad-sense heritability is the proportion of all additive and non-additive (dominance, interaction, epistasis) genetic variation in the phenotypic variation. Narrow-sense heritability is the proportion of additive genetic factors variation only. In this thesis, the heritability, estimated using twin and SNPs data, was considered. Before the human genome was decoded and information about allele frequency variation became available, twin studies were the most commonly used method to estimate the heritability [26]. Monozygotic twins (MZs) are genetically (almost) identical and dizygotic twins (DZs) share on average half of their segregating genes. Therefore, the correlations between pairs of MZs and DZs can be compared and variation of the phenotype can be decomposed into the variation in 1) additive genetic and 2) common environmental or dominant genetic and 3) unique environmental factors (which also includes error). Proportion of genetic factors will comprise twin-heritability [27]. If genotype data are available then SNP-heritability can be estimated directly from SNPs, available on a current genotyping platform or imputed, using the GREML method introduced above. If genotype data are not available, but the meta-analysis summary statistics are, the LD score regression approach is the alternative method to estimate the SNP-heritability [28]. Here, the square test statistic, obtained for a SNP in a GWAS, is regressed on the LD score, which is the sum of all LD between a particular SNP and its neighbouring SNPs. Under a polygenic model many SNPs contribute to variation in a trait. Therefore a SNP in strong LD with its neighbours has a higher likelihood of tagging a causal SNP than a SNP in weak LD. A SNP with a higher LD score is expected to have on average a higher test statistic, while population stratification is expected, on

average, to inflate the test statistic for all SNPs equally. Therefore, intercept in the LD score regression will reflect the genome wide increase in test statistics due to population stratification, whereas the slope will reflect the increase in test statistics due to true polygenic effects on the trait. Both GREML and LD score regression assume the additive model and are often compared to the twin-heritability estimates.

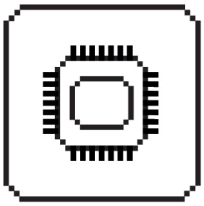
Shared and unique aetiology of human complex traits

All methods discussed above allow for estimation of heritability and also, for bivariate phenotypes, allow estimation of genetic correlations between phenotypes, i.e. the shared genetic background. This can be useful, if two phenotypes are co-morbid and it is important to know to which extent their co-morbidity is accounted for by genetic factors. The shared genetic factors may unravel new biological pathways that are relevant for both phenotypes. Genetic correlation within domains of psychological or disease phenotypes are expected, however, the co-morbidity between disease and psychological phenotypes are of particular interest. For example, epidemiological studies established the association between Type 2 Diabetes (T2D), insulin resistance and Major Depressive Disorder (MDD) [29-31], but their genetic architecture has not been studied within the context of shared biological basis. Another example is Subjective Wellbeing, which has been consistently reported to correlate with personality traits [32, 33] and requires further exploration using genome wide data. It is also possible that the same phenotype can be represented by shared and unique parts, when assessed by multiple raters. For example, in childhood psychopathology phenotypes are often assessed by different raters, such as parents, teachers or peers. These ratings usually correlate to a certain extent (genetically and phenotypically), but not perfectly, implying the presence of a shared part of the phenotype assessed by multiple raters, but also unique one, observed by different raters exclusively, or rater bias [34-36]. Therefore, the genetic factors may contribute to the variation in the shared and unique parts of each phenotype. The consequence is that in molecular studies different ratings, such as maternal and paternal, might be represented by the same or different loci in the genome. In this thesis we explored the genetic correlation between various phenotypes aiming to quantify their shared and unique aetiology.

This thesis

Chapter 2 explores the GoNL reference set as the basis to resolve platform stratification between cohorts, allowing a combination of the two Dutch cohorts genotyped on different platforms with little SNP overlap. In Chapter 2, data from two Dutch cohorts, the Netherlands Twin Register (NTR) and Generation R (GENR), were combined using imputation with the GoNL reference set and three approaches to build a Genetic Relatedness Matrix (GRM) were compared. We evaluated the performance of each approach, estimating the SNP-heritability of childhood height. In **Chapter 3** SNP-heritabilities of the child behavior problems were estimated based on the combined

cross-platform imputed data, which were described in Chapter 2. We looked at various phenotypes in the childhood psychopathology domains, namely Attention Deficit and Hyperactivity problems, Internalizing and Externalizing behavior, Pervasive Developmental Problems and non-verbal Cognition. Most of the phenotypes were rated by mothers, and Attention Problems and Externalizing were also rated by teachers. Data combination allowed for an increased sample size and thus power to estimate SNP-heritabilities. In **Chapter 4**, bivariate analysis of twin data and exploration of rater effects on heritability estimates were performed as a follow-up of SNP-heritability results, described in Chapter 3. Based on this information, future molecular studies may analyze different ratings separately or in combination. A comprehensive set of behavioral and emotional problems was analyzed in 7-year old twins whose fathers and mothers rated them on all CBCL 6-18 empirical scales, including Internalizing (Anxious/ Depressed, Withdrawn/Depressed, Somatic Complaints), Externalizing (Rule-Breaking and Aggressive Behaviors), as well as Social, Thought, Attention Problems, Dysregulation Profile and Total Problems scales. **Chapter 5** analyzed a dataset from adults, which was cross-platform imputed, as described in Chapter 2. NTR has collected genotype data across different time points and various genotyping platforms and have been cross-platform imputed against GoNL reference to allow the combination of the data. With increased sample and using recent method to estimate SNP-heritability, including family members [37], we explored the genetic correlations between Subjective Well-being (SWB) and personality traits, such as Neuroticism (NEU) and Extraversion (EXT). In **Chapter 6**, we analyzed MAGIC consortia summary statistics and computed homeostatic model assessment of β-cell function (HOMA-B) and Insulin resistance (HOMA-IR) from Fasting Insulin (FI) and Fasting Glucose (FG) meta-analyses results. Here, we employed the newly developed Genome Wide Inferred Statistics (GWIS) method [38], which allows to analytically infer the statistics of complex non-linear functions (i.e. HOMA) from its compounds (i.e. FI and FG). We compared effects of HOMA-IR and HOMA-B significant SNPs with their effect on FI and FG. Finally, we predicted the MDD status in **Chapter 7** by computing the Polygenic Risk Score based on Fasting Glucose, Fasting Insulin, HOMA-B and HOMA-IR meta-analysis summary statistics.



CHAPTER 2

ESTIMATION OF GENETIC RELATIONSHIPS BETWEEN INDIVIDUALS ACROSS COHORTS AND PLATFORMS: APPLICATION TO CHILDHOOD HEIGHT

This chapter is based on:

Iryna O. Fedko, Jouke-Jan Hottenga, Carolina Medina-Gomez, Irene Pappa, Catharina E.M. van Beijsterveldt, Erik. A. Ehli, Gareth E. Davies, Fernando Rivadeneira, Henning Tiemeier, Morris A. Swertz, Christel M. Middeldorp, Meike Bartels, and Dorret I. Boomsma. (2015). Estimation of Genetic Relationships Between Individuals Across Cohorts and Platforms: Application to Childhood Height. *Behavior Genetics*, *45*(5), 514-528.

Abstract

Combining genotype data across cohorts increases power to estimate the heritability due to common SNPs, based on analyzing a Genetic Relationship Matrix. However, the combination of SNP data across multiple cohorts may lead to stratification, when for example, different genotyping platforms are used. In the current study, we address issues of combining SNP data from different cohorts, the Netherlands Twin Register (NTR) and the Generation R (GENR) study. Both cohorts include children of Northern European Dutch background (n=3,102 + 2,826 respectively) who were genotyped on different platforms. We explore imputation and phasing as a tool and compare three GRM-building strategies, when data from two cohorts are: 1) just combined, 2) pre-combined and cross-platform imputed and 3) cross-platform imputed and post-combined. We tested these three strategies with data on childhood height for unrelated individuals (n = 3,124, average age 6.7 years) to explore their effect on SNP-heritability estimates and compare results to those obtained from the independent studies. All combination strategies resulted in SNP-heritability estimates with a standard error smaller than those of the independent studies. We did not observe significant differences in estimates of SNP-heritability based on various cross-platform imputed GRMs. SNP-heritability of childhood height was on average estimated as 0.50 (SE=0.10). Introducing a cohort as a covariate resulted in ≈2% drop. PCs adjustment resulted in SNP-heritability estimates of about 0.39 (SE = 0.11). Strikingly, we did not find significant difference between cross-platform imputed and combined GRMs. All estimates were significant regardless of the use of PCs adjustment. Based on these analyses, we conclude that imputation with a reference set helps to increase power to estimate SNP-heritability by combining cohorts of the same ethnicity genotyped on different platforms. However, important factors should be taken into account such as remaining cohort stratification after imputation and /or phenotypic heterogeneity between and within cohorts. Whether one should use imputation or just combine the genotype data depends on the number of overlapping SNPs in relation to the total number of genotyped SNPs for both cohorts, and their ability to tag all the genetic variance related to the specific trait of interest.

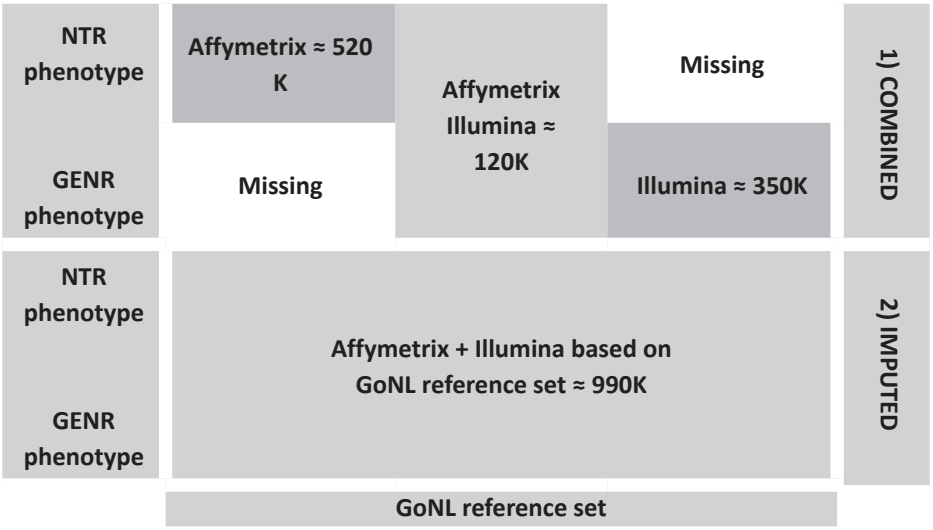
Introduction

Before embarking on Genome Wide Association (GWA) projects, the heritability of complex traits is often assessed in twin and family studies, or, more recently, assessed based on common single nucleotide polymorphisms (SNPs). Such SNP-based heritability can be estimated when genetic similarities between distantly related individuals are summarized in a genetic relatedness matrix (GRM), which then is used to predict their phenotype similarity [39-42]. This technique, known as genomic-relatedness-matrix restricted maximum likelihood [19] (GREML), is implemented, for example, in the software package GCTA [20] (Genome-wide Complex Trait Analysis). Estimating the heritability based on measured SNPs requires the availability of raw genotype and phenotype data. Therefore, these analyses are usually performed in one, or a few separate cohorts that contribute to a meta-analysis GWAS. However, in single studies, these SNP-based heritability estimates tend to have large standard errors due to small cohorts sample sizes. The large standard errors also result in variation in estimates between different studies for the same trait.

Here we investigated the possibility of combining individual-level genotype data across cohorts in order to obtain a larger and better GRM. A cross-cohort GRM will allow inclusion of all possible combinations of pairs of individuals, both within and between cohorts, and estimation of the genetic variance explained by common variants (SNP heritability) will likely improve. However, this requires sharing and pooling of raw phenotype and genotype data from multiple cohorts. For genotype data this likely means that data of multiple genotyping platforms need to be combined, which might lead to biased results due to "platform stratification", when relationships between individuals of different cohorts are estimated based on overlapping SNPs only. In case of GWA meta-analyses, each individual cohort performs its own imputation using a reference set (e.g. HapMap or 1000 Genome) and statistical analysis prior to the combination of results. This way, the confounding effects of genotyping platforms are avoided. SNPs showing platform stratification effects will be detected with heterogeneity testing and meta-analysis QC. With GREML analyses, the genotyped data of cohorts need to be combined at the SNP level. If different genotype platforms have been used for genotyping, a cross-platform imputation is required in order to combine genotypes from several cohorts and ensure that all individuals have the same SNP information to estimate the relationship between them.

In this paper, we compared approaches that combine autosomal genotype data from different cohorts and genotype platforms into a single GRM. We aim to address and resolve problems of stratification when cohorts differ in genotyping strategies and phenotype characteristics. Therefore, this study has two aims: 1) to allow the combination of genetic data from two cohorts, where participants are genotyped on different platforms with little overlap and 2) to explore the effect of three different strategies of combining such data on SNP-heritability estimates, when two cohorts are either cross-platform imputed (post- or pre-combined) or just combined (Figure 1).

Figure 1: Strategies of combing two cohorts genotyped on different platforms, when two cohorts are either: 1) combined or 2) cross-platform imputed.



We based our analysis on genotype data from two Dutch cohorts, the Netherlands Twin Register [43, 44] (NTR) and the Generation R study (GENR) [45, 46]. NTR recruits twin families across the Netherlands, whereas GENR targets a birth cohort from Rotterdam. The cohorts have genotyped their participants on different Affymetrix and Illumina platforms, respectively. We illustrate the imputation approaches and tested their performance using Principal Components Analysis (PCA) to check for stratification due to genotyping platform. Subsequently, we demonstrated the differences of using cross-platform imputation versus just combining datasets for childhood height.

The methods, considered to pre-combine and cross-platform impute the NTR and GENR genotype data, include combining both genotype data sets at the SNP level and then phasing (i.e. estimating haplotypes) the combined data as a single dataset. We phase combined data without - and with a reference imputation set using MaCH [7] and MaCH-Admix [47] and inherently impute. When a reference set was used, the data were imputed with reference to data from the Genome of the Netherlands (GoNL) project [16]. The GoNL imputation reference set is a resource of sequenced data from the Netherlands, where a group of 250 trio's from all Dutch provinces was sequenced at a depth of ~12-13X. We chose this reference panel, because this set is the closest to both cohorts with respect to their genetic background [48]. Our results show that phasing without a reference set is not able to eliminate differences between platforms. However, phasing together with a reference set helps to bring the two cohorts together with minimum platform stratification left. Strict imputation quality control (pre- and post- QC) as well as GCTA specific quality control is required to eliminate remaining platform stratification in a cross-platform imputed dataset.

Materials and Methods

Sample

Two population based cohorts, comprised of Dutch children, supplied genotype information and data on height [43, 46, 49, 50]. Genotype data were available for 3,102 children from the NTR and 2,826 children from GENR (Table 1).

Table 1 Cohort description

Sample	N	Sex		N families	N independent observations
		Males	Females		
GENR	2826	1450	1376	171	2508[a]
NTR	3102	1381	1721	1709	1644[a]

[a]Based on the list of distantly related individuals, which were selected using GCTA cut-off 0.025 independently in each cohort.

All children were of Northwestern European Dutch background as was checked by principal components analysis (PCA). Among them, 2,226 subjects had height measurements in GENR and 2,072 in NTR (Table 2, Figure 2).

Figure 2: Distributions of height across cohorts after correction for age and sex. Figure 2a shows the distribution of height for all individuals. Figure 2b shows the distribution of height for distantly related individuals.

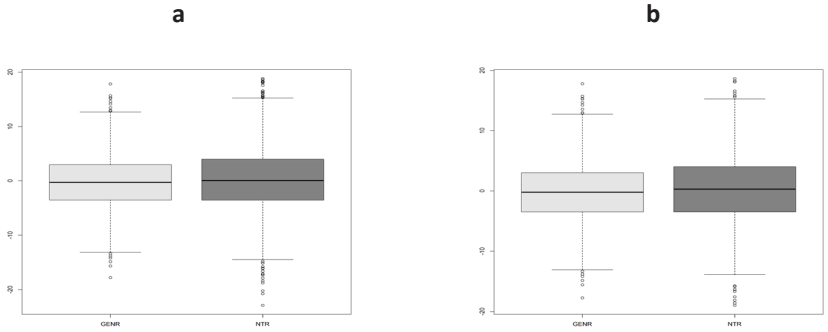


Table 2 Height measurements of all individuals

Sample	N	Sex		Age (mean, SD)	Height, sm (mean, SD)
		Males	Females		
GENR	2226	1124 (50.5%)	1102 (49.5%)	6 (0.4)	119.6 (5.6)
NTR	2072	948 (45.8%)	1124 (54.2%)	7.7 (1.4)	129.6 (9.8)

After applying a cut-off of 0.025 for genetic relatedness recommended in GREML analyses [21], there were 1,134 and 1,990 individuals left in NTR and GENR, respectively, with height measurements. The NTR cohort comprised of 528 males and 606 females at ages 4.6 - 11 years old. The GENR cohort comprised of 998 males and 992 females at ages 4.8 - 9 years old (Table 3, Figure 2). All parents gave informed consent. Study protocols were approved by Medical Ethics Committee of the VU University Medical Center, Amsterdam for NTR and by Medical Ethical Committee of the Erasmus Medical Centre, Rotterdam for GENR.

Table 3 Height measurements of distantly related individuals

Sample	N	Sex		Age (mean, SD)	Height, sm (mean, SD)
		Males	Females		
GENR	1990	998 (50.2%)	992 (49.8%)	6.1 (0.4)	119.6 (5.6)
NTR	1134	528 (46.6%)	606 (53.4%)	7.7 (1.4)	129.7 (9.8)
GENR + NTR	3124	1526 (48.8%)	1598 (51.2%)	6.7 (1.2)	123.2 (8.8)

Within sample pre-imputation SNP QC

The 3,107 subjects in the NTR cohort were genotyped for 692,694 SNPs on Affymetrix 6.0 chip [51]. The 2,830 subjects in the GENR cohort were genotyped for 489,878 SNPs on two Illumina chips (660W, 610K) [52]. Outliers were excluded from the GENR sample (4 individuals) and from the NTR sample (5 individuals) based on visual inspection of PC1 vs. PC2 plots prior to analysis. As a result, individuals cluster within -0.06 > PC1 < 0.05 and -0.05 > PC2 < 0.07 intervals in GENR and -0.06 > PC1 < 0.06 and -0.05 > PC2 < 0.04 intervals in NTR. For GENR, the overlapping SNPs between the two platforms were used as input for imputation as reported before [53]. Standard quality control steps were applied to the separate data sets using Plink 1.07 [54]. A sample call rate > 0.975, and a SNP call rate > 0.950 were applied for both cohorts. SNPs with minor allele frequency (MAF) < 0.001 and SNPs with Hardy-Weinberg equilibrium (HWE) p-value < 10^{-5} were excluded. Individuals were checked for excess heterozygosity and subjects with an inbreeding coefficient, as estimated in Plink, F <= -0.05 or F > 0.05, were excluded. Identical by state (IBS), identical by Descent (IBD) and gender mismatch were checked and samples not fitting the expected relations and/or gender were removed.

The next quality control step was a cross-check of alleles and SNP positions between the two cohorts as well as the GoNL reference set v.4 (build 37). SNPs that did not match by strand were flipped to the reference set strand. SNPs with discordant alleles or those that were not present in the reference set were excluded. Genotyped data from the NTR and GENR cohorts have 120,568 overlapping autosomal SNPs, of which 255 (0.2%) SNPs were significantly different in frequency across cohorts (p-value < 10^{-5}, one-sided test). Pairwise comparison between the SNPs overlapping in NTR and GoNL, in GENR and GoNL and in NTR and GENR combined identified 4,001 SNPs, which were significantly different in

allele frequency (p-value < 10^{-5}, 1,969 between NTR and reference set, 2,012 between GENR and reference set and 255 between NTR and GENR combined). All SNPs differing in allele frequency were removed. The resulting set of SNPs was either present on both platforms and in the reference set, or in a single platform and in the reference set. In order to minimize the amount of imputation stratification between samples, we selected the SNPs from the GoNL reference set that were present either on one or both genotype platforms (Illumina or Affymetrix, N=989,757) using VCFtools [55].

After QC was performed there were 3,102 NTR (1,381 males, 1,721 females) and 2,826 GENR (1,450 males, 1,376 females) individuals left. These individuals were genotyped for 641,554 and 468,259 SNPs in NTR and GENR, respectively. The two data sets were merged in Plink for pre-combined imputation.

Imputation strategies

First explorations of pre-combined cross-platform imputation approaches were done for chromosome 22. Genotype data comprising 13,712 SNPs were extracted, phased and imputed using the three methods described below, with the aim to determine the one to apply to the autosomal genome. The first approach uses MaCH phasing (selected because GCTA can read MaCH dosage files) and inherently also imputation of the missing genotypes. No reference set was involved. The second approach uses MaCH phasing, but with the GoNL reference set. Here, the haplotypes are predicted and genotypes imputed based on the GoNL reference set, which contains the full SNP haplotypes representing the Dutch population regardless of the platform. The third approach uses MaCH-Admix. This approach uses a new piecewise reference selection method [47] with GoNL as a reference set. This method, which is implemented in MaCH-Admix, breaks a genomic region into small fragments and searches for haplotypes in the reference set for matches. In all three approaches, we imputed missing genotypes as dosage scores. We have not considered only using the SNPs that were present on both platforms, because the final data set would comprise of only ≈120K SNPs after a genome-wide QC.

After an imputation approach for the pre-combined dataset was chosen, we evaluated the effect of the two possible scenarios of imputation on platform stratification and SNP-heritability estimates. In the first case, we pre-combined datasets and then imputed using a chosen approach, whereas in the second case we imputed datasets independently using the same software and reference set as for the pre-combined dataset and post-combined.

Post-imputation SNP QC

Post imputation QC aimed to examine the stratification between NTR and GENR due to genotyping platform after imputation on chromosome 22 first and on the autosomal genome afterwards. A comparison between all imputation approaches was done based on the imputation quality metric (R^2) calculated by the MaCH tools. The R^2 measures imputation quality and ranges between 0 and 1 with higher values indicating better

imputation accuracy, hence better genotype prediction. We used R^2 to inspect whether filtering on this measure helps to reduce platform stratification. Subsequently, a case-control analysis of the imputed sample with cohort as the phenotype was done using the Mach2dat software [7] for dosages and Plink for best-guess to check if there were differences in allele frequencies after imputation. Note, that in order to pool two independently imputed samples we had to: 1) convert dosage files to best-guess and 2) merge using Plink. The latter should be taken into account when comparing N of SNPs different in frequency between cohorts based on dosages and best-guess. The threshold for significance chosen was a genome-wide suggestive p-value of 10^{-5}.

Genetic pairwise relationships estimation

Genetic relationship matrices (GRMs) were built from pre-combined cross-platform imputed dosages of the three approaches for chromosome 22 using GCTA. Different SNP filter criteria can be used to build these GRMs, which might affect the results of the outcome. Therefore, we employed the criteria from three filters to estimate the matrices resulting in 9 GRMs. These criteria were: 1) without any filtering options on SNPs, 2) filtering on the imputation quality of $R^2 > 0.8$, leaving only the high quality imputed SNPs and 3) filtering with $R^2 > 0.8$ and MAF > 0.01, additionally excluding alleles with low minor allele frequency. To estimate the effects of stratification by SNP platforms after imputation we examined the GRMs using principal components analysis (PCA) in GCTA tool. We performed PCA on data from unrelated individuals. As PCs can be confounded by inversions of long LD regions of chromosomes, which are observed in the Dutch population [17, 56], we pruned GoNL for LD with standard Plink options (--indep 50 5 2), excluded 24 long LD regions [23] and repeated PCA for each GRM selecting GoNL pruned set of SNPs. The method that showed the least stratification due to genotyping platform and higher imputation quality was chosen for the pre-combined cross-platform imputation of the autosomal genome. To explore the effect of cross-platform imputed pre-combined, cross-platform imputed post-combined and combined GRMs on SNP-heritability estimate of childhood height, we built: 1) a GRM with MAF > 0.01 and $R^2 > 0.8$ filters from the total cross-platform imputed data set; 2) a GRM with MAF > 0.01 and $R^2 > 0.8$ filters from NTR and GENR cohorts imputed independently; 3) a GRM with a MAF > 0.01 from QC-ed NTR and GENR genotypes combined, merged in Plink. Additionally, to check the effect of QC, we built the GRM with MAF > 0.01 and $R^2 > 0.8$ filters from the total cross-platform imputed dataset excluding SNPs significantly different in frequency between cohorts after imputation. To distinguish between combination approaches throughout the paper we will refer to these GRMs as “imputed”, “imputed independently”, “combined” and “imputed clean”. Finally, SNP-heritability of height was estimated in NTR and GENR, after building two separate GRMs with MAF > 0.01 filter from QC-ed NTR and GENR samples. We performed PCA for each of the autosomal GRM based on GoNL pruned set of SNPs and included these PCs in the analysis of height.

Statistical analysis

Estimation of variance due to genetic effect of childhood height

Using GCTA, we estimated SNP-heritability of height using GRMs based on the autosomal genome. Imputation, SNP quality control, as well as employing the different imputation approaches, all determine the GRM relatedness of individuals. Therefore, for fair comparison between different ways of combining the genotype data in a GRM, we used the same unrelated individuals for each analysis. These were selected using the relatedness cut-off of 0.025 for individuals with height measurements from the combined and imputed GRMs (N = 3,124). The difference in relatedness selection between the combined and imputed GRM was 22 individuals, which were excluded from the analyses. For the independent study analyses, however, we selected unrelated individuals, as one would have based on the GRM of the single study alone, using the same GRM cut-off of 0.025. Hence, if there are samples with family relations between NTR and GENR studies, they were still included in these separate study analyses.

In the SNP-heritability analyses, age and sex were included as covariates. To test whether there is still a platform effect present after imputation, we included cohort as an extra covariate in addition to sex and age and compared results of both analyses. To detect and account for possible genetic stratification in relation to height [23], we included the first 10 PCs obtained from each GRM for unrelated individuals excluding long LD regions. Finally, we ran association analysis of height for imputed, combined, NTR and GENR datasets with age and sex as covariates for unrelated individuals. In addition, we built QQ-plots to check for possible inflation of test statistics before and after pooling cohorts together without using 10 PCs and cohort as covariates.

Results

Imputation method

Three imputation approaches, aimed to pre-combine and cross-platform impute two cohorts, were tested on chromosome 22: the first was MaCH without a reference set (i.e. the two datasets were only phased and imputed against each other), the second was MaCH with the GoNL reference set and the third was MaCH-Admix with the GoNL reference set. The comparison of the post-imputation quality control measures for these approaches is shown in Figures 3 and 4.

Figure 3: Comparison of imputation quality for chromosome 22. Figures 3.1 – 3.3 show PC1 vs. PC2 plots of GRM based on MaCH without reference set, MaCH with reference set and MaCH-Admix with reference set respectively. Figures 3.a – Fig3.b show PCs plots including and excluding long LD regions (a. including, b. excluding). All PC plots are based on GRMs filtered with $R^2 > 0.8$ and MAF > 0.01, where black color represents NTR and grey color represents GENR.

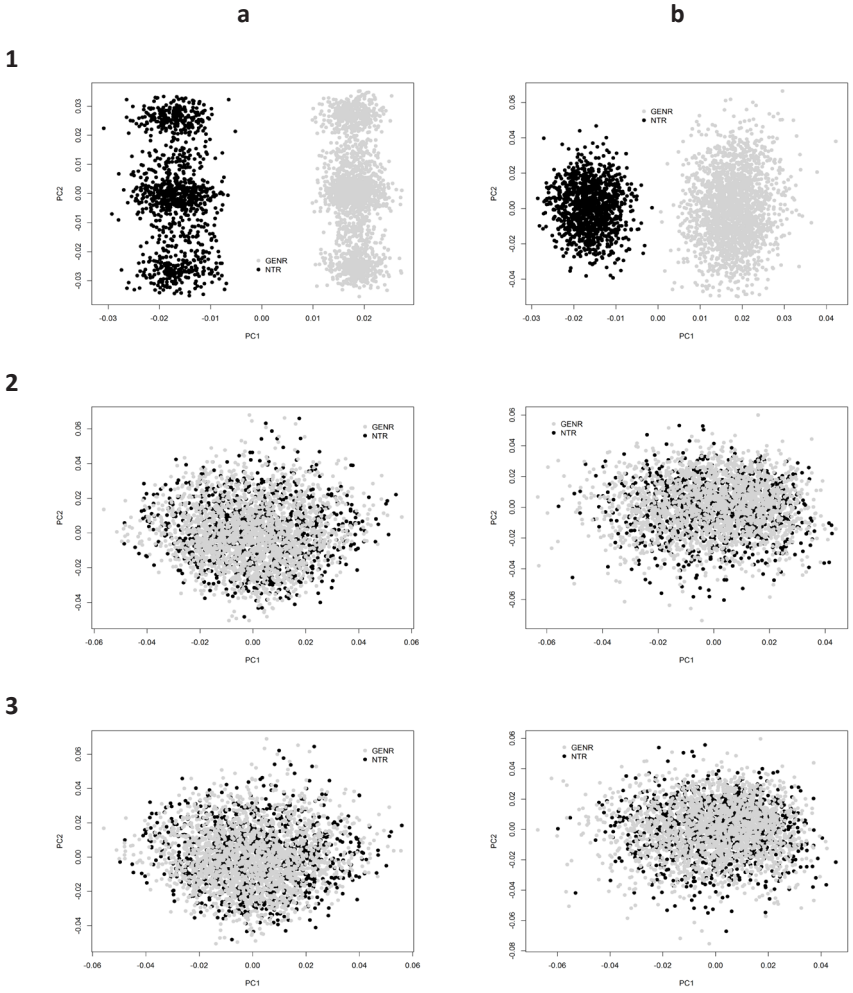
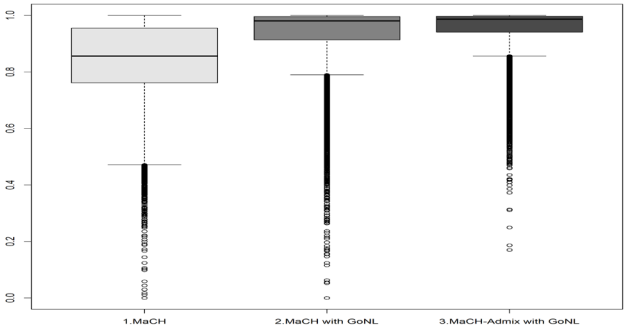


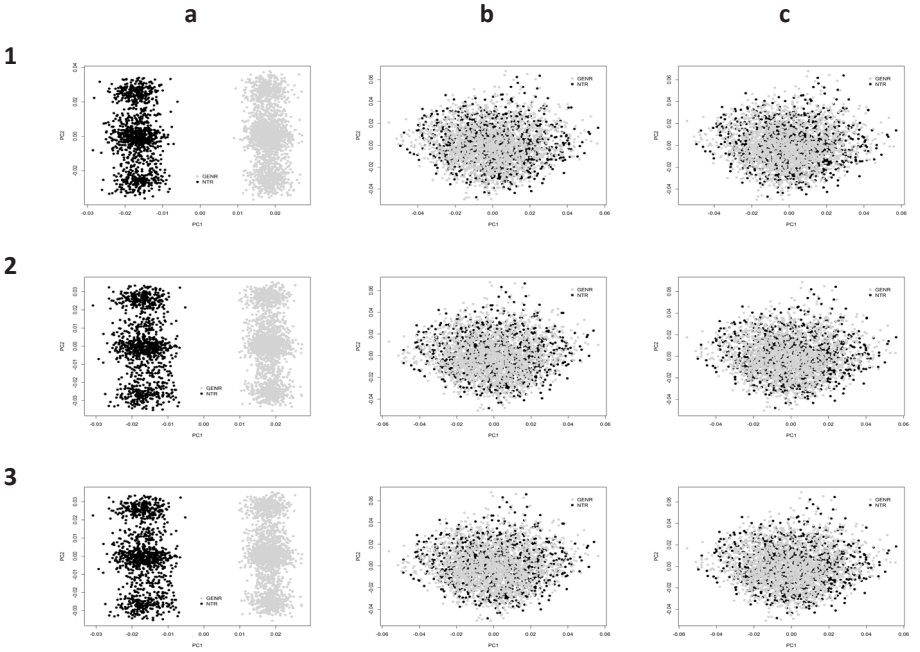
Figure 4: Comparison of R^2 distribution of three methods for chromosome 22.



A NTR vs. GENR case-control analysis after imputation showed that 4535, 203, and 93 SNPs were significantly different in frequency for the first, second and third method, respectively ($p < 10^{-5}$, Wald test). The R^2 measure also demonstrated different imputation quality: mean = 0.83 and median = 0.86 for the first, mean = 0.93 and median = 0.98 for the second and mean = 0.95 and median = 0.99 for the third method.

We plotted the first (PC1) and second (PC2) principal components for each imputed GRM matrix in R [57]. In Figure 3 the GRMs based on the $R^2 > 0.8$ and MAF > 0.01 filters are shown. As expected given the median quality of SNPs, filtering on R^2 and MAF (4,611 and 46, 1684 and 106, 1186 and 105 SNPs were excluded in the first, second and third approach respectively) did not affect the outcome of the imputation results (Figure 5).

Figure 5: Chromosome 22 PC plots based on GRMs, each with three filtering options. Figure 5.a shows the performance of MaCH without reference set, Figure 5.b shows performance of MaCH with reference set and Figure 5.c demonstrates performance of MaCH-Admix with reference set. Plots 5.1-5.3 show application of different filter criteria (1.none, 2. $R^2 > 0.8$, 3. $R^2>0.8$ and MAF>0.01) for the corresponding imputation method.



As shown in Figure 3.1a, PC1 clearly captures the cohort differences due to the genotyping platform. GENR and NTR are separated into two clusters with the first PC. For the PC2 component we observe three blocks that disappear after eliminating the long LD regions as shown on Figure 3.1b. Figures 3.2a and 3.2b show homogeneity is reached when using MaCH phasing with a reference set, with and without excluding long LD regions. Similarly, Figures 3.3a and 3.3b using MaCH-Admix instead of MaCH also shows no population stratification due to the genotyping platform. Finally, as presented in Figure 4, it becomes clear that Mach-Admix outperforms MaCH with overall imputation quality.

When examining imputation differences for individual SNPs by comparing the allele frequencies between cohorts, we identified some significantly different SNPs, as was noted above. We computed squared LD correlations between each significant SNP that resulted from post-imputation QC analysis of the chromosome 22 imputation with MaCH-Admix and all neighboring SNPs within a 1 Mega-basepairs (Mb) region in Plink. The majority of these estimates were low (IQR = 0.0009, mean = 0.005, median = 0.0003), indicating regions with weak LD around significant SNPs. Therefore, we can hypothesize that these SNP differences may arise from imperfect phasing and imputation for these SNPs with low LD.

Repeating the same MaCH-Admix imputation procedure of chromosome 22, 1) the NTR and GENR pre-combined sample was cross-platform imputed for all autosomal chromosomes and subsequently an "imputed" GRM was completed; 2) the NTR and GENR samples were imputed independently for all autosomal chromosomes, post-combined and an "imputed independently" GRM was built. Figures 6 and 7 demonstrate QC results after imputation of the whole sample: Figure 6 shows PC1 and PC2 plot with and without exclusion of long LD regions and Figure 7 demonstrates the R^2 distribution for imputed (mean = 0.97, median = 0.99), imputed clean (mean = 0.97, median = 0.99), NTR imputed independently (mean = 0.97, median = 1.0) and GENR imputed independently (mean = 0.96, median = 1.0) samples.

Figure 6: PCA results of combined, imputed, imputed clean and imputed independent datasets respectively. PC1 vs. PC2 plots are made from GRM with $R^2 > 0.8$ and MAF > 0.01 filters in case of imputed and with MAF > 0.01 filters in case of combined GRMs. Fig6.a – Fig6.b shows PCs plots including and excluding long LD regions (a. including, b. excluding).

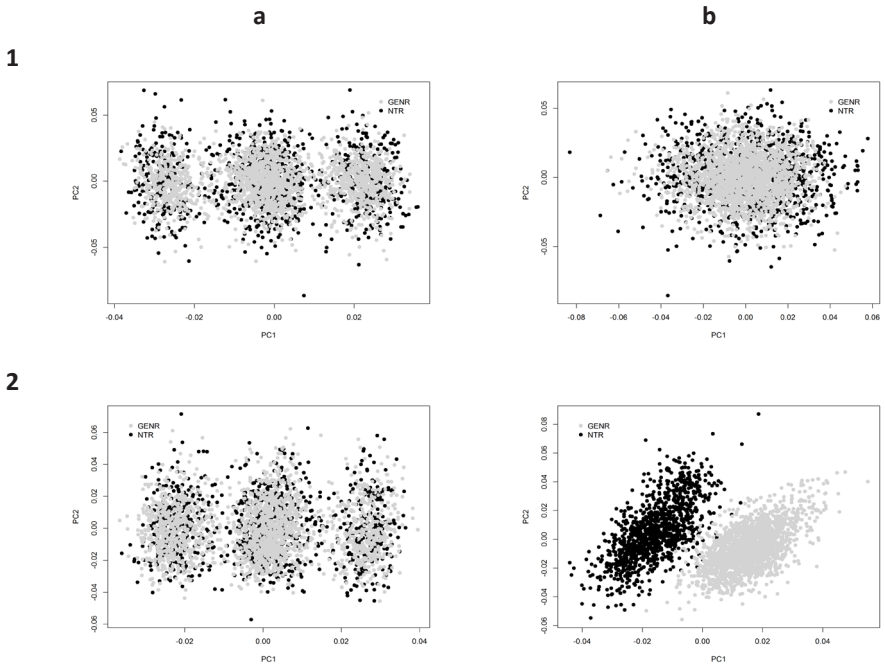


Figure 6 Continued

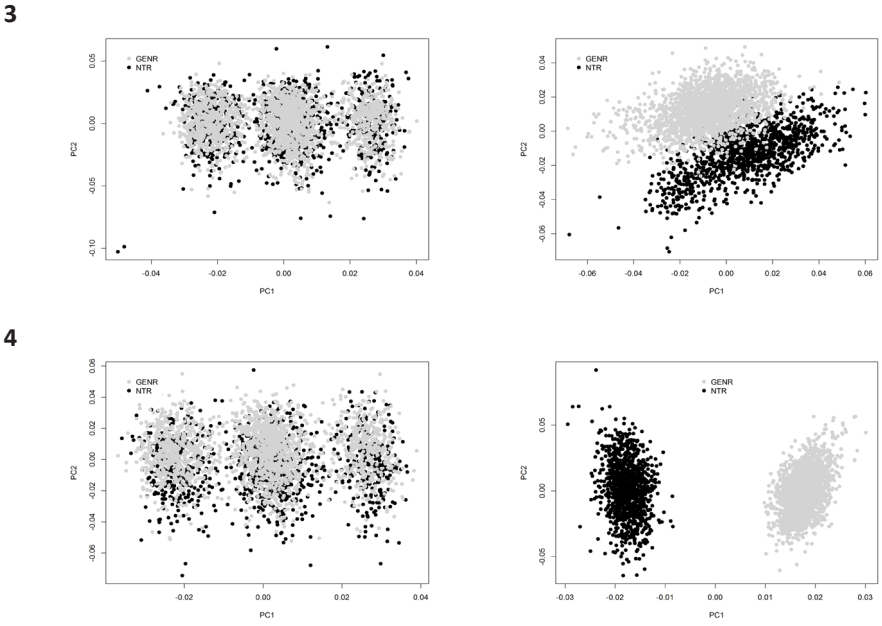
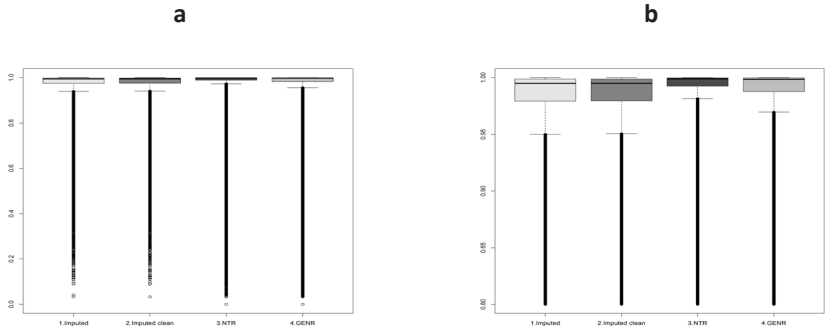


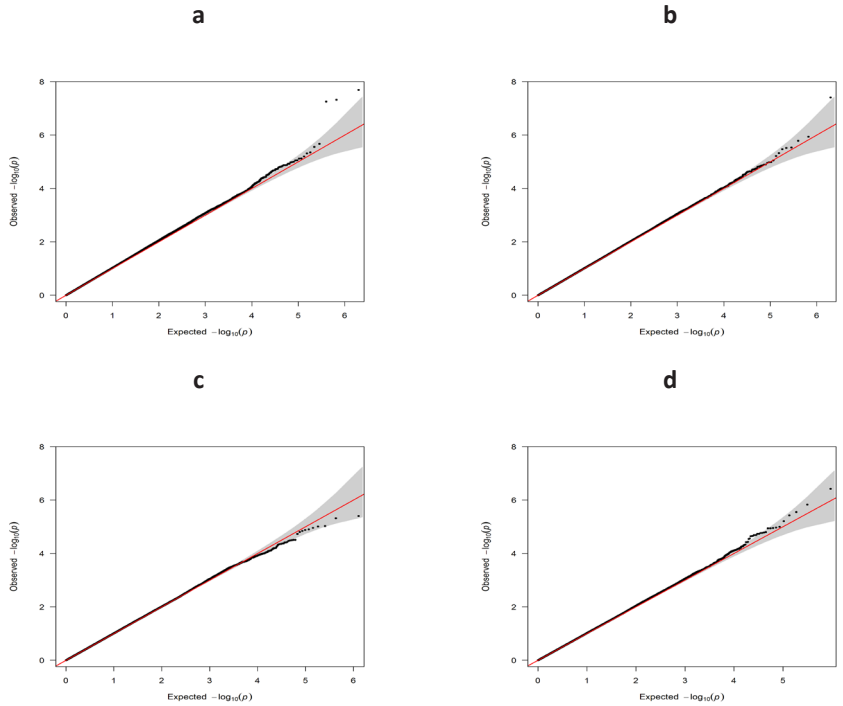
Figure 7: Comparison of R^2 distribution of imputed, imputed clean, independently imputed NTR and GENR datasets. Fig7.a represent all SNPs, Fig7.a represents SNPs with $R^2 > 0.8$



The quality of imputation in NTR seems slightly better than in GENR, which showed 203 monomorphic SNPs after imputation. These SNPs were excluded from calculation of mean and median of R^2 for GENR. They also did not contribute to further analysis, as they have MAF = 0 and were filtered out with MAF > 0.01 option. As shown in Figures 6.1 – 6.4, PC2 captures three blocks that are inversions of long LD regions of chromosomes and we did not observe any cohort differences due to the genotyping platform for

any of GRMs after different combination approaches. After the exclusion of long LD regions, PC1 and PC2 captured population structure for each of the approaches. Figure 8 demonstrates quantile-quantile (QQ) plots of GWAS test-statistics for imputed (λ = 1.04), combined (λ = 1.02), NTR (λ = 1.01) and GENR (λ = 1.02) datasets. NTR vs. GENR case-control analysis showed a total of 4,340 and 18,306 SNPs, respectively, that significantly differ in frequency after imputation, when datasets were pre-combined and imputed and imputed and post-combined. We excluded 4,430 SNPs from GRM "imputed" to build GRM "imputed clean".

