



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 kindercohorten, 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 Genetische 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 $\approx 2\%$ in de SNP erfelijkheid voor alle combinatie methoden. Correctie met Nederlandse principle components berekend uit de SNPs resulteerde in een SNP erfelijkheid die $\approx 11\%$ 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 kwantiel-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 studiegroepen 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 gedragsproblemerfelijkheid overlapt is het verstandig om de beoordeling van beide ouders mee te nemen in een Genome Wide Association 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 ($N \approx 9,000$). 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 Insuline 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 genom 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 preserven. 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.

