

SAMENVATTING



Dit proefschrift bestaat uit vier delen die elk het verslag van twee onderzoeken bevatten. Het eerste deel gaat over de erfelijkheid van een verzameling psychiatrische symptomen die betrekking hebben op verminderde organisatie van je gedachten (“Thought Problems” in het Engels). De erfelijkheid, dat wil zeggen de mate waarin verschillen tussen mensen worden beïnvloed door genetische verschillen, werd geschat met een tweelingendesign. Hierin wordt de gelijkenis tussen een- en twee-eiige tweelingen voor de eigenschap (het ‘fenotype’) gebruikt om het belang van genetische en omgevingsinvloeden te schatten. Dit design berust op de aanname dat eeneiige tweelingen genetisch identiek zijn, terwijl twee-eiige tweelingen ongeveer 50% delen van alle genen die in de populatie verschillen. In het tweede deel van het proefschrift wordt gezocht naar zeldzame mutaties die genetische verschillen tussen eeneiige tweelingen veroorzaken, en wordt bekeken of dergelijke mutaties ook tot fenotypische verschillen kunnen leiden. In het derde deel worden patronen van genetische variatie in Nederland in kaart gebracht en gebruikt om gevolgen van de evolutionaire geschiedenis bloot te leggen. In deel vier tenslotte wordt beschreven hoe het gedrag van (voor)ouders (met name via partnerkeuze en migratie) genetische variatie van de huidige bevolking beïnvloedt.

Deel I - Tweelingstudies: Beoordelaar effecten, meetinvariantie, en longitudinale erfelijkheid van “Thought Problems”

“Thought Problems” (TP) wordt gemeten door een verzameling vragen die statistisch met elkaar samenhangen en die symptomen meten uit verscheidene psychiatrische aandoeningen: hallucinaties, obsessief-compulsieve symptomen, vreemde gedachten en gedragingen, en neigingen tot zelfbeschadiging.

In hoofdstuk 2 zien we dat ongeveer de helft van de individuele verschillen in TP in een groep van ~9000 7-jarige tweelingparen verklaard kunnen worden door genetische factoren. De erfelijkheid werd geschat door ouderbeoordelingen van de symptomen (zowel van vaders als van moeders) voor eeneiige tweelingen te vergelijken met die voor twee-eiige tweelingen. Gegeven de aanname dat eeneiige tweelingen genetisch identiek zijn, moeten eeneiige tweelingen voor erfelijke fenotypes meer gelijkenissen vertonen dan twee-eiige tweelingen. TP werd opgedeeld in twee componenten waarvoor genetische en omgevingsinvloeden werden geschat: het gedeelte van het fenotype waarover beide ouders het eens zijn (wat 67% van de variantie verklaarde, waarvan 76% erfelijk), en het gedeelte dat uniek door beide ouders wordt gerapporteerd (33% van de variantie: voor de moeder is 61% hiervan erfelijk, voor de vader 65%). Het feit dat het deel van de symptomen waar de ouders het niet over eens waren ook erfelijk is, geeft aan dat het unieke perspectief van beide ouders mede gebaseerd is op de symptomen van het kind, en niet op rater bias of op meetfout. Het unieke perspectief van de ouders wordt ook beïnvloed door “gedeelde omgevingsinvloeden”, die wel kunnen

samenhangen met rater bias (moeder: 13%; vader: 13%). De resterende individuele verschillen konden verklaard worden door unieke omgevingsinvloeden, die ook meetfouten kunnen bevatten.

TP wordt in hoofdstuk 3 verder geanalyseerd in tweelingen en familieleden van 12–59 jaar oud. Een exploratieve factor analyse bevestigt dat er één onderliggend construct wordt gemeten door TP. De vragenlijst is meetinvariant over leeftijd en sekse, wat betekent dat hetzelfde construct wordt gemeten in mannen en vrouwen van verschillende leeftijden. Met het tweelingmodel, uitgebreid met informatie van broers en zussen, werd de longitudinale erfelijkheid geschat. Individuele verschillen in TP worden in adolescenten (12–18 jaar), jong volwassenen (19–27 jaar) en oudere volwassenen (28–59 jaar) voor ongeveer 37% beïnvloed door genetische invloeden (lager dan in kinderen). Vanaf de adolescentie beïnvloeden dezelfde genen TP in alle drie de leeftijdsgroepen, met uitzondering van een additionele genetische component, die pas bij volwassenen een rol begint te spelen. De resterende individuele verschillen kunnen verklaard worden door unieke omgevingsfactoren (die deels ook meetfouten kunnen bevatten).