Figure 8: Quantile-quantile plots based on test-statistic from association analysis of height of: a) imputed, b) combined, c) NTR and d) GENR datasets



Heritability of childhood height

The pooled data set comprised a total of 3,124 distantly related individuals, where 1,526 were males and 1,598 were females. Childhood mean height in the pooled dataset was 123.2 cm (SE = 8.8) at mean age of 6.7 years (SE = 1.2) (Table 3). GREML analysis of height yields a SNP-heritability estimate of 0.43 (SE = 0.10) when combining (not imputing) the data from both cohorts (Table 4).

Table 4 Results of analyses of height based on imputed, imputed clean, imputed independently and combined GRMs including results of specific analysis of NTR and GENR selected individuals.

Data set	V(G)/Vp	SE	N	Pval
Imputed[a]	0.51	0.10	3124	1×10^{-7}
Imputed clean[b]	0.49	0.10	3124	2.9×10^{-7}
Imputed independently[c]	0.52	0.10	3124	8.8×10^{-8}
Combined[d]	0.43	0.10	3124	2×10^{-6}
NTR imputed[a]	0.42	0.29	1134	0.07
NTR imputed clean[b]	0.39	0.29	1134	0.09
NTR imputed independently[c]	0.45	0.29	1134	0.07
NTR combined[d]	0.50	0.28	1134	0.04
NTR independent [e]	0.47	0.27	1173	0.04
GENR imputed[a]	0.52	0.16	1990	3.7×10^{-4}
GENR imputed clean[b]	0.52	0.16	1990	3.9×10^{-4}
GENR imputed independently[c]	0.53	0.16	1990	3.4×10^{-4}
GENR combined[d]	0.58	0.17	1990	2×10^{-4}
GENR independent[e]	0.57	0.17	1994	2.2×10^{-4}

[a] GRM based on data cross-platform imputed SNPs

[b] GRM based on data cross-platform imputed SNPs, excluding SNPs significantly different in frequency

[c] GRM based on SNPs imputed separately and combined afterwards

[d] GRM based on the combined SNP data without imputation

[e] GRM based on each genotyped sample separately

The estimates of the SNP-heritability based on GRMs of the imputed data are 0.51 (SE = 0.10), and 0.49 (SE = 0.10) after cleaning SNPs that were significantly different between the two cohorts. The estimates of the SNP-heritability based on GRM data imputed independently are 0.52 (SE = 0.10). When considering only NTR individuals or GENR participants in the various GRM matrices, NTR gives estimates of 0.42 (SE = 0.29), 0.39

(SE = 0.29), 0.45 (SE = 0.29) and 0.50 (SE = 0.28) for the imputed GRM, imputed clean GRM, imputed independently and combined GRMs respectively; GENR gives estimates of 0.52 (SE = 0.16), 0.52 (SE = 0.16), 0.53 (SE = 0.16), 0.58 (SE = 0.17) for the imputed GRM, imputed clean GRM, imputed independently and combined GRMs, respectively. The variances explained by the independent cohorts were 0.47 (SE = 0.27) for NTR and 0.57 for GENR (SE = 0.17), if one would conduct two separate GCTA studies. These results show that for each of the individual cohorts (NTR or GENR) selected, either from the imputed GRMs or from combined, the amount of variance explained by the SNPs remains the same given the large standard errors. Strikingly, cross-platform imputed GRMs shows suggestive, if any, increase of the variance explained by the SNPs in comparison to the combined (not imputed) GRM. If cohort is taken into account as a covariate, results show a ≈2% reduction of explained variance in the cross-platform imputed GRMs, while the combined GRM estimate remains the same (Table 5).

Table 5 Results of analyses of height with correction for cohort as a covariate based on imputed, imputed clean, imputed independently and combined datasets

Data set	V(G)/Vp	SE	n	Pval
Imputed[a]	0.49	0.10	3124	3×10^{-7}
Imputed clean[b]	0.47	0.10	3124	7×10^{-7}
Imputed independently[c]	0.50	0.10	3124	3.6×10^{-7}
Combined[d]	0.43	0.10	3124	3.8×10^{-6}

[a] GRM based on data cross-platform imputed SNPs

[b] GRM based on data cross-platform imputed SNPs, excluding SNPs significantly different in frequency

[c] GRM based on SNPs imputed separately and combined afterwards

[d] GRM based on the combined SNP data without imputation

This indicates that there is still little stratification left by platform. Repeating the comparison procedure including the first 10 PCs resulted in SNP-heritability estimates that were on average ≈11% lower for all pooled GRMs, ≈13% for NTR and ≈7% for GENR (Table 6).

Table 6 Results of analyses of height based on imputed, imputed clean, imputed independently and combined datasets adjusted for age, sex and 10 PCs, but not for cohort as covariate. Additionally results of analysis of height in NTR and GENR independent cohorts adjusted for age, sex and 10 PCs.

Data set	V(G)/Vp	SE	N	Pval
Imputed[a]	0.41	0.11	3124	4.6×10^{-5}
Imputed clean[b]	0.38	0.11	3124	1.2×10^{-4}
Imputed independently[c]	0.39	0.11	3124	1.2×10^{-4}
Combined[d]	0.33	0.10	3124	7.2×10^{-4}
NTR independent [e]	0.34	0.28	1173	0.12
GENR independent[e]	0.50	0.17	1994	1.6×10^{-3}

[a] GRM based on data cross-platform imputed SNPs
[b] GRM based on data cross-platform imputed SNPs, excluding SNPs significantly different in frequency
[c] GRM based on SNPs imputed separately and combined afterwards
[d] GRM based on the combined SNP data without imputation
[e] GRM based on each genotyped sample separately

When cohort was used as a covariate together with 10 PCs (Table 7) there was no effect on SNP-heritability estimates in comparison to the effect of 10 PCs alone. The comparison of results shows that all heritability estimates, given the standard errors, are not significantly different from each other. However, by combining the two cohorts, the standard errors were largely reduced as the sample size increased, thereby allowing the SNP heritability to reach significance.

Table 7 Results of analysis of height based on imputed, imputed clean, imputed independently and combined datasets adjusted for age, sex and 10 PCs, as well as for cohort as covariate.

Data set	V(G)/Vp	SE	N	Pval
Imputed[a]	0.41	0.11	3124	5×10^{-5}
Imputed clean[b]	0.38	0.11	3124	1.4×10^{-4}
Imputed independently[c]	0.39	0.11	3124	1.2×10^{-4}
Combined[d]	0.32	0.10	3124	9×10^{-4}

[a] GRM based on data cross-platform imputed SNPs
[b] GRM based on data cross-platform imputed SNPs, excluding SNPs significantly different in frequency
[c] GRM based on SNPs imputed separately and combined afterwards
[d] GRM based on the combined SNP data without imputation

Discussion

GREML estimates the narrow-sense heritability from all common SNPs genotyped or imputed in a sample. However, often sample sizes are small, for example when closely related individuals are excluded. In this paper, we examined imputation-phasing approaches to create a GRM that combines genotype data across genotype platforms and cohorts and explored the effect of using different GRM build strategies, when cohorts are: 1) pre-combined and cross-platform imputed, 2) cross-platform imputed and post-combined and 3) just combined (Figure 1). Imputed GRM genetic relationships between individuals are estimated within studies as well as between studies based on all Illumina and Affymetrix SNPs. Combined GRM genetic relationships are estimated in three groups: within cohort pairs of NTR which all have Affymetrix SNPs, within cohort pairs of GENR which all have Illumina SNPs, and between cohort pairs that only have the overlapping SNPs. Therefore, cross-platform imputation is required to supply individuals genotyped on one platform with SNPs genotyped on another platform. Note that we did not aim to impute a large number of additional (rare) SNPs from the reference set to increase number of SNPs. Instead the total number of SNPs in a cross-platform imputed dataset remains approximately the same (Affymetrix SNPs + Illumina SNPs), but all individuals from both cohorts pooled together have complete information from the same SNPs. This way we tried to minimize the possible differences between platforms, while also trying to retain as much information of the genotyping platforms as possible. Because the quality of cross-platform imputation depends on LD-phase information, which correctly represents the Dutch population from which GENR and NTR cohorts were drawn, the Dutch GoNL reference set was used.

Based on the chromosome 22 analyses of pre-combined cross-platform imputation approaches, we showed that phasing and imputation of missing genotypes with a reference dataset that contains all SNPs and LD information between these SNPs does not substantially increase cohort stratification due to genotyping platform within the GRM, while phasing without a reference set, lacking this essential LD information, does. Using only the SNPs that are overlapping between genotyping platforms as an imputation backbone is insufficient, as was evident from the subsequent PC analyses. Given that one could consider two cohorts with different platforms as a stratified population, the use of MaCH-Admix seems to have helped to improve the imputation quality. However, this effect was much weaker in comparison to the use of a reference set. Analysis based on PCs also showed that post imputation filtering on MAF and R^2 did not largely seem to influence the cohort stratification, mainly because the quality of the imputed SNPs was generally high. Imputation of the autosomal genome followed by PC analysis showed that to some extend there is still platform stratification present after imputation (Figure 6). Interestingly, the combined GRM did not show platform stratification, which may indicate that backbone of ≈120K SNPs is enough to estimate the genetic relationships between individuals from different cohorts.

The analysis of childhood height yielded relatively the same estimates of SNP-heritability for cross-platform imputed GRMs, suggesting a slight increase in estimates in comparison to combined GRM. Adjusting for 10 PCs with or without study as covariate results in ≈11% reduction of SNP-heritability for all GRMs, including combined. However, there was only ≈2% reduction in SNP heritability when study was used as covariate for imputed GRMs and not for the combined. PC adjustment of independent cohorts results in a SNP-heritability drop of ≈13% for NTR and ≈7% for GENR. Drop in NTR SNP-heritability estimate in contrast to GENR is more pronounced, as individuals in NTR spread across the Netherlands resulting in more diverse cohort. Given that λ estimates obtained from association analysis are not inflated it is possible that PCs may capture true variation of height along with platform stratification and may overcorrect the estimates. On the other hand, PCs may help to capture and correct for other sources of stratification within cohorts. Interestingly, SNP-heritability estimates resulting from GRM imputed and GRM imputed independently are approximately the same for all conditions. Moreover, SNP-heritability estimates from combined GRM are just slightly lower in comparison to imputed GRMs, which may support the conclusion that relationships between individuals across cohorts, estimated from SNPs overlap of ≈ 120K, is sufficient to explain the substantial proportion of variation in childhood height.

In this study we estimated SNP-heritability of childhood height using different GRM building strategies. These GRMs yielded significant estimates of SNP-heritability that range from 0.33 to 0.52 depending on various correction options. Height is a highly heritable trait with heritability estimates ranging from 0.89 to 0.93 in adults [58]. A SNP-heritability of 60% has been estimated based on all common SNPs together in the recent GWA meta-analysis study of adult height [59]. In children, heritability estimates vary during growth. Mook-Kanamori et al. showed that heritability increases from 26% and 27% at birth to 63% and 72% at 36 months in twins from the NTR study and in singletons from GENR study (parent-child trio's design) [60]. Notably, heritability estimates for singletons and twins were very similar, justifying the pooling of data from these cohorts. In this study, we used height, as it is a highly heritable GCTA benchmark trait and can be easily measured. For other traits, which are less heritable and less easily measured, additional increase of sample size may be required in order to increase power to accurately estimate SNP-heritability. To calculate the power given a sample size, one can use the GCTA-GREML Power Calculator [61].

Strategies aiming to detect and correct for platform stratification after cross-platform imputation were considered in this study for cohorts with the same ethnicity. However, when combining cohorts from different ethnicities this approach is unlikely going to be appropriate for several reasons [62]. First, SNP-heritability of combined multi-ethnic dataset depends on heritability of the trait in each population, which can differ. Second, different LD-patterns may imply that causal SNPs in one population will be tagged better than in the other population. Third, if cohorts with different ancestry are genotyped on different platforms it might be difficult to distinguish the two confounding

factors: platform and population stratification. Finally, informative SNPs that are common in one population and are rare in another, will be eliminated from analysis after QC and the effect of remaining SNPs, reflecting ancestry, will be corrected with PCs. Thus, the estimate would reflect part of SNP-heritability, which is based on causal SNPs shared across ethnicities. The extent to which causal SNPs are shared between different ethnicities depends on the genetic architecture of the trait in each population. For example, a recent study has provided evidence that genetic variation in schizophrenia is largely shared between two different ethnic cohorts, African and European [62]. There are also other statistical methods that can be applied to combine cohort information to estimate the SNP heritability of traits, such as the Density Estimation (DE) method [63]. The DE method does not require the raw genotype data, as it uses summary statistics from GWAS or meta-analysis GWAS. However, it requires LD-pruning to obtain a list of relatively independent SNPs to estimate their effect, which may result in variability of estimates depending on the pruning threshold and on SNP density in a single GWAS [64]. In addition, Van Beek et al. suggested that SNP-heritability can be underestimated due to genotypic heterogeneity or phenotypic differences between cohorts in meta-analysis GWAS and summary statistics correction, such as for multiple testing and genomic control inflation factor.

In conclusion, using the complete information of a reference set for phasing and imputation of all SNPs on two different genotyping platforms, allows the combination of cohort data genotyped on both of these platforms. When combining genotype data across platform or cohort, thorough pre - and post QC is required, which can be tested with association and principal component analyses. For our approach, we assumed that the cohorts have a similar ethnicity and genetic background. To account for platform stratification or phenotypic differences in the dataset, cohort should always be included as a covariate. Whether one should use imputation, or just combine the genotype data, depends on the number of overlapping SNPs in relation to the total number of genotyped SNPs for both cohorts and their ability to tag all the genetic variance related to the specific trait of interest.

CHAPTER 3

SINGLE NUCLEOTIDE POLYMORPHISM HERITABILITY OF BEHAVIOR PROBLEMS IN CHILDHOOD: GENOME-WIDE COMPLEX TRAIT ANALYSIS

This chapter is based on:

Irene Pappa, Iryna O. Fedko, Viara R. Mileva-Seitz, Jouke-Jan Hottenga, Marian J. Bakermans-Kranenburg, Meike Bartels, Catharina E.M. van Beijsterveldt, Vincent W.V. Jaddoe, PhD, Christel M. Middeldorp, Ralph C.A Rippe, Fernando Rivadeneira, Henning Tiemeier, Frank C. Verhulst, Marinus H. van IJzendoorn, Dorret I. Boomsma (2015). Single nucleotide polymorphism heritability of behavior problems in childhood: Genome-wide complex trait analysis. *Journal of the American Academy of Child & Adolescent Psychiatry, 54*(9), 737-744.

Abstract

Objective: Genetic factors contribute to individual differences in behavior problems. In children, genome-wide association studies (GWAS) have yielded the first suggestive results in identifying genetic variants that explain heritability, but the proportion of genetic variance that can be attributed to common single nucleotide polymorphisms (SNPs) remains to be determined, as only a few studies have estimated SNP heritability, with conflicting results.

Method: Genomic-relationship-matrix restricted maximum likelihood (GREML), as implemented in the software Genome-wide Complex Trait Analysis (GCTA), was used to estimate SNP heritability (SNP h^2) for multiple phenotypes within four broad domains of children's behavioral problems (attention deficit/ hyperactivity symptoms, internalizing, externalizing, and pervasive developmental problems) and cognitive function. We combined phenotype and genotype data from two independent, population-based Dutch cohorts, yielding a total number of 1,495 to 3,175 of three, seven and nine-year-old children.

Results: Significant SNP heritability estimates were found for attention deficit/ hyperactivity symptoms (SNP h^2 =0.37-0.71), externalizing problems (SNP h^2 = 0.44), and total problems (SNP h^2 = 0.18), rated by mother or teacher. Sensitivity analyses with exclusion of extreme cases and quantile normalization of the phenotype data decreased SNP h^2 as expected under genetic inheritance, but they remained statistically significant for most phenotypes.

Conclusion: We provide evidence of the influence of common SNPs on child behavior problems in an ethnically homogenous sample. These results support the continuation of large GWAS collaborative efforts, to unravel the genetic basis of complex child behaviors.

Introduction

Complex behaviors are shaped by both genetic and environmental influences [65, 66]. Numerous twin, family and adoption studies have estimated significant contributions of genetic factors to individual differences in behavioral and psychiatric traits [67-69]. In addition, longitudinal population-based studies provide evidence of the genetic stability of common behavioral problems (e.g. anxiety and depression symptoms [70], attention problems [71]) across the lifespan, with higher heritability estimates in childhood (e.g. for attention problems, heritability estimates decreased from 0.70 in childhood to 0.40 in adulthood [71]).

In adult samples, Genome-Wide Association Studies (GWAS) identified genes and pathways related to complex traits [72, 73]. This approach has also yielded positive findings in studies of important traits in children (e.g., birth weight [74] and length [75]). For childhood psychiatric traits and problem behaviors, successes have been limited [76-79], which can be ascribed to the very modest sample sizes in these studies [80]. The relatively small or absent genetic associations with complex traits of interest in GWAS [76-79] may seem in contrast to the large heritability estimates from twin and family studies, but are indeed in line with recent evidence that the small effect sizes of individual SNPs may be responsible for the non-replicability of these associations [81]. To assess whether GWA studies of child behavior problems can be expected to yield important findings regarding biological pathways, we address the question what part of the heritability of childhood behavior problems is captured by common (minor allele frequency > 1%) single-nucleotide polymorphisms (SNPs) included in standard genotyping arrays.

The genetic variance explained by genome-wide SNPs [20] can be estimated by using the genetic similarity among unrelated individuals as a predictor of their phenotypic resemblance. When individual level genotype data are available, these can be used to obtain a measure of genetic similarity between all possible pairs of (unrelated) individuals in the study. In a second step, this genetic relatedness matrix (GRM) is used to predict the phenotype similarity between individuals just as the different similarity of monozygotic (MZ) and dizygotic (DZ) twin pairs predicts their different phenotype resemblances. This approach has been implemented in the software package Genome-wide Complex Trait Analysis (GCTA) [20]. The heritability estimates from GCTA (SNP h^2) are commonly considered an indicator of the upper limit of the variance that can be explained by current GWAS efforts. Power estimations have indicated that for quantitative traits, a sample size of 3,000 individuals is required to detect a SNP h^2 of 0.30 with 80% power [82]. Thus, large sample sizes are required to reliably estimate the SNP heritability of complex behavioral traits, which can imply the need to pool data from multiple studies.

To date, few SNP heritability estimates are available for behavioral problems in childhood. Some studies indicate substantial additive genetic heritability of normative differences in children's social communication difficulties [83] and in clinical cases of Attention Deficit/Hyperactivity Disorder (ADHD) [84, 85] and childhood onset Obsessive-

Compulsive Disorder (OCD) [86]. However, other studies indicate modest, statistically non-significant SNP heritability estimates for children's internalizing problems [76], anxiety [87], and callous-unemotional (CU) traits [77] in population-based samples. A study from the Twins Early Development Study (TEDS) indicated no significant SNP heritability for parent-, teacher-, and self-reported behavioral problems (i.e. attention problems, internalizing and externalizing problems) in contrast to cognitive and anthropomorphic traits in a population-based sample (*N* =2,500) of 12-year-old children [88].

Here, we focus on four domains of children's behavioral problems: attention deficit problems, externalizing, internalizing, and pervasive developmental problems. Genetic influences on non-verbal cognitive abilities were also estimated. To obtain sufficient power, we combined genotype and phenotype data from two independent, population-based Dutch cohorts: the Generation R Study (GEN-R) and the Netherlands Twin Register (NTR). Genotyped SNP data from both studies were used to construct a genetic relatedness matrix (GRM) [102]. For both studies, behavior problems of a total *N* = 1,495 to 3,175 of three, seven and nine year old children were rated by mothers and / or teachers. We estimated the SNP heritability in each of these traits and compared our findings to the SNP heritability estimates previously reported.

Methods

Participants

This study included data from children from two population-based Dutch cohorts, the Generation R Study (GEN-R) and the Netherlands Twin Register (NTR). GEN-R is a prospective cohort based in Rotterdam. The characteristics of the study have been previously described in detail [89]. NTR is a nationwide longitudinal sample of twins and their family members followed from birth onwards after voluntary registration [44]. In both studies, parents gave informed consent for participation and also to approach the teachers of the children. Study protocols were approved by the local ethics committees.

Measures

All phenotypes analyzed in this study have been described in detail in previous publications of GEN-R and NTR, and twin-based heritabilities in the Dutch population were reported for these traits (see Supplementary Material, Table S1).

Conners' Parent Rating Scale (CPRS-R)

ADHD symptoms and related co-morbid symptoms were assessed using the CPRS-R [90] completed by the mothers. Four scales of the CPRS-R were used: (i) ADHD combined, (ii) ADHD Inattentive, (iii) ADHD Hyperactive-Impulsive, and (iv) Oppositional Defiant Disorder (ODD) scale.

Child Behavior Checklist (CBCL; behavior problems)

We assessed child behavior problems using the well-validated Child Behavior Checklist (CBCL) [91], completed by the mother. Internalizing, externalizing and total problems were assessed using the appropriate CBCL syndrome scales. For the CBCL internalizing, externalizing, and total problems scores, the GEN-R study used the CBCL for ages 1½-5 [92] and NTR used the CBCL for ages 6-18 years [93]. In the CBCL for ages 1½-5 years, the Internalizing scale consists of four scales (Emotionally Reactive, Anxious/ Depressed, Somatic Complaints, and Withdrawn) and the Externalizing scale consists of two scales (Attention Problems and Aggressive Behavior). In the CBCL for ages 6-18 years, the Internalizing scale consists of three scales (Anxious/ Depressed, Withdrawn/Depressed and Somatic Complaints) and the Externalizing scale consists of two scales (Rule-Breaking Behavior and Aggressive Behavior). The Total Problems score was computed by summing the ratings of all problem items included in the CBCL. To avoid phenotypic heterogeneity in the combined data set due to differences in the items between the two CBCL versions, we selected only overlapping items to compute the scores (see Supplementary Material, Table S2).

Child Behavior Checklist (CBCL; pervasive developmental problems)

We assessed pervasive developmental problems using the Pervasive Developmental Disorder (PDD) subscale of the CBCL 1½-5 years [92]. The PDD subscale has been shown to be a valid screening tool for autism spectrum disorders (ASD) [94].

Teacher's Rating Form (TRF; ADHD-related symptoms and behavior problems)

The TRF for ages 6 to 18 years [93] was used to assess attention problems (Attention Problems scale) and behavioral problems (Externalizing scale) rated by the teacher. We used the teachers' ratings of externalizing and not internalizing problems, since it has been previously shown that they can better identify children with externalizing than internalizing problems [95]. The teacher reports were also selected to assess behavior in a different environment, and to avoid informant effects which could bias estimates of genetics contribution to common child behavior problems [96, 97].

Non-verbal cognitive abilities

Non-verbal cognitive abilities were assessed with the Snijder-Oomen nonverbal intelligence test [98] (SON-R 2.5-7 years) in the GEN-R study and the non-verbal subtest of the Revised Amsterdam Children Intelligence Test [99] (RAKIT) in the NTR. Both measurements have been well-validated and correlate substantially with the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R) [100] and the Wechsler Intelligence Scale for Children (WISC) [101]. The non-verbal cognition scores in both samples were transformed to *Mean* = 100 and *SD* = 15.

Genotyping and Imputation

Caucasian children from the GEN-R study (*N* = 3,102) and NTR (*N* = 2,826) were genotyped on Illumina (660W, 610K) and Affymetrix 6.0 platforms, respectively. As the number of overlapping SNPs between platforms was small (*N* = 123,953), both cohorts were cross-platform imputed using MaCH-Admix imputation software [47] as described in Fedko et al [102]. Cross-platform imputation supplies all participants from both cohorts with genetic information from all SNPs genotyped on both platforms. To avoid population stratification between samples due to a genotyping platform, the Genome of the Netherlands reference set [103] was used to phase and subsequently impute missing genotypes into both cohorts. The final dataset consisted of 5,928 individuals, where each individual had information for *N* = 989,757 SNPs expressed in dosage scores. Post-imputation quality control (QC) was performed on imputed datasets to check and control for possible residual imputation stratification due to a genotyping platform or true genetic differences between cohorts. The overall imputation quality measure (R2) was high (*mean* = 0.97, *median* = 0.99). Case-control analysis of the imputed sample, where GEN-R children were assigned as cases and NTR children as controls showed n = 4,340 SNPs that were significantly different in frequency ($p < 10^{-5}$). These SNPs were excluded from further analysis.

Genome-Wide Complex Trait Analysis (GCTA)

We built a genetic relationship matrix (GRM) based on cross-platform imputed data using GCTA version 1.20 [20]. Data for the GRM was filtered based on the following two criteria: 1) $R^2 > 0.8$ to allow SNPs with high imputation quality and 2) MAF > 0.01 to exclude SNPs with low minor allele frequency. We performed Principal Components Analysis (PCA) on the resulting GRM to check for possible residual stratification due to the genotyping platform. We used the GREML (Genomic-relatedness-matrix restricted maximum likelihood) method to estimate SNP heritability in distantly related individuals from all genotyped and imputed SNPs in the dataset. The convention excludes those subjects whose genetic relatedness exceeds the 0.025 threshold in GRM, which corresponds to relationships of third-fourth degree cousins. We applied such a cut-off while performing GREML analysis and one of each pair of closely related individuals was excluded from analysis, which resulted in a number range from 1,495 to 3,175 depending on phenotype (see Supplementary Material, Table S4). For all phenotypes we included age and sex as covariates. We also adjusted for the cohort of origin (GEN-R or NTR) to control for residual imputation stratification due to genotyping platform, true genetic differences, and possible phenotype differences.

Statistical Analyses

In both GEN-R and NTR, non-response analysis indicated no differences in the baseline characteristics of children whose assessment of child behavior problems was not completed at ages 7 and 9 years.

To explore the effect of extreme cases, often found in ratings of children's behavior problems, we winsorized phenotypes used in this study when it was required. If the corresponding absolute z-score was more than 3.29 for a phenotype, we replaced the raw score with the less extreme value, i.e. the next highest score plus one unit [104]. Additionally we checked for possible population stratification in a combined dataset, adjusting for 10 principal components (PCs) in analysis of each scale.

Sensitivity analyses were also performed to explore the influence of exclusion of extreme cases, skewness on SNP-heritability estimates, and the influence of study of origin. First, we estimated SNP-heritability by excluding extreme scores below or above three standard deviations from the mean. If such cases represent extremes, such as due to measurement errors, we expect SNP heritability to increase after exclusion of these cases. However, if they represent genuine outliers, we expect them to be also outliers for heritable traits and consequently the SNP heritability will decrease after exclusion of these cases. Second, we transformed the data to the quantile normalized scale, using the Van der Waerden transformation. This transformation reduces the extreme influence that outliers could have by ranking them as low or high within a normal distribution [105], although it results in some loss of phenotypic information. In addition, we performed GCTA separately on the two participating studies (i.e. GEN-R and NTR) to explore possible effects of the specific study. All transformations were conducted in SPSS 21.0 [106].

Results

Genotypic and Phenotypic Sample Characteristics

The sample characteristics of the children participating in each study and in the combined dataset, before and after exclusion of related individuals, are presented in Supplementary Tables S3 and S4. The distribution of age, sex, and behavior problems did not significantly differ between the two studies.

Estimates of SNP heritability

Table 1 summarizes the SNP heritability estimates using the combined GRM, adjusting for age, sex and sample of origin. For the mother ratings of child problem behavior, estimates were substantial and statistically significant for the ADHD Combined scale (SNP h^2=0.40, *SE*=.14, *p* = .001), the ADHD Inattentive scale (SNP h^2=0.37, *SE*=.14, *p*=.003) and the Hyperactive-Impulsive scale (SNP h^2=0.45, *SE*=.14, *p*=.0006) measured by the Conners' Parent Rating Scale (CPRS). We also found significant SNP heritability estimates for the CBCL Total problems score (SNP h^2=0.18, *SE*=.10, *p*=.03). For the teacher ratings, we obtained significant SNP heritability estimates for both the Attention problems scale (SNP h^2=0.71, *SE*=.22, *p*=.0006) and the Externalizing scale (SNP h^2=0.44, *SE*=.22, *p*=.03). No significant estimates were found for the CPRS ODD scale, CBCL PDD subscale, CBCL Internalizing and Externalizing scales and non-verbal cognition. When 10 PCs were used as covariates, we found nonsignificant difference in SNP heritability estimates (1-3%

drop or 1-2% increase, data not shown) for all scales. The level of significance remained the same, except for the CBCL Total problems score (p-value = 0.07 and p-value = 0.03 with and without PCs adjustment accordingly).

Table 1. SNP heritability estimates in child behavior problems.

	Age (*SD*)	**SNP h^2**	***SE***	**95% *CI*[a]**	***N***	***P*-value**
Parent ratings						
CPRS ADHD Combined scale	8.34 (0.7)	0.40	0.14	(0.13, 0.67)	2,262	<0.01**
CPRS ADHD Inattentive scale	8.34 (0.7)	0.37	0.14	(0.10, 0.64)	2,262	<0.01**
CPRS Hyperactive-Impulsive scale	8.34 (0.7)	0.45	0.14	(0.18, 0.72)	2,260	<0.001***
CPRS ODD scale	8.34 (0.7)	0.20	0.14	(0.00, 0.47)	2,262	0.07
CBCL Internalizing scale	6.57 (0.83)	0.12	0.10	(0.00, 0.32)	3,175	0.11
CBCL Externalizing scale	6.57 (0.83)	0.12	0.10	(0.00, 0.32)	3,174	0.13
CBCL Total problems score	6.57 (0.83)	0.18	0.10	(0.00, 0.38)	3,175	<0.05*
CBCL PDD subscale	3.15 (0.23)	0.16	0.11	(0.00, 0.33)	3,015	0.07
Teacher ratings						
TRF Attention problems scale	6.82 (2.35)	0.71	0.22	(0.28, 1.00)	1,495	<0.001***
TRF Externalizing scale	6.82 (2.35)	0.44	0.22	(0.01, 0.87)	1,495	<0.05*
Observational ratings						
Non-verbal cognition	6.14 (0.42)	0.11	0.16	(0.00, 0.42)	1,974	0.23

Note: All analyses were performed with the combined GRM and were adjusted for age, sex, and sample of origin (GEN-R or NTR) on winsorized scores. CPRS = Conners' Parent Rating Scale, ADHD= Attention Deficit/ Hyperactivity Disorder, ODD= Oppositional Defiant Disorder, CBCL= Child Behavior Checklist, PDD= Pervasive Developmental Disorder, TRF=Teacher's Rating Form, SE= standard error

[a]Note: SNP heritability estimates are limited to (0.00-1.00)

*p < 0.05; **p<0.01; ***p<0.001

Sensitivity Analyses

To examine the influence of extreme cases on SNP heritability estimates and to enable comparison with the TEDS results [88] that were based on deletion of extreme cases, we also performed GCTA analyses excluding individuals above or below three standard deviations from the mean. Overall, the exclusion of extreme cases decreased the SNP heritability estimates almost by half in most of the complex problem behaviors, suggesting that these children are genuine outliers and that their extreme phenotype values do not represent measurement errors or other artifacts. Even after removal of outliers, SNP heritability estimates for all scales of CPRS (ADHD combined, ADHD Inattentive, ADHD Hyperactive-Impulsive and ODD scale) were still substantial. The CBCL Total problems score and the teacher-reported Attention problems and Externalizing problems also remained significant after exclusion of extreme cases. The results are summarized in Table 2.

Table 2. Impact of extreme cases on SNP heritability estimates in child behavior problems.

	Age (*SD*)	SNP h^2	*SE*	95% *CI*[a]	*N*	*P*-value
Parent ratings						
CPRS ADHD Combined scale	8.34 (0.7)	0.22	0.14	(0.00, 0.49)	2,240	0.05*
CPRS ADHD Inattentive scale	8.34 (0.7)	0.24	0.14	(0.00, 0.51)	2,229	<0.05*
CPRS Hyperactive-Impulsive scale	8.34 (0.7)	0.33	0.15	(0.04, 0.62)	2,231	0.01*
CPRS ODD scale	8.34 (0.7)	0.28	0.14	(0.01, 0.55)	2,246	<0.05*
CBCL Internalizing scale	6.57 (0.83)	0.04	0.10	(0.00, 0.24)	3,139	0.36
CBCL Externalizing scale	6.57 (0.83)	0.06	0.10	(0.00, 0.26)	3,136	0.28
CBCL Total problems score	6.57 (0.83)	0.16	0.10	(0.00, 0.36)	3,143	0.05*
CBCL PDD subscale	3.15 (0.23)	0.14	0.11	(0.00, 0.36)	2,999	0.10
Teacher ratings						
TRF Attention problems scale	6.82 (2.35)	0.49	0.22	(0.06, 0.92)	1,470	0.01*
TRF Externalizing scale	6.82 (2.35)	0.46	0.23	(0.01, 0.91)	1,463	<0.05*
Observational ratings						
Non-verbal cognition	6.14 (0.42)	0.11	0.16	(0.00, 0.42)	1,968	.23

Note: In all analyses, statistical outliers (mean ±3sd) were excluded. All analyses were performed with the combined GRM and were adjusted for age, sex, and sample of origin (GEN-R or NTR). CPRS = Conners' Parent Rating Scale, ADHD= Attention Deficit/ Hyperactivity Disorder, ODD= Oppositional Defiant Disorder, CBCL= Child Behavior Checklist, PDD= Pervasive Developmental Disorder, TRF=Teacher's Rating Form, SE= standard error

[a]Note: SNP heritability estimates are limited to (0.00-1.00)

*$p < 0.05$

Furthermore, similar to the TEDS study [88], we performed GCTA analyses on the quantile normalized scales using the Van der Waerden transformation to examine the potential influence of skewness on SNP heritability estimates. The SNP heritability with the transformed scales were similar to those with untransformed scales, with substantial genetic effects contributing to the ADHD Combined scale (SNP h^2=0.30, *SE*=.14, *p*=.01), ADHD Inattentive scale (SNP h^2=0.30, *SE*=.14, *p*=.01), and Hyperactivity-Impulsive scale (SNP h^2=0.37, *SE*=.14, *p*=.004) rated by the mother using the CPRS. Also, the CBCL PDD subscale (SNP h^2=0.18, *SE*=.11, *p*=.05) and the teacher-reported Attention problems scale (SNP h^2=0.64, *SE*=.22, *p*=.002) and Externalizing scale (SNP h^2=0.60, *SE*=.22, *p*=.004) yielded significant SNP heritability estimates. The results are summarized in Supplementary Table S5.

Finally, we also provided SNP heritability estimates of the two samples independently. As expected, the smaller NTR sample shows estimates with larger SE values. Although variable, the SNP heritability estimates of the two samples did not differ significantly from each other. These results are summarized in Supplementary Table S6.

Discussion

The aim of this study was to provide estimates of SNP heritability of normative differences in attention deficit problems (measured at 7 and 9 years), externalizing and internalizing problems (measured at 7 years), pervasive developmental problems (measured at 3 years) and non-verbal cognitive function (measured at 7 years) in population-based samples. Our study provides evidence of significant SNP heritability for attention deficit/hyperactivity problems, externalizing and total problems rated by mother or teacher. We identified nonsignificant SNP heritability estimates for pervasive developmental and internalizing problems. These results are parallel to twin heritabilities previously reported on the same phenotypes, i.e. higher twin heritabilities were associated with higher SNP heritabilities. Sensitivity analyses showed that SNP heritability estimates decreased but remained significant for most phenotypes after exclusion of the extreme cases.

Previous studies on the heritability captured by common SNPs have yielded significant SNP heritability estimates for normative differences in autistic-like traits [83, 107], in clinical cases of childhood-onset OCD [86] and ADHD [84, 85]. Quantifiable, although non-significant SNP heritability has also been reported for internalizing problems in population-based samples of preschoolers [76]. Surprisingly, however, a recent TEDS study by Trzaskowski et al [88] indicated no additive genetic effects for common child behavior problems. This discrepancy may be due to several factors. First, there are methodological differences between the two studies. In the present study, we removed ethnic outliers instead of correcting for them using principal components analysis. Moreover, we estimated heritability with and without extreme cases, showing that in some cases, treatment of outliers results in substantially different findings.

Extreme cases might be biologically significant extremes or they might constitute statistical outliers. Our results indicate that extremes were more likely to be genetic extremes rather than statistical outliers and they suggest that in the TEDS study [88] the exclusion of extreme cases may have resulted in an underestimation of SNP heritability [108]. Winsorizing the extreme cases instead of excluding them may address the problem of extremely skewed distributions while still retaining information for all subjects. It should be noted, however, that even after exclusion of the extreme cases, we found significant additive genetic heritability for ADHD-related symptoms and children's behavior problems.

Secondly, the two studies involved different samples. The TEDS sample [88] involved 12-year-old children, whereas in our sample we analyzed data on behavior problems at three, seven and nine year olds. Estimations of genetic effects may differ developmentally, although the direction depends on the phenotype of interest. For example, SNP heritability of autistic-like traits was shown to be low (SNP h^2=0.24, *SE*=.07, *N* = 5,204) but strongest in early childhood [83], whereas it increased from ages 7-to 12-years in the case of general cognitive ability [109]. Previous twin studies have also indicated an increase in the heritability for general cognitive ability [110], as well as for non-verbal IQ [111] from childhood to adulthood. In line with the low heritability estimates in 5 to 6-year-old twins for non-verbal IQ, we found no significant SNP heritability of non-verbal cognitive ability at 6 years, in a subsample of 1,974 unrelated individuals. Similarly, our study indicated non-significant SNP heritability of pervasive developmental problems in 3-year old children (n = 3,015). Given the low overall heritability, larger samples may be needed to estimate modest SNP heritability of non-verbal cognition and pervasive developmental problems in early childhood. The perception of genetic heritability as time and age-dependent [112] could explain discrepancies between samples and between measurements at different time-points (e.g. CBCL measures at 7 years and CPRS measures at 9 years) and suggests that SNP heritability estimates cannot be easily generalized across age.

Thirdly, SNP heritability, as an estimation of the fraction of phenotypic variation explained by common SNPs, is dependent of sample characteristics [113]. Thus, as a population property, SNP heritability estimates can differ between samples, because environmental factors are different. Environmental influences may play a more important role in a sample derived from multiple, culturally diverse sites in the United Kingdom (UK), while genetic effects would be more prominent in the geographically restricted and rather homogeneous Dutch society, in terms of socio-economic conditions. Parental reports of child problem behaviors might partly be determined by subjective criteria for what parents consider to be problem behavior and these criteria may be dependent on cultural norms or socio-economic circumstances, such as crowding [114].

In this study, we found non-significant SNP heritability for parent-reported internalizing problems in 7-year-old children. One reason for this finding could be the difficulty in assessing internalizing symptoms in early childhood. Internalizing symptoms are often

not overtly expressed in young children and thus not easily observed by the parents [115]. Another reason could be that since the prevalence of internalizing symptoms typically increases in middle-to-late adolescence [116, 117], we are not yet able to identify all children who will develop internalizing symptoms later in life. The particularly high heterogeneity and the distinctive genetic architecture of internalizing problems have also been addressed in previous work [118].

