

Genetic risk and architecture of Amyotrophic Lateral Sclerosis

Proefschrift Wouter van Rheenen

Nederlandse samenvatting

Amyotrofische laterale sclerose (ALS) is een snel progressieve neurodegeneratieve aandoening, waarmee in Nederland jaarlijks 400 tot 500 mensen worden gediagnosticeerd. De *motorische neuron*en in de hersenen en het ruggenmerg sterven af, waardoor patiënten toenemend verlamd raken. ALS is een ongeneeslijke ziekte en het enige beschikbare medicijn, riluzole, heeft maar een beperkt ziekte remmend effect. Gemiddeld is de levensverwachting van patiënten met ALS ongeveer 3 jaar.

De exacte oorzaak van ALS is niet bekend, maar het is duidelijk dat meerdere factoren een rol spelen. Uit tweelingonderzoek blijkt dat *genetische risicofactoren* een belangrijke bijdrage leveren: ongeveer 61% van het risico op ALS is genetisch bepaald. Daarnaast komt bij ongeveer 10 - 15% van de patiënten met ALS de ziekte in de familie voor, hetgeen duidt op een belangrijke genetische component.

De afgelopen jaren zijn er verschillende *genen* ontdekt waarin *mutaties of polymorfismen* het risico op ALS kunnen verhogen. Genetische variatie in sommige van die genen leidt bijna altijd tot ALS, waardoor deze variatie vaak wordt gezien in de familiale vorm. Voorbeelden hiervan zijn mutaties in het *SOD1*, *TARDBP* en *FUS* gen. Echter, bij de meeste ALS patiënten komt de ziekte niet in de familie voor en zijn er geen mutaties in deze genen. De gedachte is dat bij deze patiënten een combinatie van genetische en niet-genetische risicofactoren tot de ziekte hebben geleid. Een enkele risicofactor maakt niet direct ziek, maar de combinatie van factoren leidt wel tot ALS. Daarom wordt ALS een complexe genetische ziekte genoemd.

Het in kaart brengen van deze risicofactoren is van groot belang om de ziekte ALS beter te begrijpen en in de toekomst te kunnen behandelen. Elk van deze risicofactoren vormt namelijk een potentieel aangrijpingspunt voor nieuwe therapieën. Ik richt me daarom in mijn proefschrift op het zoeken naar genetische risicofactoren van ALS.

Kandidaatgen studies

In de afgelopen jaren zijn er reeds enkele genen gevonden waarin mutaties het risico op ALS verhogen. Een van de belangrijkste genen is *C9orf72*. Het *locus* waar dit gen in ligt was reeds jaren bekend voordat de pathogene *repeat expansie* werd gevonden die daadwerkelijk tot ALS kan leiden. In **hoofdstuk 2** beschrijf ik het vóórkomen van de pathogene repeat expansie in *C9orf72* in de Nederlandse ALS patiënten, maar ook in patiënten met PSMA en PLS,

waarbij enkel het respectievelijk perifeer en centraal motorisch neuron zijn aangedaan. We vonden de repeat expansie in 42.3% van de families met familiale ALS in Nederland. Daarnaast draagt 6.4% van de sporadische ALS patiënten, 1.6% van de patiënten met PSMA en 0.9% van de patiënten met PLS deze mutatie. Daarentegen kwam de repeat expansie niet voor in de 768 gescreende gezonde controlepersonen. Hiermee is de repeat expansie in *C9orf72* de belangrijkste oorzaak van ALS in Nederland. Het is opvallend dat de repeat expansie zo vaak voorkomt bij patiënten zonder dat er ALS in de familie voorkomt. Hiervoor zijn verschillende verklaringen. Naast incomplete *penetrantie* zouden de *pleiotrope effecten* van de repeat expansie het hoge aantal sporadische ALS patiënten kunnen verklaren. Familieleden die de mutatie dragen kunnen namelijk niet alleen ALS krijgen, maar ook frontotemporale dementie. Dit wordt ondersteund door het feit dat we beschrijven dat familieleden van patiënten met de *C9orf72* repeat expansie vaker dement zijn, in vergelijking met familieleden van ALS patiënten zonder de repeat expansie of gezonde controles. Doordat de ziekte zich op verschillende manieren kan uiten binnen een familie, wordt deze vaak niet herkend als familiair voorkomende ziekte.