Deel II - Copy Number Varianten: Mutaties na de splitsing van eeneiige tweelingen

In hoofdstuk 4 wordt de DNA sequentie van 50 eeneiige tweelingparen bestudeerd die geselecteerd zijn op aandachtsproblemen uit een groep van ~3200 eeneiige tweelingparen: 17 concordante paren met aandachtsproblemen, 22 concordante paren zonder aandachtsproblemen en 11 discordante paren. Aandachtsproblemen zijn op meerdere leeftijden gemeten met een schaal die een voorspellende waarde heeft voor ADHD en net als ADHD een hoge erfelijkheid heeft. In deze groep tweelingen werden *Copy Number Varianten* (CNVs) gemeten (bij 25 paren ook in de ouders). CNVs zijn DNA segmenten die een variabel aantal keren aanwezig zijn (“normaal” aantal kopieën is twee: één van allebei de ouders, maar bij CNVs kan het aantal kopieën ook nul, één, drie of meer zijn). Er is gezocht naar CNV mutaties die voor of na de splitsing van de bevruchte eicel plaats vonden, en er is een analyse uitgevoerd om te testen of CNVs geassocieerd zijn met aandachtsproblemen. Er is één mutatie gevonden die voor de splitsing van de eicel plaats vond (aanwezig bij beide tweelingen, maar niet bij de ouders) in een tweelingpaar zonder aandachtsproblemen. Er zijn twee mutaties gevonden die na de splitsing hebben plaatsgevonden (aanwezig in een van de tweelingen) in een concordant tweelingpaar met aandachtsproblemen (een deletie op chromosoom 4) en in de persoon met aandachtsproblemen in een discordant tweelingpaar (een duplicatie op chromosoom 17). Deze CNV mutaties liggen op plekken die overlappen met genen die eerder in verband zijn gebracht met psychiatrische aandoeningen. Behalve meer gedragsproblemen hadden de dragers van de mutatie

ook een lager geboortegewicht dan hun tweelingbroer. Een analyse bij alle 50 tweelingparen liet zien dat kinderen met meer aandachtsproblemen, verspreid over het hele genoom gemiddeld grotere CNVs hadden, vooral bij CNVs die met genen overlaptten.

In hoofdstuk 5 is het hele genoom gescand voor CNV verschillen binnen eeneiige tweelingen bij ~1100 niet geselecteerde eeneiige tweelingparen. Van ongeveer de helft van de tweelingparen kwam het DNA uit wangcellen (voornamelijk kinderen), en van de andere helft uit bloed (voornamelijk volwassenen). Er zijn 153 mogelijke CNV mutaties gevonden, waarvan de meerderheid uit dezelfde instabiele regio: 15q11.2. De meerderheid hiervan werd geobserveerd in DNA uit bloed (dus bij de volwassenen). De 15q11.2 mutaties die in bloed zijn waargenomen werden significant vaker in oudere tweelingparen gemeten. Een eerste selectie van 11 CNV mutaties (bij kinderen en volwassenen) zijn met qPCR ter validatie opnieuw gemeten, waarvan er uiteindelijk 2 uit wangcellen door qPCR zijn geconfirmeerd binnen hetzelfde gezonde 13-jarige tweelingpaar. Er zijn geen grote fenotypische verschillen binnen het tweelingpaar waargenomen in longitudinale vragenlijst gegevens die werden verzameld van 1 tot 21 jaar.