A limitation of this study is the sample size. GCTA power calculations indicate that even with large sample sizes, the SEs of the SNP heritability estimates are large [119]. Thus, even larger samples are needed to estimate modest additive genetic effects. However, the sample size of the current study is, for most phenotypes, comparable to the study of Trzaskowski et al. [88], indicating that sample size is not exclusively responsible for the discrepancies between the two studies. An inherent parameter to most behavior problems research is the skewed distribution of the phenotypes. Nevertheless, sensitivity analyses with transformed distributions and winsorized extreme cases did not reduce the significant SNP heritability estimates to non-significance. This study is based on data from two longitudinal studies (GEN-R and NTR). Systematic attrition is a limitation inherent to longitudinal studies [120], potentially leading to selective dropout of high-risk individuals, and thus to underestimation of the heritability of common behavior problems in children. However, previous research has shown that psychopathology of the participants has a small to moderate effect on attrition rates [121, 122] and estimations from longitudinal studies are robust and generalizable [123, 124]. Finally, the results of this study are derived from population-based samples of children. Although it has been shown that additive effects of hundreds of SNPs are responsible for observed normal variation in most quantitative traits [125], it is possible that the genetic architecture of children diagnosed with severe behavioral problems differs from that of children in population-based samples (e.g. increased role of rare variants, *de novo* mutations and dominance genetic effects).

In summary, this study provides molecular genetic evidence of additive genetic influences on specific child behavior problems in an ethnically and socio-economically homogeneous sample. SNP-heritability for other common behavior problems in children, or for the p factor as proposed by Caspi et al. (2014) [126], remains to be estimated. SNP heritability estimates may be influenced by diversity in a socioeconomic environment, developmental stage, and study design, arguing for approaches that model gene-by-environment interactions, developmental information, and possibly data from population-based and clinical samples in GCTA research. Our results provide support for and encourage the continuation of GWAS efforts by genetics consortia focusing on complex behavioral traits in search of elusive heritability.

Supplementary Material

Table S1. Twin-Based Heritability Estimates in Netherlands Twin Register for the Phenotypes Analyzed in This Study (Attentional Deficit Problems, Externalizing Problems, Internalizing Problems, and Autistic-Like Traits), as Previously Published.

Phenotypes	**Age (y)**	**n** *twin pairs*	**Estimations**	**Reference**
Parent Ratings				
CPRS ADHD Combined scale	7	1,595	0.78	[127]
CPRS ADHD Inattentive scale	9	3,470	0.38-0.78	[128]
CPRS Hyperactive-Impulsive scale	9	3,470	0.72-0.80	[128]
CPRS ODD scale	7	1,595	0.55	[129]
CBCL Internalizing scale	7	1,940	0.38	[130]
CBCL Externalizing scale	7	1,940	0.52	[130]
CBCL Total Problems score	3	1,358	0.38	[131]
CBCL PDD subscale	NA	NA	NA	NA
Teacher Ratings				
TRF Attention Problems scale	7	2,259	0.39	[132]
TRF Externalizing scale	7	215	0.43	[133]
Observational Ratings				
Nonverbal cognition	5	237	0.31	[134]

Note: When both mother and father reported estimates were available, the mother ratings were selected. ADHD = attention-deficit/hyperactivity disorder; CBCL = Child Behavior Checklist; CPRS = Conners' Parent Rating Scale; NA = not available; ODD = oppositional defiant disorder; PDD = pervasive developmental disorder; TRF = Teacher's Rating Form.

Table S2. List of Common Child Behavior Checklist (CBCL) 1 ½-5 Years and CBCL 6-18 Years Items Used to Assess Children's Behavior Problems in the Two Participating Studies (the Generation R Study or the Netherlands Twin Register).

Scale	Common Items in CBCL 1 ½- 5 and CBCL 6-18
Internalizing	sulks a lot worries nervous, tense self-conscious or easily embarrassed too fearful or anxious unhappy, sad, or depressed aches or pains (without medical cause) constipated, doesn't move bowels headaches (without medical cause) nausea, feels sick (without medical cause) stomachaches or cramps (without medical cause) vomiting, throwing up (without medical cause) withdrawn, doesn't get involved with others
Externalizing	wanders away destroys things belonging to his/her family or other children disobedient doesn't seem to feel guilty after misbehaving gets in many fights hits others physically attacks people screams a lot stubborn, sullen, or irritable sudden changes in mood or feelings wants a lot of attention temper tantrums or hot temper
Total Problems	aches or pains (without medical cause) acts too young for age can't concentrate, can't pay attention for long can't sit still, restless, or hyperactive clings to adults or too dependent constipated, doesn't move bowels cries a lot cruel to animals destroys his/her own things

Table S2 Continued

Scale	Common Items in CBCL 1 ½- 5 and CBCL 6-18
	destroys things belonging to his/her family or other children
	disobedient
	doesn't eat well
	doesn't get along with other children
	doesn't seem to feel guilty after misbehaving
	easily jealous
	fears certain animals, situations, or places
	gets hurt a lot, accident prone
	gets in many fights
	headaches (without medical cause)
	hits others
	nausea, feel sick (without medical cause)
	nervous movements or twitching
	nervous or tense
	nightmares
	overeating
	overtired
	physically attacks people
	picks nose, skin, or other parts of body
	plays with own sex parts too much
	poorly coordinated or clumsy
	problems with eyes (without medical cause)
	rashes or other skin problems (without medical cause)
	screams a lot
	self-conscious or easily embarrassed
	shows little interest in things around him/her
	too shy or timid
	sleeps less than most kids during day and/or night
	speech problems
	stares into space or seems preoccupied
	stomachaches or cramps (without medical cause)
	strange behavior
	stubborn, sullen, or irritable
	sudden changed in mood or feelings

Table S2 Continued

Scale	Common Items in CBCL 1 ½- 5 and CBCL 6-18
	sulks a lot
	talks or cries out in sleep
	temper tantrums or hot temper
	too fearful or anxious
	underactive, slow moving, or lacks energy
	unhappy, sad, or depressed
	unusually loud
	vomiting, throwing up (without medical cause)
	wakes up often at night
	wanders away
	wants a lot of attention
	whining
	withdrawn, doesn't get involved with others
	worries

Table S3. Descriptive Statistics for the Distribution of Common Children's Problem Behavior, in the Two Genotyped Samples and the Combined Dataset.

	GEN-R	**NTR**	**Combined**
		CPRS ADHD Combined Scale	
n	1,971	854	2,825
Age (SD)	8.15 (0.21)	9.03 (1.24)	8.41 (0.81)
Sex (% girls)	49	54	51
Mean (SD)	7.27 (6.70)	8.16 (8.01)	7,53 (7.13)
Skewness (SE)	1.29 (0.06)	1,27 (0.08)	1,32 (0.05)
Kurtosis (SE)	1.49 (0.11)	1.14 (0.17)	1.52 (0.09)
		CPRS ADHD Inattentive Scale	
n	1,974	854	2,828
Age (SD)	8.15 (0.21)	9.03 (1.24)	8.41 (0.81)
Sex (% girls)	49	54	51
Mean (SD)	3.09 (3.58)	3.86 (4.30)	3.32 (3.83)
Skewness (SE)	1.54 (0.06)	1.28 (0.08)	1.48 (0.05)
Kurtosis (SE)	2.30 (0.11)	1.03 (0.17)	1.91 (0.09)
		CPRS Hyperactive-Impulsive Scale	
n	1,973	852	2,825
Age (SD)	8.15 (0.21)	9.03 (1.24)	8.41 (0.81)
Sex (% girls)	49	54	51
Mean (SD)	2.08 (2.71)	3.00 (3.76)	2.36 (3.10)
Skewness (SE)	1.91 (0.06)	1.67 (0.08)	1.94 (0.05)
Kurtosis (SE)	4.10 (0.11)	2.54 (0.17)	4.12 (0.09)

Table S3 Continued

	GEN-R	NTR	Combined
		CPRS ODD Scale	
n	1,973	856	2,829
Age (SD)	8.15 (0.21)	9.03 (1.24)	8.41 (0.81)
Sex (% girls)	49	54	51
Mean (SD)	3.47 (2.87)	4.63 (3.80)	3.82 (3.22)
Skewness (SE)	1.15 (0.06)	1.05 (0.08)	1.23 (0.05)
Kurtosis (SE)	1.57 (0.11)	0.81 (0.17)	1.70 (0.09)
		CBCL PDD Problems Scale	
n	2,030	2,084	4,144
Age (SD)	3.04 (0.09)	3.32 (0.26)	3.18 (0.24)
Sex (% girls)	49	54	51
Mean (SD)	1.71 (1.89)	3.26 (2.86)	2.49 (2.55)
Skewness (SE)	1.70 (0.05)	1.17 (0.05)	1.50 (0.04)
Kurtosis (SE)	4.76 (0.11)	1.53 (0.11)	2.86 (0.08)
		CBCL Internalizing Scale	
n	2,175	2,126	4,301
Age (SD)	5.99 (0.38)	7.49 (0.44)	6.74 (0.85)
Sex (% girls)	49	54	52
Mean (SD)	1.44 (1.94)	2.17 (2.33)	1.80 (2.17)
Skewness (SE)	2.15 (0.05)	1.77 (0.05)	1,94 (0.04)
Kurtosis (SE)	6.35 (0.11)	4.48 (0.11)	5.29 (0.08)

Table S3 Continued

	GEN-R	NTR	Combined
		CBCL Externalizing Scale	
n	2,174	2,126	4,300
Age (SD)	5.99 (0.38)	7.49 (0.44)	6.74 (0.85)
Sex (% girls)	49	54	52
Mean (SD)	2.43 (2.85)	3.49 (3.55)	2.95 (3.26)
Skewness (SE)	2.15 (0.05)	1.48 (0.05)	1.78 (0.04)
Kurtosis (SE)	6.45 (0.11)	2.49 (0.11)	3.93 (0.08)
		CBCL Total Problems Score	
n	2,175	2,127	4,302
Age (SD)	5.99 (0.38)	7.49 (0.44)	6.74 (0.85)
Sex (% girls)	49	54	52
Mean (SD)	9.15 (8.51)	12.31 (10.06)	10.71 (9.44)
Skewness (SE)	2.04 (0.05)	1.44 (0.053)	1.70 (0.04)
Kurtosis (SE)	6.71 (0.11)	2.70 (0.11)	4.12 (0.08)
		TRF Attention Problems Scale	
n	1,419	358	1,777
Age (SD)	6.72 (2.42)	7.32 (0.42)	6.83 (2.18)
Sex (% girls)	47	48	47
Mean (SD)	3.03 (4.71)	5.69 (6.32)	3.56 (5.19)
Skewness (SE)	2.17 (0.07)	1.52 (0.13)	2.02 (0.06)
Kurtosis (SE)	4.84 (0.13)	2.44 (0.26)	4.37 (0.12)

Table S3 Continued

	GEN-R	NTR	Combined
		TRF Externalizing Scale	
n	1,419	358	1,777
Age (SD)	6.72 (2.42)	7.32 (0.42)	6.83 (2.18)
Sex (% girls)	47	48	47
Mean (SD)	2.39 (5.12)	4.75 (7.38)	2.86 (5.73)
Skewness (SE)	3.55 (0.07)	2.43 (0.13)	3.24 (0.06)
Kurtosis (SE)	15.25 (0.13)	6.59 (0.26)	12.52 (0.12)
		Nonverbal Cognition	
n	2,059	202	2,261
Age (SD)	6.09 (0.39)	6.73 (0.33)	6.15 (0.43)
Sex (% girls)	50	56	51
Mean (SD)	100 (15)	100 (15)	100 (15)
Skewness (SE)	-0.16	-0.15	-0.16
Kurtosis (SE)	0.39 (0.11)	0.26 (0.44)	0.35 (0.10)

Note: ADHD = attention-deficit/hyperactivity disorder; CBCL = Child Behavior Checklist; CPRS = Conners' Parent Rating Scale; GEN-R = the Generation R Study; NA = not available; NTR = Netherlands Twin Register; ODD = oppositional defiant disorder; PDD = pervasive developmental disorder; TRF = Teacher's Rating Form.

Table S4. Descriptive Statistics for the Distribution of Common Children's Problem Behavior, in the Two Genotyped Samples and the Combined Dataset After Applying Cut-Off 0.025 of Relatedness to Combined Sample

	GEN-R	**NTR**	**Combined**
	CPRS ADHD Combined Scale		
n	1,797	465	2,262
Age (SD)	8.15 (0.22)	9.05 (1.24)	8.34 (0.7)
Sex (% girls)	49	57	51
Mean (SD)	7.27 (6.7)	8.23 (8.11)	7.47 (7.02)
Skewness (SE)	1.33 (0.06)	1.3 (0.11)	1.35 (0.05)
Kurtosis (SE)	1.66 (0.12)	1.27 (0.23)	1.72 (0.1)
	CPRS ADHD Inattentive Scale		
n	1,799	465	2,264
Age (SD)	8.15 (0.22)	9.05 (1.24)	8.34 (0.7)
Sex (% girls)	49	56	51
Mean (SD)	3.08 (3.6)	3.89 (4.38)	3.25 (3.78)
Skewness (SE)	1.59 (0.06)	1.32 (0.11)	1.55 (0.05)
Kurtosis (SE)	2.48 (0.12)	1.12 (0.23)	2.21 (0.1)
	CPRS Hyperactive-Impulsive Scale		
n	1,798	462	2,260
Age (SD)	8.15 (0.22)	9.05 (1.24)	8.33 (0.7)
Sex (% girls)	49	57	51
Mean (SD)	2.09 (2.74)	2.97 (3.77)	2.27 (3.00)
Skewness (SE)	1.92 (0.06)	1.76 (0.11)	1.99 (0.05)
Kurtosis (SE)	4.13 (0.12)	3.02 (0.23)	4.49 (0.1)

Table S4 Continued

	GEN-R	NTR	Combined
		CPRS ODD Scale	
n	1,798	464	2,262
Age (SD)	8.15 (0.22)	9.05 (1.24)	8.34 (0.7)
Sex (% girls)	49	56	51
Mean (SD)	3.46 (2.86)	4.74 (3.91)	3.73 (3.15)
Skewness (SE)	1.13 (0.06)	1.1 (0.11)	1.25 (0.05)
Kurtosis (SE)	1.51 (0.12)	0.92 (0.23)	1.91 (0.1)
		CBCL PDD Problems Scale	
n	1,823	1,192	3,015
Age (SD)	3.04 (0.1)	3.32 (0.26)	3.15 (0.23)
Sex (% girls)	49	54	51
Mean (SD)	1.73 (1.9)	3.29 (2.87)	2.35 (2.45)
Skewness (SE)	1.67 (0.06)	1.17 (0.07)	1.57 (0.04)
Kurtosis (SE)	4.65 (0.11)	1.54 (0.14)	3.3 (0.09)
		CBCL Internalizing Scale	
n	1,977	1,198	3,175
Age (SD)	6.01 (0.38)	7.5 (0.44)	6.57 (0.83)
Sex (% girls)	49	54	51
Mean (SD)	1.46 (1.97)	2.23 (2.36)	1.75 (2.16)
Skewness (SE)	2.16 (0.06)	1.78 (0.07)	1.99 (0.04)
Kurtosis (SE)	6.34 (0.11)	4.62 (0.14)	5.57 (0.09)
		CBCL Externalizing Scale	
n	1,976	1,198	3,174
Age (SD)	6.01 (0.38)	7.5 (0.44)	6.57 (0.83)
Sex (% girls)	49	54	51
Mean (SD)	2.46 (2.88)	3.5 (3.57)	2.86 (3.2)
Skewness (SE)	2.17 (0.06)	1.51 (0.07)	1.88 (0.04)
Kurtosis (SE)	6.56 (0.11)	2.7 (0.14)	4.57 (0.09)

Table S4 Continued

	GEN-R	NTR	Combined
		CBCL Total Problems Score	
n	1,977	1,198	3,175
Age (SD)	6.01 (0.38)	7.5(0.44)	6.57 (0.83)
Sex (% girls)	49	54	51
Mean (SD)	9.28 (8.6)	12.39 (10.2)	10.46 (9.36)
Skewness (SE)	2.05 (0.06)	1.52 (0.07)	1.81 (0.04)
Kurtosis (SE)	6.78 (0.11)	3.00 (0.14)	4.81 (0.09)
		TRF Attention Problems Scale	
n	1,295	200	1,495
Age (SD)	6.74 (2.51)	7.3 (0.43)	6.82 (2.35)
Sex (% girls)	47	52	48
Mean (SD)	3.05 (4.73)	5.9 (6.65)	3.43 (5.12)
Skewness (SE)	2.17 (0.07)	1.54 (0.17)	2.1 (0.06)
Kurtosis (SE)	4.89 (0.14)	2.68 (0.34)	4.86 (0.13)
		TRF Externalizing Scale	
n	1,295	200	1,495
Age (SD)	6.74 (2.51)	7.3 (0.43)	6.82 (2.35)
Sex (% girls)	47	52	48
Mean (SD)	2.41 (5.13)	5.2 (8.01)	2.78 (5.68)
Skewness (SE)	3.52 (0.07)	2.28 (0.17)	3.31 (0.06)
Kurtosis (SE)	15.1 (0.14)	5.4 (0.34)	12.94 (0.13)
		Nonverbal Cognition	
n	1,856	118	1,974
Age (SD)	6.1 (0.4)	6.72 (0.35)	6.14 (0.42)
Sex (% girls)	51	54	51
Mean (SD)	99.85 (15.04)	100.26 (14.96)	99.87 (15.03)
Skewness (SE)	-0.12 (0.06)	-0.14 (0.22)	-0.12 (0.06)
Kurtosis (SE)	0.32 (0.11)	0.26 (0.44)	0.31 (0.11)

Note: ADHD = attention-deficit/hyperactivity disorder; CBCL = Child Behavior Checklist; CPRS = Conners' Parent Rating Scale; GEN-R = the Generation R Study; NA = not available; NTR = Netherlands Twin Register; ODD = oppositional defiant disorder; PDD = pervasive developmental disorder; TRF = Teacher's Rating Form.

Table S5. Impact of Skewness on Single Nucleotide Polymorphisms (SNP) Heritability Estimates in Child Behavior Problems

	SNP h^2	***SE***	**95% *CI*[a]**	***n***	***p* Value**
Parent Ratings					
CPRS ADHD Combined scale	0.30	0.14	(0.03, 0.57)	2,262	0.01*
CPRS ADHD Inattentive scale	0.30	0.14	(0.03, 0.57)	2,264	0.01*
CPRS Hyperactive-Impulsive scale	0.37	0.14	(0.10, 0.64)	2,260	<0.01**
CPRS ODD scale	0.19	0.14	(0.00, 0.46)	2,262	0.09
CBCL Internalizing scale	0.07	0.10	(0.00, 0.27)	3,175	0.26
CBCL Externalizing scale	0.12	0.10	(0.00, 0.32)	3,174	0.13
CBCL Total Problems score	0.13	0.10	(0.00, 0.33)	3,175	0.10
CBCL PDD subscale	0.18	0.11	(0.00, 0.40)	3,015	0.05*
Teacher Ratings					
TRF Attention Problems scale	0.64	0.22	(0.21, 1.00)	1,495	<0.01**
TRF Externalizing scale	0.60	0.22	(0.17, 1.00)	1,495	<0.01**
Observational Ratings					
Nonverbal cognition	0.11	0.16	(0.00, 0.42)	1,974	0.23

Note: All scales have been normalized using the Van der Waerden method. All analyses were performed with the combined genetic relationship matrix and adjusted for age, sex, and sample of origin (the Generation R Study or the Netherlands Twin Register). ADHD = attention-deficit/hyperactivity disorder; CBCL = Child Behavior Checklist; CPRS = Conners' Parent Rating Scale; ODD = oppositional defiant disorder; PDD = pervasive developmental disorder; SE = standard error; TRF = Teacher's Rating Form.

[a] SNP heritability estimates are limited to (0.00-1.00).

*p<.05; **p<.01; ***p<.001

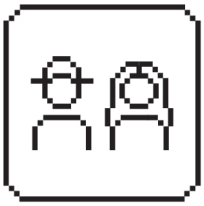
Table S6. Comparison of Single Nucleotide Polymorphisms (SNP) Heritability Estimates in Child Behavior Problems, in the Two Participating Samples

	GEN-R Sample					NTR Sample				
	SNP h^2	***SE***	**95% *CI*[a]**	***n***	***p* Value**	**SNP h^2**	***SE***	**95% *CI*[a]**	***n***	***p* Value**
Parent Ratings										
CPRS ADHD Combined scale	0.38	0.18	(0.03, 0.73)	1,798	0.01*	0.11	0.70	(0.00, 1.00)	477	0.44
CPRS ADHD Inattentive scale	0.36	0.17	(0.03, 0.69)	1,800	0.01*	0.01	0.70	(0.00, 1.00)	477	0.49
CPRS Hyperactive-Impulsive scale	0.52	0.18	(0.17, 0.87)	1,799	<0.01**	0.00	0.70	(0.00, 1.00)	474	0.50
CPRS ODD scale	0.33	0.18	(0.00, 0.68)	1,799	<0.05*	0.10	0.66	(0.00, 1.00)	476	0.44
CBCL Internalizing scale	0.33	0.17	(0.00, 0.66)	1,979	<0.05*	0.21	0.26	(0.00, 0.72)	1,223	0.20
CBCL Externalizing scale	0.23	0.17	(0.00, 0.56)	1,978	0.08	0.00	0.26	(0.00, 0.51)	1,223	0.50
CBCL Total Problems score	0.35	0.17	(0.02, 0.68)	1,979	<0.05*	0.29	0.27	(0.00, 0.82)	1,223	0.13
CBCL PDD scale	0.29	0.18	(0.00, 0.64)	1,824	0.06	0.00	0.27	(0.00, 0.53)	1,217	0.50
Teacher Ratings										
TRF Attention Problems scale	0.73	0.25	(0.24, 1.00)	1,296	0.001**	.00	1.53	(0.00, 1.00)	203	0.50
TRF Externalizing scale	0.59	0.25	(0.10, 1.00)	1,296	<0.01**	.00	1.50	(0.00, 1.00)	203	0.50
Observational Ratings										
Nonverbal cognition	0.14	0.17	(0.00, 0.47)	1,856	0.19	.00	2.80	(0.00, 1.00)	122	0.50

Note: All analyses are adjusted for age and sex. ADHD = attention-deficit/hyperactivity disorder; CBCL = Child Behavior Checklist; CPRS = Conners' Parent Rating Scale; GEN-R = the Generation R Study; NTR = Netherlands Twin Register; ODD = oppositional defiant disorder; PDD = pervasive developmental disorder; SE = standard error; TRF = Teacher's Rating Form.

[a]SNP heritability estimates are limited to (0.00-1.00).

*p<.05; **p<.01; ***p<.001



CHAPTER 4

HERITABILITY OF BEHAVIORAL PROBLEMS IN 7-YEAR OLDS BASED ON SHARED AND UNIQUE ASPECTS OF PARENTAL VIEWS

This chapter is based on:

Iryna O. Fedko, Laura W. Wesseldijk, Michel G Nivard, Jouke-Jan Hottenga, Catharina E.M. van Beijsterveldt, Christel M. Middeldorp, Meike Bartels, and Dorret I. Boomsma. Heritability of behavioral problems in 7-year olds based on shared and unique aspects of parental views *(as accepted by Behavior Genetics).*

Abstract

In studies of child psychopathology, phenotypes of interest are often obtained by parental ratings. When behavioral ratings are obtained in the context of a twin study, this allows for the decomposition of the phenotypic variance, into a genetic and a non-genetic part. If a phenotype is assessed by a single rater, heritability is based on the child's behavior as expressed in the presence of that particular rater, whereas heritability based on assessments by multiple raters allows for the estimation of the heritability of the phenotype based on rater agreement, as well as the heritability of the rater specific view of the behavior. The aim of this twin study was to quantify the rater-common and rater specific contributions to the variation in children's behavioral problems. We estimated the heritability of maternal and paternal ratings of the Child Behavior Checklist (CBCL) 6-18 empirical emotional and behavioral problem scales in a large sample of 12,310 7-year old Dutch twin pairs. Between 30% and 59% of variation in the part of the phenotype parents agree upon was explained by genetic effects. Common environmental effects that make children in the same family similar explained less variance, ranging between 0% and 32%. For unique views of their children's behavioral problems, heritability ranged between 0% and 20% for maternal and between 0% and 22% for paternal views. Between 7% and 24% of the variance was accounted for by common environmental factors specific to mother's and father's views. The proportion of rater shared and rater specific heritability can be translated into genetic correlations between parental views and inform the design and interpretation of results of molecular genetic studies. Genetic correlations were nearly or above 0.7 for all CBCL based psychopathology scales. Such large genetic correlations suggest two practical guidelines for Genome-Wide Association Studies (GWAS): when studies have collected data from either fathers or mothers, the shared genetic aetiology in parental ratings indicates that is possible to analyze paternal and maternal assessments in a single GWAS or meta-analysis. Secondly, if a study has collected information from both parents, a gain in statistical power may be realized in GWAS by the simultaneous analysis of the data.

Introduction:

To assess children's behavioral and emotional problems, researchers often rely on parental ratings. However, parents are not always in agreement on the behavior of their child. Maternal and paternal ratings on the Child Behavior Checklist (CBCL) 6-18, for example, correlate around 0.75, which is lower than the average test-retest reliability of the instrument, which is 0.89 for the empirical subscales [135-137]. Differences in parental normative standards or perception of child's behavior could explain why the correlations between parents are below the test retest reliability; an alternative or additional explanation involves the existence of specific parental views on the child's behaviors if a child behaves differently in the presence of each parent [34-36]. Maternal ratings are the most common single informant assessment found in the literature. However, as children interact with both parents, adding paternal observations may provide additional information about a child's behavior.

The Child Behavior Checklist 6-18 (CBCL 6-18) assesses child behavioral and emotional problems on a number of scales that indicate problems in the Internalizing (INT) domain (Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints) and the Externalizing (EXT) domain (Rule-Breaking, and Aggressive Behavior) as well as Social, Thought, Attention Problems, Dysregulation, which sums Anxious/Depressed, Aggressive Behavior and Attention Problems [138], and Total Problems. The contribution of genetic (twin heritability) and environmental effects to the variation in rater agreement and disagreement of some of these scales were explored for children of age 7 years and showed that the common part of multi-informant assessments was the most heritable, ranging from 24% to 51% [34, 139-142], free of possible rater bias and specific parental views. Specific parental views usually were less heritable, ranging from 4% to 24% across the studies, scales and domains, but still provided information about child behavior. Phenotypes such as Somatic Complaints, Rule-Breaking Behavior, Social Problems and the Dysregulation Profile received less attention.

In molecular genetic studies, heritability as estimated in the twin model is often contrasted with SNP-heritability, the phenotypic variance explained by a large subset of all common genetic variants (single nucleotyde polymorphisms, SNPs). SNP-heritability can be obtained from Genomic-Relatedness-matrix restricted Maximum Likelihood (GREML) analysis [19, 20] where the effect of individual genetic variants can be estimated in Genome Wide Association Studies (GWAS). In general, SNP heritability and twin-heritability are correlated across traits, i.e. traits with high twin heritability tend to have a high SNP-heritability. The power to detect genetic variants in a GWAS in turn is also related to, among other factors, the SNP and twin heritability estimates. If child behavioral problems assessed by multiple informants, for example mother, father or teacher, are more heritable, due to the focus on the part of the behavior on which all raters agree and with reduction of measurement error, rater bias or specific rater view, power will be increased in a GWAS by combining information from different raters. Alternatively, a substantial rater specific heritability might indicate that ratings

from particular informants should be analyzed separately in GWAS, to identify variants contributing to that part of the behavior that is only seen by a specific rater in a specific context. Results obtained from twin studies with multiple informants may address these questions and convey additional information, which can aid in the design of molecular genetic studies and results interpretation.

The aim of this study was to estimate the relative contribution of genetic factors (twin heritability) to the raters agreement and disagreement of the all empirical scales of CBCL 6-18 in a large sample (N = 12,310 pairs) of twins around age 7 years and in this way inform molecular studies. These twins participate in an ongoing longitudinal data collection for the Netherlands Twin Register [43, 44]. We investigated agreement and disagreement between parents in the psychometric model [34, 36]. The large sample size allowed us to exploit the liability threshold model [143] and consider data as categorical, as in population based samples CBCL scales tend to be skewed. It has been shown that this approach has an advantage over various data transformations for skewed data [144]. The large sample further allowed for the assessment of quantitative and qualitative sex differences.

Methods:

Participants and Measures:

The data analyzed in this study are obtained by the Netherlands Twin Register (NTR), which is a population-based longitudinal study of the health and life style of twins and their families. Participants are voluntarily registered with the NTR and the data collection protocol was approved by the Medical Research Ethics Committee of the VU University Medical Center. For 12,629 twin pairs, born between 1986 and 2006, maternal and paternal ratings were available. Data from 312 pairs were excluded since one or both twins had an illness or handicap that interfered with daily functioning. For the same-sex twin pairs zygosity was determined by blood group (n pairs = 194), DNA polymorphisms (n pairs = 1,558) or by parental zygosity questionnaire (n pairs = 6,661). Twins for whom zygosity was unknown (n pairs = 7) were also excluded from the analysis. The final sample comprised 12,310 twin pairs: 2,079 monozygotic male (MZM), 2,086 dizygotic male (DZM), 2,324 monozygotic female (MZF), 1,924 dizygotic female (DZF) and 3,897 opposite-sex pairs (DOS). CBCL data were collected when the twin pair was about 7 years old (mean = 7.45, sd = 0.40, N = 24,620). Maternal questionnaires were available for 12,086 pairs, paternal questionnaires for 8,555 pairs. Either the CBCL 4-18 [145] or the CBCL 6–18 [137] were used, depending on the year in which the questionnaire was sent to participants. The sum scores for each scale were computed based on syndrome scale (version Achenbach and Rescorla, 2001) [137]. Means, standard deviations, and information on skewness and kurtosis for all scales are provided in Supplementary Table 1. The scale scores are the sum of all items, where a lower score indicates less or no behavior problems and higher scores indicate the presence of behavioral problems. Because twin studies represent population samples, the distribution of CBCL data is often

skewed (L-shaped). This could lead to biased parameter estimates [144]. Therefore, we categorized the data in 3 categories (0,1,2) and carried out the analyses using a liability threshold model. The two thresholds approximately divided the dataset with both parental ratings into 3 equal parts. The liability threshold model assumes an underlying normal distribution, which we scaled with a mean of 0 and unit variance. In this context thresholds reflect the prevalences of childhood psychopathology rated by mother and father. Descriptive statistics were calculated with SPSS [146]. Relationships between raw data and categories can be found in Supplementary Table 2.

Genetic epidemiological analyses:
For each CBCL scale, a 4x4 polychoric correlation matrix was estimated in all zygosity by sex groups (MZM, DZM, MZF, DZF and DOS). It contained parental twin1-twin2 correlations, the parental cross-correlations between twins (e.g. father rating of twin1 and mother rating of twin2) and the parental agreement correlations (Table 1).

Table 1. 4x4 correlation matrix for 5 zygosity by sex groups

	Mother twin 1	**Father twin 1**	**Mother twin 2**	**Father twin 2**
Mother twin 1	1	Parental agreement correlation	Mother correlation twin1-twin2	Mother(twin1)-father(twin2) cross-correlation
Father twin 1	Parental agreement correlation	1	Father(twin1)-mother(twin2) cross-correlation	Father correlation twin1-twin2
Mother twin 2	Mother correlation twin1-twin2	Father(twin1)-mother(twin2) cross-correlation	1	Parental agreement correlation
Father twin 2	Mother(twin1)-father(twin2) cross-correlation	Father correlation twin1-twin2	Parental agreement correlation	1

We constrained the correlations, such that 1) parental agreement correlations across sex and zygosity were equal, and 2) parental twin1-twin2 correlations across sex within MZ and DZ pairs were equal. The most parsimonious models, in terms of the constraints outlined above, were used in the subsequent genetic analyses. A psychometric genetic model, as described by Hewitt et al. (1992) [36] and Bartels et al. (2007) [34] was fitted to the data to estimate heritability and to disentangle shared and specific aspects of the parental ratings of the child's behavior. The model specifies a common component to the phenotype, as assessed by both parents and a unique component of the child's phenotype reflected in the assessments of each parent. The total variance of mother's

ratings (Vmother) is decomposed into common (Vshared) and unique (Vunique,mother) parts. The total variance of father's rating (Vfather) is decomposed in the same way. Vshared is decomposed into variance components representing additive genetic (Va,shared), common environment (Vc,shared) or dominant genetic (Vd,shared), and unique environment (Ve,shared) components. The additive genetic variance (Va, shared) represents the part of the heritability of the phenotype that is assessed by both parents. Likewise (Vunique,mother) is decomposed into (Va,unique,mother), (Vc,unique,mother) or (Vd,unique,mother), and (Ve, unique,mother). Vunique, father is decomposed in the same way. The additive genetic variance of the unique component of mother or father ratings (Va,unique) represents the part of the heritability of the trait that is uniquely expressed in the presence of each parent. Whether parents truly rated the specific aspect of the child behavior was tested by constraining the genetic variance of the specific view (Va,unique) to 0. We also tested if common environmental variance of the shared aspect of the phenotype (Vc,shared), which is free of bias and specific parental view, equaled 0. The rater bias is reflected in the proportion of common environmental variance of raters disagreement (Vc,unique). The genetic correlation between maternal and paternal ratings was computed based on the estimates of the additive genetic components of the most parsimonious model based on the formula:

$$r_g = Va,shared/(\sqrt{(Va,shared+Va,unique,mother)}\times\sqrt{(Va,shared+Va,unique,father)}).$$

The level of significance was 0.05/12 = 0.0042 to account for multiple testing of 12 CBCL scales. Analyses were performed in OpenMx 2.2.6 [147].

Results

Descriptive statistics

Means and standard deviations for boys and girls for mother and father ratings are Means and standard deviations for boys and girls for mother and father ratings are given in Table 2, which also gives the thresholds. For all CBCL scales, the means of the sum scores were higher for maternal than for paternal ratings and ratings for boys and girls were significantly different. Both mothers and fathers rated girls higher for the Anxious/ Depressed and Somatic Complaints subscales and the Internalizing scale. For all other scales boys scored higher than girls, with the exception of the Withdrawn/Depressed scale, for which they scored similarly. Differences in prevalences between boys and girls are reflected in the significant loss of fit of the model (Supplementary Material, Table 3) when constraining the thresholds to be the same across sexes.

Table 2. Means, and standard deviations of the untransformed data and thresholds estimated for categorical transformation of data

	Anxious / Depressed		Withdrawn / Depressed		Somatic Complaints		Rule-Breaking Behavior		Aggressive Behavior		Social Problems		Thought Problems		Attention Problems		INT		EXT		Dysregulation Profile		Total Problems	
	Mo	Fa	Mo	Fa	Mo	Fa	Mo	Fa	Mo	Fa	Mo	Fa	Mo	Fa	Mo	Fa	Mo	Fa	Mo	Fa	Mo	Fa	Mo	Fa
Boys																								
Mean	2.19	1.68	1.19	0.96	1.13	0.83	1.59	1.37	5.87	5.03	2.28	1.86	1.71	1.33	3.58	3.16	4.49	3.47	7.45	6.40	11.63	9.86	23.11	19.13
SD	2.58	2.12	1.68	1.51	1.60	1.30	2.04	1.87	5.37	4.83	2.59	2.29	2.19	1.86	3.27	3.02	4.61	3.86	6.92	6.23	9.28	8.30	17.70	15.56
Threshold 1	0.06	0.27	-0.11	0.07	-0.05	0.16	-0.24	-0.14	-0.43	-0.29	0.02	0.17	-0.34	-0.17	-0.34	-0.25	-0.19	0.06	-0.56	-0.43	-0.53	-0.34	-0.56	-0.34
Threshold 2	0.48	0.70	0.59	0.75	0.59	0.84	0.37	0.45	0.37	0.52	0.44	0.62	0.27	0.47	0.49	0.59	0.53	0.81	0.29	0.44	0.32	0.50	0.23	0.45
Girls																								
Mean	2.36	1.80	1.14	0.92	1.29	0.92	1.12	0.97	4.48	3.89	1.98	1.66	1.32	0.94	2.62	2.32	4.78	3.63	5.59	4.86	9.47	8.01	19.41	15.88
SD	2.61	2.19	1.59	1.42	1.74	1.39	1.61	1.49	4.42	4.03	2.29	2.02	1.83	1.50	2.83	2.65	4.72	3.92	5.60	5.07	8.02	7.25	15.69	13.84
Threshold 1	-0.03	0.19	-0.10	0.08	-0.15	0.09	-0.02	0.07	-0.19	-0.06	0.11	0.27	-0.14	0.09	0.01	0.10	-0.26	-0.01	-0.33	-0.21	-0.28	-0.10	-0.33	-0.09
Threshold 2	0.39	0.63	0.64	0.80	0.48	0.73	0.62	0.71	0.66	0.79	0.57	0.72	0.49	0.74	0.83	0.91	0.47	0.78	0.59	0.70	0.62	0.80	0.49	0.71

Correlations between twins and raters

Correlations, estimated in 4x4 matrix for each of the five zygosity by sex groups are summarized in Table 3. For all scales parental agreement correlations were similar between boys and girls as well as between MZ and DZ twins. Parental agreement correlations were constrained to be equal across sex and zygosity, this did not lead to significant worsening of fit of the model to the data (Supplementary Table 3). We detected sex effects for the Aggressive Behavior, Externalizing scale and Dysregulation Profile reflected by significant sex differences in parental agreement, but since the differences were small, we decided not to model sex specific effects in the variance decomposition. Parental twin correlations and cross-twin-cross-rater-correlations were higher for MZ twins, than for DZ twins, indicating that rater disagreement partly reflects a rater specific or context specific view and not only rater bias. Parental twin correlations were similar for boys and girls within MZ and DZ pairs, and therefore were constrained to be the equal across sex in subsequent submodels (Supplementary Table 3).

Table 4 summarizes the correlations obtained from the constrained model. For all scales, except Attention problems, parental correlations in MZ twin pairs were lower than one and twice the DZ correlations, or less, indicating contributions of Additive Genetic (V_A), Shared Environmental (V_C) and Unique Environmental (V_E) variation to the total phenotypic variation. For Attention problems MZ correlations were, lower than one and larger than twice the DZ correlations, indicating Additive Genetic (V_A), Dominant Genetic (V_D) and Unique Environmental (V_E) variation. Thus, for all traits a V_{ACE} variance decomposition model was fitted, except for Attention Problems, for which an V_{ADE} model was used.

Table 3. Correlations estimated from a saturated model: parental agreement, twin correlations and twin cross-correlations

	Anxious/ Depressed		Withdrawn / Depressed		Somatic Complaints		Rule-Breaking Behavior		Aggressive Behavior		Social Problems		Thought Problems		Attention Problems		INT		EXT		Dysregulation Profile		Total Problems	
*Parental agreement **																								
MZM	0.65		0.61		0.68		0.65		0.75		0.67		0.63		0.74		0.65		0.75		0.75		0.74	
DZM	0.66		0.62		0.65		0.64		0.74		0.69		0.64		0.74		0.68		0.74		0.75		0.74	
MZF	0.67		0.61		0.66		0.60		0.70		0.64		0.63		0.73		0.65		0.70		0.71		0.71	
DZF	0.64		0.62		0.66		0.62		0.70		0.66		0.64		0.73		0.63		0.69		0.68		0.71	
Twin correlations for mother and father ratings																								
	Mo	Fa	Mo	Fa	Mo	Fa	Mo	Fa	Mo	Fa	Mo	Fa	Mo	Fa	Mo	Fa	Mo	Fa	Mo	Fa	Mo	Fa	Mo	Fa
MZM	0.71	0.74	0.72	0.74	0.69	0.71	0.90	0.91	0.88	0.90	0.80	0.82	0.78	0.83	0.81	0.81	0.74	0.76	0.89	0.89	0.87	0.88	0.90	0.90
DZM	0.42	0.43	0.34	0.31	0.44	0.43	0.68	0.69	0.60	0.62	0.52	0.54	0.48	0.48	0.29	0.32	0.51	0.48	0.63	0.64	0.62	0.64	0.69	0.71
MZF	0.74	0.76	0.74	0.76	0.72	0.70	0.88	0.89	0.86	0.88	0.78	0.83	0.80	0.81	0.77	0.81	0.77	0.77	0.89	0.88	0.87	0.88	0.89	0.90
DZF	0.46	0.50	0.34	0.46	0.45	0.41	0.70	0.73	0.58	0.63	0.50	0.54	0.53	0.52	0.24	0.26	0.55	0.57	0.62	0.70	0.62	0.66	0.71	0.77
DOS	0.49	0.47	0.41	0.44	0.46	0.43	0.68	0.67	0.57	0.59	0.48	0.54	0.46	0.49	0.30	0.36	0.57	0.55	0.62	0.63	0.58	0.63	0.68	0.70
Mother – Father cross-correlations																								
MZM	0.45		0.44		0.46		0.60		0.67		0.54		0.51		0.61		0.49		0.67		0.66		0.67	
DZM	0.22		0.12		0.23		0.42		0.44		0.34		0.29		0.17		0.30		0.46		0.45		0.50	
MZF	0.50		0.46		0.45		0.52		0.61		0.51		0.51		0.57		0.51		0.62		0.63		0.64	
DZF	0.26		0.20		0.22		0.43		0.39		0.33		0.34		0.11		0.32		0.44		0.40		0.52	
DOS	0.28		0.21		0.26		0.40		0.38		0.31		0.29		0.20		0.33		0.41		0.40		0.47	

* In DOS twin pairs scores were organized in such a way, that boys are first and girls are second. Therefore in 4x4 matrix parental agreement for twin 1 is a correlation for a boy and parental agreement for twin 2 is a correlation for a girl and DZM and DZF correlations reflects DOS correlations.