In **hoofdstuk 3** beschrijf ik *polyalanine repeat* expansies in *NIPA1* als risicofactor voor ALS. Het *NIPA1* gen was reeds bekend als oorzaak van een aanverwante neurodegeneratieve aandoening, *erfelijke spastische paraplegie type 6*, en deleties van dit gen verhogen het risico op ALS. In 2,292 ALS patiënten en 2,777 gezonde controlepersonen werden aminozuurveranderende mutaties en de lengte van de polyalanine repeat in *exon 1* onderzocht. Het bleek dat een polyalanine repeat met 14 of 15 alanines vaker voorkomt bij patiënten met ALS. Daarnaast begint de ziekte bij deze patiënten op jongere leeftijd en is de overleving gemiddeld korter. Vanuit het onderzoek naar erfelijke spastische paraplegie type 6 is bekend dat *NIPA1* een rol speelt bij *axonaal transport* en groei van de *neuromusculaire overgang* hetgeen wordt gemedieerd door de *BMP-cascade*. Dit maakt de BMP-cascade een interessant proces om de effecten van de polyalanine repeat expansie in *NIPA1* op het functioneren van motorische neuronen te bestuderen.

Kandidaatgen studies die een enkel gen bestuderen hebben belangrijke beperkingen. Er wordt zelden voldaan aan strikte statistische criteria om te bewijzen dat een gen daadwerkelijk het risico op een ziekte verhoogt. Daarnaast is het aantal patiënten en controlepersonen dat onderzocht wordt vaak beperkt. De betrouwbaarheid van de resultaten en daarmee de kans op een fout positieve associatie, hangt daarom sterk af van het bewijs dat reeds voor aanvang van de studie is geleverd voor het kandidaatgen in kwestie. In **hoofdstuk 4** toon ik aan dat de vermeende associatie tussen het H63D polymorfisme in *HFE* berust op verschillende fout positieve studies. In een grote groep ALS patiënten en controlepersonen bleek er geen relatie te bestaan tussen het H63D polymorfisme en het risico op ALS. Ook de meta-analyse van alle

reeds gepubliceerde studies met onze eigen resultaten toonde geen statistisch significante associatie. Noch had dit polymorfisme een effect op het ziektebeloop.

In **hoofdstuk 5** en **6** houd ik het bewijs voor de associatie tussen *CHCHD10* mutaties en het risico op ALS tegen het licht. Het is bekend dat één bepaalde *CHCHD10* mutatie een ziekte geeft die sterk kan lijken op ALS, maar tevens andere verschijnselen zoals een *myopathie* en doofheid kent. Na de beschrijving van deze mutatie verschenen er verschillende publicaties waarin andere mutaties in *CHCHD10* in andere families als oorzaak van ALS werden beschreven. In **hoofdstuk 5** laat ik zien dat het geleverde bewijs dat wordt geleverd in deze vervolgstudies, methodologisch van onvoldoende kwaliteit is. In **hoofdstuk 6** blijkt dat er voor de meeste beschreven mutaties onvoldoende bewijs voor *pathogeniciteit* is. De mutaties die mogelijk pathogeen zijn, zijn beschreven in ALS patiënten met atypische kenmerken. Derhalve dienen *CHCHD10* mutaties, in de afwezigheid van deze atypische kenmerken, niet zonder meer te worden beschouwd als een oorzaak van ALS.

Genoomwijde studies

In complexe genetische ziekten zijn genoomwijde associatiestudies zeer succesvol geweest in het identificeren van nieuwe genetische risicofactoren. Deze studies vergelijken het voorkomen van genetische polymorfismen in patiënten en controles over het gehele *genoom*. De loci waar bepaalde polymorfismen vaker voorkomen bij patiënten dan bij controles helpen zo bij het identificeren van nieuwe genetische risicofactoren.

Waar deze studies in ziekten zoals schizofrenie, de ziekte van Alzheimer en auto-immuunziekten tientallen tot honderden loci hebben beschreven, kenden we er voor ALS maar drie (*C9orf72*, *UNC13A*, *SARM1*). Het aantal patiënten en controles dat in deze studies in ALS tezamen is geïncludeerd is relatief klein (~6,000 patiënten en ~15,000 controles). **Hoofdstuk 7** beschrijft de tot op heden grootste genoomwijde associatiestudie in ALS met 15,156 patiënten en 26,242 controles. In totaal werden, na *imputatie* met een ALS specifiek imputatie panel, 8,697,640 polymorfismen vergeleken. Zo vonden we drie nieuwe loci (*C21orf2*, *MOBP* en *SCFD1*).