Deel III - Populatiegenetica: De genetische opmaak van Nederland

Het autochtone deel van de Nederlandse bevolking wordt over het algemeen als een genetisch homogene populatie gezien. In hoofdstuk 6 bekijken we hoe homogeen deze populatie daadwerkelijk is en worden patronen van genetische variatie in Nederland in kaart gebracht met een principale componenten analyse (PCA) op 500,000 *Single Nucleotide Polymorfismen* (SNPs). SNPs zijn vaak voorkomende genetische varianten en bestaan uit een verandering in een enkele nucleotide (DNA is opgebouwd uit vier verschillende nucleotiden met de nucleobasen adenine, thymine, guanine en cytosine, afgekort als A, T, G en C). SNP varianten in een populatie zijn verspreid over het hele genoom. Een PCA is een statistische methode die in een groot aantal gemeten variabelen (SNPs in dit geval) de grootste patronen van variatie samenvat in zogenaamde principale componenten (PCs). Deze analyses zijn uitgevoerd bij 4441 ongerelateerde Nederlandse individuen en 1014 ongerelateerde individuen uit 14 verschillende populaties uit de hele wereld (de 1000 Genomes dataset). Het filteren van SNPs die hoog met elkaar zijn gecorreleerd (oftewel hoog in linkage disequilibrium [LD] met elkaar zijn) had niet veel invloed op de PCs van de 1000 Genomes dataset (waarin de populaties relatief sterker van elkaar verschillen), maar zorgde er in de Nederlandse dataset voor dat de PCs aanzienlijk beter de genetische verschillen tussen Nederlanders oppikte. Het minimaliseren van LD verhoogde de correlaties tussen PCs en geografie binnen Nederland significant en resulteerde in drie PCs die genetische afkomst reflecteren: 1) de Noord-Zuid PC, die de verschillen

oppikt tussen de noordelijke provincies en de provincies ten zuiden van de drie grote rivieren (en in de Randstad zijn die noord-zuid verschillen meer met elkaar vermengd), 2) de Oost-West PC, die de verschillen tussen het noordoosten en de rest van het land reflecteert, en 3) de Midden-Strook PC, waaruit verschillen tussen een strook door het midden van het land en de rest van het land te zien zijn. De Noord-Zuid PC liet verscheidene overeenkomsten zien met Europese verschillen in genetische afkomst tussen noord en zuid: 1) een correlatie van .66 met de 1000 Genomes PC die Noord-Europa van Zuid-Europa van elkaar onderscheidt, 2) een significante correlatie met genoom-wijde homozygositeit (noorden = meer homozygoot), 3) een significante correlatie met lichaamslengte (noord = langer), 4) een signaal van selectiedruk op de SNP die bepalend is voor bruin/blauwe oogkleur (noord = meer blauwe ogen). De PCs konden ook gebruikt worden om SNPs te detecteren die onder selectiedruk hebben gestaan. Onder selectiedruk komen genetische varianten meer te verschillen tussen de Nederlandse subpopulaties dan het overgrote gedeelte van het genoom. Naast het signaal uit HERC2 (het gen verantwoordelijk voor bruin/blauw oogkleur, en het sterkste signaal) waren er nog 544 SNPs (uit 184 genen) die een signaal van selectiedruk suggereerden. Genen die een rol spelen in het brein waren significant oververtegenwoordigd in deze signalen.

In hoofdstuk 7 wordt de variatie van zeldzamer en moeilijker te meten genetische varianten verkend: indels (inserties en deleties < 20 baseparen) en grotere deleties (20 – 10,000 baseparen). Deze zijn in kaart gebracht met Next Generation Sequencing en zijn afkomstig uit een viertal Nederlandse biobanken in het *Genome of the Netherlands* (GoNL) project. Er zijn PCAs uitgevoerd op indels en op grotere deleties bij 490 ongerelateerde Nederlandse individuen met het doel variatie binnen Nederland in kaart te brengen. Indels lieten drie PCs zien die genetische afkomst reflecteren, en grotere deleties vijf. De indel PC die de meeste variatie verklaarde in genetische afkomst liet geografisch dezelfde Noord-Zuid distributie zien als de SNPs in hoofdstuk 6. De andere twee indel PCs lieten een geografische verdeling zien die sterk leek op de Oost-West SNP PC. De vijf PCs van de PCA op grotere deleties lieten significante maar lagere correlaties met de SNP PCs uit hoofdstuk 6 zien (significante correlaties tussen .16 en .30). Er was geen duidelijke geografische verdeling zien op de Nederlandse kaart (al lieten ze wel significante correlaties tussen .18 en .26 zien met geografie gebaseerd op geboorteplaats). PCs van grotere deleties zijn ook de enige PCs die geen significante correlatie binnen ouderparen laten zien. Het is nog niet duidelijk of de grotere deleties andere genetische afkomstverschillen oppikken, of dat de signalen zwakker zijn vanwege de relatief weinig betrouwbaar gemeten grotere deleties. Een aantal indel en deletie PCs lieten net als de Noord-Zuid SNP PC overeenkomsten zien met Europese Noord-Zuid verschillen (significante correlaties met de Europese Noord-Zuid PC

uit de 1000 Genomes dataset, homozygositeit, lichaamslengte en haarkleur).