Table 4. Correlations estimated from the most parsimonious model

	Zygosity	Twin correlation		Mother – Father cross-correlations	Parental Agreement	
		Mo	Fa		Phenotypic correlation	Genetic correlation
Anxious/ Depressed	MZ	0.73	0.75	0.48	0.66	0.87
	DZ	0.46	0.47	0.26		
Withdrawn/ Depressed	MZ	0.73	0.75	0.46	0.62	0.72
	DZ	0.37	0.41	0.18		
Somatic Complaints	MZ	0.70	0.71	0.45	0.66	0.78
	DZ	0.45	0.42	0.24		
Rule Breaking Behavior	MZ	0.89	0.90	0.56	0.63	0.74
	DZ	0.69	0.69	0.41		
Aggressive Behavior	MZ	0.87	0.89	0.64	0.72	0.86
	DZ	0.58	0.61	0.40		
Social Problems	MZ	0.79	0.83	0.54	0.67	0.77
	DZ	0.50	0.54	0.32		
Thought Problems	MZ	0.79	0.82	0.52	0.64	0.68
	DZ	0.48	0.49	0.30		
Attention Problems	MZ	0.79	0.81	0.59	0.74	0.42 / 1.0 / 0.74*
	DZ	0.29	0.32	0.17		
INT	MZ	0.75	0.77	0.51	0.65	0.89
	DZ	0.55	0.53	0.32		
EXT	MZ	0.89	0.89	0.64	0.72	0.82
	DZ	0.62	0.65	0.43		
Dysregulation Profile	MZ	0.87	0.88	0.64	0.72	0.89
	DZ	0.60	0.64	0.42		
Total Problems	MZ	0.89	0.90	0.66	0.73	0.90
	DZ	0.69	0.72	0.49		

* For Attention Problems both the additive genetic correlation, the correlation between genetic dominance factors and the total genetic correlation (the correlation between the summed additive and dominant genetic effects) are given.

Genetic Psychometric model

For all scales, the common component of the phenotype assessed by both parents was substantial, with parental ratings correlations varying between 0.62 and 0.74 (Figure 1, Table 4), however, a contribution of specific aspects of the child's phenotype was present as well. For all scales, a substantial amount of the total variance ranging from 34% to 65% (Table 5) was accounted for by additive genetic variation, of which 15% to 48% was shared between parents (agreement) and 0% to 22% was unique to each parent. For all scales, except Attention Problems, between 7% to 56% of total variation was accounted for by common environmental factors. The proportion of such factors that contribute to variation in parental agreement, i.e. free of rater bias and specific parental view, ranged from 0% to 32%. The proportion that is unique to each parent's perspective ranged from 7% to 24%. The contribution of dominant genetic effects to the total variation in the Attention Problems was 44% and was reflected in the part of the phenotype that parents agreed upon. Genetic correlations, computed based on additive genetic components of the most parsimonious psychometric model, ranged from 0.42 to 0.90 (Figure 1, Table 4). The additive genetic correlation of 0.42 is an outlier, which was observed for Attention Problems. This is the only scale for which dominant effects were found and the dominance genetic correlation was one. The total genetic correlation (the correlation between the summed additive and dominant genetic effects) was 0.74.

Figure 1: Genetic and phenotypic correlations between maternal and paternal ratings across all CBCL 6-18 scales. For all scales, except Attention Problems, the genetic correlations between additive genetic factors are depicted. For Attention Problems scale the *total* genetic correlation (between summed additive and dominant effects) is shown.

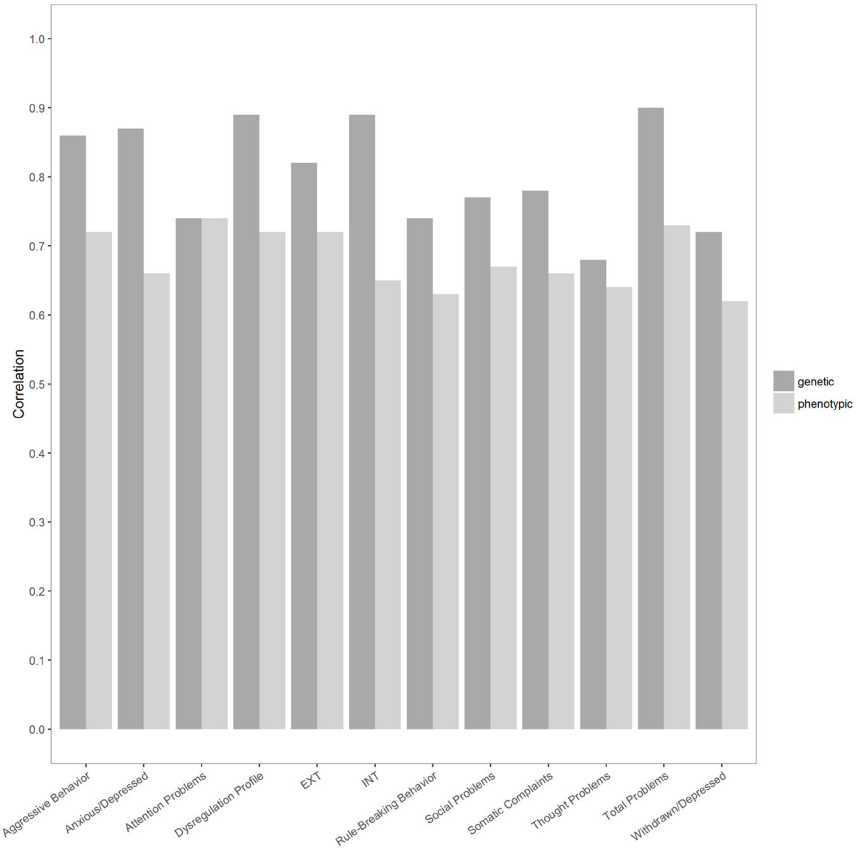


Table 5. Heritability (A), Shared (C) and Unique (E) Environmental effects, estimated from most parsimonious psychometric model for each of the empirical scale of CBCL 6-18

		Phenotype, parents agree upon, %	Unique mother's assessment of the phenotype, %	Unique father's assessment of the phenotype, %	Total mother's assessment, %	Total father's assessment, %
Anxious/ Depressed	A	48 (46 - 50)	-	16 (11 - 21)	48	64
	C	-	23 (21 -25)	12 (10 - 16)	23	12
	E	17 (15 - 19)	12 (10 - 13)	7 (5 - 9)	29	24
Withdrawn/ Depressed	A	45 (43 - 47)	20 (15 - 26)	15 (12 - 21)	65	60
	C	-	7 (3 - 13)	15 (10 - 20)	7	15
	E	17 (16 - 19)	11 (9 - 12)	8 (7 - 10)	28	25
Somatic Complaints	A	45 (42 - 47)	9 (3 - 14)	16 (12 - 22)	54	61
	C	-	17 (13 - 21)	10 (6- 16)	17	10
	E	21 (19 - 23)	8 (7 - 10)	8 (6 - 10)	29	29
Rule Breaking Behavior	A	30 (27 - 34)	11 (10 - 14)	10 (7 - 14)	41	40
	C	26 (22 - 29)	22 (19 - 25)	23 (20 - 27)	48	49
	E	7 (6 - 8)	4 (3 - 5)	4 (3 - 5)	11	11
Aggressive Behavior	A	48 (44 - 52)	9 (6 - 13)	7 (5 - 11)	57	55
	C	16 (12 - 20)	14 (10 - 17)	17 (14 - 21)	30	33
	E	8 (7 - 9)	5 (4 - 6)	3 (2 - 4)	13	11
Social Problems	A	45 (40 - 51)	14 (9 - 20)	13 (7 - 18)	59	58
	C	9 (5 - 14)	11 (6 - 15)	16 (11 - 20)	20	25
	E	13 (11 - 14)	8 (6 - 10)	5 (3 - 6)	21	18

Table 5 Continued

		Phenotype, parents agree upon, %	Unique mother's assessment of the phenotype, %	Unique father's assessment of the phenotype, %	Total mother's assessment, %	Total father's assessment, %
Thought Problems	A	43 (38 - 48)	18 (17 - 24)	22 (17 - 25)	61	65
	C	8 (4 - 12)	9 (7 - 14)	8 (4 - 13)	17	16
	E	12 (11 - 13)	9 (7 - 11)	6 (5 - 8)	21	18
Attention Problems	A	15 (9 - 24)	20 (19 - 22)	22 (21 - 24)	35	37
	D	44 (36 - 50)	-	-	44	44
	E	15 (14 - 17)	6 (4 - 7)	4 (2 - 5)	21	19
INT	A	39 (34 - 43)	-	10 (5 - 15)	39	49
	C	12 (8 - 16)	24 (22 - 26)	16 (12 - 21)	36	28
	E	15 (13 - 16)	11 (9 - 12)	8 (7 - 10)	26	23
EXT	A	41 (38 - 45)	12 (8 - 15)	6 (4 - 10)	53	47
	C	22 (19 - 26)	13 (10 - 16)	19 (15 - 22)	35	41
	E	8 (7 - 9)	3 (2 - 4)	3 (3 - 4)	11	11
Dysregulation Profile	A	45 (41 - 49)	12 (9 - 16)	-	57	45
	C	19 (16 - 23)	11 (8 - 15)	23 (22 - 25)	30	42
	E	8 (7 - 9)	4 (3 - 6)	5 (4 - 6)	12	13
Total Problems	A	34 (31 - 38)	8 (5 - 11)	-	42	34
	C	32 (28 - 35)	16 (13 - 19)	24 (22 - 25)	48	56
	E	7 (6 - 8)	3 (2 - 4)	3 (3 - 4)	10	10

Discussion

In this study we employed a psychometric model to determine to what extend parental assessments of a child's behavioral problems around age 7 reflect common and parent specific aspects of the child behavior or if parents disagree due to rater bias. We observe interparental phenotypic correlations between 0.62 and 0.74, reflecting substantial but incomplete agreement between parents. Incomplete agreement may result in different heritability estimates between a single phenotype as assessed by different informants. Different informants provide information about child's behavior and it is important to identify, prior to large GWAS efforts, whether the additive genetic effects on a trait strictly are found in the phenotypic variation which correlates between raters. Our analyses showed these were fairly highly correlated and that genetic correlations ranged from moderate to high (Figure 1, Table 4), that is from 0.68 to 0.90 for all problem scales, with the exception of Attention problems. The Attention Problem scale, as observed in numerous studies, has a different genetic architecture with non-additive genetic influences explaining a substantial part of the heritability.

Comparison to previous results

In the NTR exploration of parental rater bias effect were conducted earlier for Anxious/Depressed [148], Attention Problems [149], Withdrawn behavior [150], Aggression [151], Thought Problems [142], Internalizing and Externalizing domains [152-154]. The larger collection of NTR data in the current paper allowed for analysis of categorical data under a threshold model. Several new scales were analyzed for the first time using multiple rater assessments at age 7, such as Somatic Complaints, Rule-Breaking Behavior, Social Problems, Dysregulation Profile and Total Problems Score. In addition, the earlier papers had focus on separate scales and domains, whereas all CBCL scales were explored simultaneously in the current study, allowing for comparison between scales. Our results showed that heritability estimates of Internalizing, Externalizing, Dysregulation Profile and Total Problems score in comparison to subscales comprising them, varies. The estimates of the unique aspect of the child behavior rated by mother are more variable across the Internalizing scale subscales, than they are for father. This trend is not reflected in the Internalizing scale, where all three phenotypes are combined. In addition, the absence of genetic effects estimated for unique aspect of maternal rating of the child's behavior is likely driven by the Anxious/Depressed scale and not by others. In contrast, for the Externalizing scale estimates of the contribution of genetic and environmental components to the variation of the phenotype shared by both parents were more variable across subscales. Only for the Dysregulation Profile and Total problems scales a specific paternal contribution was accounted for by rater bias reflected by a significant Vc,unique component. A possible explanation is the heterogeneity of these measures in comparison to homogeneous single scales. We did not observe any sex differences in genetic architecture or in parental agreement for behavior rated for girls and boys, except for Aggressive Behavior, Externalizing and the

Dysregulation Profile, but observed the well-known differences between boys and girls for mean scores. Also, we observed that mothers rated the behavioral and emotional problems in their offspring higher than fathers.

Results obtained in our study are in line with earlier studies of CBCL 6-18 scales in twins aged 7. Both studies of single or multiple raters reported genetic influence on variability in behavioral and emotional problems [155-159]. In Brendgen et al. (2005) [156] peers' and teachers' assessments were used to study genetic influences on social and physical aggression in 6 year olds, and heritability estimates were similar in magnitude between raters. The phenotypic correlation between teachers and mother ratings of aggressive behavior was moderate (r = 0.20) in the study of Haberstick et al. (2006) [157] and heritability estimates differed in magnitude between raters for children at age 7 and the authors suggested that parents and teacher provide unique information that can be specific to the settings. In Eley et al. (1999) [158] sex-differences in aggressive antisocial behavior were reported for boys and girls, which were also detected in our study. To our knowledge there is limited research on Somatic Complaints, Rule-Breaking Behavior, Social Problems scales and the Dysregulation Profile of CBCL 6-18 at this specific age. For the latter, the agreement and disagreement between raters were reported in an American non-twin cohort [138]. Report based on an Italian sample of twins (N = 398 pairs), rated by mothers, in age range from 8 to 17 years showed no additive genetic effect, but 54% of common and 46% of unique environment effects for Social Problems [159]. Because heritability might change as a function of age [26, 160] previous reports on younger and older twins are not directly comparable to the current study. These findings have implications for molecular genetic studies.

Implications of our findings for molecular genetic studies

In molecular genetic studies, the distinction between rater bias and rater specific assessment of child's behavior may have implications for the estimation of the SNP-heritability of behavioral and emotional problems. GWAS and GREML analyses will benefit from the determination to what extend two different sources of disagreement contribute to the phenotypic variance and affect the covariance. For example, differences in mother and father ratings suggest using rater as a covariate, if raters information is combined. In the recent study of Pappa et al. (2015) [161] SNP-heritability of a range of children's behavior problems were estimated. Attention Deficit Hyperactivity Disorder (ADHD) related scales and Externalizing behavior were assessed by both mother and teacher. Estimates of SNP-heritability of Attention Problems for teacher's ratings and of ADHD Combined scale for Conner's Parent Rating scale were 0.71 (s.e. = 0.22, n = 1,495, p-value < 0.001) and 0.40 (s.e. = 0.14, n = 2,262, p < 0.01) respectively. For Externalizing behavior scale estimates of SNP-heritability were 0.44 (s.e. = 0.22, n= 1,495, p < 0.05) for teacher's ratings and 0.12 (s.e. = 0.10, n = 3,174, p = 0.13) for maternal ratings. The differences in SNP-heritability estimates are consistent with the findings obtained from twin studies, which account for rater specific effects. As was suggested in a study of

Attention Problems by Derks et al. (2006) [132] both mother and teacher provide valid, but specific information about a child's behavior in addition to a commonly assessed part. Therefore, variation explained by SNPs in teachers and mothers ratings may be represented by different, partly overlapping, genetic loci.

Our investigation of rater common and rater specific contributions to phenotypic variation serves as an indication of whether or not combined analyses of different informant ratings are likely to be fruitful. The substantial genetic correlations between different raters as evident from our results, suggest two practical guidelines: when studies have collected data from either fathers or mothers, the shared genetic aetiology in parental ratings indicates that is possible to analyze paternal and maternal assessments in a single GWA study or meta-analysis. Secondly, if a study has collected information from both parents, a gain in statistical power should be realized in a GWA study by simultaneous analysis of the data.

The power of various ways of modeling bivariate phenotype information, including analyses based on sum and factor scores, exploratory factor analysis (EFA), MANOVA, and combined multivariate analyses (CMV) were explored by Van Der Sluis et al. (2010), Medland and Neale (2010) and Minica et al. (2010) [162-164]. Each of these approaches was evaluated in terms of power to discover genetic loci. Based on results of these studies, if the genetic correlation between different raters is very high, implying that genetic loci, that influence parental ratings, overlap almost completely, combining the ratings in a single trait, using sum score is, is justifiable [164]. If the correlations are moderate to high, one might prefer a technique that has high power when loci are expected to influence the shared, as well as unique part of the phenotype as assessed by the different raters [163]. Finally, if the genetic correlations between raters is low to moderate, one might prefer to perform separate analysis in either rater and combine the resulting p-values by using Trait-based Association Test that uses Extended Simes procedure (TATES) [165].

In current study we considered parental ratings and did not make an attempt to analyze rater effects based on teachers ratings or on self-assessments of children. Inclusion of other raters will convey additional information about possible combined or separate analysis of problem behaviors assessed by multiple raters.

Based on the results reported in this paper, we conclude that aggregating multiple raters' in genetic studies of childhood psychopathology potentially will improve power. At age 7, our study showed that heritability of phenotypes reflecting a shared perspective on the child's problem behavior is substantially higher than that of unique view. These results suggest a model in which genome wide analysis of different raters are combined into a single trait, accounting for genetic correlation, and differences in heritability, could prove optimal. For traits with a (somewhat) lower genetic correlation or if including further raters, for which substantial rater specific genetic effects are present (e.g. self ratings, teacher ratings, clinician ratings), a multitude of multivariate genetic analysis tools exist.

Supplementary Material

Table 1. Descriptive statistics of raw CBCL 6-18 scales.

	Mother				Father			
	N	**Mean (SD)**	**Skewness (SE)**	**Kurtosis (SE)**	**N**	**Mean (SD)**	**Skewness (SE)**	**Kurtosis (SE)**
Anxious/Depressed	24031	2.25 (2.58)	1.91 (0.02)	5.07 (0.03)	17042	1.73 (2.15)	2.10 (0.02)	6.53 (0.04)
Withdrawn/ Depressed	24001	1.14 (1.60)	2.26 (0.02)	6.73 (0.03)	16988	0.92 (1.44)	2.48 (0.02)	8.11 (0.04)
Somatic Complaints	23794	1.20 (1.65)	2.23 (0.02)	8.11 (0.03)	16931	0.87 (1.34)	2.48 (0.02)	10.10 (0.04)
Rule-Breaking Behavior	24028	1.33 (1.83)	2.06 (0.02)	5.88 (0.03)	16999	1.16 (1.69)	2.30 (0.02)	8.27 (0.04)
Aggressive Behavior	24006	5.10 (4.89)	1.45 (0.02)	2.60 (0.03)	16995	4.40 (4.40)	1.45 (0.02)	2.55 (0.04)
Social Problems	24017	2.08 (2.39)	1.88 (0.02)	4.78 (0.03)	17028	1.72 (2.11)	2.03 (0.02)	5.83 (0.04)
Thought Problems	23898	1.49 (1.98)	2.28 (0.02)	7.52 (0.03)	16938	1.11 (1.66)	2.66 (0.02)	11.16 (0.04)
Attention Problems	24041	3.04 (3.03)	1.25 (0.02)	1.69 (0.03)	17028	2.69 (2.81)	1.30 (0.02)	1.93 (0.04)
Internalizing	23599	4.58 (4.61)	1.87 (0.02)	5.05 (0.03)	16804	3.51 (3.84)	2.12 (0.02)	7.11 (0.04)
Externalizing	23962	6.43 (6.26)	1.58 (0.02)	3.30 (0.03)	16950	5.56 (5.64)	1.61 (0.02)	3.46 (0.04)
Dysregularion Profile	23841	10.39 (8.59)	1.38 (0.02)	2.60 (0.03)	16908	8.81 (7.68)	1.43 (0.02)	2.69 (0.04)
Total Problems	23936	20.92 (16.45)	1.51 (0.02)	3.30 (0.03)	16957	17.26 (14.50)	1.67 (0.02)	4.50 (0.04)

Table 2. Relationship between the categories and raw CBCL 6-18 scores. The raw maternal and paternal scores for each scale (rows) were categorized into three categories. Ranges of raw scores in each category are shown in 'min' and 'max' columns.

	Mother						Father					
	Category 1		Category 2		Category 3		Category 1		Category 2		Category 3	
	min	max	min	max	min	max	min	max	min	max	min	max
Anxious/Depressed	0	1	2	2	3	25	0	1	2	2	3	21
Withdrawn/ Depressed	0	0	1	1	2	15	0	0	1	1	2	13
Somatic Complaints	0	0	1	1	2	22	0	0	1	1	2	18
Rule-Breaking Behavior	0	0	1	1	2	17	0	0	1	1	2	20
Aggressive Behavior	0	2	3	6	7	35	0	2	3	6	7	30
Social Problems	0	1	2	2	3	20	0	1	2	2	3	19
Thought Problems	0	0	1	1	2	19	0	0	1	1	2	18
Attention Problems	0	1	2	4	5	20	0	1	2	4	5	19
Internalizing	0	2	3	5	6	45	0	2	3	5	6	40
Externalizing	0	2	3	7	8	48	0	2	3	7	8	43
Dysregularion Profile	0	5	6	12	13	70	0	5	6	12	13	63
Total Problems	0	11	12	22	23	143	0	11	12	22	23	142

Table 3. Models fitting to CBCL 6-18 empirical scales: 4x4 zygosity by sex correlations matrix and 8 thresholds with submodels, psychometric model with submodels.

CBCL 6-18 scale		Model	Estimated parameters	-2 LL	df	Compared to model	Δ LL	Δ df	p
Anxious/Depressed	1	Saturated model	27	72112.62	41046	-	-	-	-
	2	*Equal thresholds boys and girls*	23	72150.41	41050	1	37.79	4	1.2×10^{-07}
	3	*Parental agreement across zygosity*	25	72114.75	41048	1	2.12	2	0.35
	4	*Parental agreement across sex*	24	72115.01	41049	3	0.26	1	0.61
	5	*MZM=MZF & DZM = DZF*	18	72121.16	41055	4	6.15	6	0.41
	6	*DZ=DOS*	15	72128.16	41058	5	7.00	3	0.07
	7	Psychometric model	17	72128.16	41058	-	-	-	-
	8	*no C*	16	72129.44	41059	7	1.28	1	0.26
	9	*no Am*	15	72136.16	41060	8	6.72	1	0.01
	10	*no Af*	15	72149.50	41060	8	20.06	1	7.5×10^{-06}
	11	*no Am, no Af*	14	72164.78	41061	8	35.34	2	2.1×10^{-08}
Withdrawn/Depressed	1	Saturated model	27	77275.25	40962	-	-	-	-
	2	*Equal thresholds boys and girls*	23	77289.28	40966	1	14.03	4	0.01
	3	*Parental agreement across zygosity*	25	77275.45	40964	1	0.20	2	0.91
	4	*Parental agreement across sex*	24	77275.49	40965	3	0.05	1	0.83
	5	*MZM=MZF & DZM = DZF*	18	77291.10	40971	4	15.61	6	0.02

Table 3 Continued.

CBCL 6-18 scale		Model	Estimated parameters	-2 LL	df	Compared to model	Δ LL	Δ df	p
	6	*DZ=DOS*	15	77300.52	40974	5	9.42	3	0.02
	7	Psychometric model	17	77316.88	40974	-	-	-	-
	8	*no C*	16	77316.88	40975	7	0.00	1	1.00
	9	*no Am*	15	77367.63	40976	8	50.75	1	1.0×10^{-12}
	10	*no Af*	15	77342.44	40976	8	25.56	1	4.3×10^{-07}
	11	*no Am, no Af*	14	77414.57	40977	8	97.69	2	6.1×10^{-22}
Somatic Complaints	1	Saturated model	27	75333.91	40698	-	-	-	-
	2	*Equal thresholds boys and girls*	23	75388.41	40702	1	54.50	4	4.1×10^{-11}
	3	*Parental agreement across zygosity*	25	75336.10	40700	1	2.19	2	0.34
	4	*Parental agreement across sex*	24	75336.19	40701	3	0.09	1	0.77
	5	*MZM=MZF & DZM = DZF*	18	75340.12	40707	4	3.93	6	0.69
	6	*DZ=DOS*	15	75341.72	40710	5	1.60	3	0.66
	7	Psychometric model	17	75341.56	40710	-	-	-	-
	8	*no C*	16	75344.69	40711	7	3.13	1	0.08
	9	*no Am*	15	75355.51	40712	8	10.82	1	1.0×10^{-03}
	10	*no Af*	15	75372.68	40712	8	27.99	1	1.2×10^{-07}
	11	*no Am, no Af*	14	75398.35	40713	8	53.66	2	2.2×10^{-12}

Table 3 Continued.

CBCL 6-18 scale		Model	Estimated parameters	-2 LL	df	Compared to model	Δ LL	Δ df	p
Rule Breaking Behavior	1	Saturated model	27	72083.57	41000	-	-	-	-
	2	*Equal thresholds boys and girls*	23	72442.12	41004	1	358.55	4	2.5×10^{-76}
	3	*Parental agreement across zygosity*	25	72084.30	41002	1	0.74	2	0.69
	4	*Parental agreement across sex*	24	72090.07	41003	3	5.77	1	0.02
	5	*MZM=MZF & DZM = DZF*	18	72099.72	41009	4	9.64	6	0.14
	6	*DZ=DOS*	15	72104.29	41012	5	4.57	3	0.21
	7	Psychometric model	17	72104.29	41012	-	-	-	-
	8	*no C*	16	72289.73	41013	7	185.44	1	3.1×10^{-42}
	9	*no Am*	16	72142.36	41013	7	38.07	1	6.8×10^{-10}
	10	*no Af*	16	72135.86	41013	7	31.57	1	1.9×10^{-08}
	11	*no Am, no Af*	15	72185.32	41014	7	81.03	2	2.5×10^{-18}
Aggressive Behavior	1	Saturated model	27	73373.97	40974	-	-	-	-
	2	*Equal thresholds boys and girls*	23	73754.93	40978	1	380.96	4	3.6×10^{-81}
	3	*Parental agreement across zygosity*	25	73374.20	40976	1	0.22	2	0.89
	4	*Parental agreement across sex*	24	73390.07	40977	3	15.88	1	6.8×10^{-05}
	5	*MZM=MZF & DZM = DZF*	18	73395.31	40983	4	5.24	6	0.51
	6	*DZ=DOS*	15	73400.25	40986	5	4.94	3	0.18

Table 3 Continued.

CBCL 6-18 scale		Model	Estimated parameters	-2 LL	df	Compared to model	Δ LL	Δ df	p
	7	Psychometric model	17	73400.25	40986	-	-	-	-
	8	*no C*	16	73463.21	40987	7	62.96	1	2.1×10^{-15}
	9	*no Am*	16	73424.82	40987	7	24.57	1	7.2×10^{-07}
	10	*no Af*	16	73416.47	40987	7	16.22	1	5.6×10^{-05}
	11	*no Am, no Af*	15	73454.24	40988	7	53.99	2	1.9×10^{-12}
Social Problems	1	Saturated model	27	70539.94	41018	-	-	-	-
	2	*Equal thresholds boys and girls*	23	70605.89	41022	1	65.95	4	1.6×10^{-13}
	3	*Parental agreement across zygosity*	25	70541.34	41020	1	1.40	2	0.50
	4	*Parental agreement across sex*	24	70546.81	41021	3	5.46	1	0.02
	5	*MZM=MZF & DZM = DZF*	18	70548.85	41027	4	2.05	6	0.92
	6	*DZ=DOS*	15	70551.79	41030	5	2.94	3	0.40
	7	Psychometric model	17	70551.79	41030	-	-	-	-
	8	*no C*	16	70565.87	41031	7	14.07	1	1.8×10^{-04}
	9	*no Am*	16	70580.23	41031	7	28.43	1	9.7×10^{-08}
	10	*no Af*	16	70574.55	41031	7	22.76	1	1.8×10^{-06}
	11	*no Am, no Af*	15	70615.59	41032	7	63.79	2	1.4×10^{-14}
Thought Problems	1	Saturated model	27	76500.53	40809	-	-	-	-
	2	*Equal thresholds boys and girls*	23	76792.95	40813	1	292.42	4	4.7×10^{-62}

Table 3 Continued.

CBCL 6-18 scale		Model	Estimated parameters	-2 LL	df	Compared to model	Δ LL	Δ df	p
	3	*Parental agreement across zygosity*	25	76501.05	40811	1	0.52	2	0.77
	4	*Parental agreement across sex*	24	76501.24	40812	3	0.18	1	0.67
	5	*MZM=MZF & DZM = DZF*	18	76508.16	40818	4	6.93	6	0.33
	6	*DZ=DOS*	15	76512.26	40821	5	4.09	3	0.25
	7	Psychometric model	17	76512.26	40821	-	-	-	-
	8	*no C*	16	76525.54	40822	7	13.28	1	2.7×10^{-04}
	9	*no Am*	16	76559.62	40822	7	47.37	1	5.9×10^{-12}
	10	*no Af*	16	76578.25	40822	7	66.00	1	4.5×10^{-16}
	11	*no Am, no Af*	15	76640.10	40823	7	127.84	2	1.7×10^{-28}
Attention Problems	1	Saturated model	27	74865.33	41042	-	-	-	-
	2	*Equal thresholds boys and girls*	23	75444.63	41046	1	579.30	4	4.7×10^{-124}
	3	*Parental agreement across zygosity*	25	74865.62	41044	1	0.29	2	0.87
	4	*Parental agreement across sex*	24	74866.87	41045	3	1.25	1	0.26
	5	*MZM=MZF & DZM = DZF*	18	74878.10	41051	4	11.23	6	0.08
	6	*DZ=DOS*	15	74882.93	41054	5	4.83	3	0.18
	7	Psychometric model	17	74910.91	41054	-	-	-	-
	8	*no Dm*	16	74910.90	41055	7	-0.01	1	1.00
	9	*no Df*	15	74910.90	41056	8	0.00	1	1.00

Table 3 Continued.

CBCL 6-18 scale		Model	Estimated parameters	-2 LL	df	Compared to model	Δ LL	Δ df	p
	10	*no Am*	14	75471.23	41057	9	560.33	1	7.1×10^{-124}
	11	*no Af*	14	75545.44	41057	9	634.54	1	5.1×10^{-140}
	12	*no Am, Af*	13	76191.07	41058	9	1280.17	2	1.0×10^{-278}
Internalizing	1	Saturated model	27	75160.06	40376	-	-	-	-
	2	*Equal thresholds boys and girls*	23	75187.11	40380	1	27.05	4	1.9×10^{-05}
	3	*Parental agreement across zygosity*	25	75162.65	40378	1	2.59	2	0.27
	4	*Parental agreement across sex*	24	75169.23	40379	3	6.58	1	0.01
	5	*MZM=MZF & DZM = DZF*	18	75181.43	40385	4	12.20	6	0.06
	6	*DZ=DOS*	15	75186.48	40388	5	5.05	3	0.17
	7	Psychometric model	17	75187.24	40388	-	-	-	-
	8	*no C*	16	75221.28	40389	7	34.04	1	5.4×10^{-09}
	9	*no Am*	16	75189.21	40389	7	1.97	1	0.16
	10	*no Af*	16	75198.34	40389	7	11.10	1	8.6×10^{-04}
	11	*no Am, no Af*	15	75202.25	40390	7	15.00	2	5.5×10^{-04}
Externalizing	1	Saturated model	27	72875.88	40885	-	-	-	-
	2	*Equal thresholds boys and girls*	23	73296.23	40889	1	420.34	4	1.1×10^{-89}
	3	*Parental agreement across zygosity*	25	72876.47	40887	1	0.58	2	0.75
	4	*Parental agreement across sex*	24	72892.77	40888	3	16.30	1	5.4×10^{-05}

Table 3 Continued.

CBCL 6-18 scale		Model	Estimated parameters	-2 LL	df	Compared to model	Δ LL	Δ df	p
	5	*MZM=MZF & DZM = DZF*	18	72899.02	40894	4	6.25	6	0.40
	6	*DZ=DOS*	15	72904.88	40897	5	5.86	3	0.12
	7	Psychometric model	17	72905.00	40897	-	-	-	-
	8	*no C*	16	73038.14	40898	7	133.14	1	8.4×10^{-31}
	9	*no Am*	16	72951.51	40898	7	46.51	1	9.1×10^{-12}
	10	*no Af*	16	72917.44	40898	7	12.44	1	4.2×10^{-04}
	11	*no Am, no Af*	15	72977.68	40899	7	72.68	2	1.6×10^{-16}
Dysregulation Profile	1	Saturated model	27	72916.78	40722	-	-	-	-
	2	*Equal thresholds boys and girls*	23	73345.75	40726	1	428.97	4	1.5×10^{-91}
	3	*Parental agreement across zygosity*	25	72919.22	40724	1	2.44	2	0.30
	4	*Parental agreement across sex*	24	72945.21	40725	3	25.98	1	3.4×10^{-07}
	5	*MZM=MZF & DZM = DZF*	18	72948.13	40731	4	2.93	6	0.82
	6	*DZ=DOS*	15	72952.54	40734	5	4.41	3	0.22
	7	Psychometric model	17	72952.54	40734	-	-	-	-
	8	*no C*	16	73051.38	40735	7	98.84	1	2.7×10^{-23}
	9	*no Am*	16	72987.24	40735	7	34.70	1	3.8×10^{-09}
	10	*no Af*	16	72958.13	40735	7	5.59	1	0.02
	11	*no Am, no Af*	15	72999.75	40736	7	47.21	2	5.6×10^{-11}

Table 3 Continued.

CBCL 6-18 scale		Model	Estimated parameters	-2 LL	df	Compared to model	Δ LL	Δ df	p
Total problems score	1	Saturated model	27	71530.16	40866	-	-	-	-
	2	*Equal thresholds boys and girls*	23	71929.17	40870	1	399.01	4	4.6×10^{-85}
	3	*Parental agreement across zygosity*	25	71530.17	40868	1	0.00	2	1.00
	4	*Parental agreement across sex*	24	71537.02	40869	3	6.85	1	0.01
	5	*MZM=MZF & DZM = DZF*	18	71545.51	40875	4	8.49	6	0.20
	6	*DZ=DOS*	15	71551.16	40878	5	5.65	3	0.13
	7	Psychometric model	17	71551.16	40878	-	-	-	-
	8	*no C*	16	71851.20	40879	7	300.04	1	3.2×10^{-67}
	9	*no Am*	16	71573.65	40879	7	22.50	1	2.1×10^{-06}
	10	no Af	16	71553.24	40879	7	2.08	1	0.15
	11	*no Am, no Af*	15	71578.84	40880	7	27.68	2	9.8×10^{-07}

df = degrees of freedom, -2LL = - 2 log likelihood, C = shared environmental variance of the common part of total variance, Am = additive genetic variance of mother specific component of total variance, Af = additive genetic variance of father specific component of total variance, Dm = dominant genetic variance of mother specific component of total variance, Df = dominant genetic variance of father specific component of total variance, MZM = monozygotic twins males, MZF = monozygotic twins females, DZM = dizygotic twins males, DZF = dizygotic twins females, DOS – opposite sex twins.

CHAPTER 5

A GENETIC LIABILITY TO HIGHER SUBJECTIVE WELL-BEING IS MORE INDICATIVE OF LOWER LEVELS OF NEUROTICISM THAN HIGHER LEVELS OF EXTRAVERSION

This chapter is based on:

Iryna O. Fedko, Jouke-Jan Hottenga, Erik A. Ehli, Gareth E. Davies, Dorret I. Boomsma, and Meike Bartels. A genetic liability to higher Subjective Well-being is more indicative of lower levels of Neuroticism than higher levels of Extraversion *(as to be submitted).*

Abstract

There is ample evidence regarding the significant role of genes in explaining individual differences in Subjective Well-being (SWB), Neuroticism (NEU) and Extraversion (EXT). Twin data suggests that genetic influences on personality traits and SWB are correlated. The genetic correlation for SWB and NEU based on Single Nucleotide Polymorphisms (SNPs) has been found to be high, but has not been studied for SWB and EXT. In the current study, we applied a bivariate genetic model to estimate the SNP-heritability for the SWB and personality traits, and compared the genetic correlations between them. We used both the information from distantly and closely related individuals, thereby estimating total trait heritability, heritability explained by SNPs, the total genetic correlation and the part of the genetic correlation that can be attributed to SNPs. We found that 7%, 10% and 16% of the variance in SWB, NEU and EXT is accounted for by SNPs present on current genotyping platforms. The magnitude of the SNP-based genetic correlation between SWB and NEU was in line with previous work (r_g = -.80, SE = .25), but higher than the SNP-based genetic correlation between SWB and EXT (r_g = .18, SE = .26), while the phenotypic correlations are largely comparable. This indicates that a higher genetic liability to SWB is related to a lower NEU levels and not to higher EXT levels. Also, environmental influences explain a larger part of the phenotypic correlation between SWB and EXT than between SWB and NEU.

Introduction

There has been a great deal of evidence that shows the significant role of genes explaining individual differences in Subjective Well-being (SWB), Neuroticism (NEU) and Extraversion (EXT). Earlier studies reported personality being the important predictor of the SWB [32, 33]. Twin-family studies revealed that approximately 32% [166], 27%, and 24% [167] of the variances in SWB, NEU and EXT are accounted for by additive genetic variance, respectively. The recently reported SNP-based heritability, that is the variance in the phenotype accounted for by Single Nucleotide Polymorphisms (SNPs), was 4-10% for SWB [168, 169], 6-15% for NEU [168, 170-173] and 0-12% for EXT [170, 173, 174].

Personality traits, such as Extraversion and Neuroticism, were found to be related to SWB. Observed correlations were estimated to be in a range from -.14 to -.25 between SWB and NEU and from .17 to .27 between SWB and EXT [32]. Twin data suggested the genetic correlation as .58 between SWB and NEU and .66 between SWB and EXT [33]. A study in a German twin and non-twin sample [175] detected larger association between SWB and NEU than SWB and EXT. A recent large-scale collaborative effort reported SNP-based genetic correlation between SWB and NEU to be -.75 (SE = 0.034) [168]. SNP-based genetic correlation between SWB and EXT has not been explored yet.

Here we applied a powerfull bivariate model, in which the variance of two traits and the covariance between them is modeled using bivariate restricted maximum likelihood (REML) method of Genome-wide Complex Trait Analysis (GCTA) tool [20]. Besides estimates of SNP heritability of the two traits (e.g. SWB and EXT) this bivariate model also provides the genetic correlation, which is an estimate of the additive genetic component that is common to both traits. We applied this method to SWB, NEU, and EXT within the Netherlands Twin Register, a general Dutch population based sample, to investigate the shared genetic aetiology.

Methods

Participants

For this study we selected ≈ 9,000 participants, which were registered at the Netherlands Twin Register (NTR), established by the Department of Biological Psychology at the VU University in Amsterdam. Around 60% of the sample was female and the average age was 38 years (SD = 16). Seventy five percent of the sample regularly participated in the Adult Netherlands Twin Register survey studies (ANTR) [176], while 25% were derived from the Young Netherlands Twin Register (YNTR) [177].

Measurements

SWB was assessed with the Satisfaction with Life Scale [178]. The scale consists of 5 items which had to be answered on a 7-point scale ranging from 1 = *'strongly disagree'* to 7 = *'strongly agree'*. An example item is "My life is going more or less as I wished". Internal consistency of the scale was good with a Chronbach's Alpha of .86. SWB have been assessed longitudinally. To maximize the sample size for the current analyses, we took the last valid assessment and replaced missing values with assessment scores from earlier time points.

NEU and EXT were based on the Item response scores from the Genetics of Personality Consortium meta-analysis [167].

DNA collection, Genotyping and Imputation

Genotyping was done on several genotyping platforms; including the Perlegen-Affymetrix platform, the Affymetrix 6.0 platform, the Illumina Human Quad-Beadchip 660K, and the Illumina Omni 1M. Genotyped data were cross-platform imputed against GONL reference set to infer the SNPs missing per platform in the combined data [102]. Pre-imputation Quality Control (QC) included aligning the alleles to the plus strand, excluding alleles with frequencies differences more than 10%, SNPs with MAF < 0.005, significant deviation from Hardy-Weinberg Equilibrium (HWE) $p < 10^{-12}$ and call rate < 0.95. Samples with genotype call rate < 0.90, heterozygosity falling outside of the interval (F < -0.075; F > 0.075), Affymetrix CQC < 0.40 if applicable, Mendelian error rate > 5 standard deviations (SDs) from the mean, gender and Identity-by-State (IBS) status mismatch between known status and genotypic assessment were excluded. Phasing and imputation was performed with MaCH-Admix [47] software and probabilities of inferred genotypes were converted to best guess format using Plink 1.90 [179]. After imputation, SNPs that were significantly associated with genotyping platform ($p < 10^{-5}$) and had allele frequencies difference > 10% with GoNL reference set, HWE $p < 10^{-5}$, Mendelian error rate > 5sd of mean over all markers and imputation quality $R^2 < 0.90$, were excluded. Part of the NTR sample was sequenced using the GONLseq platform and these persons were added to the dataset *after* imputation. We then performed a Principal Components Analysis (PCA) to check for possible ethnic and platform stratification.

Statistical analysis

Phenotypes were corrected for sex, age z-score, age z-score squared (for SWB), Dutch population structure based on the Principal Components Analysis (PCA) and chip effects using linear regression in SPSS v.21. The standardized residuals were analyzed. The variance in each trait explained by common SNPs and the genetic correlation between traits were estimated using the bivariate restricted maximum likelihood (REML) method of Genome-wide Complex Trait Analysis (GCTA) tool [20], by specifying a genetic relationship matrix (GRM) among all unrelated individuals in the study and a second matrix among relatives [37]. The advantage of using GCTA is that it overcomes the obstacle that SNPs might have too little effect to detect them in conventional GWAS analysis on one dataset. Genetic Relationship Matrix (GRM) was calculated based on best guess imputed genotypes. Ethnic outliers (n = 710) were excluded. SNPs with MAF < 0.01 were excluded as well to calculate the GCTA matrix. Total N of SNPs that passed the QC was 1,228,124.