Het polymorfisme in *C21orf2* betrof een relatief zeldzame genetische variant welke resulteerde in een aminozuur verandering in het *C21orf2* gen. Vervolgens toonden we aan dat er naast dit ene polymorfisme meer aminozuur veranderingen in *C21orf2* voorkomen bij ALS patiënten vergeleken met controles. Dit maakt *C21orf2* tot een nieuw "ALS gen". Alhoewel er niet veel bekend is over de functie van *C21orf2* is wel bekend dat het direct

interacteert met een ander recent ontdekt ALS gen: *NEK1*. Samen spelen zij een rol in diverse cellulaire processen zoals DNA-reparatie en axonaal transport.

Binnen het *MOBP* locus liggen meerdere genen en er is nog niet duidelijk welk gen hier precies het risico op ALS verhoogt. Opvallend genoeg wordt exact hetzelfde locus gevonden in een genomwijde associatiestudie naar *progressieve supranucleaire paralyse* (PSP) en *corticobasale degeneratie* (CBD). Daarnaast zijn de polymorfismen in dit locus geassocieerd met de overleving van patiënten met frontotemporale dementie. Blijkbaar speelt genetische variatie in dit locus een belangrijke rol in verschillende neurodegeneratieve ziekten.

Naast het feit dat genomwijde associatiestudies nieuwe genetische risicofactoren identificeren, kunnen genomwijde analyses de genetische architectuur van een aandoening beschrijven. De genetische architectuur van een ziekte wordt in engere zin bepaald door het aantal, de frequentie en de effectgrootte van genetische risicofactoren. In bredere zin kan hierbij ook gedacht worden aan de verschillende vormen van genetische variatie, *genetische heterogeniteit* en pleiotropie.

In **hoofdstuk 7** tonen we middels verschillende methoden aan dat ALS een *polygene architectuur* kent. De genetische architectuur verschilt echter van andere polygene aandoeningen, zoals schizofrenie en multipele sclerose, aangezien het risicoprofiel van ALS met name bepaald wordt door laagfrequente (1 - 10%) en mogelijk zeldzame (< 1%) genetische variatie. Al met al kan 8.5% van het risico op ALS worden verklaard door genetische variatie met een allelfrequentie > 1%. Opvallend genoeg wordt maar 0.2% verklaard door de 6 genomwijd significante loci. Er zijn dus nog vele loci die bijdragen aan het risico op ALS en geïdentificeerd kunnen worden in nog grotere genomwijde associatiestudies.

De genetische architectuur van ALS wordt verder beschreven in **hoofdstuk 8**, waarin we de genetische overlap tussen ALS en verschillende psychiatrische ziektebeelden beschrijven. Uit drie verschillende analyses blijkt dat er een *genetische correlatie* bestaat tussen ALS en schizofrenie. Deze correlatie wordt exclusief gevonden voor ALS en schizofrenie, maar niet voor andere psychiatrische aandoeningen. Daarnaast kan deze correlatie niet worden verklaard door misdiagnoses waarbij ALS patiënten ten onrechte met schizofrenie worden gediagnosticeerd of andersom. Tenslotte vinden we 5 loci (*CNTN6*, *TNIP1*, *PPP2R2D*, *NCKAP5L*, *ZNF295-AS1*) die bijdragen aan deze genetische correlatie.

Alhoewel het bestuderen van genetische polymorfismen in grotere genomwijde associatiestudies meer genetische risicofactoren voor ALS zal opleveren, wordt de zeldzame genetische variatie, die in ALS een belangrijke rol speelt, niet direct bestudeerd. Dit gebeurt