Deel IV – Runs of Homozygosity: De invloed van het gedrag van (voor)ouders op de huidige genetische samenstelling

In hoofdstuk 8 worden opeenvolgende reeksen van homozygote varianten onderzocht, ook wel *Runs of Homozygosity* (ROHs) genoemd. Een stuk genoom is homozygoot als er voor dat stuk identieke kopieën van genetische varianten op beide chromosomen aanwezig zijn. Omdat dit betekent dat een kind van beide ouders dezelfde varianten heeft geerfd, is de proportie van het genoom dat uit ROHs bestaat (ook F_{roh} genoemd) groter naarmate ouders meer verwant zijn. F_{roh} laat een significante associatie met religiositeit zien, wat waarschijnlijk verklaard kan worden door demografische en historische factoren. De geografische verdeling van de Noord-Zuid PC in hoofdstuk 6 (het sterkste patroon van verschillen in genetische afkomst binnen Nederland) komt overeen met de geografische verdeling van de twee religieuze groepen in Nederland (katholieken en protestanten) die ongeveer vier eeuwen vrij stabiel is gebleven. In de huidige samenleving is dezelfde distributie nog steeds zichtbaar, maar met een toename van niet-religieuze individuen door de toenemende secularisatie in de afgelopen halve eeuw. De correlatie binnen ouderparen voor religie is zeer sterk (.73), in lijn met de hoge correlatie tussen ouders voor de Noord-Zuid PC (.56) en de historische documentatie van de partnerkeuze van katholieken en protestanten (“*twee geloven op één kussen, daar slaapt de duivel tussen*”). Post-hoc analyses wezen uit dat de associatie tussen religie en F_{roh} verklaard kan worden door het feit dat het niet-religieuze deel van de bevolking minder homozygote varianten heeft. Dit komt hoogstwaarschijnlijk omdat niet-religieuzen een minder beperkte partnerkeuze hebben dan katholieken en protestanten, waardoor genetische verschillen met hun partner groter kunnen zijn. In de niet-religieuze groep waren significant meer mensen met een klinische depressie, waardoor er een significante associatie leek te zijn tussen F_{roh} en depressie. Deze indirecte associatie verdwijnt na corrigeren voor religie. De PCs zoals beschreven in hoofdstuk 6 waren niet voldoende om te corrigeren voor deze oneigenlijke correlatie.

In hoofdstuk 9 wordt een significant verband tussen F_{roh} en opleidingsniveau gerapporteerd. Deze associatie is niet te wijten is aan causale ROHs die de kans verkleinen op een hoger opleidingsniveau, maar aan het gedrag van de (voor)ouders. F_{roh} in het nageslacht is veel sterker geassocieerd met het opleidingsniveau van hun ouders dan van het nageslacht zelf: ouders met een hoger opleidingsniveau hebben kinderen met een lagere F_{roh} . Voor ouders met een hoger opleidingsniveau was er een grotere afstand tussen hun eigen geboorteplaats en de geboorteplaats van hun kinderen of echtgeno(o)t(e). De afstand tussen de geboorteplaats van de vader en die van de moeder liet ook een significante associatie zien met F_{roh} (grotere

afstand = lagere F_{roh}). Als er gecorrigeerd wordt voor deze afstand, verdwijnt ook het significante verband tussen het (ouderlijke) opleidingsniveau en F_{roh} . Dit geeft aan dat de associatie tussen F_{roh} en (ouderlijke) opleidingsniveau ontstaat omdat hoger opgeleide ouders vaker en verder migreren. Hoger opgeleide ouders kiezen vaker een partner die ook hoger opgeleid is, waardoor de kans groter wordt op een partner die zelf ook meer mobiel is en uit een geografische regio komt met een andere genetische achtergrond. Een andere aanwijzing hiervoor is dat kinderen van hoger opgeleide ouders een lagere correlatie laten zien tussen de PCs die genetische afkomst reflecteren en geografie.