Results

Bivariate pedigree-based analysis yielded estimates of total heritability of 32% for SWB, 37% for NEU and 42% for EXT (Table 1) with 7%, 10% and 16% of the variance in SWB, NEU and EXT, respectively, accounted for by SNPs

The phenotypic correlations (Table 2) between SWB and NEU (r = -.46) and between NEU and EXT (r = -.44) were negative, whereas a positive correlation was observed between SWB and EXT (r = .32). The directions of genetic correlations mirrored the observed ones, detecting negative pedigree-based (r_g = -.70 and r_g = -.53) and SNP-based (r_g = -.80 and r_g = -.34) correlations between SWB and NEU and between NEU and EXT, respectively, and positive genetic correlation between SWB and EXT in pedigree-based (r_g = .48) and SNP-based analysis (r_g = .18).

Table 1

Bivariate pedigree (total h^2) and SNP-heritability (SNP h^2) estimates and genetic correlations (r_G); N represents the sample size for each phenotype; *N1 + N2* represents the number of data points in bivariate analysis (sum of sample sizes for both phenotypes).

Pedigree-based heritability and genetic correlation			
	N	*total h2*	*SE*
SWB	9,141	.32	.02
NEU	9,020	.37	.02
EXT	9,018	.42	.02
	$N_1 + N_2$	r_G	*SE (p-value*)*
SWB-NEU	18,161	-.70	.03 (< .001)
SWB-EXT	18,159	.48	.03 (< .001)
NEU-EXT	18,038	-.53	.03 (< .001)
Heritability and genetic correlation that can be attributed to SNPs			
	N	*SNP* h^2	*SE*
SWB	9,141	.07	.04
NEU	9,020	.10	.04
EXT	9,018	.16	.04
	$N_1 + N_2$	r_G	*SE (p-value*)*
SWB-NEU	18,161	-.80	.25 (.01)
SWB-EXT	18,159	.18	.26 (.26)
NEU-EXT	18,038	-.34	.20 (.09)

*One-tailed test, when rG fixed at 0.000

Table 2

Phenotypic correlations (*r*) between SWB, NEU and EXT. N represents the sample size for each phenotype, complete pairs of observations were used to compute correlations.

Phenotypic correlation			
	N	*r*	*p-value*
SWB-NEU	7,935	-.46	< .001
SWB-EXT	7,934	.32	< .001
NEU-EXT	9,018	-.44	< .001

Discussion

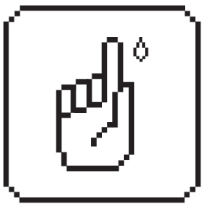
Bivariate GCTA analyses revealed that a significant part of the variance in SWB, NEU and EXT was accounted for by the additive effect of SNPs. The SNP-based heritability for SWB was estimated to be 7%, which is similar to that previously reported [169]. The SNP-based heritability for NEU and EXT was estimated at 10% and 16%, respectively. The overlap in genetic influences between SWB and NEU was substantial (r_g = -.80, SE = .25), while the genetic overlap between SWB and EXT was only moderate (r_g = .18, SE = .26), which is in sharp contrast to the phenotypic correlation (r = -.46 for SWB-NEU and r = .32 for SWB-EXT). Thus, the SNPs that influence SWB and NEU overlap substantially, with opposing effects, while much less genetic overlap is observed for SWB and EXT, while their phenotypic correlation is similar to that of SWB and NEU. This finding is in contrast to previous results, in which a larger common genetic variance between SWB and NEU, than between SWB and EXT, was found [175], but similar genetic correlations between SWB and NEU/EXT were reported [33].

The results of our powerful bivariate SNP-based design indicate that the similar overlap between SWB and NEU and between SWB and EXT has different underlying sources. The main source of overlap between SWB and NEU has repeatedly been found to be genetic, while our novel results indicate that genetic influences are much less important in explaining the observed association between SWB and EXT. This implies that EXT would not pass the test of being an informative proxy phenotype for SWB, but could be an attractive target for environmental based prevention or intervention stratiegies. Our results need to be interpreted in light of the following limitations.

First, the study of Hanh et al (2013) [175] showed that the SWB and NEU/EXT shared both additive and non-additive genetic factors, which cannot be detected using GREML, as it assumes the additive genetic model. Moreover, they share common environmental factors. High and significant pedigree-based genetic correlation between SWB and NEU (r_g = -.70, SE = .03) and SWB and EXT (r_g = .48, SE = .03) support this hypothesis. Regardless of the magnitudes of shared additive, non-additive genetic and common environmental influences, they should be taken into account, while interpreting the results of genetic studies of SWB [169].

Second, although we applied a bivariate model in a relative large sample, our sample size is still small for restricted maximum likelihood (REML) methods. In current study, we detected significant SNP-heritability, by employing a new method that includes family members, to increase power. However, it is possible, that a larger sample size is required for more reliable estimates of genetic correlations.

To conclude, in this study we showed that genetic overlap between SWB and NEU is larger than between SWB and EXT. Based on our results, the detection of loci common to both phenotypes is likely between SWB and NEU, rather than SWB and EXT. The role of EXT in explaining the inter-individual differences in SWB remains to be explored. Genetic association studies of SWB and its relation to other personality traits will add important information to reveal new biological pathways, which can then serve as a strong foundation for the development of multidisciplinary health, social, and economic policies.



CHAPTER 6

HOMEOSTATIC MODEL ASSESSMENT OF β-CELL FUNCTION AND INSULIN RESISTANCE: A GENOME WIDE INFERRED STATISTICS ASSOCIATION STUDY

This chapter is based on:

Iryna O. Fedko, Michel G. Nivard, Jouke-Jan Hottenga, Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) Investigators, Reedik Mägi, Inga Prokopenko, and Dorret I. Boomsma. Homeostatic Model Assessment of β-cell Function and Insulin Resistance Genome Wide Inferred Statistics Association Study *(as to be submitted).*

Abstract

We applied a recently developed approach called GWIS (Genome Wide Inferred Statistics) to discover genetic variants influencing Homeostatic Model Assessment of β-cell function (HOMA-B) and Insulin Resistance (HOMA-IR).

We achieved a boost in power by approximating HOMA-B/-IR functions summary statistics from Fasting Glucose (FG) and Fasting Insulin (FI) from recent meta-analysis results. Earlier GWAS conducted with HOMA as an outcome suffered from the missing values in either FI or FG, thus leading to decrease in sample size and power compared to single FI or FG meta-analysis. GWIS allows overcoming this issue and approximates the summary statistics for the non-linear function of HOMA phenotypes. The GWIS analysis revealed eleven loci, including four novel, for HOMA-B and five loci, including three novel, for HOMA-IR ($p < 5\times10^{-8}$). Previously, seven HOMA-B and one HOMA-IR locus were also reported for an association with T2D risk. Significant genetic correlation was detected between HOMA-IR and T2D (r_g=0.53, SE=0.08, $p=2.94\times10^{-10}$), suggesting that in the future, better powered genome-wide association studies will show large locus overlap between these two related phenotypes. Shared genetic risk factors between BMI and both HOMA-B (r_g = 0.39, SE = 0.05, p-value = 6.58×10^{-15}) and HOMA-IR (r_g = 0.62, SE = 0.05, p-value = 9.26×10^{-35}) as well as between HOMA-IR and T2D (r_g = 0.53, SE = 0.08, p-value = 2.94×10^{-10}) highlight the complex interplay between obesity, insulin resistance and β-cell dysfunction risk to T2D. Our results show that FI loci are predictive of insulin resistance loci and FG loci are predictive of β-cell function loci, indicating a role of the FG and FI in pathophysiology of T2D through measures of HOMA-B and HOMA-IR.

Introduction

Genome-wide association studies (GWAS) for fasting glucose (FG) and insulin (FI) in non-diabetic European individuals have allowed identification of over 72 genetic loci [180]. While FG can be considered in the diagnosis of type 2 diabetes (T2D) and FI level is indicative for the body insulin sensitivity, they do not provide informative mechanistic insights about physiological measures of insulin secretion and action [181]. Homeostatic Model Assessment of β-cell function (HOMA-B) and Insulin Resistance (HOMA-IR) are commonly used glycaemic indices studied in relation to T2D [182-190] and both are calculated from FG and FI values. HOMA-B reflects the function of the β-cell with respect to insulin secretion, whereas HOMA-IR is the estimate of insulin sensitivity.

FI and FG have to be assessed in the same individual to derive each of HOMA indices. Missing values in original estimates cause a decrease in the sample size at study and meta-analysis levels for each of the derived measures. Whilst published HOMA-B/HOMA-IR GWAS meta-analyses were undertaken at the same time with FG/FI, they featured a much smaller sample size compared to the original phenotypes, for instance, they were based on up to 36,466/37,037 individuals in contrast to 46,186/38,238 for FG and FI, respectively [191], in the discovery stage. The loss in power meant that compared to 15 FG/0 FI loci, the HOMA-based GWAS reported only 4 HOMA-B/1 HOMA-IR loci reaching genome-wide significance ($p < 5\times10^{-8}$) in the discovery sample.

The recent FG and FI GWAS meta-analyses were based on up to 88,320 and 64,090 individuals, respectively [192], and offer a unique opportunity to improve our knowledge about genetic variants influencing HOMAs, even though the samples used in those meta-analyses do not necessarily overlap. The aim of this study was three-fold: (i) to infer analytically the GWAS summary statistics for HOMA-B/-IR using FG/FI recent GWAS meta-analysis results from the MAGIC [192] using GWIS [38], (ii) to evaluate the gain in power through the comparison to previously reported HOMA GWAS results and (iii) to define the effects of glycaemic loci on insulin secretion and action through their effects on HOMA-B/HOMA-IR. Throughout this paper, we will refer to the previous meta-analysis of HOMA [191] as **'published'** and the results of the present study and analysis as **'inferred'** meta-analysis of HOMA.

Methods

Phenotype definitions

Homeostatic Model Assessment of β-cell function (HOMA-B) and Insulin resistance (HOMA-IR) are calculated from the Fasting Glucose and Fasting Insulin measures using the following formulas:

$$\text{HOMA-B} = \frac{20 \times \text{FI}}{\text{FG-3.5}}, \qquad \text{HOMA-IR} = \frac{\text{FG} \times \text{FI}}{22.5},$$

where FG is measured in mmol/l and FI is in mU/l units [193].

Method

We applied a recently developed approach by Nieuwboer et al [38] for the approximation of a GWAS summary statistics for HOMA, which is a function of FI and FG. GWIS allows conducting the *in silico* GWAS meta-analysis of HOMA-B and HOMA-IR based on summary statistics for FI and FG, if the means of the FI and FG, the effect allele frequency, the correlation between the two traits and the sample overlap between them are known (note, that GWIS can be computed based on overlapping, as well as on non-overlapping samples; in case of overlapping samples, the GWIS results can be corrected. In the absence of accurate knowledge of the phenotypic correlation between traits, this correlation can be inferred using LD score regression).

We used the summary statistics from the latest GWAS meta-analysis of FG and FI performed by the Meta-Analysis of Glucose and Insulin-related traits Consortium (MAGIC) in up to 88,320/64,090 individuals and 40/32 studies, respectively (Supplementary Material). In the MAGIC meta-analysis, FI was natural log transformed and measured in pmol/l units, whereas FG was measured in mmol/l with a cut-off at 7mmol/l. The standard HOMA formulas require untransformed FG/FI measures and use mU/l units for FI. We adapted an approximation to compute HOMA-IR and HOMA-B given ln(FI). In addition, to account for previous modeling of both HOMAs for the meta-analyses that used natural log transformation, we further implemented the approximation to result in summary statistics for ln(HOMA-IR) and ln(HOMA-B). Therefore, the relationships between the FG and FI phenotypes as used in GWAS and the HOMA-B and HOMA-IR phenotypes are:

$$\text{HOMA-B} = \ln \frac{20 \times \frac{e^{\text{FI}}}{6.945}}{\text{FG-3.5}}, \qquad \text{HOMA-IR} = \ln \frac{\text{FG} \times \frac{e^{\text{FI}}}{6.945}}{22.5},$$

where division by 6.945 was required to convert FI from pmol/l to mU/l units. The GWIS method requires, in addition to genome wide summary statistics for FG and FI, the mean FG and FI values, the phenotypic correlation between FG and FI and sample overlap across the studies, included in both meta-analyses studies, to correct for dependence

between the FG and FI GWAS.

Study overlap and genetic correlation between published and inferred GWAS results
As sample overlap differs per SNP, e.g. because SNPs can be missing in individual cohorts, we computed sample overlap for each SNP by the following: 1) we ran the LD score regression and estimated the correlation intercept (gcov_int) between FG and FI; 2) we computed *r*, which is the phenotypic correlation obtained from the LD score cross trait intercept and knowledge of Ns1 and Ns2 (see Supplementary Material for the procedure), where Ns1 and Ns2 are the minimum sample sizes in the FI and FG individual cohort GWAS, and this quantity is equal to the expected correlation between the standard errors obtained in the FI and FG GWAS; 3) we obtained the vector of correlation intercepts for each SNP from *r* and information about sample sizes per SNP. The LD score regression correlation intercept (gcov_int) from the analysis of genetic correlation between FI and FG was 0.2607. The values of computed local correction vector ranged from 0.02847 to 0.306 with the mean = 0.2635 and median = 0.2607. The mean across studies in the meta-analysis was 5.08 mmol/l for FG and 58.56 pmol/l for FI. The approximate sample size, for the inferred HOMA meta-analysis results, was computed as the geometric mean of the sample sizes for FG (N1) and FI (N2) $N_{approx}=\sqrt{(N1 \times N2)}$. We calculated the effect allele frequency as a weighted mean of the FI and FG effect allele frequencies

Quality Control (QC)
We applied the following QC criteria to the FG and FI summary statistics: first, the effect alleles for FG and FI effect estimates were aligned; second, the effect allele frequencies were set to be less or equal to 10% for the difference between FG and FI (4,299 SNPs were excluded) and not larger 20% with CEU HapMap 2 reference set [12] (36 SNPs for FI and 66 SNPs for FG were removed). We computed the effect estimates and SEs and applied the LD score regression to compute the r_g between published and inferred HOMA-B/-IR. The expectation, when the same trait is analyzed, is that the genetic correlation equals 1. To evaluate the power gained by using GWIS, we compared meta-analysis results of the inferred and published HOMA-B/-IR. A possible inflation of summary statistics for variants with a low N is the limitation of the GWIS method. We applied the post-calculation QC (Supplementary Material Figures 2,7-8) and excluded 278,534/ 278,577 SNPs with N < 35,000 for HOMA-B/-IR, among which 113/113 were filtered out based on with MAF < 0.01 filter. The total number of SNPs after QC was 2,432,775 for both HOMA-B and HOMA-IR. In addition, we estimated the LD score intercept (Table 3), which was then used to adjust the summary statistics for any residual population stratification or misspecification of the sample dependence. We built the Manhattan and QQ plots from inferred results for HOMA-B and HOMA-IR.

Cross-SNP cross-trait comparison

We assumed signals reaching p-value = 5×10^{-8} as GWAS significant for each of FG/FI/ HOMA-B/HOMA-IR traits. Guided by epidemiological correlations between FG and HOMA-B, and FI and HOMA-IR [194], we investigated whether we observed effects on HOMA-B reaching GWAS significance in our study for FG loci and similarly on HOMA-IR analyses for FI loci. We additionally compared the effect estimates between the published and inferred HOMA GWAS meta-analyses. We also obtained publicly available meta-analysis summary statistics for HDL, LDL, TC, TG [195], BMI [196], T2D [197] (see URLs) and performed the LD score regression to compute the genetic correlations between these phenotypes and inferred HOMA-B/-IR and FI, FG.

Results

Analytically inferred GWAS for HOMA-B/-IR

We inferred the HOMA-B/HOMA-IR GWAS summary statistics for 2,432,775 variants for up to 88,320/64,090, individuals from the MAGIC FG/FI GWAS [192]. We observed significant associations in seven previously established loci at *G6PC2, ADCY5, DGKB, GCK, GLIS3, FADS1/2/3, MTNR1B* and identified four loci *SLC30AB, TCF7L2, ARAP1, FOXA2* reaching the genome-wide significance for HOMA-B (Figure 1 and 3, Table 1, Supplementary Figures 13-16). We observed three novel loci at *LYPLAL1/SLC30A10, PER4, PPP1R3B* and confirmed established loci at *GCKR* and *IGF1* were genome-wide significant in the inferred HOMA-IR GWAS (Figure 2 and 3, Table 2, Supplementary Figures 13-16). In total eight loci (*ADCY5, DGKB, GCK, SLC30AB, GLIS3, TCF7L2, ARAP1, MTNR1B*) out of eleven for HOMA-B and one (*GCKR*) out of five for HOMA-IR were also associated with T2D in previous studies [180].

Figure 1. Manhattan plot for an inferred HOMA-B with FG loci (indicated in green)

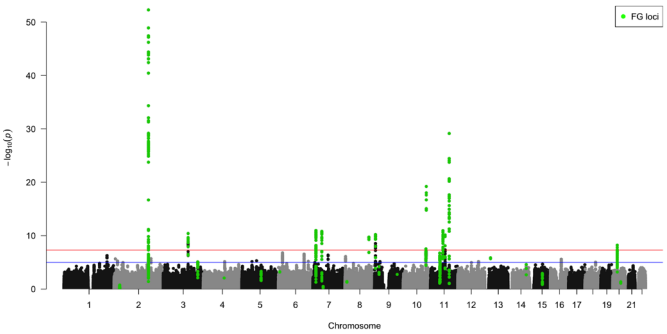


Figure 2. Manhattan plot for an inferred HOMA-IR with FI SNPs (indicated in green)

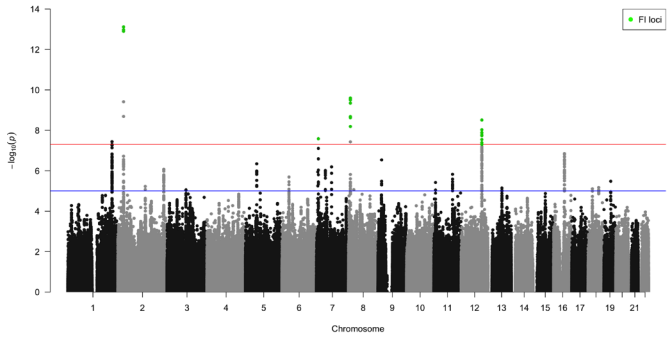
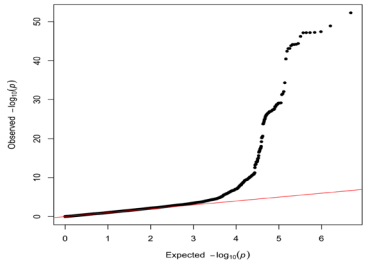


Figure 3. QQ plots for inferred a) HOMA-B and b) HOMA-IR

a



b

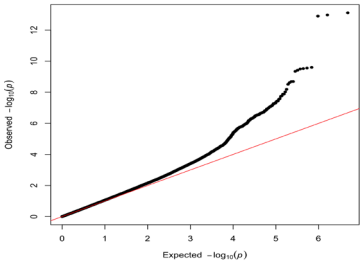


Table 1. Comparison of HOMA-B and FG effects across genome-wide significant loci

SNP	CHR	Locus name	HOMA-B			FG	
			Alleles (effect / other)	Effect (SE), p-value	N	Effect (SE), p-value	N
rs560887	2	*G6PC2*	C/T	-0.050 (0.0033), 5.33×10^{-53}	75,173	0.069 (0.0027), 2.41×10^{-144}	88,202
rs11708067	3	*ADCY5*	A/G	-0.025 (0.0038), 3.93×10^{-11}	75,227	0.022 (0.0031), 5.55×10^{-13}	88,301
rs10258074	7	*DGKB*	A/T	0.021 (0.0031), 1.07×10^{-11}	75,202	-0.028 (0.0025), 1.18×10^{-27}	88,240
rs2908289	7	*GCK*	G/A	0.029 (0.0042), 1.57×10^{-11}	74,672	-0.059 (0.0034), 8.36×10^{-66}	87,756
rs3802177	8	*SLC30AB*	G/A	-0.023 (0.0036), 1.85×10^{-10}	68,044	0.029 (0.0030), 1.24×10^{-21}	76,819
rs7034200	9	*GLIS3*	A/C	-0.020 (0.0030), 6.20×10^{-11}	74,669	0.016 (0.0025), 2.32×10^{-10}	87,749
rs7903146	10	*TCF7L2*	C/T	0.032 (0.0035), 6.19×10^{-20}	75,212	-0.022 (0.0028), 1.04×10^{-14}	88,262
rs10830963	11	*MTNR1B*	C/G	0.044 (0.0039), 7.28×10^{-30}	68,612	-0.073 (0.0032), 2.37×10^{-113}	77,412
rs1552224	11	*ARAP1*	C/A	0.027 (0.0041), 7.45×10^{-11}	75,207	-0.023 (0.0034), 5.56×10^{-12}	88,250
rs174555	11	*FADS1/2/3*	T/C	-0.022 (0.0033), 1.15×10^{-11}	75,237	0.021 (0.0027), 2.17×10^{-14}	88,322
rs6048205	20	*FOXA2*	G/A	0.043 (0.0073), 6.10×10^{-09}	70,262	-0.042 (0.0061), 8.03×10^{-12}	80,759

Table 2. Comparison of HOMA-IR and FI effects across genome-wide significant loci

			HOMA-IR			FI	
SNP	**CHR**	**Locus name**	**Alleles (effect / other)**	**Effect (SE), p-value**	**N**	**Effect (SE), p-value**	**N**
rs2605101	1	*LYPLAL1/ SLC30A10*	A/T	0.019 (0.0034), 3.65×10^{-08}	75,238	0.017 (0.0032), 1.46×10^{-07}	64,091
rs780093	2	*GCKR*	C/T	0.024 (0.0032), 7.70×10^{-14}	75,226	0.018 (0.0030), 1.73×10^{-09}	64,091
rs10224545	7	*PER4*	T/C	-0.031 (0.0056), 2.61×10^{-08}	71,951	-0.029 (0.0052), 2.92×10^{-08}	64,091
rs4240624	8	*PPP1R3B*	A/G	-0.035 (0.0056), 2.52×10^{-10}	68,617	-0.030 (0.0051), 4.25×10^{-09}	60,824
rs2114912	12	*IGF1*	T/G	0.026 (0.0044), 3.09×10^{-09}	74,634	0.024 (0.0041), 3.76×10^{-09}	63,539

GWIS power gain

The genetic correlation between the published and inferred HOMA-B and HOMA-IR were at or above 1 (Tables 3). The LD score intercepts were estimated as 1.05 for HOMA-B and 1.03 for HOMA-IR. The comparison of summary statistics distributions between published and inferred HOMAs demonstrated that the standard errors (SEs) decreased in current analyses compared to the previously determined, supporting the gain in power over previous GWAS efforts (see Supplementary Table 1 – 12) [191]. The comparison of the variants with low N and MAF between the published and inferred GWAS showed a lower concordance between SEs for low N and MAF (Supplementary Figures 2, 6-8, 11) thus supporting the QC filters for minimal sample size and MAF. We also compared the magnitude and direction of the effects in genome-wide significant loci in the published and inferred analysis and confirmed a power gain, particularly for HOMA-IR and FI (Supplementary Table 2 and 3).

Table 3. Genetic correlation (r_g) between published and inferred HOMA-B/-IR, LD score regression intercept, SNP-heritability (h^2) and their standard errors (SE).

	Published	**Inferred**
	HOMA-B	
r_g (SE), p-value	-	1.14 (0.07), 2.22×10^{-56}
Intercept (SE)	0.99 (0.008)	1.05 (0.008)
h^2 (SE)	0.08 (0.015)	0.056 (0.01)
	HOMA-IR	
rg (SE), p-value	-	1.18 (0.08), 9.11×10^{-53}
Intercept (SE)	1.004 (0.007)	1.03 (0.008)
h2 (SE)	0.07 (0.01)	0.059 (0.008)

Effect of glycaemic traits loci on HOMA-B/-IR

As expected from epidemiological data, in our study, the largest shared content was observed between genetic effects on FG and HOMA-B, and between FI and HOMA-IR (Table 1 - 2).

The genetic correlation between FI and FG was 0.32 (SE = 0.1, p-value = 0.0008). The genetic correlation between FI and HOMA-IR was almost complete, i.e. 0.98 (SE = .005, p-value < 0.001), consistent with observation that higher levels of FI are associated with lower insulin sensitivity. Counter intuitively, HOMA-B and FI showed substantial degree of genetic correlation (r_g =0.76, SE = 0.05, p-value = 1.76×10^{-55}), which was higher than that between FG and HOMA-B (r_g = -0.38, SE = 0.12, p-value = 0.002) or HOMA-IR (r_g = 0.49, SE = 0.07, p-value = 4.2×10^{-12}). A negative genetic correlation between FG and HOMA-B suggested that increased FG is related to the reduced β-cell function. We detected significant genetic correlation between HOMA-B/-IR and BMI (r_g =0.39, SE = 0.05, p-value = 6.58×10^{-15} and r_g = 0.62, SE = 0.05, p-value = 9.26×10^{-35}). In contrast, the genetic correlation with T2D was found only with HOMA-IR (r_g = 0.53, SE = 0.08,

p-value = 2.94×10^{-10}), while there was no correlation genome-wide between T2D and HOMA-B [198]. For both HOMA indices, genetic correlations with HDL and TG were statistically significant.

Table 4. Genetic correlation (r_g) between inferred HOMA-B and HOMA-IR and other phenotypes, with their standard errors (SE) and p-values.

	HOMA-B			HOMA-IR		
	r_g	SE	P-value	r_g	SE	P-value
FI	0.76	0.05	1.76×10^{-55}	0.98	0.005	< 0.001
FG	-0.38	0.12	0.002	0.49	0.07	4.2×10^{-12}
HDL	-0.30	0.06	1.6×10^{-7}	-0.46	0.05	9.1×10^{-19}
LDL	0.11	0.06	0.05	0.07	0.06	0.22
TC	0.07	0.05	0.18	-0.005	0.05	0.92
TG	0.30	0.06	1.4×10^{-6}	0.37	0.09	4.6×10^{-5}
BMI	0.39	0.05	6.58×10^{-15}	0.62	0.05	9.26×10^{-35}
T2D	-0.02	0.10	0.85	0.53	0.08	2.94×10^{-10}

Discussion

In this study, we presented a new and the largest to date GWAS meta-analysis of HOMA-B and HOMA-IR glycaemic indices based on an analytical inference from FI and FG meta-analysis summary statistics. The gain in power was obtained from the doubled sample size of HOMA GWAS compared to the previously published meta-analysis [191]. This GWIS-based analysis revealed four novel HOMA-B and three novel HOMA-IR loci and allowed to characterize the effect of the established FG and FI loci on HOMA-B/-IR.

We achieved the boost in power using an analytical approach, which potentially can save large amounts of analytical time at individual study level. The ability to compute the summary statistics in partially overlapping samples demonstrated an advantage of the GWIS inference-based method over direct analytical GWAS for composite phenotypes. Additionally, this approach does not suffer from the power losses due to missing values in original phenotypes. With this study, we pave the way for future studies to explore other traits and many other glycaemic indices, defined as a function of other phenotypes. Novel mechanistic insights in T2D pathophysiology [199] could be brought through comprehensive characterization of a number of indices of insulin secretion, action and sensitivity, including Insulin Sensitivity Index, Insulin-Glucose Ratio, Insulin Ratio, Insulinogenic Index, Corrected Insulin Response. To calculate these indices, pairs of measures of Insulin and Glucose are required at basal level, after 30 min or 2 hours glucose load intake [200]. Such measures may not be available in the same person or in large overlapping samples, thus, studies have been underpowered [191]. GWIS allows usage of the summary statistics from just Insulin and Glucose measured at 3 time

points and approximating summary statistics for all glycaemic indices irrespective of the sample overlap. In case of no overlap between cohorts with different measures, computation of GWIS is in fact the only way to approximate summary statistics, as GWAS cannot be performed in that case.

High insulin and glucose levels, insulin resistance and β-cell dysfunction characterize T2D pathophysiology, however, the exact mechanism of genetic interrelationships between these measures remains unclear. *ADCY5, DGKB, GCK, SLC30AB, GLIS3, TCF7L2, ARAP1, MTNR1B* HOMA-B loci and the *GCKR* HOMA-IR locus were previously reported for an association with T2D [180]. It has been shown that insulin resistance and β-cell function may have a distinct impact on susceptibility to T2D, and mechanistically T2D loci can be related to a specific biological process affecting insulin secretion, resistance or processing [201].

Previous studies established both positive and negative effects of HOMA-B in T2D loci [202]. Opposite direction of the effects at these loci is confirmed by the non-significant genetic correlation between HOMA-B and T2D genome-wide in this study. Our results are in line with previously reported genetic correlations between HOMA-B and FI/FG/BMI/T2D and HOMA-IR and FI/FG/BMI [198], however, the significant genetic correlation between HOMA-IR and T2D we observed has not been reported previously. The role of obesity and adiposity [203, 204], in particular, in risk of T2D is reflected in a significant genetic correlations between HOMA-B/-IR and BMI observed in our study. A substantial genetic correlation between HOMA-B and FG/FI and a perfect correlation between HOMA-IR and FI (Table 4) suggests that the fasting glycaemic trait loci for FG and FI contribute to HOMA-B/-IR as well (as would be expected based on the formulas).

To conclude, this study implemented a novel GWIS method and conducted the largest to date GWAS meta-analysis for HOMA-B/-IR indices. We gained power in comparison with previously published GWAS and reported four novel HOMA-B and three novel HOMA-IR loci. We enriched our knowledge about the role of the relationships between FG and β-cell function, FI and insulin resistance, thus, enabling further mechanistic characterization of genetic effects on T2D pathophysiology [201].

URLs: HDL, LDL, TC, TG [195] (Global Lipids Genetic Consortium, http://csg.sph.umich.edu//abecasis/public/lipids2013/), BMI [196] (GIANT, http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files), T2D [197] (DIAGRAM: http://diagram-consortium.org/downloads.html)

Supplementary Material

Study samples

FG: AGES, ALSPAC, AMISH, ARIC, ASCOT, BLSA, BSN, CHS, COLAUS, CROATIA, DECODE, DGI, ERF, FAMHS, FENLAND, FHS, FRENCHADULTCONTROL, FRENCHADULTOBESE, FRENCHYOUNGCONTROL, FRENCHYOUNGOBESE, FUSION, GENOA, GENOMEUTWINS, HABC, HEALTH2000, INCHIANTI, KORA, KORCULA, LIFELINES, NFBC66, ORCADES, PREVEND, PROCARDIS, RS, SARDINIA, SORBS, SPLIT, SUVIMAX, TWINSUK, TYROL.

FI: AGES, AMISH, ARIC, BLSA, BSN, CHILD, CHS, COLAUS, CROATIA, DECODE, DGI, ERF, FAMHS, FENLAND, FHS, FRENCHADULTCONTROL, FRENCHADULTOBESE, FUSION, GENOA, GENOMEUTWINS, HABC, HEALTH2000, INCHIANTI, KORCULA, NFBC66, ORCADES, PREVEND, PROCARDIS, RS, SARDINIA, SORBS, TWINSUK.

GWIS

As sample overlap can differ per SNP, we computed the correlation intercept per SNP as described in Methods. However, we also computed the version of GWIS using the correlation intercept, estimated for FI and FG in the LD score regression directly, i.e. we corrected the results by sample overlap uniformly across SNPs. Based on comparison of the two versions, one was selected for further analysis. We called them 'global' and 'local' correction throughout the Supplementary Material.

To compute the 'local' correction factor we 1) computed *r*, which is the phenotypic correlation obtained from the LD score cross trait intercept and knowledge of Ns1 and Ns2, using formula

$$r = gcov\_int \times \frac{\sqrt{(min(Ns1) \times min(Ns2))}}{min(min(Ns1), min(Ns2))},$$

where Ns1 and Ns2 are the sample sizes in the FI and FG single cohort GWAS. We based *r* calculation on the sample excluding low N (Ns1 < 42,727, Ns2 < 58,876) from FI and FG meta-analysis, as was done by default in LD score regression for FI and FG; 2) we obtained the vector of correlation intercepts using formula

$$\frac{Nmin}{\sqrt{(N1 \times N2)}} \times r,$$

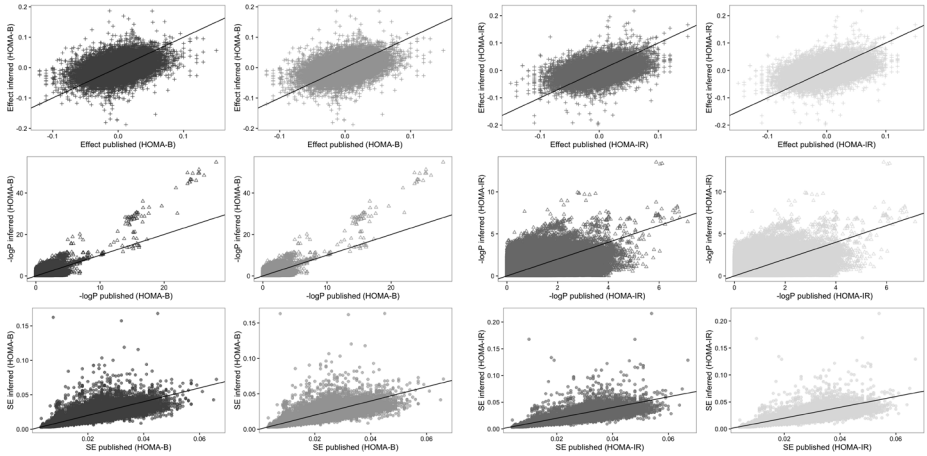
where N_{min} is the minimum of the two sample sizes per SNP, *N1* is the FI sample size *N2* is the FG sample size per SNP.

Post-GWIS QC

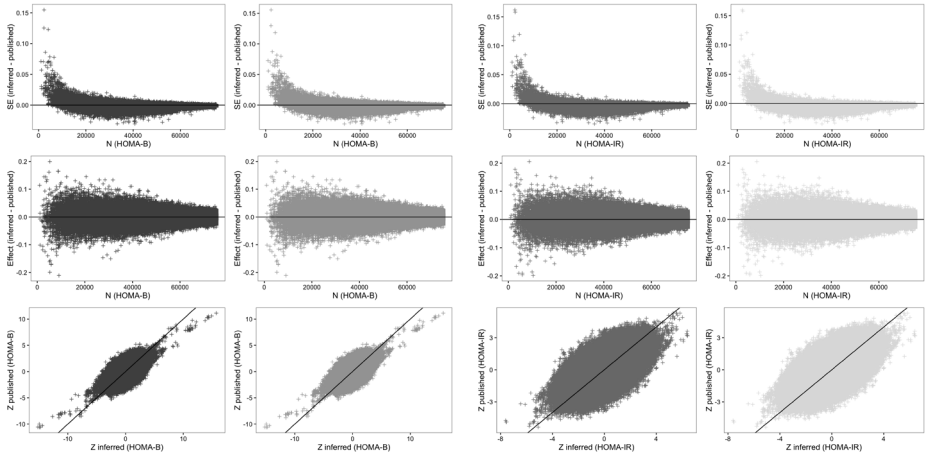
We compared the genetic correlation between published and inferred 'global' and 'local' HOMA and found similar results (Supplementary Table 1). We also compared the distribution of summary statistics of two versions of inferred and published HOMAs aiming to select 'global' or 'local' corrected version. The distributions between 'global' and 'local' correction versions of HOMA results were also very similar (Supplementary Figures 1 – 11); therefore, we proceeded with 'local' correction as it is assumed to be more precise. However, to re-iterate we did not observe the difference between 'global' and 'local' correction approaches in current analysis and given the similarity of summary statistics, they both could have been used in further analysis (Supplementary Figure 12). As was mentioned in Methods, the difference in SEs between inferred and published meta-analysis results becomes larger when N

decreases (Supplementary Figures 7-8, 11). We built the Manhattan and QQ plots for the published and inferred HOMA-B/-IR meta-analysis results (Supplementary Figures 15-18) and demonstrated the novel signals at genome-wide significant level.

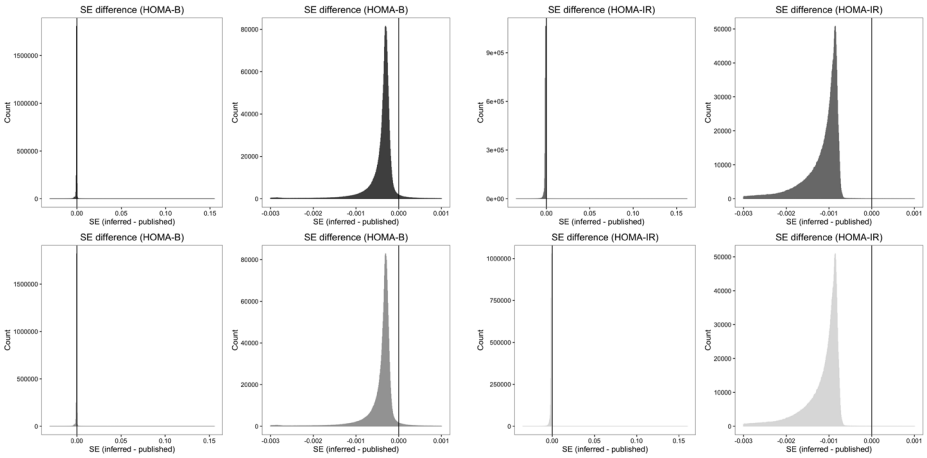
Supplementary Figure 1. Effect size, $-\log_{10}$P and SE comparison between published and inferred meta-analysis results for 'global' (dark color) and 'local' (light color) correction.



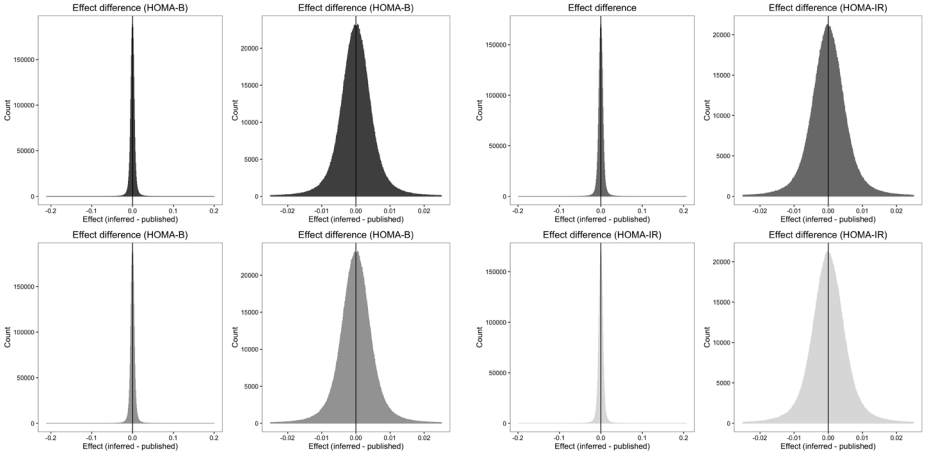
Supplementary Figure 2. Difference in SE, effect size and Z-score plotted against N for 'global' (dark color) and 'local' (light color) correction.



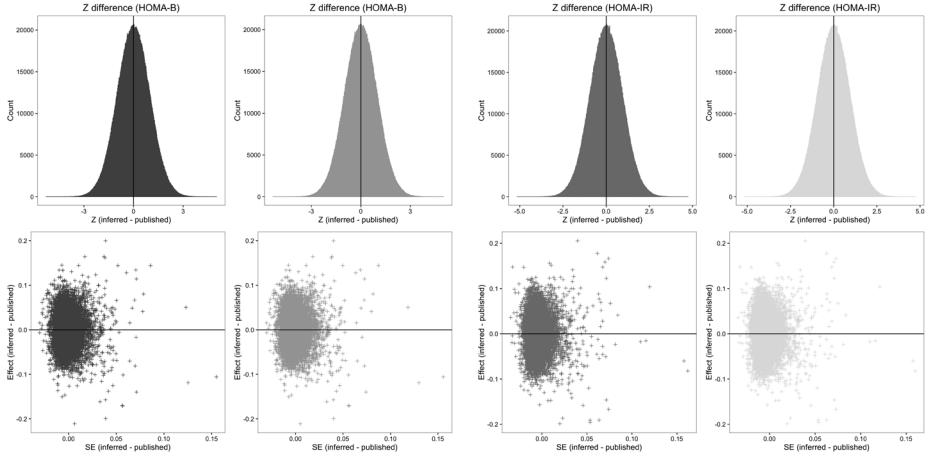
Supplementary Figure 3. Distributions of the difference between inferred and published meta-analysis SE, for all SNPs (left) and for SNPs with smaller difference (right)



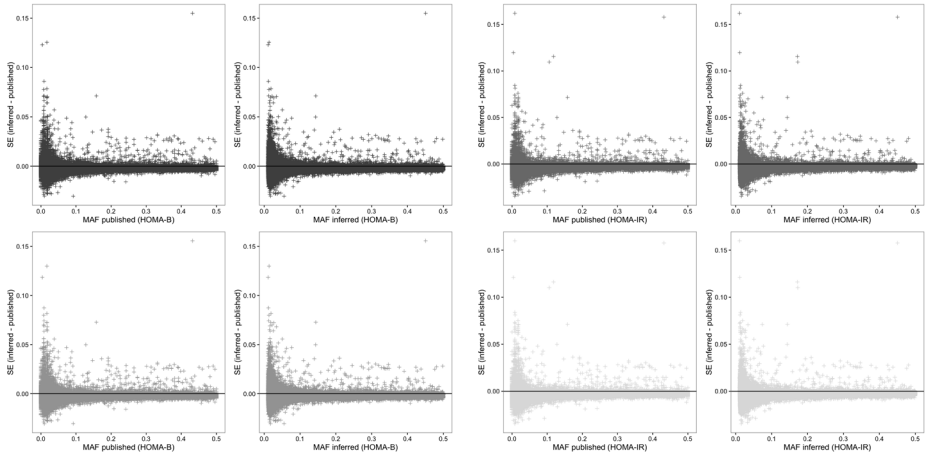
Supplementary Figure 4. Distributions of the difference between inferred and published meta-analysis effect sizes, for all SNPs (left) and for SNPs with smaller difference (right)



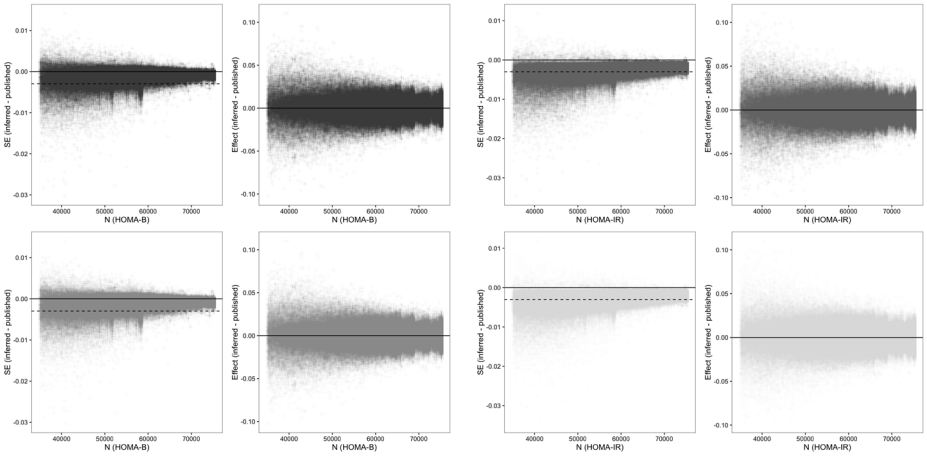
Supplementary Figure 5. Distributions of the difference between inferred and published meta-analysis Z scores and difference in effect sizes plotted against difference in SE's for 'global' (dark color) and 'local' (light color) versions.