wel in *whole genome sequencing* studies waarbij alle genetische variatie over het hele frequentiespectrum in kaart wordt gebracht. Daarom is het doel van Project MinE om het hele genoom van 15,000 ALS patiënten en 7,500 controlepersonen te bestuderen middels *whole genome sequencing* en zo de zeldzame genetische risicofactoren van ALS in kaart te brengen. In **hoofdstuk 9** beschrijf ik het ontwerp van Project MinE en de eerste analyses van 1,264 ALS patiënten en 611 controlepersonen uit Nederland. Uit deze analyses blijkt dat het van groot belang is om te beschikken over een adequate ICT-infrastructuur om de honderden terabytes aan data te analyseren. Daarnaast blijkt dat de zeldzame genetische variatie zeer geografisch geclusterd voorkomt en het daarom van belang is dat de ALS patiënten en gezonde controlepersonen geografisch gelijk zijn verdeeld. Tenslotte tonen we aan dat de technische en analytische aspecten van het *whole genome sequencing* het includeren van externe controles, zoals gebruikelijk in de genomwijde associatiestudies, kunnen bemoeilijken. Op basis van deze bevindingen komen we met een voorstel hoe de volledige data van Project MinE in de toekomst te geanalyseerd kan worden.

De verschillende genetische risicofactoren die over de afgelopen jaren zijn ontdekt hebben talloze vervolgstudies geïnitieerd welke belangrijke ziekteprocessen in ALS beschrijven. Daarnaast is de genetische counseling van ALS patiënten en hun families sterk verbeterd. Bij het stellen van de diagnose en de keuze van een behandeling speelt genetica nog een beperkte rol. In **hoofdstuk 10** beschrijf ik de ontwikkeling van een diagnostische *biomarker*, gebaseerd op *genexpressie* profielen. Alhoewel het mogelijk is om op basis van *genexpressie* modellen een onderscheid te maken tussen ALS patiënten en gezonde controles, blijken deze modellen niet te kunnen discrimineren tussen ALS patiënten en patiënten die een ziekte hebben die lijkt op ALS, zoals die worden gezien in de klinische praktijk. Daarnaast zijn er veel versturende, technische artefacten die de externe validiteit van de ontwikkelde modellen beperken. Desalniettemin bevatten *genexpressie* profielen in het bloed van ALS patiënten informatie die, wellicht met technieken zoals *RNA-seq*, wel zou kunnen bijdragen aan een valide biomarker voor ALS.

In **hoofdstuk 11** vat ik mijn bevindingen samen, bediscussieer ik de inzichten die deze hebben opgeleverd en de richting waarin het genetisch onderzoek naar ALS zich beweegt.

Conclusies

Met de onderzoeken die ik heb beschreven in dit proefschrift kom ik tot de volgende conclusies:

1. Alhoewel deze niet volledig penetrant is, is de repeat expansie in *C9orf72* de belangrijkste oorzaak van ALS in Nederland.
2. Polyalanine repeat expansies in *NIPA1* verhogen het risico op ALS en geven een agressiever ziektebeloop.
3. Het H63D polymorfisme in *HFE* is niet geassocieerd met het risico op ALS en beïnvloedt het ziektebeloop niet.
4. *CHCHD10* mutaties dragen veel minder bij aan het risico op ALS dan is beschreven in de literatuur.
5. Laagfrequente, coderende, variatie in *C21orf2* verhogen het risico op ALS.
6. Naast het *UNC13A* en *SARM1* locus dragen het *MOBP* en *SCFD1* locus bij aan het risico op ALS.
7. De genetische architectuur van ALS kent een polygene component met een belangrijke rol voor laagfrequente genetische variatie.
8. Deze polygene component wordt gedeeltelijk gedeeld tussen ALS en schizofrenie.
9. Genexpressie profielen uit bloed zijn nog niet geschikt als een diagnostische biomarker voor ALS.

Verklarende woordenlijst

Axonaal transport Het transport door de lange uitlopers van (motorische) neuronen.

Biomarker. Een test waarmee bijvoorbeeld een diagnose of prognose voorspeld kan worden.

BMP-cascade. Groep van eiwitten dat in zenuwcellen zorg voor het onderhoud van de neuromusculaire overgang en axonaal transport. BMP = bone morphogenetic proteins, naar de oorspronkelijke ontdekking dat deze eiwitten een rol spelen bij de botformatie.

Corticobasale degeneratie Neurodegeneratieve aandoening die klinische kenmerken deelt met de ziekte van Parkinson.

Erfelijke spastische paraplegie. Erfelijke aandoening waarbij patiënten met name aan de benen toenemen verlamd en spastisch raken. Elk type wordt veroorzaakt door een andere mutatie.

Exon. Deel van een gen dat (in tegenstelling tot het intron) codeert voor een eiwit.

Genen. Deel van het *genoom* dat codeert voor een eiwit. Het hele genoom telt ongeveer 25,000 genen.