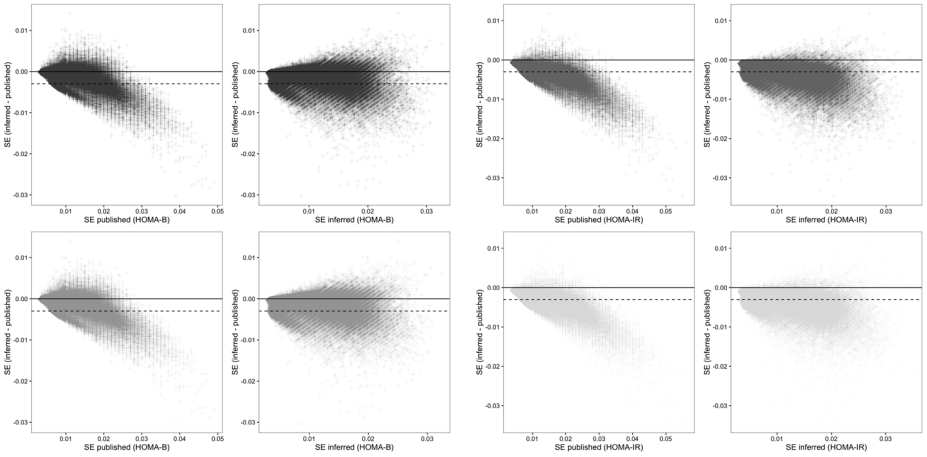
Supplementary Figure 6. SE differences plotted against MAF in published and inferred meta-analysis for 'global' (dark color) and 'local' (light color) correction



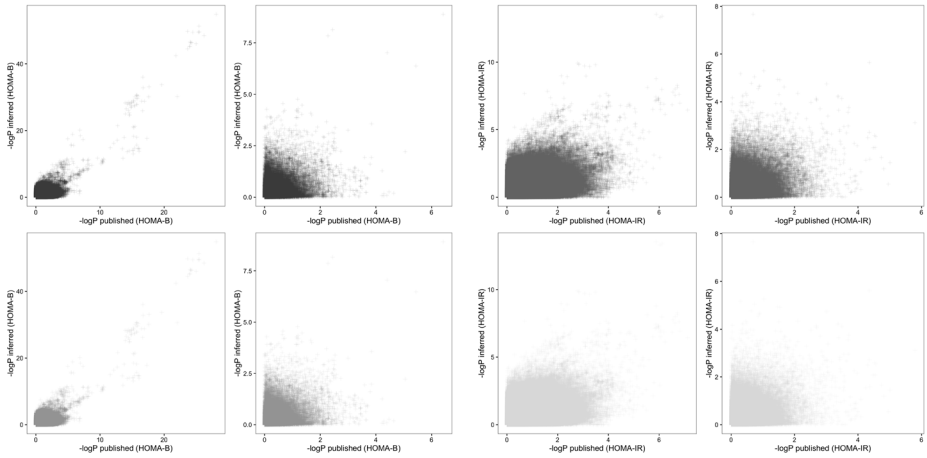
Supplementary Figure 7. Difference in SE (left) and effect sizes (right) between inferred and published GWAS plotted against N. Red line indicated difference in SE > -0.003. Datasets are filtered with N > 35,000 and MAF > 0.01



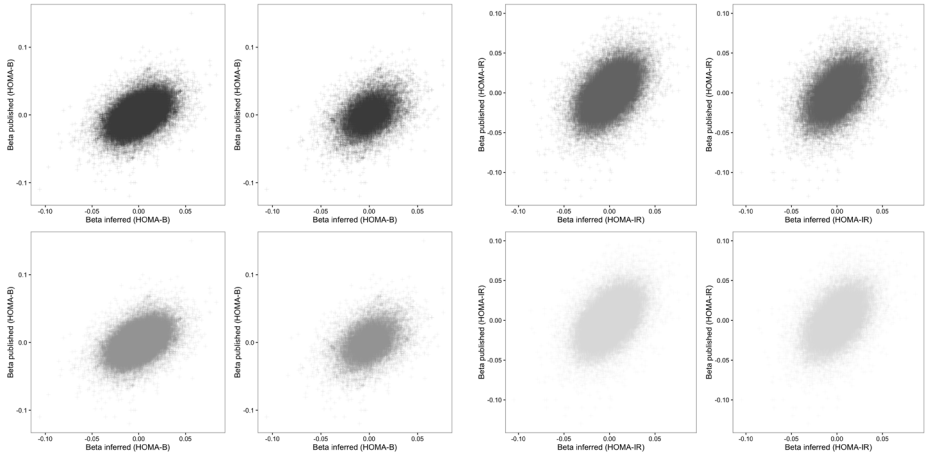
Supplementary Figure 8. Difference in SE between inferred and published GWAS plotted against SE published (left) and inferred (right). Red line indicated difference in SE > -0.003. Datasets are filtered with N > 35,000 and MAF > 0.01



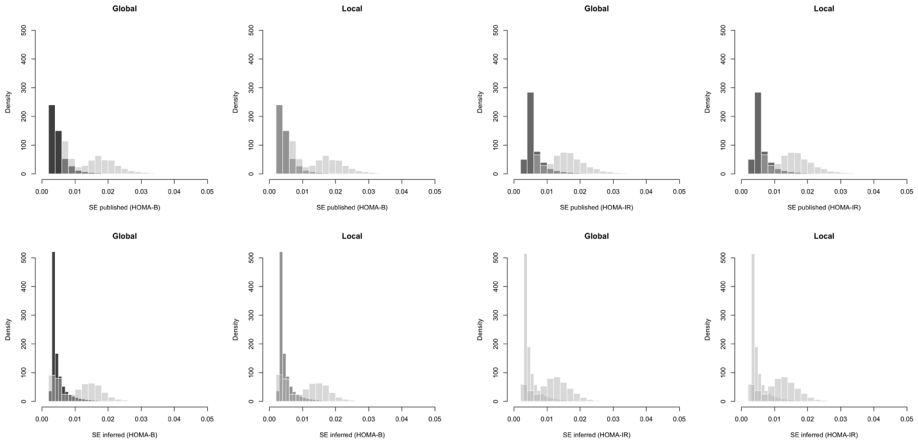
Supplementary Figure 9. -log10P of inferred against published meta-analysis results plotted for all SNPs (left) and SNPs, which SE difference between inferred and published GWAS less than -0.003 (right).



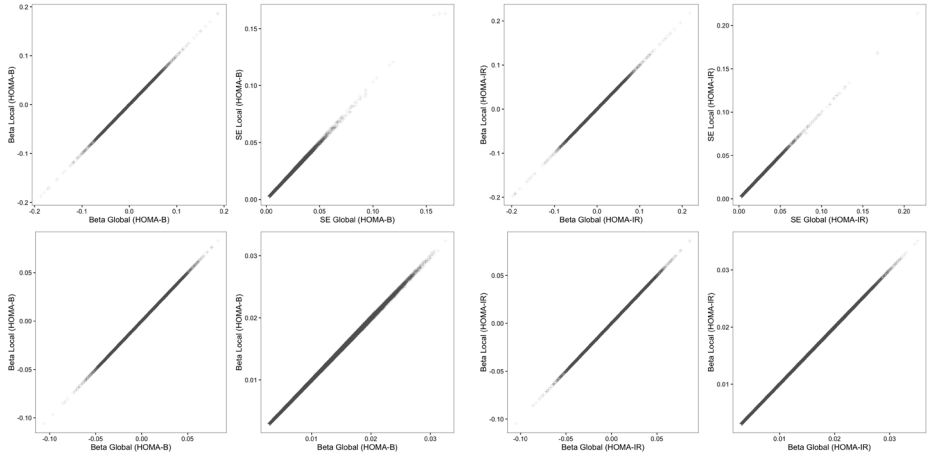
Supplementary Figure 10. Effect size of inferred against published meta-analysis results plotted for all SNPs (left) and SNPs, which SE difference between inferred and published GWAS less than -0.003 (right).



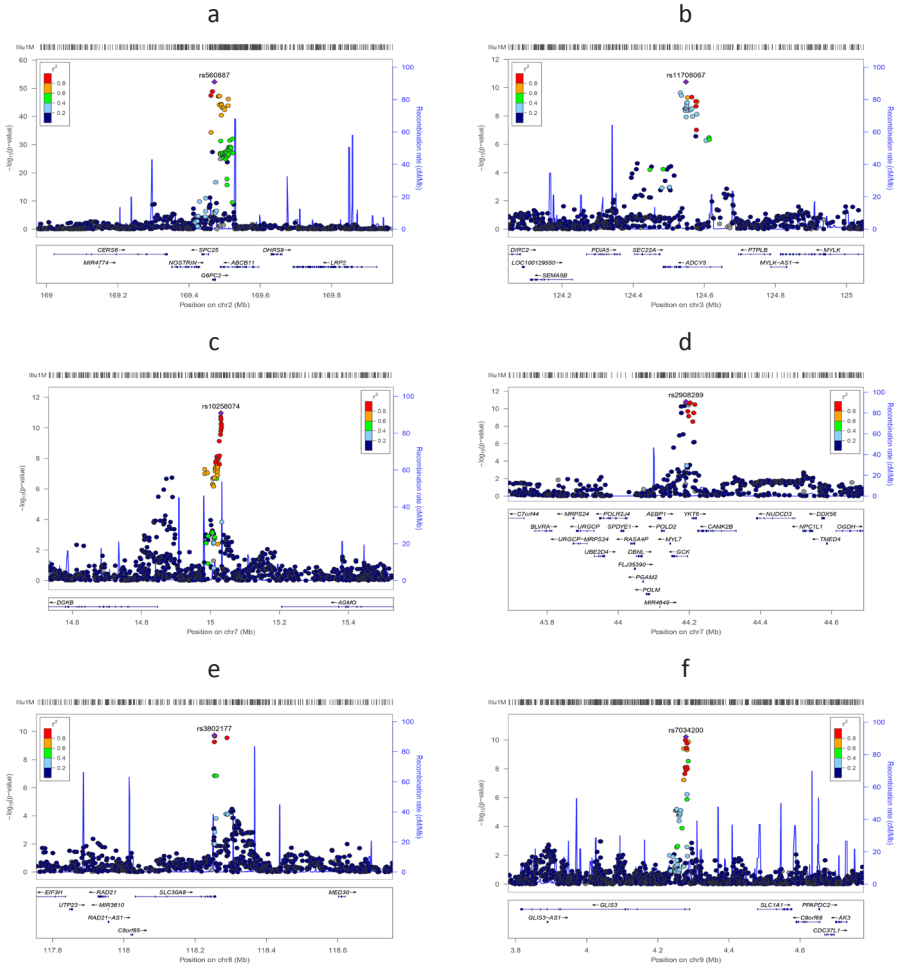
Supplementary Figure 11. Distribution of SE's of the published (upper panel) and inferred (lower panel) results stratified by SE difference < -0.003. Low (blue or green color) vs. large (gray) SEs difference between inferred and published GWAS.



Supplementary Figure 12. Comparison of SE's and effect sizes between 'global' and 'local' correction versions.

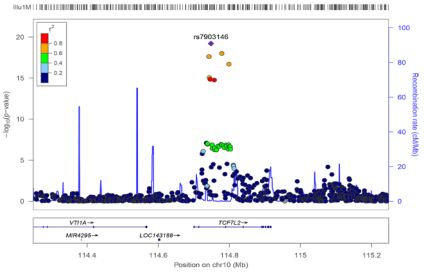


Supplementary Figure 13. Region plots for each of the leading SNPs in HOMA-B GWIS results

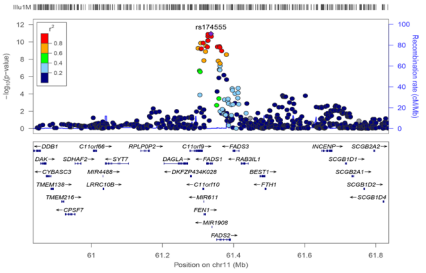


Supplementary Figure 13 Continued

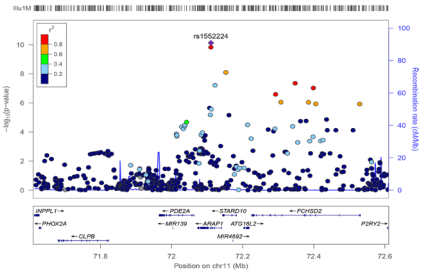
g



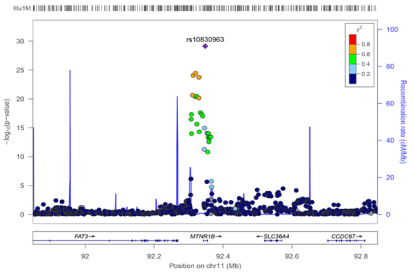
h



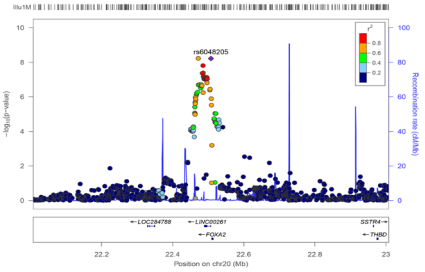
i



j

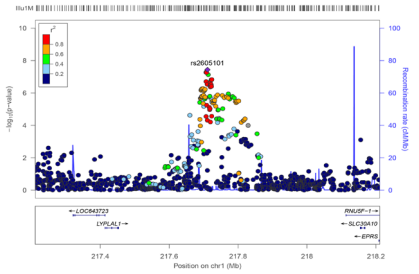


k

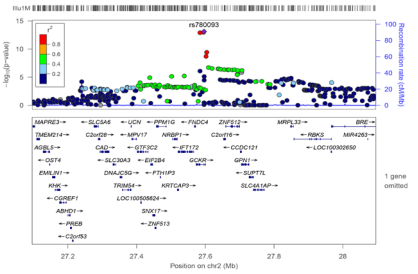


Supplementary Figure 14. Region plots for each of the leading SNPs in HOMA-IR GWIS results.

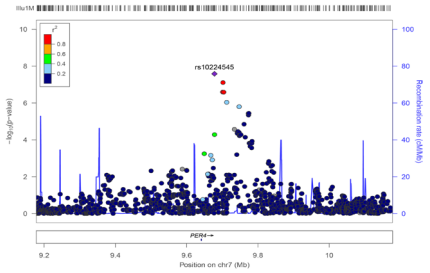
a



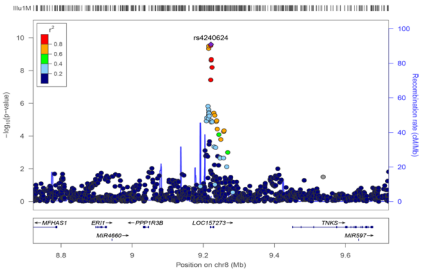
b



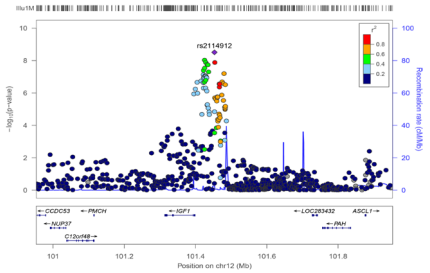
c



d



e



Supplementary Table 1. Genetic correlation (rg) between inferred and published HOMA-B and HOMA-IR, the LD score regression intercepts, SNP-heritability and their standard errors (SE).

	Previous	**Inferred Global Correction**	**Inferred Local Correction**
		HOMA-B	
r_g (SE), p-value	-	$1.12\,(0.07), 2.17 \times 10^{-59}$	$1.14\,(0.07), 2.22 \times 10^{-56}$
Intercept (SE)	0.99 (0.008)	1.044 (0.008)	1.05 (0.008)
h^2 (SE)	0.08 (0.015)	0.059 (0.01)	0.056 (0.01)
		HOMA-IR	
r_g (SE), p-value	-	$1.19\,(0.08), 2.21 \times 10^{-52}$	$1.18\,(0.08), 9.11 \times 10^{-53}$
Intercept (SE)	1.004 (0.007)	1.04 (0.008)	1.03 (0.008)
h^2 (SE)	0.07 (0.01)	0.058 (0.008)	0.059 (0.008)

Supplementary Table 2. Comparison of HOMA-B effects across genome-wide significant loci between inferred and published GWAS

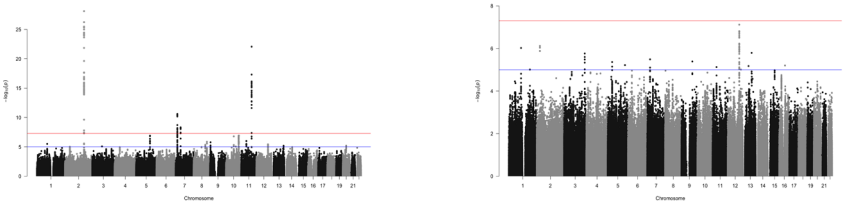
			HOMA-B inferred			HOMA-B published	
SNP	**CHR**	**Locus name**	**Alleles (effect / other)**	**Effect (SE), p-value**	**N**	**Alleles (effect / other)**	**Effect (SE), p-value**
rs560887	2	*G6PC2*	C/T	-0.050 (0.0033), 5.33×10^{-53}	75,173	T/C	0.040 (0.0036), 7.67×10^{-29}
rs11708067	3	*ADCY5*	A/G	-0.025 (0.0038), 3.93×10^{-11}	75,227	A/G	-0.016 (0.0043), 1.77×10^{-04}
rs10258074	7	*DGKB*	A/T	0.021 (0.0031), 1.07×10^{-11}	75,202	A/T	0.022 (0.0033), 2.99×10^{-11}
rs2908289	7	*GCK*	G/A	0.029 (0.0042), 1.57×10^{-11}	74,672	A/G	-0.026 (0.0046), 6.00×10^{-09}
rs3802177	8	*SLC30AB*	G/A	-0.023 (0.0036), 1.85×10^{-10}	68,044	A/G	0.016 (0.0038), 1.96×10^{-05}
rs7034200	9	*GLIS3*	A/C	-0.020 (0.0030), 6.20×10^{-11}	74,669	A/C	-0.016 (0.0033), 1.67×10^{-06}
rs7903146	10	*TCF7L2*	C/T	0.032 (0.0035), 6.19×10^{-20}	75,212	T/C	-0.020 (0.0038), 1.39×10^{-07}
rs10830963	11	MTNR1B	C/G	0.044 (0.0039), 7.28×10^{-30}	68,612	C/G	0.039 (0.0040), 8.60×10^{-23}
rs1552224	11	ARAP1	C/A	0.027 (0.0041), 7.45×10^{-11}	75,207	A/C	-0.017 (0.0043), 9.39×10^{-05}
rs174555	11	FADS1-2-3	T/C	-0.022 (0.0033), 1.15×10^{-11}	75,237	T/C	-0.015 (0.0034), 1.32×10^{-05}
rs6048205	20	FOXA2	G/A	0.043 (0.0073), 6.10×10^{-09}	70,262	A/G	-0.033 (0.0085), 1.16×10^{-04}

Supplementary Table 3. Comparison of HOMA-IR effects across genome-wide significant loci between inferred and published GWAS

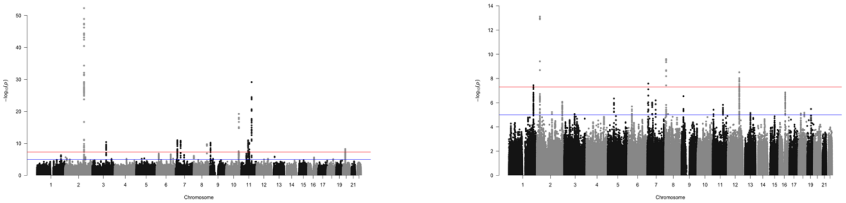
SNP	CHR	Locus name	HOMA-IR inferred			HOMA-IR published	
			Alleles (effect / other)	Effect (SE), p-value	N	Alleles (effect / other)	Effect (SE), p-value
rs2605101	1	*LYPLAL1/ SLC30A10*	A/T	0.019 (0.0034), 3.65×10^{-08}	75,238	A/T	0.016 (0.0042), 1.24×10^{-04}
rs780093	2	*GCKR*	C/T	0.024 (0.0032), 7.70×10^{-14}	75,226	T/C	-0.019 (0.0040), 1.31×10^{-06}
rs10224545	7	*PER4*	T/C	-0.031 (0.0056), 2.61×10^{-08}	71,951	T/C	-0.015 (0.0069), 2.51×10^{-02}
rs4240624	8	*PPP1R3B*	A/G	-0.035 (0.0056), 2.52×10^{-10}	68,617	A/G	-0.021 (0.0066), 1.66×10^{-03}
rs2114912	12	*IGF1*	T/G	0.026 (0.0044), 3.09×10^{-09}	74,634	T/G	0.026 (0.0053), 1.33×10^{-06}

HOMA-B HOMA-IR

Supplementary Figure 15. Manhattan plots of published meta-analysis results

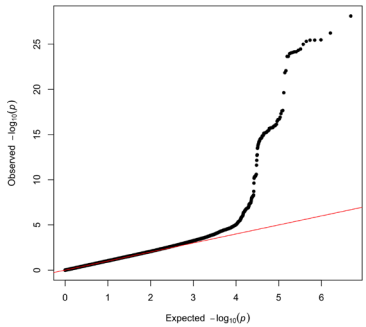


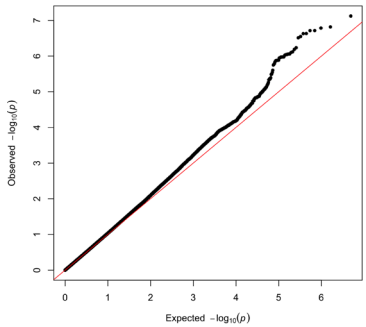
Supplementary Figure 16. Manhattan plots of inferred meta-analysis results



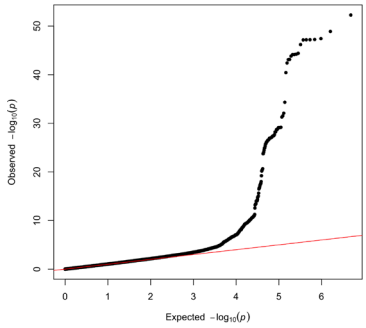
HOMA-B HOMA-IR

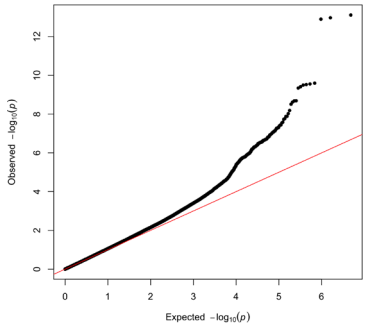
Supplementary Figure 17. QQ plots of published meta-analysis results





Supplementary Figure 18. QQ plots of inferred meta-analysis results





CHAPTER 7

DISSECTING THE ROLE OF GLUCOSE HOMEOSTASIS THROUGH MEASURES OF β-CELL FUNCTION AND INSULIN RESISTANCE IN SUSCEPTIBILITY TO MAJOR DEPRESSIVE DISORDER

This chapter is based on:

Iryna O. Fedko, Jouke-Jan Hottenga, Yuri Milaneschi, Reedik Mägi, Meike Bartels, Gonneke Willemsen, Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) Investigators, B.W.J.H. Penninx, Dorret I. Boomsma and Inga Prokopenko. Dissecting the role of glucose homeostasis through measures of beta-cell function and insulin resistance in susceptibility to Major Depressive Disorder *(as to be submitted).*

Abstract

Co-morbidity between type 2 Diabetes (T2D) and Major Depressive Disorder (MDD) is consistently observed, with epidemiological studies suggesting a bi-directional relationship between T2D and MDD. Glycaemic traits are the biomarkers of T2D and insulin resistance and have been associated with depression. The aim of this study was to evaluate the shared genetic effects on glycaemic trait and MDD risk. We employed Polygenic Risk Score (PRS) approach to predict MDD status in two Dutch cohorts (N cases = 1,687, N controls = 2,847), Netherlands Twin Register (NTR) and Netherlands Study of Depression and Anxiety (NESDA), using Fasting Insulin, Fasting Glucose, Homeostatic Model Assessment Insulin Resistance (HOMA-IR) and β-cell function (HOMA-B) PRS profiles, adjusted and not adjusted for BMI. Finally, we used LD score regression to estimate the overall genetic correlation between four fasting glycaemic traits, MDD, Depressive Symptoms and Neuroticism as MDD predictive factors. Results of PRS analyses indicated that glycaemic traits did not significantly predict MDD status (OR ≈ 1). This finding was confirmed by results of LD score regression, as we did not find a significant genetic correlation between T2D/FI/FG/HOMA-B/HOMA-IR and MDD, Depressive Symptoms and Neuroticism. Our results suggest that fasting glycaemic traits and MDD and its symptoms have distinct genetic aetiology. Comorbidity between T2D and MDD may be influenced by other than genetic shared external risk factors.

Introduction

Both, type 2 diabetes (T2D) and major depressive disorder (MDD) are complex diseases with high prevalence and an ever increasing burden on the society [31]. T2D and MDD are also co-morbid disorders [205]. Patients with T2D have higher prevalence of MDD than non-diabetic individuals (17.6% and 9.8% respectively) [206]. Although comorbidity between T2D and MDD is consistently reported, the pathophysiological relationships between them remain unclear and controversial [207]. Studies suggested that the bidirectional relationship between these two disorders involves a range of biological processes, such as hypothalamic-pituitary-adrenal axis (HPA) dysregulation, inflammatory response system, reduced neuroplasticity and metabolic abnormalities [208, 209]. Fasting glucose (FG) and fasting insulin (FI) levels are indicators of body glucose homeostasis and can be considered for the definition of Metabolic Syndrome or T2D. FG and FI levels are also used for defining homeostasis model assessment of β-cell function and insulin resistance traits (HOMA-B and HOMA-IR). Glycaemic trait variability in non-diabetic individuals is associated with risk of T2D, but has also been linked to MDD as compared to non-affected controls [210]. Previous research detected associations between insulin resistance and depression [30]. Genetic studies have identified over 70 loci associated with glycaemic traits and over 80 with T2D risk, and have also demonstrated that the overlap between T2D and glycemic trait loci is incomplete and the magnitude of effects in shared loci differ [180].

The main aim of this study was to investigate whether FG, FI, HOMA-B and HOMA-IR traits have a shared genetic background and are associated with MDD. In addition, we explored the genetic association of glycaemic traits with MDD subtypes identified according to direction of change (decreased vs. increased) in the symptoms of appetite and weight, which have previously been shown to have partially distinct, neuro-functional patterns [211] and polygenic signatures [212]. To evaluate the shared genetic component, we performed Polygenic Risk Score (PRS) profiling [213], using summary statistics from the Meta-Analysis of Glucose and Insulin-related traits Consortium (MAGIC) and the Psychiatric Genetic Consortium (PGC) to predict MDD in two population-based cohorts, the Netherlands Twin Registry (NTR) [43] and the Netherlands Study of Depression and Anxiety (NESDA) [214], with and without taking BMI into account. In addition, we applied LD score regression [28] to the summary statistics mentioned above and those based on the large Neuroticism and Depressive Symptoms meta-analysis from the Social Science Genetic Association Consortium (SSGA) and the DIAGRAM T2D consortium.

Methods

Discovery datasets

FI and FG summary statistics were available from the latest HapMap-based genome-wide association study (GWAS) meta-analysis of the MAGIC consortium (N ≈ 64K/88K for FI/FG, respectively) [215]. As required by the PRS approach, the meta-analyses have been

performed with NTR and NESDA cohorts excluded, compared to the respective primary study. Summary statistics from MDD GWAS mega-analysis (N cases ≈ 8K and N controls ≈ 8K) [216] were available from the PGC, and have been used previously, excluding NTR and NESDA as well [212]. In addition to checking the overall genetic correlation between glycaemic traits and MDD, summary statistics for proxy phenotypes, namely Depressive Symptoms (N ≈ 161K) and Neuroticism (N ≈ 171K), were downloaded from the Social Science Genetic Association Consortium website, which were reported in a recent paper by Okbay et al (2016) [168]. We used the GWAS meta-analysis summary statistics for T2D (N cases ≈ 12K and N controls ≈ 57K) in Europeans from the DIAGRAM consortium [197] web-site to estimate the genetic correlation with MDD, Depressive Symptoms and Neuroticism.

Target dataset.

The target dataset comprised of two Dutch population based cohorts, NTR and NESDA. NTR is the Netherlands Twin Register, which is an ongoing longitudinal study of health, personality and lifestyle of Dutch twins and their families [217]. NESDA is an longitudinal cohort study with participants recruited from general population, general practice and mental care organizations [214].

MDD cases were provided by NESDA (n = 1,687, mean age = 42.3, mean BMI = 25.8, women = 68.3%, n_{T2D} = 95). Controls were mostly provided by NTR (n = 2,505, mean age = 37,3, mean BMI = 24.2, women = 61.8%) and partly by NESDA (n = 342, mean age = 43.3, mean BMI = 25.3, women = 59.1%, n_{T2D} = 15). We excluded individuals with T1D and possible T2D (N = 121 and N = 9 in NTR and NESDA, respectively). If there were individuals with clearly defined T2D, but not T1D, they were left in the analysis. In NESDA DSM-IV lifetime diagnosis of MDD was assessed using the Composite Interview Diagnostic Instrument (CIDI) [218]. Assessment took place at baseline and/or one or more of the biannual follow-up interviews. MDD cases were stratified based on the direction of change of the symptoms of appetite and weight as previously described [212]: decreased appetite/weight (n = 645, mean age = 41.5, women = 61.5%) and increased appetite/weight (n = 424, mean age = 42.9, women = 68.5%). Fasting Glucose was available in both NTR (n = 3,813, mean age = 37.5, mean BMI = 24.4 (n_{BMI} = 3,790), women = 65.7%) and NESDA (n = 2,240, mean age = 42.4, mean BMI = 25.7, women = 65.8%). Fasting Insulin was available in NTR only (n = 3,737, mean age = 37.5, mean BMI = 24.4, women = 65.9%). From both datasets we selected individuals genotyped on the same platform (Affymetrix 600), which were imputed to 1000 Genome project Phase 1 v3 Mixed reference set using Michigan Imputation Server (see URLs).

Statistical Analysis

Polygenic Risk Score

SNPs previously reported in the literature to be associated with T2D and related traits, namely FG, FI, T2D, lipids, waist-to-hip ratio or with their overlap [180, 195, 204, 219,

220], were collected and collapsed into one SNP set. Because multiple SNPs were reported per locus, we applied the following criteria to select one independent SNP per locus. First, we selected the SNPs which were confirmed as significant at the threshold of p-value < 0.05 in the current MAGIC FI and FG meta-analyses and also have reported effect size in MDD mega-analysis. If summary statistics were missing in MDD dataset for a selected SNP, we chose a proxy SNP with LD > 0.8 with selected SNP (30 SNPs) using the SNP Annotation and Proxy Search (SNAP) [221]. Second, SNPs with the lowest p-value in discovery dataset and with imputation r^2 > 0.8 in target dataset were selected per locus. Third, if a glycaemic SNP was reported previously in literature for association with FI, FG or T2D, it was preferred over others. Finally, if application of these criteria did not result in one SNP per locus, then the selection was performed randomly. SNPs selected using the literature review were weighted by MDD effect sizes and formed a MDD SNP set (N = 134).

The other SNP sets were computed using FI/FG/HOMA-B/HOMA-IR summary statistics by clumping SNPs in the target dataset around index variants with p-values *0.00001, 0.001, 0.01,* and *1* using Plink 1.9 [179] command (--clump-p1 option with default r^2 threshold 0.50 and 250kb window) across all SNPs, independent of prior knowledge of association with glycaemic traits, T2D, lipids or waist-to-hip ratio. HOMA-IR and HOMA-B summary statistics were calculated using GWIS, a method [222] recently developed by Nieuwboer et al. 2016 [38], and reported by Fedko et al. (2016) [223] *(in preparation).* Here, the *'in silico'* GWAS can be performed and summary statistics of HOMA-B/-IR, which are the functions of FI and FG phenotypes [193], can be approximated from FI and FG summary statistics.

The PRS was calculated for each individual in the target sample (NTR and NESDA), by summing up alleles in each of PRS SNP sets, weighted by their effect sizes from the summary statistics of the discovery sample (FI, FG, HOMA-B, HOMA-IR, MDD) using Plink1.9 [179] software.

Prediction of FI and FG

We used the generated PRS profiles to perform a Generalized Estimating Equations (GEE) linear regression to predict levels of FG and FI in NTR and FG in NESDA. Sex, age, cohort (for FG), T2D status and Dutch population PCs [23] were used as covariates. PRS for all SNP sets were standardized. We included all available subjects from target samples in the analysis and corrected for relatedness using a sandwich correction of standard errors, which increases power, while correcting for inflation of statistics [224].

Prediction of MDD

We computed phenotypic correlation between FG and MDD, adjusted and not adjusted for BMI, in combined NTR/NESDA dataset, selecting unrelated individuals (1 person per family) controlling for age, sex, T2D status, cohort and Dutch population structure. FI measure was only present for NTR MDD controls and did not allow for comparison

between MDD cases and controls. With GEE logistic regression, we tested whether PRS constructed from the various SNPs sets of FG, FI, HOMA-IR, HOMA-B and MDD summary statistics can predict MDD status in NTR and NESDA cohorts. We standardized PRSs and included age, sex, T2D status and Dutch population PCs as covariates. In all analyses we included all available subjects from the target dataset and corrected for relatedness, as described above. In addition, we alternated between correcting for BMI or not, because obesity as assessed by BMI may partly explain the co-morbidity between T2D and MDD. Level of significance was set to 0.05/4 = 0.0125 to adjust for multiple testing (4 SNP sets per each glycaemic trait).

Finally, we performed the evaluation of the whole-genome shared variability using the method of LD score regression (LDSR) [28] and compared the LDSR analysis outcomes to the PRS estimates for their ability to help inferences about relationships between glycaemic and mood trait variability.

Results

PRS profiles clearly predicted the FI in the NTR and FG in the NTR and NESDA datasets, confirming the validity of selected SNP sets (Supplementary Material Table 1). However, we did not find a statistically significant phenotypic correlation between FG and MDD (r = -0.026, p-*value* = 0.143, N = 3,142 and r = -0.018, p-*value* = 0.305, N= 3,154 for adjusted/not adjusted for BMI, respectively). None of the glyceamic SNP sets, whether selected based on prior knowledge of association with T2D related traits and weighted by MDD effect sizes, or based on all SNPs clumped around four association significance thresholds using FI/FG/HOMA-B/HOMA-IR summary statistics, predicted the MDD status in NTR and NESDA (Figure 1, Table 1). Any adjustment for BMI did not have a large effect on the estimates. Results obtained from the analysis of MDD with increased and decreased appetite symptoms were non-significant (Figure 2, Table 2).

Table 1. Results of Polygenic Risk prediction of MDD from various subsets of SNPs with and without adjustment for BMI

	No BMI adjustment		With BMI adjustment	
SNP set	**OR**	**95% CI**	**OR**	**95% CI**
Glycaemic SNPs MDD weights	1.02	(0.97, 1.07)	1.02	(0.97, 1.07)
		Fasting Glucose SNP sets		
FG SNPs p-value < 0.00001	1.00	(0.95, 1.05)	1.00	(0.95, 1.05)
FG SNPs p-value < 0.001	1.01	(0.96, 1.06)	1.00	(0.95, 1.05)
FG SNPs p-value < 0.01	1.01	(0.96, 1.06)	1.00	(0.95, 1.05)
FG SNPs p-value < 1	1.00	(0.95, 1.05)	0.99	(0.94, 1.04)
		Fasting Insulin SNP sets		
FI SNPs p-value < 0.00001	1.02	(0.97, 1.08)	1.02	(0.97, 1.07)
FI SNPs p-value < 0.001	1.00	(0.95, 1.05)	0.99	(0.94, 1.04)
FI SNPs p-value < 0.01	1.01	(0.96, 1.06)	0.99	(0.95, 1.04)
FI SNPs p-value < 1	1.00	(0.95, 1.05)	0.99	(0.94, 1.04)
		HOMA-B SNP sets		
HOMA-B SNPs p-value < 0.00001	0.96	(0.92, 1.01)	0.96	(0.91, 1.01)
HOMA-B SNPs p-value < 0.001	0.99	(0.94, 1.04)	0.98	(0.93, 1.03)
HOMA-B SNPs p-value < 0.01	0.98	(0.93, 1.03)	0.97	(0.93, 1.02)
HOMA-B SNPs p-value < 1	0.99	(0.94, 1.04)	0.98	(0.93, 1.03)
		HOMA-IR SNP sets		
HOMA-IR SNPs p-value < 0.00001	1.01	(0.96, 1.07)	1.01	(0.96, 1.06)
HOMA-IR SNPs p-value < 0.001	1.01	(0.96, 1.06)	0.99	(0.94, 1.04)
HOMA-IR SNPs p-value < 0.01	1.01	(0.96, 1.06)	1.00	(0.95, 1.05)
HOMA-IR SNPs p-value < 1	0.99	(0.94, 1.05)	0.98	(0.93, 1.04)

Figure 1. Results of Polygenic Risk prediction of MDD from various subsets of SNPs with and without adjustment for BMI

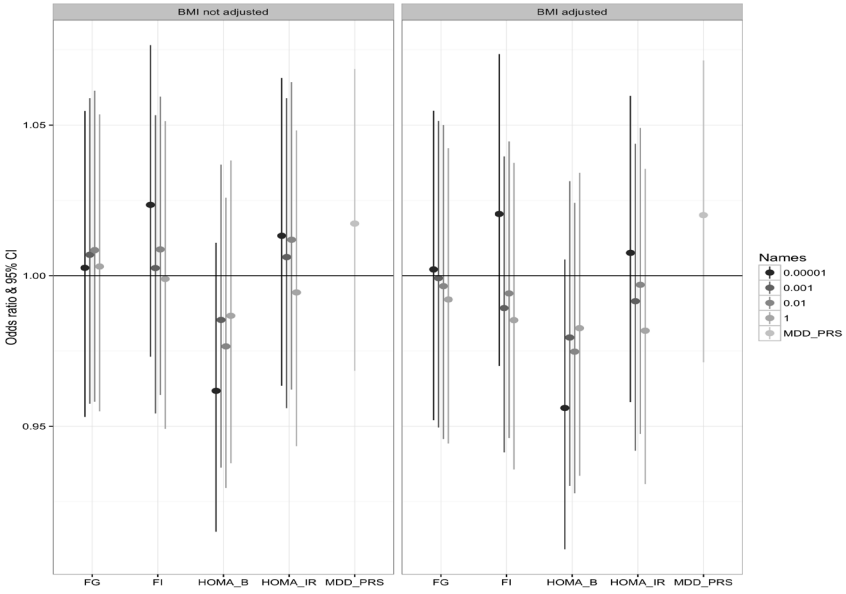
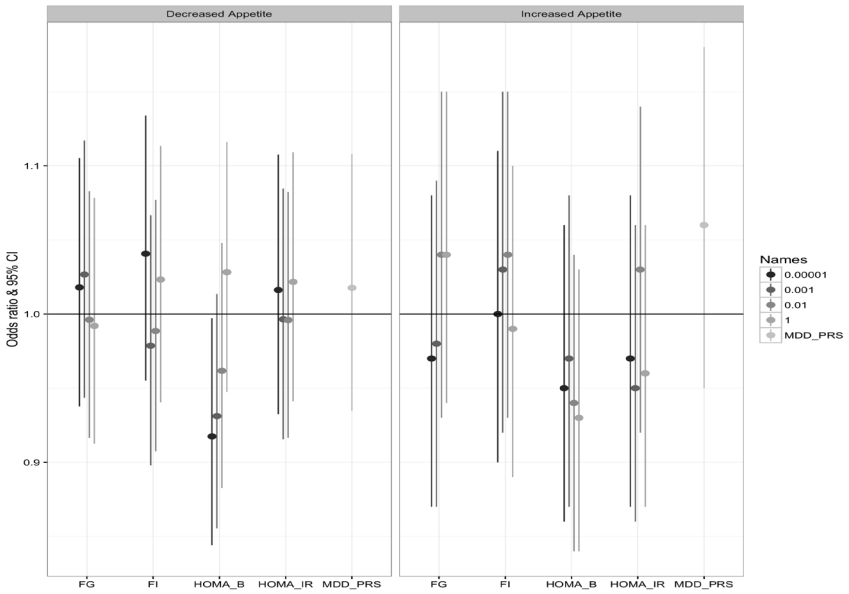


Table 2. Results of Polygenic Risk prediction of MDD with symptoms of increased and decreased appetite/weight from various subsets of SNPs

	Decreased BMI		Increased BMI	
SNP set	**OR**	**95% CI**	**OR**	**95% CI**
Glycaemic SNPs MDD weights	1.02	(0.93, 1.11)	1.06	(0.95, 1.18)
		Fasting Glucose SNP sets		
FG SNPs p-value < 0.00001	1.02	(0.94, 1.11)	0.97	(0.87, 1.08)
FG SNPs p-value < 0.001	1.03	(0.94, 1.12)	0.98	(0.87, 1.09)
FG SNPs p-value < 0.01	1.00	(0.92, 1.08)	1.04	(0.93, 1.15)
FG SNPs p-value < 1	0.99	(0.91, 1.08)	1.04	(0.94, 1.15)
		Fasting Insulin SNP sets		
FI SNPs p-value < 0.00001	1.04	(0.96, 1.13)	1.00	(0.90, 1.11)
FI SNPs p-value < 0.001	0.98	(0.90, 1.07)	1.03	(0.92, 1.15)
FI SNPs p-value < 0.01	0.99	(0.91, 1.08)	1.04	(0.93, 1.15)
FI SNPs p-value < 1	1.02	(0.94, 1.11)	0.99	(0.89, 1.10)
		HOMA-B SNP sets		
HOMA-B SNPs p-value < 0.00001	0.92	(0.84, 1.00)	0.95	(0.86, 1.06)
HOMA-B SNPs p-value < 0.001	0.93	(0.86, 1.01)	0.97	(0.87, 1.08)
HOMA-B SNPs p-value < 0.01	0.96	(0.88, 1.05)	0.94	(0.84, 1.04)
HOMA-B SNPs p-value < 1	1.03	(0.95, 1.12)	0.93	(0,84, 1.03)
		HOMA-IR SNP sets		
HOMA-IR SNPs p-value < 0.00001	1.02	(0.93, 1.11)	0.97	(0.87, 1.08)
HOMA-IR SNPs p-value < 0.001	1.00	(0.92, 1.08)	0.95	(0.86, 1.06)
HOMA-IR SNPs p-value < 0.01	1.00	(0.92, 1.08)	1.03	(0.92, 1.14)
HOMA-IR SNPs p-value < 1	1.02	(0.94, 1.11)	0.96	(0.87, 1.06)

Figure 2. Results of Polygenic Risk prediction of MDD with symptoms of increased and decreased appetite/weight from various subsets of SNPs



Results of the LD score regression showed no significant genetic correlations between FI/FG/HOMA-IR/HOMA-B/T2D and MDD (Table 3). A small, however, non-significant (after correction for multiple testing) genetic correlation was detected between FI/HOMA-IR and Depressive Symptoms (r_g = 0.17/0.16 respectively, p-value = 0.02 for both traits); and between FI/HOMA-IR and Neuroticism (r_g = 0.11, p-value = 0.08/0.07 respectively).

Table 3. LDScore regression results of genetic correlation (r_g) between mood and glycaemic traits.

		MDD	***Depressive Symptoms***	***Neuroticism***
		SNP-h²= 0.15 (0.03)	***SNP-h²= 0.05 (0.005)***	***SNP-h2²= 0.09 (0.008)***
	SNP-h²	***r_g (SE, p-value)***	***r_g (SE, p-value)***	***r_g (SE, p-value)***
FI	0.06 (0.01)	-0.07 (0.11, 0.5)	0.17 (0.07, 0.02)	0.11 (0.06, 0.08)
FG	0.09 (0.02)	0.07 (0.09, 0.41)	0.04 (0.05, 0.44)	0.02 (0.04, 0.64)
HOMA-B	0.05 (0.01)	-0.11 (0.10, 0.27)	0.12 (0.07, 0.07)	0.09 (0.06, 0.14)
HOMA-IR	0.06 (0.01)	-0.05 (0.11, 0.66)	0.16 (0.07, 0.02)	0.11 (0.06, 0.07)
T2D	0.09 (0.01)	0.01 (0.10, 0.89)	0.06 (0.07, 0.35)	-0.04 (0.06, 0.45)

Note: *SNP-h²* denotes the SNP-heritability

Discussion

Following various independent reports from epidemiological studies on the bidirectional relationships between T2D and MDD [208, 209] accompanied by evidence supporting the role of insulin resistance [30] in association with both T2D and MDD, we explored whether the risk variants of glycaemic traits such as FI, FG, HOMA-IR, HOMA-B share genetic risk factors with MDD. In addition, we explored the shared genetic aetiology with phenotypes known to be proxies to MDD, such as Depressive Symptoms and Neuroticism. None of the PRS SNP sets crossed the threshold of significance for association with MDD after correction for multiple testing.

To our knowledge there are only a few studies that explored the shared aetiology between glycaemic traits and MDD based on SNPs. Bulik-Sullivan et al. (2015) detected non-significant genetic correlations between T2D, FG and MDD [198] as well as Lubke et al (2012) between FG and MDD [40] and Samaan et al (2014) between impaired fasting glucose/impaired glucose tolerance/T2D/dysglycemia and MDD [207]. For T2D and Neuroticism, the recent paper by Gale et al (2016) [225] reported the non-significant genetic correlation in the UK biobank sample size of 108,038 individuals after correction for multiple testing. The only study that detected genetic correlation between MDD and T2D to date, was the recent study in Swedish and Danish twins [226]. Twin-based genetic correlations were reported to be significant in Swedish women (r_g = 0.23) and Danish twins, irrespective of sex (r_g = 0.25 in men and r_g = 0.18 in women). Authors also reported qualitative sex differences in the comorbidity of T2D and MDD, i.e. different genetic risk factors can operate in men and women [226].

When SNP-heritability (SNP-h2) of Fasting Glucose and BMI was partitioned across specific cell type groups [227], the FG SNP-h2 was significantly enriched with adrenal or pancreas cell type groups, whereas BMI SNP-h2 was significantly enriched with cell type group of Central Nervous System (CNS), suggesting BMI as the potential mediator of the relationship between Insulin Resistance and T2D with MDD in epidemiological studies. Negligible enrichment of CNS cell type group in Fasting Glucose SNP-heritability may explain the non-significant genetic correlation between T2D and MDD and their related traits in current and few previous genetic studies. In a study of Danish and Swedish twins, genetic correlation has been reported between T2D and MDD cases and controls and could be due to elevated BMI in T2D cases. In contrast, in our study we corrected for T2D status. Because T2D correlates with BMI as well as FI/FG/HOMA-B/HOMA-IR [198, 223], we therefore possibly corrected for some variation due to BMI.

BMI is a biomarker of T2D and also has been associated with MDD [228]. In our study BMI did not change the results of MDD prediction for any of the PRS profiles. This is probably because most of the sample comprised of a non-diabetic population. Therefore most of the sample has a BMI, fasting insulin, fasting glucose levels and metabolism processes in the normal range. A recent report suggested that introducing a high-fat diet leads to T1D/T2D in mice through apoptosis of β-cells [229]. Unhealthy diets have

been reported to be a risk factor for depression [230], whereas a healthy diet in turn was associated with reduced risk of depressive symptoms in both T2D and non-diabetic individuals [231]. Another possibility could be that, for instance, depression behavioral consequences, which include also poor diet, reduced physical activity, increased alcohol consumption and smoking may determine an increased risk for the development of diabetes-related alterations.

No phenotypic correlation was observed in our study between FG and combined MDD sample, although previous work in other samples has indicated a relation between T2D and MDD [226]. Previous studies started from a sample of T2D patients, whereas we combined a MDD patient cohort (NESDA) and a population based cohort (NTR). Previous NESDA studies also did not observe a phenotypic correlation between FG and MDD and its subtypes (atypical/melancholic, which roughly corresponds to MDD with increased and decreased appetite/weight symptoms) [40, 232]. However, the current study was motivated by the evidence for association between Insulin Resistance and MDD [30]. FI/FG/HOMA-IR/HOMA-B loci differ in their effect on pathophysiology of T2D [199] and so it could be in pathopysiology of MDD. A recent study in the prospective cohort in the United States (US) reported somatic-vegetative depressive symptoms as a predictor of deteriorating insulin resistance and therefore risk of T2D development in adults aged 50-70 through increasing BMI [233]. Thus PRS profiles constructed using summary statistics from largest up to date MAGIC meta-analysis could have pinpointed the biological pathways not detected previously.

In conclusion, our results suggest that glycaemic traits, namely Fasting Glucose, Fasting Insulin, indices of Insulin resistance (HOMA-IR) and β-cell function (HOMA-B) have distinct genetic aetiology with MDD and its symptoms. Comorbidity between T2D and MDD may be influenced by other shared environmental risk factors, such as diet, smoking or other demographic and socio-economic factors, suggesting possible intervention and warrants further research [234].

URLs: Michigan Imputation Server: https://imputationserver.sph.umich.edu/index.html

Supplementary Material.

Supplementary Table 1. Results of Polygenic Risk prediction of Insulin from various subsets of SNPs

	Fasting Insulin			Fasting Glucose		
	Beta	**SE**	**P-value**	**Beta**	**SE**	**P-value**
Glycaemic SNPs with MDD weights	-0.10	0.09	0.31	1.3×10^{-03}	0.01	0.88
			Fasting Glucose SNP sets			
FG SNPs p-value < 0.00001	0.09	0.09	0.37	0.09	0.01	1.3×10^{-27}
FG SNPs p-value < 0.001	0.06	0.10	0.53	0.08	0.01	3.3×10^{-23}
FG SNPs p-value < 0.01	0.02	0.10	0.82	0.07	0.01	1.5×10^{-18}
FG SNPs p-value < 1	0.08	0.10	0.39	0.06	0.01	4.4×10^{-12}
			Fasting Insulin SNP sets			
FI SNPs p-value < 0.00001	0.21	0.11	0.05	0.01	0.01	0.24
FI SNPs p-value < 0.001	0.27	0.12	0.02	-5.9×10^{-04}	0.01	0.94
FI SNPs p-value < 0.01	0.29	0.12	0.01	2.6×10^{-03}	0.01	0.76
FI SNPs p-value < 1	0.50	0.11	5.5×10^{-06}	0.02	0.01	0.04
			HOMA-B SNP sets			
HOMA-B SNPs p-value < 0.00001	0.11	0.11	0.34	-0.06	0.01	1.8×10^{-13}
HOMA-B SNPs p-value < 0.001	0.22	0.11	0.04	-0.05	0.01	1.2×10^{-10}
HOMA-B SNPs p-value < 0.01	0.30	0.13	0.02	-0.04	0.01	9.3×10^{-07}
HOMA-B SNPs p-value < 1	0.50	0.11	4.7×10^{-06}	-0.02	0.01	0.05
			HOMA-IR SNP sets			
HOMA-IR SNPs p-value < 0.00001	0.30	0.11	0.01	0.03	0.01	2.2×10^{-05}
HOMA-IR SNPs p-value < 0.001	0.43	0.13	8.3×10^{-04}	0.02	0.01	0.01
HOMA-IR SNPs p-value < 0.01	0.51	0.12	1.5×10^{-05}	0.03	0.01	1.1×10^{-03}
HOMA-IR SNPs p-value < 1	0.58	0.11	1.2×10^{-07}	0.03	0.01	1.7×10^{-03}

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 ≈ 2% drop in the SNP-heritability estimates for each combination approach. Correction for the Dutch PCs alone resulted in ≈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.

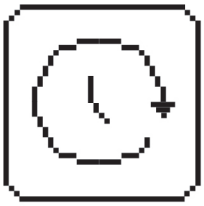
Following the results of chapter 3 and for the sake of comparison between heritability estimates, resulting from different raters' perspectives, we explored the rater shared and unique contribution to the variation of the child behavior problems in chapter 4. We estimated the heritability of maternal 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 ≈ 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.



CHAPTER 9

GENERAL DISCUSSION

This thesis encompasses studies on three partly overlapping themes. *The first theme* concerns the imputation against the GoNL reference set and its ability to align separate cohorts genotyped on different platforms, to one set of SNPs based on LD patterns from the same ethnicity. Combining datasets on the genotype level is crucial to increase the sample size and estimate the SNP-heritability for behavior phenotypes. Datasets, combined based on the GoNL reference set, were used to estimate the SNP-heritability for child behavior problems. *The second theme* concerns the heritability comparisons between different raters, as a follow-up of results of SNP-based heritability estimates from the first section. When different raters assess behavior, it may affect heritability estimates, as raters may partly assess the same and partly different aspects of the phenotype. Loci, whose effects contribute to the phenotype variation, shared between raters can be potentially detected in molecular studies if ratings are combined. Likewise, if two generally distinct phenotypes genetically correlate to a large extend, the loci, with effects that are correlated, can be detected in future GWAS studies, if such phenotypes are used in combination or as a proxy. Therefore, the genetic correlations between SWB and personality traits NEU/EXT were explored, as well as between glycaemic traits and MDD. For the latter the HOMA-B and HOMA-IR summary statistics were inferred from FI and FG meta-analysis results. This was the focus of *theme three*: implementation and evaluation of the performance of a new developed method (GWIS) and further exploration of the effect of HOMA-B/-IR loci on FI and FG, in which variation are important risk factors for T2D.

Imputation as a tool to combine the genotype data for estimation of genetic relatedness

In GWAS, imputation is the widely used approach to overcome the platform stratification to combine the summary statistics from each cohort in further meta-analysis. If combining the genotypic data is required, the overlapping SNP will be selected. If the SNP overlap is small, I showed that cross-platform imputation with the reference set from the same ethnicity reduces the stratification bias in combined data. Note, that we imputed only SNPs absent either on one or another platform, aiming to fill in the gaps. Imputation to a large reference set, i.e. 1000G can be done in the next stage, when all individuals 'genotyping' rate (> 99%) is appropriate according to quality control procedures. GoNL showed the ability to impute the rare variants better [18] and Dutch specific SNPs associated with cholesterol levels were detected in a recent study [235]. Improvement in overall imputation accuracy was reported for the homogeneous founder Sardinian population, when imputation was performed against local reference sets [236]. More accurate imputation of the rare variants is possible due to closer relationships to the common ancestor, therefore, the LD is stronger between SNPs and rare variants that are tagged better. The feature of stronger LD was used in the current study to combine the datasets, genotyped on different platforms. We also estimated the SNP-heritability of childhood height based on combined and cross-platform imputed data. The estimates

were similar and the results suggested that even a small number of SNPs could reflect the relationships between individuals, although SNP-heritability for combined sample was slightly lower, thus possibly underestimating the SNP-heritability.

Twin- and SNP-heritability across raters and their effect on expected variation to be explained by SNPs in molecular studies

For heritable traits, such as height, an underestimation of the relationships in GRM maybe not be as critical for behavioral phenotypes, such as childhood behavior problems. A study of Trzaskowski et al (2013) in the TEDS cohort in the UK, did not detect genetic influences for a range of child behavior problems [237], whereas the NTR-GENR study detected the significant SNP-heritability for attention deficit/hyperactivity, externalizing problems and total problems, rated by mother or teacher. One of the reasons for this discrepancy can be the age of the children. Genetic influences have been described to change with age as well as environmental influences. In TEDS children were aged 12 years in comparison to NTR and GENR children of three, seven and nine years old. As heritability is the proportion of the total phenotypic variance, the change in environment may modify the heritability estimate. The SNP-heritability estimates differed across raters in our study and exploration of rater effect on heritability estimates was performed in the next chapter of this thesis.

It has been shown that heritability can be rater dependent and we compared the maternal and paternal contribution to the variation in the childhood behavior problems in chapter 4. In general, the heritability of the phenotype that parents agree upon was large for all of the empirical scales in CBCL 6-18, suggesting the possibility to combine the paternal and maternal ratings in molecular genetic studies. Our results also showed that mothers rated their children higher than fathers and inclusion of the appropriate covariate is necessary if the ratings are combined. The difference in the rater's assessment partly explains the difference in SNP-heritability estimates, based on maternal and teachers ratings, described in the chapter 3. The loci, which effects are shared between both parents' assessments, are likely to be detected in these settings; however, loci that contribute to the variability of the specific maternal and paternal part of the phenotype may not be detected. It also has been suggested that rater-specific loci may not be possible to detect in a GWAS of a single phenotype, because in GWAS raters assess their children in a different context [35].

Shared aetiology of comorbid traits may guide future molecular study designs

As shared genetic factors may influence the two assessments of mother and father, they can influence two different, but correlated phenotypes. In this case the focus on one trait may lead to the discovery in another trait, as was recently shown in a meta-analysis of Subjective Well-being, Depressive Symptoms and Neuroticism [168]. The loci detected to be associated for SWB were taken for the follow up analysis for DS and NEU and resulted in new detected loci. The results described in chapter 5 showed substantial

genetic correlation between SWB and NEU, but not between SWB and EXT. Therefore, the focus on the SWB-NEU pair will likely be more successful, than focus on SWB and EXT.

SWB and personality are behavioral phenotypes, which generally are more likely to share the genetic influences. In contrast, in chapter 7 we aimed to estimate the shared risk factor between psychiatric phenotype and disease, MDD and T2D through the measures of glycaemic traits, FI, FG, HOMA-B/-IR. MDD status was not predicted from polygenic scores. LDscore regression also did not detect a significant genetic correlation, using meta-analysis results with the largest sample size to date. Although comordibity between T2D and MDD has been reported as well as association with Insulin Resistance, the mechanism behind it remains unclear [238]. The absence of overlapping genetic risk factors suggests that other influences play a role, including environmental exposures or lifestyle. Among these, BMI seems to be a plausible candidate for future studies in relation to T2D and MDD. Both increased and decreased BMI is associated with MDD [228] and relationships are not linear. Therefore, it is also possible that complex mechanisms behind the association of T2D and MDD are difficult to pinpoint when applying linear models. Another possible explanation lies in cultural differences. This study was conducted in a Dutch population, which in general has a lower BMI, than in US or UK (World Health Organization, Global Database on Body Mass Index, http://apps.who.int/bmi/). It does not explain, however, a lack of association between glycaemic traits and MDD/Depressive symptoms in a LD score regression analysis, where data from international collaborations were used. Power would be one of the possible explanations, as for glacaemic traits and T2D, there were a number of loci detected, whereas MDD is still lacking statistically significant hits. Lack of association should not hinder work in that direction, but rather should encourage a reconsideration of the approach by increasing sample size, using proxy phenotypes (Depressive Symptoms and Neuroticism) and MDD subtypes, including environmental factors and exploring non-linear relationships between T2D and MDD.

Leveraging existing data to get new insight into pathophysiology of complex diseases

The exploration between MDD and Insulin resistance and β-cell function would not be possible without computation of HOMA-B and HOMA-IR meta-analysis summary statistics of sufficient sample size to detect new loci. These results were used as weight in PRS and in LD score regression. The newly developed method GWIS (Genome Wide Inferred Statistics) was used to infer the summary statistics of the complex non-linear functions (HOMA-B/-IR) from FI and FG. Insulin resistance and β-cell function are both biomarkers for T2D diabetes. Both are difficult to compute as FI and FG should be available in the same person, and also at the same time point. GWIS allows to overcome this obstacle, as any function can be computed from the statistics of its components for overlapped and also for not overlapped sample. This feature may pave the way for the future research, where obtaining measurements from the same individuals are required, but not always possible.

Implications for the future studies and conclusion

In this thesis a combination of several new methods and approaches were applied to optimize the analysis of gene-phenotype data to facilitate research on the genetic influences on human complex traits, including psychiatric and somatic diseases.

As data are usually collected cross-sectionally at different time points, they may have been genotyped on a series of different genotyping platforms. When genotyped data are available from earlier arrays and need to be combined, but re-genotyping may not be possible, data need to be cross-platform imputed to a local reference set. This local reference set, describing the variation within the boundaries of a relatively homogeneous population, will allow characterization of the migration history, detection of rare variants and structural variations and investigation of the subtle differences in allele frequencies and population substructure with a greater resolution than an international reference set [18, 23, 239, 240]. All this information will form the solid foundation for the imputation into existing SNP-based genotype data. While whole-genome sequencing is thought to be the next 'big thing' in genetic research, many research questions could exploit the existing SNP-based genotype data far more time and cost-efficiently. This requires the use of imputation as a tool to infer the missing genotypes, increase the coverage, or combine the data.

There is a vast amount of GWAS summary statistics available from large international collaborations. Over the last few years a range of methods emerged, for example LD score regression or Genome Wide Inferred Statistics (GWIS), which aimed to gain biological insights from the publicly available summary statistics data without the (privacy-sensitive) disclosure of raw genotyped data [28, 222]. Other techniques use summary statistics across many studies in combination with the raw genotype data from a target study to apply e.g. Polygenic Risk Score [241] or LD pred [242]. These approaches often rely on the LD information available from reference sets representative of the population. Here too, it can be beneficial to employ LD information from local reference set in addition to an international set.

In this thesis, I utilised the broader perspective on complex human traits. For complex biological traits, different (endo)phenotypes and for complex behavioral traits, different raters as well as different (endo)phenotypes can be used jointly to detect the genetic loci common to these different (endo)phenotypes. Exploring such shared genetic aetiology may lead to an increase in statistical power, but also can lead to important conclusions about non-genetic, i.e. environmental risk factors' role and serve as a starting point for future studies detecting biological and environmental risk factors. This will, in the long run, pave the way for preventions and interventions at both levels.

Unique genetic influences, that is, influences pertaining to one trait, but not to another, largely remain to be discovered. The effect of the loci, which contribute to the specific part of the phenotype, can be explored in the multivariate twin model, involving correlated phenotypes. Genetic information can be expressed in various forms, such as SNPs, SNPs from a gene or a genomic region, PRS or gene expression profile. An

interesting addition, for both shared and specific parts of the phenotype, would be the longitudinal modeling as genetic influences changes with age, and the effect of loci may also change across time. Models incorporating environmental and gene×environment effects could be employed to explain the observed association between T2D and MDD as well as between SWB and EXT through genetic and environmental risk factors. It has been suggested that part of the detected GWAS loci could reflect the modifiable effect of the environmental exposure or certain behavior and could be explored using Mendelian randomization approach [234].

With more and more genetic variants for various phenotypes being discovered, the next step would be to investigate the pathophysiology of the diseases by bringing the genetic and environmental variables in the same model. Here, the twin data can be utilized together with information about genetic variation [243]. Increasing data collection in biobanks (metabolic, proteomic, etc.) allows exploration of the genetic effect on complex (disease) traits through these intermediate biological phenotypes [217]. The ongoing characterization of the genome expands the genomic databases not only in size but also in the type of genetic variation available [244]. Other types of variation, rather than SNPs exclusively, should be considered in the future, such as copy-number-variation (CNVs) [245].

To conclude, the future perspective is one in which genomic, biological, behavioral, and environmental data are combined to explore the aetiology of complex traits.

BIBLIOGRAPHY

1 Feero, W.G., A.E. Guttmacher, and T.A. Manolio, *Genomewide association studies and assessment of the risk of disease.* New England Journal of Medicine, 2010. **363**(2): p. 166-176.

2 Browning, S.R., *Missing data imputation and haplotype phase inference for genome-wide association studies.* Human genetics, 2008. **124**(5): p. 439-450.

3 Li, Y., et al., *Genotype imputation*. Annual review of genomics and human genetics, 2009. **10**: p. 387.

4 Marchini, J. and B. Howie, *Genotype imputation for genome-wide association studies.* Nature Reviews Genetics, 2010. **11**(7): p. 499-511.

5 Anderson, C.A., et al., *Evaluating the effects of imputation on the power, coverage, and cost efficiency of genome-wide SNP platforms.* The American Journal of Human Genetics, 2008. **83**(1): p. 112-119.

6 Howie, B.N., P. Donnelly, and J. Marchini, *A flexible and accurate genotype imputation method for the next generation of genome-wide association studies.* PLoS Genet, 2009. **5**(6): p. e1000529.

7 Li, Y., et al., *MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes.* Genetic epidemiology, 2010. **34**(8): p. 816-834.

8 de Bakker, P.I., et al., *Practical aspects of imputation-driven meta-analysis of genome-wide association studies.* Human molecular genetics, 2008. **17**(R2): p. R122-R128.

9 Spencer, C.C., et al., *Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip.* PLoS Genet, 2009. **5**(5): p. e1000477.

10 Roshyara, N.R. and M. Scholz, *Impact of genetic similarity on imputation accuracy.* BMC genetics, 2015. **16**(1): p. 1.

11 Burdick, J.T., et al., *In silico method for inferring genotypes in pedigrees.* Nature genetics, 2006. **38**(9): p. 1002-1004.

12 Frazer, K.A., et al., *A second generation human haplotype map of over 3.1 million SNPs.* Nature, 2007. **449**(7164): p. 851-861.

13 Consortium, I.H., *Integrating common and rare genetic variation in diverse human populations.* Nature, 2010. **467**(7311): p. 52-58.

14 Consortium, G.P., *A map of human genome variation from population-scale sequencing.* Nature, 2010. **467**(7319): p. 1061-1073.

15 Consortium, G.P., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.

16 Boomsma, D.I., et al., *The Genome of the Netherlands: design, and project goals.* European Journal of Human Genetics, 2014. **22**(2): p. 221-227.

17 McEvoy, B.P., et al., *Geographical structure and differential natural selection among North European populations.* Genome research, 2009. **19**(5): p. 804-814.

18 Deelen, P., et al., *Improved imputation quality of low-frequency and rare variants in European samples using the 'Genome of The Netherlands'.* European Journal of Human Genetics, 2014. **22**(11): p. 1321-1326.

19 Benjamin, D.J., et al., *The genetic architecture of economic and political preferences*. Proceedings of the National Academy of Sciences, 2012. **109**(21): p. 8026-8031.

20 Yang, J., et al., *GCTA: a tool for genome-wide complex trait analysis.* The American Journal of Human Genetics, 2011. **88**(1): p. 76-82.

21 Yang, J., et al., *Common SNPs explain a large proportion of the heritability for human height.* Nature genetics, 2010. **42**(7): p. 565-569.

22 Novembre, J., et al., *Genes mirror geography within Europe.* Nature, 2008. **456**(7218): p. 98-101.

23 Abdellaoui, A., et al., *Population structure, migration, and diversifying selection in the Netherlands.* European Journal of Human Genetics, 2013. **21**(11): p. 1277-1285.

24 Price, A.L., et al., *Principal components analysis corrects for stratification in genome-wide association studies.* Nature genetics, 2006. **38**(8): p. 904-909.

25 Falconer, D. and T. Mackay, *Introduction to Q uantitative Genetics.* Longman, 1995. **19**(8): p. 1.

26 Polderman, T.J., et al., *Meta-analysis of the heritability of human traits based on fifty years of twin studies.* Nature genetics, 2015.

27 Neale, M. and L. Cardon, *Methodology for genetic studies of twins and families.* Vol. 67. 1992: Springer Science & Business Media.

28 Bulik-Sullivan, B.K., et al., *LD Score regression distinguishes confounding from polygenicity in genome-wide association studies.* Nature genetics, 2015. **47**(3): p. 291-295.

29 Nouwen, A., et al., *Type 2 diabetes mellitus as a risk factor for the onset of depression: a systematic review and meta-analysis.* Diabetologia, 2010. **53**(12): p. 2480-2486.

30 Kan, C., et al., *A systematic review and meta-analysis of the association between depression and insulin resistance.* Diabetes care, 2013. **36**(2): p. 480-489.

31 Egede, L.E. and C. Ellis, *Diabetes and depression: global perspectives.* Diabetes research and clinical practice, 2010. **87**(3): p. 302-312.

32 DeNeve, K.M. and H. Cooper, T*he happy personality: a meta-analysis of 137 personality traits and subjective well-being.* Psychological bulletin, 1998. **124**(2): p. 197.

33 Weiss, A., T.C. Bates, and M. Luciano, *Happiness is a personal (ity) thing the genetics of personality and well-being in a representative sample.* Psychological Science, 2008. **19**(3): p. 205-210.

34 Bartels, M., et al., *Twins and the study of rater (dis) agreement.* Psychological Methods, 2007. **12**(4): p. 451.

35 Kan, K.-J., et al., *Assessing genetic influences on behavior: informant and context dependency as illustrated by the analysis of attention problems.* Behavior genetics, 2014. **44**(4): p. 326-336.

36 Hewitt, J.K., et al., *The analysis of parental ratings of children's behavior using LISREL.* Behavior genetics, 1992. **22**(3): p. 293-317.

37 Zaitlen, N., et al., *Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits.* PLoS Genet, 2013. **9**(5): p. e1003520.

38 Nieuwboer, H., et al., *GWIS: Genome Wide Inferred Statistics for non-linear functions of multiple phenotypes.* bioRxiv, 2015: p. 035329.

39 Visscher, P.M., J. Yang, and M.E. Goddard, *A commentary on 'common SNPs explain a large proportion of the heritability for human height'by Yang et al.(2010)*. Twin Research and Human Genetics, 2010. **13**(06): p. 517-524.

40 Lubke, G.H., et al., *Estimating the genetic variance of major depressive disorder due to all single nucleotide polymorphisms.* Biological psychiatry, 2012. **72**(8): p. 707-709.

41 Lee, S.H., et al., *Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood.* Bioinformatics, 2012. **28**(19): p. 2540-2542.

42 Zaitlen, N., et al., *Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits.* PLoS genetics, 2013. **9**(5): p. e1003520.

43 Boomsma, D.I., et al., *Netherlands Twin Register: from twins to twin families.* Twin Research and Human Genetics, 2006. **9**(06): p. 849-857.

44 van Beijsterveldt, C.E., et al., *The Young Netherlands Twin Register (YNTR): longitudinal twin and family studies in over 70,000 children.* Twin Research and Human Genetics, 2013. **16**(01): p. 252-267.

45 Tiemeier, H., et al., *The Generation R Study: a review of design, findings to date, and a study of the 5-HTTLPR by environmental interaction from fetal life onward.* Journal of the American Academy of Child & Adolescent Psychiatry, 2012. **51**(11): p. 1119-1135. e7.

46 Jaddoe, V.W., et al., *The Generation R Study: design and cohort update 2012.* European journal of epidemiology, 2012. **27**(9): p. 739-756.

47 Liu, E.Y., et al., *MaCH-Admix: Genotype Imputation for Admixed Populations.* Genetic epidemiology, 2013. **37**(1): p. 25-37.

48 Deelen, P., et al., *Improved imputation quality of low-frequency and rare variants in European samples using the 'Genome of The Netherlands'.* European Journal of Human Genetics, 2014.

49 Silventoinen, K., et al., *Genetic regulation of growth in height and weight from 3 to 12 years of age: a longitudinal study of Dutch twin children.* Twin Research and Human Genetics, 2007. **10**(02): p. 354-363.

50 Boomsma, D., J. Orlebeke, and G. Van Baal, *The Dutch Twin Register: Growth data on weight and height.* Behavior genetics, 1992. **22**(2): p. 247-251.

51 Scheet, P., et al., *Twins, tissue, and time: an assessment of SNPs and CNVs.* Twin Research and Human Genetics, 2012. **15**(06): p. 737-745.

52 Medina-Gomez, C., et al., *Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the Generation R Study.* European journal of epidemiology, 2015. **30**(4): p. 317-330.

53 Benke, K.S., et al., *A Genome-wide Association Meta-analysis of Preschool Internalizing Problems.* Journal of the American Academy of Child & Adolescent Psychiatry, 2014. **53**(6): p. 667-676. e7.

54 Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* The American Journal of Human Genetics, 2007. **81**(3): p. 559-575.

55 Danecek, P., et al., *The variant call format and VCFtools.* Bioinformatics, 2011. **27**(15): p. 2156-2158.

56 Price, A.L., et al., *Long-range LD can confound genome scans in admixed populations.* American journal of human genetics, 2008. **83**(1): p. 132.

57 Team, R.C., *Vienna (Austria): R foundation for statistical computing; 2012.* R: A language and environment for statistical computing, 2013.

58 Silventoinen, K., et al., *Heritability of adult body height: a comparative study of twin cohorts in eight countries.* Twin research, 2003. **6**(05): p. 399-408.

59 Wood, A.R., et al., *Defining the role of common variation in the genomic and biological architecture of adult human height.* Nature genetics, 2014. **46**(11): p. 1173-1186.

60 Mook-Kanamori, D.O., et al., *Heritability estimates of body size in fetal life and early childhood.* PloS one, 2012. **7**(7): p. e39901.

61 Visscher, P.M., et al., *Statistical power to detect genetic (co) variance of complex traits using SNP data in unrelated samples.* PLoS genetics, 2014. **10**(4): p. e1004269.

62 de Candia, T.R., et al., *Additive genetic variation in schizophrenia risk is shared by populations of African and European descent.* The American Journal of Human Genetics, 2013. **93**(3): p. 463-470.

63 So, H.C., M. Li, and P.C. Sham, *Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study.* Genetic epidemiology, 2011. **35**(6): p. 447-456.

64 van Beek, J.H., et al., *Heritability of liver enzyme levels estimated from genome-wide SNP data.* European Journal of Human Genetics, 2014.

65 Bouchard, T.J., Jr. and M. McGue, *Genetic and environmental influences on human psychological differences.* J Neurobiol, 2003. **54**(1): p. 4-45.

66 McGue, M. and T.J. Bouchard, Jr., *Genetic and environmental influences on human behavioral differences.* Annu Rev Neurosci, 1998. 21: p. 1-24.

67 Hicks, B.M., et al., *Family transmission and heritability of externalizing disorders - A twin-family study.* Archives of General Psychiatry, 2004. **61**(9): p. 922-928.

68 Bartels, A., et al., *Genetic and environmental mechanisms underlying stability and change in problem behaviors at ages 3, 7, 10, and 12.* Developmental Psychology, 2004. **40**(5): p. 852-867.

69 Bezdjian, S., L.A. Baker, and C. Tuvblad, *Genetic and environmental influences on impulsivity: A meta-analysis of twin, family and adoption studies.* Clinical Psychology Review, 2011. **31**(7): p. 1209-1223.

70 Nivard, M.G., et al., *Stability in symptoms of anxiety and depression as a function of genotype and environment: a longitudinal twin study from ages 3 to 63 years.* Psychol Med, 2014: p. 1-11.

71 Kan, K.J., et al., *Genetic and Environmental Stability in Attention Problems Across the Lifespan: Evidence From the Netherlands Twin Register.* Journal of the American Academy of Child and Adolescent Psychiatry, 2013. **52**(1): p. 12-25.

72 Visscher, P.M., et al., *Five Years of GWAS Discovery.* American Journal of Human Genetics, 2012. **90**(1): p. 7-24.

73 Rietveld, C.A., et al., *GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment.* Science, 2013. **340**(6139): p. 1467-1471.

74 Horikoshi, M., et al., *New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism.* Nature Genetics, 2013. **45**(1): p. 76-U115.

75 van der Valk, R.J., et al., *A novel common variant in DCST2 is associated with length in early life and height in adulthood.* Hum Mol Genet, 2014. **24**(4): p. 1155-68.

76 Benke, K.S., et al., *A genome-wide association meta-analysis of preschool internalizing problems*. J Am Acad Child Adolesc Psychiatry, 2014. **53**(6): p. 667-676 e7.

77 Viding, E., et al., *Genetics of callous-unemotional behavior in children.* PLoS One, 2013. **8**(7): p. e65789.

78 Lasky-Su, J., et al., *Genome-wide association scan of quantitative traits for attention deficit hyperactivity disorder identifies novel associations and confirms candidate gene associations.* Am J Med Genet B Neuropsychiatr Genet, 2008. **147B**(8): p. 1345-54.

79 Mick, E., et al., *Genome-wide association study of the child behavior checklist dysregulation profile.* J Am Acad Child Adolesc Psychiatry, 2011. **50**(8): p. 807-17 e8.

80 Wray, N.R., et al., *Research Review: Polygenic methods and their application to psychiatric traits.* Journal of Child Psychology and Psychiatry, 2014. **55**(10): p. 1068-1087.

81 Rietveld, C.A., et al., *Replicability and Robustness of Genome-Wide-Association Studies for Behavioral Traits.* Psychological Science, 2014. **25**(11): p. 1975-1986.

82 Visscher, P.M., et al., *Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated Samples.* Plos Genetics, 2014. **10**(4): p. e1004269.

83 St Pourcain, B., et al., *Variability in the common genetic architecture of social-communication spectrum phenotypes during childhood and adolescence.* Mol Autism, 2014. **5**(1): p. 18.

84 Yang, L., et al., *Polygenic transmission and complex neuro developmental network for attention deficit hyperactivity disorder: genome-wide association study of both common and rare variants.* Am J Med Genet B Neuropsychiatr Genet, 2013. **162B**(5): p. 419-30.

85 Lee, S.H., et al., *Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs.* Nature Genetics, 2013. **45**(9): p. 984-+.

86 Davis, L.K., et al., *Partitioning the Heritability of Tourette Syndrome and Obsessive Compulsive Disorder Reveals Differences in Genetic Architecture.* Plos Genetics, 2013. **9**(10).

87 Trzaskowski, M., et al., *First genome-wide association study on anxiety-related behaviours in childhood.* PLoS One, 2013. **8**(4): p. e58676.

88 Trzaskowski, M., P.S. Dale, and R. Plomin, *No genetic influence for childhood behavior problems from DNA analysis.* J Am Acad Child Adolesc Psychiatry, 2013. **52**(10): p. 1048-1056 e3.

89 Kruithof, C.J., et al., *The Generation R Study: Biobank update 2015.* European Journal of Epidemiology, 2014. **29**(12): p. 911-927.

90 Conners, C.K., et al., *The revised Conners' Parent Rating Scale (CPRS-R): Factor structure, reliability, and criterion validity.* Journal of Abnormal Child Psychology, 1998. **26**(4): p. 257-268.

91 Ivanova, M.Y., et al., *Preschool Psychopathology Reported by Parents in 23 Societies: Testing the Seven-Syndrome Model of the Child Behavior Checklist for Ages 1.5-5.* Journal of the American Academy of Child and Adolescent Psychiatry, 2010. **49**(12): p. 1215-1224.

92 Achenbach, T.M. and L.A. Rescorla, *Manual for the ASEBA Preschool Forms and Profiles.* 2000, University of Vermont, Research Center for Children, Youth, and Families, Burlington, VT

93 Achenbach, T.M. and L.A. Rescorla, *Manual for the ASEBA School-Age Forms and Profiles.* 2001, University of Vermont, Research Center for Children, Youth, and Families, Burlington, VT

94 Muratori, F., et al., *The CBCL 1.5-5 and the identification of preschoolers with autism in Italy.* Epidemiology and Psychiatric Sciences, 2011. **20**(4): p. 329-338.

95 Molins, N.C. and J.R. Clopton, *Teachers' reports of the problem behavior of children in their classrooms.* Psychological Reports, 2002. **90**(1): p. 157-164.

96 Bartels, M., et al., *Twins and the study of rater (dis)agreement (vol 12, pg 451, 2007).* Psychological Methods, 2008. **13**(2): p. 170-170.

97 Burt, S.A., et al., *Sources of covariation among the child-externalizing disorders: informant effects and the shared environment.* Psychological Medicine, 2005. **35**(8): p. 1133-1144.

98 Tellegen, P. and J.A. Laros, *Fair assessment of children from cultural minorities: a description of the SON-R non-verbal intelligence tests.* 2005: Research Institute for Child Psychology and Pathopsychology, and Education Committee, Slovak Commission for Unesco.

99 Rietveld, M., et al., *A twin study of differentiation of cognitive abilities in childhood.* Behavior genetics, 2003. **33**(4): p. 367-381.

100 Moore, C., et al., *Concurrent validity of the Snijders-Oomen Nonverbal Intelligence Test 2 1/2-7-Revised with the Wechsler Preschool and Primary Scale of Intelligence - Revised.* Psychological Reports, 1998. **82**(2): p. 619-625.

101 Bleichrodt, N., et al., *Revisie Amsterdamse Kinder Intelligentie Test, RAKIT.* 1984, Lisse: Swets & Zeitlinger.

102 Fedko, I.O., et al., *Estimation of Genetic Relationships Between Individuals Across Cohorts and Platforms: Application to Childhood Height.* Behavior genetics, 2015. **45**(5): p. 514-528.

103 Francioli, L.C., et al., *Whole-genome sequence variation, population structure and demographic history of the Dutch population.* Nature Genetics, 2014. **46**(8): p. 818-825.