Genetische correlatie. Maat waarmee wordt uitgedrukt hoeveel van de genetische risicofactoren worden gedeeld door twee aandoeningen.

Genetische heterogeniteit. De mate waarin een ziekte wordt bepaald door verschillende genetische profielen.

Genetische risicofactoren. Variaties in het DNA, het erfelijke materiaal, die de kans op het krijgen van ALS vergroten. De gedachte is dat deze variatie, ook wel mutaties of polymorfismen, (zenuw)cellen kwetsbaarder maken.

Genexpressie. De mate waarmee het DNA van genen wordt omgezet in RNA, een tussenproduct voor het uiteindelijke eiwit. Door verschillen in genexpressie kunnen cellen van elkaar verschillen en kan een cel reageren op externe factoren zoals een ziekte.

Genoom. De volledige genetische code van een persoon (of organisme), ofwel het DNA. Deze code bestaat uit 4 basen guanine (G), cytosine (C), thymine (T) en adenine (A).

Imputatie Het voorspellen van genetische variatie die niet direct wordt geobserveerd in de studie. Hierbij maakt men gebruik van de patronen waarin genetische variatie over het hele genoom voorkomt. Deze patronen worden in een zogenaamd imputatie panel bepaald. In dit panel worden veel meer polymorfismen gezien dan in de genomwijde associatiestudie. Op

basis van deze patronen kunnen polymorfismen die niet worden gezien in de genomwijde associatiestudie worden voorspeld. Imputatie maakt het mogelijk om vele miljoenen polymorfismen te bestuderen in plaats van de 200,000 – 500,000 die er direct worden geobserveerd.

Locus. Gebied in het genoom. Een locus kan meerdere genen bevatten.

Meta-analyse. Statistische methode waarbij de resultaten van verschillende studies worden gecombineerd.

Motorische neuronen. De zenuwcellen in de hersenen en het ruggenmerg die de spieren aansturen en zo beweging mogelijk maken.

Mutaties. Verandering in het DNA, die vaak wordt geassocieerd met een ziekte.

Myopathie. Spierziekte.

Neuromusculaire overgang Punt waarop de motorische zenuwcellen contact maken met de spier en waar de activiteit in de zenuwcel wordt omgezet in een spiercontractie.

Pathogeniciteit. De eigenschap van een mutatie om een ziekte te veroorzaken.

Penetrantie. De kans dat iemand met een bepaalde mutatie daadwerkelijk de ziekte krijgt (op een bepaalde leeftijd). Incomplete penetrantie wil zeggen dat niet iedereen met de mutatie ziek wordt.

Pleiotrope effecten. Mutaties die verschillende ziekten kunnen veroorzaken, zoals ALS en/of frontotemporale dementie, hebben zogenaamde pleiotrope effecten.

Polyalanine repeat. Deel van het DNA dat codeert voor een deel van het eiwit met enkel alanine aminozuren. Zo bestaat de polyalanine repeat in *NIPA1* uit 7 of 8 GCG-elementen die coderen voor een alanine aminozuur, de verlengde repeats hebben 9 of 10 GCG-elementen.

Polygene architectuur Er zijn vele genetische risicofactoren nodig om de ziekte te veroorzaken.

Polymorfismen. Veel voorkomende verandering in het DNA die over het algemeen wordt gezien in 1% van de bevolking. Deze polymorfismen kunnen het risico op bepaalde ziekten verhogen.

Progressieve supranucleaire paralyse Neurodegeneratieve aandoening die klinische kenmerken deelt met de ziekte van Parkinson.

Repeat (expansie). Zich herhalend patroon in het DNA. Zo bestaat de *C9orf72* repeat normaal gesproken uit maximaal 30 elementen van het GGGGCC-patroon, wanneer de repeat bestaat uit meer dan 30 (vaak tot honderden) GGGGCC-elementen wordt dit een repeat expansie genoemd.

RNA-seq. Methode om genexpressie te meten die veel overeenkomsten kent met whole genome sequencing.

Whole genome sequencing Methode waarbij de volledige genetische code van een persoon wordt bepaald. Tegenwoordig gaat dit in een fractie van de tijd en kosten vergeleken met het allereerste genoom dat ooit in kaart werd gebracht (2003: 2.7 miljard dollar, 13 jaar. 2016: ~1,500 dollar, enkele uren).