104 Hampel, F.R., et al., *Robust Statistics: The Approach Based on Influence Functions* ed. I. John Wiley & Sons. 1986

105 Lehmann, E.L., *Nonparametrics: Statistical methods based on ranks.* San Franscisco, CA: Holden-Day. 1975.

106 Released, I.C., *IBM SPSS Statistics for Windows, Version 21.0* 2012
IBM Corp. : Armonk, NY

107 St Pourcain, B., et al., *Common variation contributes to the genetic architecture of social communication traits.* Mol Autism, 2013. **4**(1): p. 34.

108 Guey, L.T., et al., *Power in the Phenotypic Extremes: A Simulation Study of Power in Discovery and Replication of Rare Variants.* Genetic Epidemiology, 2011. **35**(4): p. 236-246.

109 Trzaskowski, M., et al., *DNA evidence for strong genetic stability and increasing heritability of intelligence from age 7 to 12.* Mol Psychiatry, 2014. **19**(3): p. 380-4.

110 Haworth, C.M.A., et al., *The heritability of general cognitive ability increases linearly from childhood to young adulthood.* Molecular Psychiatry, 2010. **15**(11): p. 1112-20.

111 Franic, S., et al., *Genetic and Environmental Stability of Intelligence in Childhood and Adolescence.* Twin Research and Human Genetics, 2014. **17**(3): p. 151-163.

112 Kan, K.J., et al., *Assessing Genetic Influences on Behavior: Informant and Context Dependency as Illustrated by the Analysis of Attention Problems.* Behavior Genetics, 2014. **44**(4): p. 326-336.

113 Visscher, P.M. and M.E. Goddard, *A General Unified Framework To Assess the Sampling Variance of Heritability Estimates Using Pedigree or Marker-Based Relationships.* Genetics, 2014.

114 Wachs, T.D., *Relation of maternal personality to perceptions of environmental chaos in the home.* Journal of Environmental Psychology, 2013. 34: p. 1-9.

115 Tandon, M., E. Cardeli, and J. Luby, *Internalizing Disorders in Early Childhood: A Review of Depressive and Anxiety Disorders.* Child and Adolescent Psychiatric Clinics of North America, 2009. **18**(3): p. 593-+.

116 Hankin, B.L., et al., *Development of depression from preadolescence to young adulthood: Emerging gender differences in a 10-year longitudinal study.* Journal of Abnormal Psychology, 1998. **107**(1): p. 128-140.

117 Cicchetti, D. and S.L. Toth, *The development of depression in children and adolescents.* American Psychologist, 1998. **53**(2): p. 221-241.

118 Sullivan, P.F., et al., *A mega-analysis of genome-wide association studies for major depressive disorder.* Molecular Psychiatry, 2013. **18**(4): p. 497-511.

119 Visscher, P.M., et al., *Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated Samples.* Plos Genetics, 2014. **10**(4).

120 Asendorpf, J.B., et al., *Reducing bias due to systematic attrition in longitudinal studies: The benefits of multiple imputation.* International Journal of Behavioral Development, 2014. **38**(5): p. 453-460.

121 Tambs, K., et al., *The Norwegian Institute of Public Health Twin Study of Mental Health: Examining Recruitment and Attrition Bias.* Twin Research and Human Genetics, 2009. **12**(2): p. 158-168.

122 Eaton, W.W., et al., *Psychopathology and Attrition in the Epidemiologic Catchment-Area Surveys.* American Journal of Epidemiology, 1992. **135**(9): p. 1051-1059.

123 Gustavson, K., et al., *Attrition and generalizability in longitudinal studies: findings from a 15-year population-based study and a Monte Carlo simulation study.* Bmc Public Health, 2012. **12.**

124 Wolke, D., et al., *Selective drop-out in longitudinal studies and non-biased prediction of behaviour disorders.* British Journal of Psychiatry, 2009. **195**(3): p. 249-256.

125 Zhu, Z.H., et al., *Dominance Genetic Variation Contributes Little to the Missing Heritability for Human Complex Traits.* American Journal of Human Genetics, 2015. **96**(3): p. 377-385.

126 Caspi, A., et al., *The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders?* Clin Psychol Sci, 2014. ***2***(2): p. 119-137.

127 Hudziak, J.J., et al., *The genetic and environmental contributions to attention deficit hyperactivity disorder as measured by the Conners' Rating Scales - Revised.* American Journal of Psychiatry, 2005. **162**(9): p. 1614-1620.

128 de Zeeuw, E.L., et al., *Childhood ODD and ADHD behavior: The effect of gender, age, teacher characteristics, classroom sharing and their interactions.* Behav Genet, 2015. **[in press].**

129 Hudziak, J.J., et al., *The genetic and environmental contributions to oppositional defiant behavior: A multi-informant twin study.* Journal of the American Academy of Child and Adolescent Psychiatry, 2005. **44**(9): p. 907-914.

130 van der Valk, J.C., et al., *Using shared and unique parental views to study the etiology of 7-year-old twins' internalizing and externalizing problems.* Behavior Genetics, 2003. ***33***(4): p. 409-420.

131 Van den Oord, E.J.C.G., F.C. Verhulst, and D.I. Boomsma, *A genetic study of maternal and paternal ratings of problem behaviors in 3-year-old twins.* Journal of Abnormal Psychology, 1996. **105**(3): p. 349-357.

132 Derks, E., et al., *Genetic analyses of maternal and teacher ratings on attention problems in 7-year-old Dutch twins.* Behavior genetics, 2006. **36**(6): p. 833-844.

133 Polderman, T.J.C., et al., *Genetic analyses of teacher ratings of problem behavior in 5-year-old twins.* Twin Research and Human Genetics, 2006. **9**(1): p. 122-130.

134 Polderman, T.J.C., et al., *A longitudinal twin study on IQ, executive functioning, and attention problems during childhood and early adolescence.* Acta Neurologica Belgica, 2006. **106**(4): p. 191-207.

135 Achenbach, T.M., L. Dumenci, and L.A. Rescorla, *DSM-oriented and empirically based approaches to constructing scales from the same item pools.* Journal of clinical child and adolescent psychology, 2003. **32**(3): p. 328-340.

136 Achenbach, T.M. and T.M. Ruffle, *The Child Behavior Checklist and related forms for assessing behavioral/emotional problems and competencies.* Pediatrics in Review, 2000. **21**(8): p. 265-271.

137 Achenbach, T. and L. Rescorla, *ASEBA school-age forms & profiles: Burlington.* VT: Aseba, 2001.

138 Althoff, R.R., et al., *Cross-informant agreement of the Dysregulation Profile of the Child Behavior Checklist.* Psychiatry research, 2010. **178**(3): p. 550-555.

139 Van der Valk, J., et al., *Using shared and unique parental views to study the etiology of 7-year-old twins' internalizing and externalizing problems.* Behavior genetics, 2003. **33**(4): p. 409-420.

140 Boomsma, D., C. Van Beijsterveldt, and J. Hudziak, *Genetic and environmental influences on Anxious/Depression during childhood: a study from the Netherlands Twin Register.* Genes, Brain and Behavior, 2005. **4**(8): p. 466-481.

141 Hoekstra, R.A., et al., *Genetic and environmental influences on the stability of withdrawn behavior in children: a longitudinal, multi-informant twin study.* Behavior genetics, 2008. **38**(5): p. 447-461.

142 Abdellaoui, A., et al., *Genetic influences on thought problems in 7-year-olds: A twin-study of genetic, environmental and rater effects.* Twin Research and Human Genetics, 2008. **11**(06): p. 571-578.

143 Falconer, D.S., T.F. Mackay, and R. Frankham, *Introduction to quantitative genetics (4th edn).* Trends in Genetics, 1996. **12**(7): p. 280.

144 Derks, E.M., C.V. Dolan, and D.I. Boomsma, *Effects of censoring on parameter estimates and power in genetic modeling.* Twin research, 2004. **7**(06): p. 659-669.

145 Achenbach, T.M., *Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles.* 1991: Department of Psychiatry, University of Vermont.

146 Spss, I., *IBM SPSS statistics version 21.* Boston, Mass: International Business Machines Corp, 2012.

147 Neale, M.C., et al., *OpenMx 2.0: Extended structural equation and statistical modeling*. Psychometrika, 2015: p. 1-15.

148 Boomsma, D.I., C.E. van Beijsterveldt, and J.J. Hudziak, *Genetic and environmental influences on Anxious/Depression during childhood: a study from the Netherlands Twin Register.* Genes Brain Behav, 2005. **4**(8): p. 466-81.

149 Rietveld, M.J., et al., *Heritability of attention problems in children: I. cross-sectional results from a study of twins, age 3-12 years.* Am J Med Genet B Neuropsychiatr Genet, 2003. **117B**(1): p. 102-13.

150 Hoekstra, R.A., et al., *Genetic and environmental influences on the stability of withdrawn behavior in children: a longitudinal, multi-informant twin study.* Behav Genet, 2008. **38**(5): p. 447-61.

151 Hudziak, J.J., et al., *Individual differences in aggression: Genetic analyses by age, gender, and informant in 3-, 7-, and 10-year-old Dutch twins.* Behavior genetics, 2003. **33**(5): p. 575-589.

152 Bartels, M., et al., *A study of parent ratings of internalizing and externalizing problem behavior in 12-year-old twins.* J Am Acad Child Adolesc Psychiatry, 2003. **42**(11): p. 1351-9.

153 Bartels, M., et al., *Disentangling genetic, environmental, and rater effects on internalizing and externalizing problem behavior in 10-year-old twins.* Twin Research, 2004. **7**(02): p. 162-175.

154 Bartels, M., et al., *Young Netherlands Twin Register (Y-NTR): a longitudinal multiple informant study of problem behavior.* Twin Research and Human Genetics, 2007. **10**(01): p. 3-11.

155 Hudziak, J.J., et al., *A twin study of inattentive, aggressive, and anxious/depressed behaviors*. Journal of the American Academy of Child & Adolescent Psychiatry, 2000. **39**(4): p. 469-476.

156 Brendgen, M., et al., *Examining genetic and environmental effects on social aggression: A study of 6-year-old twins.* Child development, 2005. **76**(4): p. 930-946.

157 Haberstick, B.C., et al., *Genes and developmental stabiltiy of aggressive behavior problems at home and school in a community sample of twins aged 7–12.* Behavior genetics, 2006. **36**(6): p. 809-819.

158 Eley, T.C., P. Lichtenstein, and J. Stevenson, *Sex differences in the etiology of aggressive and nonaggressive antisocial behavior: Results from two twin studies.* Child development, 1999. **70**(1): p. 155-168.

159 Spatola, C.A., et al., *A general population twin study of the CBCL/6-18 DSM-oriented scales.* Journal of the American Academy of Child & Adolescent Psychiatry, 2007. **46**(5): p. 619-627.

160 Bergen, S.E., C.O. Gardner, and K.S. Kendler, *Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood: a meta-analysis.* Twin Research and Human Genetics, 2007. **10**(03): p. 423-433.

161 Pappa, I., et al., *Single nucleotide polymorphism heritability of behavior problems in childhood: Genome-wide complex trait analysis.* Journal of the American Academy of Child & Adolescent Psychiatry, 2015. **54**(9): p. 737-744.

162 Van Der Sluis, S., et al., *Phenotypic complexity, measurement bias, and poor phenotypic resolution contribute to the missing heritability problem in genetic association studies.* PloS one, 2010. **5**(11): p. e13929.

163 Medland, S.E. and M.C. Neale, *An integrated phenomic approach to multivariate allelic association.* European Journal of Human Genetics, 2010. **18**(2): p. 233-239.

164 Minica, C.C., et al., *Genetic association in multivariate phenotypic data: power in five models.* Twin Research and Human Genetics, 2010. **13**(06): p. 525-543.

165 van der Sluis, S., D. Posthuma, and C.V. Dolan, *TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies.* PLoS Genet, 2013. **9**(1): p. e1003235.

166 Bartels, M., *Genetics of wellbeing and its components satisfaction with life, happiness, and quality of life: a review and meta-analysis of heritability studies.* Behavior genetics, 2015. **45**(2): p. 137-156.

167 van den Berg, S.M., et al., *Harmonization of Neuroticism and Extraversion phenotypes across inventories and cohorts in the Genetics of Personality Consortium: an application of Item Response Theory.* Behavior genetics, 2014. **44**(4): p. 295-313.

168 Okbay, A., et al., *Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses.* Nature genetics, 2016

169 Rietveld, C.A., et al., *Molecular genetics and subjective well-being*. Proceedings of the National Academy of Sciences, 2013. **110**(24): p. 9692-9697.

170 Vinkhuyzen, A.A., et al., *Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion.* Translational psychiatry, 2012. **2**(4): p. e102.

171 Smith, D.J., et al., *Genome-wide analysis of over 106,000 individuals identifies 9 neuroticism-associated loci.* bioRxiv, 2015: p. 032417.

172 de Moor, M.H., et al., *Genome-wide association study identifies novel locus for neuroticism and shows polygenic association with Major Depressive Disorder.* JAMA psychiatry, 2015. **72**(7): p. 642.

173 Power, R. and M. Pluess, *Heritability estimates of the Big Five personality traits based on common genetic variants.* 2015.

174 van Den Berg, S.M., et al., *Meta-analysis of Genome-Wide Association Studies for Extraversion: Findings from the Genetics of Personality Consortium.* Behavior genetics, 2015: p. 1-13.

175 Hahn, E., W. Johnson, and F.M. *Spinath, Beyond the heritability of life satisfaction–The roles of personality and twin-specific influences.* Journal of Research in Personality, 2013. **47**(6): p. 757-767.

176 Willemsen, G., et al., *The Adult Netherlands Twin Register: twenty-five years of survey and biological data collection.* Twin Res Hum Genet, 2013. **16**(1): p. 271-81.

177 van Beijsterveldt, C.E., et al., *The Young Netherlands Twin Register (YNTR): longitudinal twin and family studies in over 70,000 children.* Twin Res Hum Genet, 2013. **16**(1): p. 252-67.

178 Diener, E., et al., *The satisfaction with life scale. Journal of personality assessment,* 1985. 49(1): p. 71-75.

179 Chang, C.C., et al., *Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience,* 2015. **4**(7).

180 Marullo, L., J.S.E.-S. Moustafa, and I. Prokopenko, *Insights into the genetic susceptibility to type 2 diabetes from genome-wide association studies of glycaemic traits.* Current diabetes reports, 2014. **14**(11): p. 1-17.

181 Stumvoll, M., B.J. Goldstein, and T.W. van Haeften, *Pathogenesis of type 2 diabetes. Endocrine research,* 2007. **32**(1-2): p. 19-37.

182 Matthews, D., et al., *Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man.* Diabetologia, 1985. **28**(7): p. 412-419.

183 KULLER, L.H. and M. DRPH, *Insulin Sensitivity and Insulin Secretion Determined by Homeostasis Model Assessment and Risk of Diabetes in.* Diabetes Care, 2007. **30**(7): p. 1747-1752.

184 Tang, Q., et al., *Optimal cut-off values for the homeostasis model assessment of insulin resistance (HOMA-IR) and pre-diabetes screening: Developments in research and prospects for the future.* Drug discoveries & therapeutics, 2015. **9**(6): p. 380-385.

185 Morimoto, A., et al., *Increase in homeostasis model assessment of insulin resistance (HOMA-IR) had a strong impact on the development of type 2 diabetes in Japanese individuals with impaired insulin secretion: the saku study.* PloS one, 2014. **9**(8): p. e105827.

186 Haffner, S.M., et al., *A prospective analysis of the HOMA model: the Mexico City Diabetes Study.* Diabetes care, 1996. **19**(10): p. 1138-1141.

187 Cerf, M.E., *Beta cell dysfunction and insulin resistance.* Frontiers in endocrinology, 2013. **4.**

188 Kasuga, M., *Insulin resistance and pancreatic β cell failure.* The Journal of clinical investigation, 2006. **116**(7): p. 1756-1760.

189 Kahn, S., *The relative contributions of insulin resistance and beta-cell dysfunction to the pathophysiology of type 2 diabetes.* Diabetologia, 2003. **46**(1): p. 3-19.

190 Sladek, R. and I. Prokopenko, *Genome-Wide Association Studies of Type 2 Diabetes,* in *The Genetics of Type 2 Diabetes and Related Traits.* 2016, Springer. p. 13-61.

191 Dupuis, J., et al., *New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk.* Nature genetics, 2010. **42**(2): p. 105-116.

192 Lagou, V., et al. *Sex-dimorphism and variability of fasting glycaemic traits: a genome-wide association study in over 140,000 individuals (in preparation).* 2016.

193 Wallace, T.M., J.C. Levy, and D.R. Matthews, *Use and abuse of HOMA modeling.* Diabetes care, 2004. **27**(6): p. 1487-1495.

194 Song, Y., et al., *Insulin Sensitivity and Insulin Secretion Determined by Homeostasis Model Assessment and Risk of Diabetes in a Multiethnic Cohort of Women The Women's Health Initiative Observational Study*. Diabetes care, 2007. **30**(7): p. 1747-1752.

195 Consortium, G.L.G., *Discovery and refinement of loci associated with lipid levels*. Nature genetics, 2013. **45**(11): p. 1274-1283.

196 Locke, A.E., et al., *Genetic studies of body mass index yield new insights for obesity biology.* Nature, 2015. **518**(7538): p. 197-206.

197 Morris, A.P., et al., *Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes.* Nature genetics, 2012. **44**(9): p. 981.

198 Bulik-Sullivan, B., et al., *An atlas of genetic correlations across human diseases and traits.* Nature genetics, 2015.

199 Hanson, R.L., et al., *Evaluation of simple indices of insulin sensitivity and insulin secretion for use in epidemioiogic studies.* American journal of epidemiology, 2000. **151**(2): p. 190-198.

200 Prokopenko, I., et al., *A central role for GRB10 in regulation of islet function in man.* PLoS Genet, 2014. **10**(4): p. e1004235.

201 Dimas, A.S., et al., *Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity.* Diabetes, 2013: p. DB_130949.

202 Voight, B.F., et al., *Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis.* Nature genetics, 2010. **42**(7): p. 579-589.

203 Kahn, S.E., R.L. Hull, and K.M. Utzschneider, *Mechanisms linking obesity to insulin resistance and type 2 diabetes.* Nature, 2006. **444**(7121): p. 840-846.

204 Manning, A.K., et al., *A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance.* Nature genetics, 2012. **44**(6): p. 659-669.

205 Penckofer, S., et al., *State of the Science Depression and Type 2 Diabetes.* Western journal of nursing research, 2014: p. 0193945914524491.

206 Ali, S., et al., *The prevalence of co-morbid depression in adults with Type 2 diabetes: a systematic review and meta-analysis.* Diabetic Medicine, 2006. **23**(11): p. 1165-1173.

207 Samaan, Z., et al., *Lack of association between type 2 diabetes and major depression: epidemiologic and genetic evidence in a multiethnic population.* Translational psychiatry, 2015. **5**(8): p. e618.

208 Doyle, T., A. Halaris, and M. Rao, *Shared neurobiological pathways between type 2 diabetes and depressive symptoms: a review of morphological and neurocognitive findings.* Current diabetes reports, 2014. **14**(12): p. 1-12

209 Stuart, M.J. and B.T. Baune, *Depression and type 2 diabetes: inflammatory mechanisms of a psychoneuroendocrine co-morbidity.* Neuroscience & Biobehavioral Reviews, 2012. **36**(1): p. 658-676.

210 Domenici, E., et al., *Plasma protein biomarkers for depression and schizophrenia by multi analyte profiling of case-control collections.* PLoS One, 2010. **5**(2): p. e9166.

211 Simmons, W.K., et al., *Depression-Related Increases and Decreases in Appetite: Dissociable Patterns of Aberrant Activity in Reward and Interoceptive Neurocircuitry.* American Journal of Psychiatry, 2016.

212 Milaneschi, Y., et al., *Polygenic dissection of major depression clinical heterogeneity.* Molecular psychiatry, 2015.

213 Purcell, S.M., et al., *Common polygenic variation contributes to risk of schizophrenia and bipolar disorder.* Nature, 2009. **460**(7256): p. 748-752.

214 Penninx, B.W., et al., *The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods.* International journal of methods in psychiatric research, 2008. **17**(3): p. 121-140.

215 Lagou, V., et al. *Sex-dimorphosm and variability of fasting glyceamic traits: a genome-wide association study in over 140,000 individuals (in preparation).* 2016.

216 Ripke, S., et al., *A mega-analysis of genome-wide association studies for major depressive disorder.* Molecular psychiatry, 2013. **18**(4): p. 497-511.

217 Willemsen, G., et al., *The Adult Netherlands Twin Register: twenty-five years of survey and biological data collection.* Twin Research and Human Genetics, 2013. **16**(01): p. 271-281.

218 Organization, W.H., *Composite Interview Diagnostic Instrument (CIDI).* World Health Organization, 1990.

219 Shungin, D., et al., *New genetic loci link adipose and insulin biology to body fat distribution.* Nature, 2015. **518**(7538): p. 187-196.

220 Surakka, I., et al., *The impact of low-frequency and rare variants on lipid levels.* Nature genetics, 2015. **47**(6): p. 589-597.

221 Johnson, A.D., et al., *SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap.* Bioinformatics, 2008. **24**(24): p. 2938-2939.

222 Nieuwboer, et al. *GWIS: Genome Wide Inferred Statistics for non-linear functions of multiple phenotypes (under revision).* 2016.

223 Iryna O. Fedko, M.G.N., Jouke-Jan Hottenga, Reedik Mägi, Inga Prokopenko, and Dorret I. Boomsma, *Homeostatic Model Assessment of β-cell Function and Insulin Resistance Genome Wide Inferred Statistics Association Study.* 2016.

224 Minică, C.C., et al., *Sandwich corrected standard errors in family-based genome-wide association studies.* European Journal of Human Genetics, 2015. **23**(3): p. 388-394.

225 Gale, C., et al., *Pleiotropy between neuroticism and physical and mental health: findings from 108 038 men and women in UK Biobank.* bioRxiv, 2015: p. 031138.

226 Kan, C., et al., *Genetic overlap between type 2 diabetes and depression in Swedish and Danish twin registries.* Molecular psychiatry, 2016.

227 Finucane, H.K., et al., *Partitioning heritability by functional annotation using genome-wide association summary statistics.* Nature genetics, 2015. **47**(11): p. 1228-1235.

228 De Wit, L.M., et al., *Depression and body mass index, a u-shaped association*. BMC public health, 2009. **9**(1): p. 14.

229 Todd, J.A., *Intolerable secretion and diabetes in tolerant transgenic mice, revisited.* Nature genetics, 2016. **48**(5): p. 476-477.

230 Jacka, F.N., et al., *Dietary patterns and depressive symptoms over time: examining the relationships with socioeconomic position, health behaviours and cardiovascular risk.* PLoS One, 2014. **9**(1): p. e87657.

231 Dipnall, J.F., et al., *The association between dietary patterns, diabetes and depression.* Journal of affective disorders, 2015. **174**: p. 215-224.

232 Lamers, F., et al., *Evidence for a differential role of HPA-axis function, inflammation and metabolic syndrome in melancholic versus atypical depression.* Molecular psychiatry, 2013. **18**(6): p. 692-699.

233 Khambaty, T., et al., *Depressive symptom clusters as predictors of 6-year increases in insulin resistance: data from the Pittsburgh Healthy Heart Project.* Psychosomatic medicine, 2014. **76**(5): p. 363.

234 Gage, S.H., et al., G= E: *What GWAS can tell us about the environment.* PLoS Genet, 2016. **12**(2): p. e1005765.

235 Van Leeuwen, E.M., et al., *Genome of the Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels.* Nature communications, 2015. **6.**

236 Pistis, G., et al., *Rare variant genotype imputation with thousands of study-specific whole-genome sequences: implications for cost-effective study designs.* European Journal of Human Genetics, 2015. **23**(7): p. 975-983.

237 Trzaskowski, M., P.S. Dale, and R. Plomin, *No genetic influence for childhood behavior problems from DNA analysis.* Journal of the American Academy of Child & Adolescent Psychiatry, 2013. **52**(10): p. 1048-1056. e3.

238 Silva, N., E. Atlantis, and K. Ismail, *A review of the association between depression and insulin resistance: pitfalls of secondary analyses or a promising new approach to prevention of type 2 diabetes?* Current psychiatry reports, 2012. **14**(1): p. 8-14.

239 Consortium, G.o.t.N., *Whole-genome sequence variation, population structure and demographic history of the Dutch population.* Nature genetics, 2014. **46**(8): p. 818-825.

240 Leslie, S., et al., *The fine-scale genetic structure of the British population.* Nature, 2015. **519(**7543): p. 309-314.

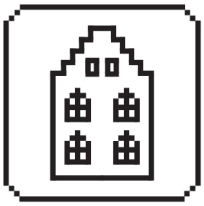
241 Euesden, J., C.M. Lewis, and P.F. O'Reilly, *PRSice: polygenic risk score software.* Bioinformatics, 2015. **31**(9): p. 1466-1468.

242 Vilhjálmsson, B.J., et al., *Modeling linkage disequilibrium increases accuracy of polygenic risk scores.* The American Journal of Human Genetics, 2015. **97**(4): p. 576-592.

243 Van Dongen, J., et al., *The continuing value of twin studies in the omics era.* Nature reviews genetics, 2012. **13**(9): p. 640-653.

244 Consortium, G.P., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.

245 Sudmant, P.H., et al., *An integrated map of structural variation in 2,504 human genomes.* Nature, 2015. **526**(7571): p. 75-81.



NEDERLANDSE SAMENVATTING

In dit proefschrift heb ik verschillende rekenmethoden toegepast op genetische Single Nucleotide Polymorfismen (SNP) en tweeling data, om het relatieve effect van genetische factoren te bepalen in een aantal complexe ziekten. Daarnaast heb ik de gedeelde genetische factoren van een aantal gecorreleerde eigenschappen onderzocht.

In hoofdstuk 2 is de GoNL genotype referentie set gebruikt om de data van twee kinder-cohorten, NTR (N = 3,102) en GENR (N = 2,826), te combineren met cross-platform genotype imputatie. De schattingen van de SNP erfelijkheid, de hoeveelheid erfelijke variantie verklaard door SNPs, van lichaamslengte bij kinderen waren hetzelfde over verschillende Gentische Relatie Matrices (GRM)s berekend uit: 1) Eerst combineren van SNP data en dan imputeren (h^2 = 51%), 2) eerst imputeren en dan combineren (h^2 = 52%) en iets lager voor alleen de data combineren h^2 = 43%). Correctie voor het cohort resulteerde in een verlies van ≈ 2% in de SNP erfelijkheid voor alle combinatie methoden. Correctie met Nederlandse principle components berekend uit de SNPs resulteerde in een SNP erfelijkheid die ≈11% lager is, in zowel de geïmputeerde - als gecombineerde data. Dit geeft aan dat de imputatie zelf niet de genetische gelijkheid van de deelnemers aanpast. De SNP erfelijkheid schattingen, gecorrigeerd voor cohort en Nederlandse populatie structuur alsook voor leeftijd en geslacht lagen tussen 32% en 41%. Onze resultaten geven aan dat slechts een beperkt aantal SNPs die overlappen op de platforms tussen cohorten voldoende is om de genetische relatie tussen mensen goed te schatten. We hebben ook laten zien dat imputatie met een referentie set de genetische platform stratificatie verminderd ten opzichte van imputatie tussen platforms zonder een referentie set. Alhoewel de imputatie met een referentie set, het dus mogelijk maakt om datasets gegenotypeerd op verschillende platforms met weinig overlappende SNPs te combineren, is het dan wel goed om cohort altijd mee te nemen als covariaat in de analyse.

In hoofdstuk 3 is de SNP erfelijkheid van een aantal gedragsproblemen bij kinderen geschat in de twee cohorten NTR en GENR. Met een groter aantal deelnemers in de gecombineerde studies waren we in staat om een significante SNP erfelijkheid te detecteren voor Attentieproblemen en Hyperactiviteit (h^2 = 0.37 - 0.71, SE = 0.14 - 0.22), externaliserende gedragsproblemen (h^2 = 0.44, SE = 0.22) en alle gedragsproblemen bij elkaar (h^2 = 0.18, SE = 0.10), aangegeven met moeder of de leraar als beoordelaar. Bij een sensitiviteitanalyse, waarbij de extreem scorende deelnemers zijn verwijderd, of een quantiel-normalisatie werd gedaan, werd de significantie van de resultaten niet anders, maar werden de erfelijkheidsschattingen wel lager. De implicatie van deze resultaten is dat het nut heeft om samen te werken tussen verschillende studiecohorten om met GWAS studies genen voor deze aandoeningen op te sporen.

In opvolging van de resultaten in hoofdstuk 3, en om de invloed op erfelijkheidsschattingen van verschillende beoordelaars van de gedragsproblemen van het kind te onderzoeken,

hebben we de gedeelde en unieke genetische variatie van de gedragsproblemen van verschillende beoordelaars bepaald in hoofdstuk 4. In een groot Nederlands cohort met 12,310 tweelingparen van rondom de 7 jaar, hebben we de erfelijkheid geschat voor de gedragsproblemen van het kind op de empirische Child Behavior CheckList (CBCL) 6-18 vragenlijsten met moeder en vader als beoordelaars. Gemiddeld schatten de moeders de gedragsproblemen van hun kinderen iets hoger in als de vaders. De ouderlijke overeenkomst was tussen de 0.62 en 0.74 over de verschillende schalen. Een groot deel van de erfelijkheid van gedragsproblemen was gedeeld tussen ouders. Dit geeft aan dat de ouders in grote mate dezelfde gedragsproblemen zien bij hun kinderen. Een ander deel van de erfelijkheid was uniek bepaald voor elke ouder. Dit kan komen omdat het kind dit gedrag alleen vertoont bij een ouder. Omdat een groot deel van de gedragsprobleemerfelijkheid overlapt is het verstandig om de beoordeling van beide ouders mee te nemen in een Genome Wide Assocociation Study (GWA). Dit zal de detectie kracht van het vinden van varianten die geassocieerd zijn met gedragsproblemen van het kind verhogen, wanneer rekening wordt gehouden met de gemiddelde verschillen tussen moeder en vader.

In hoofdstuk 5 is de genetische correlatie tussen Persoonlijk Welbevinden (SWB) en twee persoonlijkheidskenmerken, Neuroticisme (NEU) en Extraversie (EXT), geschat. Hierbij is gebruik gemaakt van de bivariate analyse die is geïmplementeerd in de GCTA software. Met behulp van de niet direct gerelateerde mensen en de direct gerelateerde mensen, is de SNP erfelijkheid van de gemeten genotype platforms, de totale erfelijkheid en de genetische correlatie in ons Nederlands Tweelingen Register studie sample geschat (N ≈ 9,000). De totale erfelijkheid was 32%, 37% en 42% voor SWB, NEU en EXT. De totale genetische correlaties waren -.70 (SE = .03) en .48 (SE = .03) tussen SWB en NEU, en tussen SWB en EXT respectievelijk. De SNP erfelijkheid voor SWB was 7%, 10% voor NEU en 16% voor EXT. De genetische correlatie voor SNPs was groter tussen SWB en NEU (r_g = -.80), dan tussen SWB en EXT (r_g = 0.18). Dit was tegengesteld aan de geobserveerde correlaties tussen de fenotypes (r = -.43 en r = .32 respectievelijk). Een grote genetische correlatie tussen SWB en NEU geeft aan dat dezelfde genen betrokken zijn bij de fenotypes, en dat deze kunnen worden gedetecteerd als we hiervan gebruik maken. Dit in tegenstelling tot de correlatie tussen SWB en EXT die vooral door komt door factoren uit het milieu.

Hoofdstuk 6 beschrijft een applicatie van een recent ontwikkelde methode GWIS, waarbij met behulp van GWAS meta-analyse resultaten van individuele fenotypes, de GWAS resultaten van een gecombineerd fenotype analytisch afgeleid kunnen worden. We hebben met de GWIS methode de MAGIC studie GWAS meta-analyse resultaten van nuchter glucose en insuline samengevoegd om zo de resultaten voor insuline resistentie HOMA-B en HOMA-IR te krijgen. Deze resultaten zijn daarna vergeleken met de originele HOMA GWAS resultaten die ook door het MAGIC zijn gedaan. Omdat niet alle studies

de HOMA analyses konden doen, was het sample dat met GWIS is geanalyseerd groter dan het originele GWAS sample, en daardoor was het mogelijk om meer genetische varianten voor HOMA te detecteren. Hierdoor hebben we 7 genetische locaties gerepliceerd die ook gevonden waren in de originele meta-analyse, maar we hebben ook nog 4 extra locaties gevonden voor HOMA-B. Voor HOMA-IR hebben we 2 locaties gevonden die eerder waren gedetecteerd, en daarnaast 3 nieuwe. Verder hebben we de genetische correlatie berekend tussen de HOMA's en een range van glucose - en metabole fenotypes namelijk nuchter Glucose, Insuline, High Density Lipoprotein (HDL), Low Density Lipoprotein (LDL), totaal cholesterol, Triglyceriden, Body Mass Index (BMI) en Type 2 Diabetes (T2D). We vonden significante correlaties tussen Inusline en HOMA-B/-IR (r_g = 0.76/0.98, SE = 0.05/0.005); Glucose en HOMA-B/-IR (r_g = -0.38/0.49, SE = 0.12/0.07); BMI en HOMA-B/-IR (r_g = 0.39/0.62, SE = 0.05/0.05) en tussen T2D en HOMA-IR (r_g = 0.53, SE = 0.08). We vonden geen significante genetische correlaties tussen de HOMAs en de lipiden LDL/TC en tussen T2D en HOMA-B. Resultaten van de analytisch afgeleide HOMA-B/-IR genoom brede associatie statistieken demonstreren het voordeel van de GWIS methode boven het doen van een extra GWAS op dit moeilijke fenotype: GWIS laat het toe dat insuline en glucose niet beiden bij iedereen aanwezig hoeven zijn voor berekening van de HOMA, dit geeft dus een grotere studiegroep en dus meer detectiekracht van genen. Het laat ook een krachtiger LD score regressie toe, omdat de afgeleide samenvatting van de associatie statistieken gebaseerd zijn op een groter sample.

In hoofdstuk 7 is de gedeelde genetische etiologie tussen twee comorbide ziekten bestudeerd, namelijk tussen Type 2 diabetes (T2D) en Depressie (MDD) met behulp van genetische risico scores berekend uit de genen die betrokken zijn bij nuchter insuline (FI), glucose (FG), β-cel functie (HOMA-B) en insuline resistentie (HOMA-IR). Deze scores werden gebruikt om de MDD status in het NTR en de NESDA studie te voorspelen. We hebben verder gekeken naar verschillende subtypes van MDD, die werden gekarakteriseerd door een verminderde of vergrote eetlust. De genetische risico scores, berekend met verschillende significantie grenswaarden, voorspelden echter niet MDD of een van de subtypes. Selectie van verschillende SNPs betrokken bij glycemische fenotypen, lipiden, middel-tot-heup breedte ratio gewogen met de GWAS Beta's van de MDD PGC mega analyse voorspelden ook geen MDD. LD score regressie werd daarna gebruikt om de genetische correlaties te berekenen tussen de glycemische fenotypes en MDD en de aanverwante fenotypes depressie symptomen en neuroticisme. Geen van de gebruikte SNP sets voorspelden echter MDD, noch de subtypes. Het meenemen van BMI als covariaat in deze analyses gaf geen opmerkelijke verschillen. LD score regressie liet wel een kleine genetische overlap zien tussen HOMA-IR en Insuline met Depressieve Symptomen (r_g = 0.16, SE = 0.07 en r_g = 0.17, SE = 0.07), maar dit was niet statistisch significant. Wat uit deze resultaten naar voren komt is dat FG en FI, en de indexen van insuline resistentie (HOME-IR) en Beta-cel functie (HOMA-B) een andere etiologie

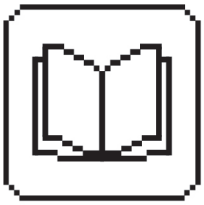
hebben als MDD en de symptomen van MDD. Het is daarom nodig om voor de overlap tussen MDD en deze factoren meer te kijken naar invloeden zoals gedrag, demografische en socio-economische factoren.

In conclusie, rekenmethoden en het referentie Genoom van Nederland (GoNL) hebben de basis gevormd van deze thesis, waarin de eerder verzamelde data van het Nederlands Tweelingen Register en data van meta-analyses van verschillende consortia zijn onderzocht. Analytische afleiding van de samenvattende statistieken van deels overlappende samples genereerde nieuwe data voor toekomstig onderzoek in Diabetes Type 2 en gerelateerde glycemische traits. Inzichten van de comorbide fenotypen, of dezelfde fenotypen gemeten binnen verschillende informanten, hebben geleid tot nieuwe strategieën om deze data te analyseren met behulp van nieuwe risicofactoren of modellen, op basis van gedeelde of niet gedeelde genetische achtergrond. In deze thesis is bij de genetische analyses gebruik gemaakt van de familieleden van mensen om de kracht van de analyses te verhogen. Als we ons richten op de moleculaire kant, dan heeft de hoge resolutie GoNL genetische referentie set geholpen om combinatie stratificatie effecten van verschillende genetische platforms te reduceren en om de hoeveelheid genetische variatie van de Nederlandse populatie te preserveren. Ten slotte, een range van verschillende analyse methoden is toegepast in deze thesis, en het heeft gezorgd dat zelfs met bestaande genotype en fenotype data, extra biologisch inzicht verkregen kan worden door her-exploratie met nieuwe technieken.

ACKNOWLEDGEMENTS

This research project was supported of Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) Rainbow project 1 and 2 (NWO 184.021.007).

This study makes use of data generated by the Genome of the Netherlands Project. Funding for the project was provided by the Netherlands Organization for Scientific Research under award number 184021007, dated July 9, 2009 and made available as a Rainbow Project of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL). Samples where contributed by LifeLines (http://lifelines.nl/lifelines-research/general), The Leiden Longevity Study (http://www.healthy-ageing.nl; http://www.langleven.net), The Netherlands Twin Registry (NTR: http://www.tweelingenregister.org), The Rotterdam studies, (http://www.erasmus-epidemiology.nl/rotterdamstudy) and the Genetic Research in Isolated Populations program (http://www.epib.nl/research/geneticepi/research.html#gip). The sequencing was carried out in collaboration with the Beijing Institute for Genomics (BGI).



LIST OF PUBLICATIONS

Fedko, I. O., Hottenga, J. J., Medina-Gomez, C., Pappa, I., van Beijsterveldt, C. E., Ehli, E. A., ... & Middeldorp, C. M. (2015). Estimation of Genetic Relationships Between Individuals Across Cohorts and Platforms: Application to Childhood Height. *Behavior genetics, 45*(5), 514-528.

Pappa, I., **Fedko, I. O.**, Mileva-Seitz, V. R., Hottenga, J. J., Bakermans-Kranenburg, M. J., Bartels, M., ... & Rivadeneira, F. (2015). Single nucleotide polymorphism heritability of behavior problems in childhood: Genome-wide complex trait analysis. *Journal of the American Academy of Child & Adolescent Psychiatry, 54*(9), 737-744.

Den Braber, A., Zilhão, N. R., **Fedko, I. O.**, Hottenga, J. J., Pool, R., Smit, D. J. A., ... & Boomsma, D. I. (2016). Obsessive–compulsive symptoms in a large population-based twin-family sample are predicted by clinically based polygenic scores and by genome-wide SNPs. *Translational psychiatry, 6*(2), e731.

Mbarek, H., Milaneschi, Y., **Fedko, I. O.**, Hottenga, J. J., de Moor, M. H., Jansen, R., ... & Penninx, B. W. (2015). The genetics of alcohol dependence: Twin and SNP-based heritability, and genome-wide association study based on AUDIT scores. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics,168*(8), 739-748.

Lin, B. D., Mbarek, H., Willemsen, G., Dolan, C. V., **Fedko, I. O.**, Abdellaoui, A., ... & Hottenga, J. J. (2015). Heritability and Genome-Wide Association Studies for Hair Color in a Dutch Twin Family Based Sample. *Genes, 6*(3), 559-576.

Minică, C. C., Dolan, C. V., Hottenga, J. J., Pool, R., **Fedko, I. O.**, Mbarek, H., ... & Genome of the Netherlands Consortium. (2015). Heritability, SNP-and Gene-Based Analyses of Cannabis Use Initiation and Age at Onset. *Behavior genetics,45*(5), 503-513.

van Den Berg, S. M., de Moor, M. H., Verweij, K. J., Krueger, R. F., Luciano, M., Vasquez, A. A., ... & Gordon, S. D. (2015). Meta-analysis of Genome-Wide Association Studies for Extraversion: Findings from the Genetics of Personality Consortium. Behavior genetics, 1-13.

De Moor, M. H., Van Den Berg, S. M., Verweij, K. J., Krueger, R. F., Luciano, M., Vasquez, A. A., ... & Gordon, S. D. (2015). Meta-analysis of genome-wide association studies for neuroticism, and the polygenic association with major depressive disorder. *JAMA psychiatry, 72*(7), 642-650.

Hibar, D. P., Stein, J. L., Renteria, M. E., Arias-Vasquez, A., Desrivières, S., Jahanshad, N., ... & Aribisala, B. S. (2015). Common genetic variants influence human subcortical brain structures. *Nature, 520*(7546), 224-229.

Middeldorp, C. M., Hammerschlag, A. R., Ouwens, K. G., Groen-Blokhuis, M. M., Pourcain, B. S., Greven, C. U., ... &Vilor-Tejedor, N. (2016). A genome-wide association meta-analysis of attention-deficit/hyperactivity disorder symptoms in population-based pediatric cohorts. *Journal of the American Academy of Child & Adolescent Psychiatry, 55*(10), 896-905.