

‘BiG Grid and beyond’ Gedrag van tweelingen onder de loep

Het lijkt alsof we alles al weten over DNA. ‘Je kunt in DNA zelfs zien uit welke provincie iemand oorspronkelijk komt’, vertelt Mathijs Kattenberg wetenschappelijk programmeur aan de Vrije Universiteit in Amsterdam. Toch is het omgekeerde waar, zo blijkt bijvoorbeeld uit het DNA van tweelingen. ‘Eeneiige tweelingen zijn genetisch identiek, maar je ziet dat hun gedrag niet altijd identiek is.’ Met data van 100.000 tweelingen pluist Kattenberg het nature-nurture debat uit.

Dit onderzoek naar de geestelijke en lichamelijke gezondheid van tweelingen, is het Nederlands Tweelingen Register (NTR) en is een initiatief van de afdeling biologische psychologie van de Vrije Universiteit in Amsterdam. Het onderzoek loopt al 25 jaar en inmiddels doen er zo’n 100.000 tweelingen aan mee. ‘Die sturen we ieder jaar een vragenlijst’, vervolgt Kattenberg. ‘Van ongeveer 13.000 individuen hebben we DNA-data en van een aantal tweelingen hebben we daarnaast gegevens over een hele lange periode. Dat is een uitzonderlijk mooie dataset.’

Genoom bepalen

Maar ook een hele grote dataset. ‘Data is een bottleneck. Met de computer die we op de VU hadden, konden we nooit grote hoeveelheden data runnen of associaties doen. Dat zou vijftig jaar duren, maar dankzij BiG Grid duurt het slechts enkele weken.’

Kattenberg onderhoudt de databases en de software voor onderzoeksadministratie, zoals de adresregistratie en het versturen van de vragenlijsten. Ook houdt hij zich bezig met de DNA-data. ‘In het laboratorium worden de data gegenotypeerd, zodat we het genoom kunnen bepalen. Daaruit komen gigantische databestanden, die wij en andere onderzoekers vervolgens analyseren. Voor het NTR kijken we naar de variatie in het genetische materiaal. Dus in hoeverre verschillen mensen van elkaar. Dat doen we door die DNA-data tegen een referentiegenoom te houden, zodat je ziet op welke punten een persoon verschillend is.’

Afwijkingen

‘Omdat we de DNA-data en vragenlijsten van een individu hebben, kunnen we zien wat hun score is op een bepaald fenotype. Dat kan biologisch zijn – bijvoorbeeld bij stoffen voor de stofwisseling – of psychologisch bij een depressie of ADHD. We houden de DNA-gegevens van de personen tegen de fenotypes. Met die associaties kunnen we zien of het fenotype correleert met bepaalde afwijkingen in het DNA.’

Controversieel

Een ander belangrijk onderdeel is de imputatie van data; de dataprocessing voordat analyse kan plaatsvinden. ‘In de inputgegevens zitten gaten die niet gemeten worden. Met een referentieset van dezelfde populatie proberen we die missende gegevens in te vullen met schattingen. Het controversiële daaraan is dat je vanuit die imputatie zeldzame DNA-varianten kunt vinden. Die heb je dus niet gemeten, maar zijn wel de uitkomst van de imputatie en kunnen een rol spelen bij bepaalde erfelijke aandoeningen.’

High-end publicaties

Via verschillende consortia wereldwijd neemt het NTR ook deel aan meta-analyses. ‘Er zijn publicaties waar wel 300 onderzoekers aan mee doen, waarvan veel uit Australië en de Verenigde Staten. Gezien de grote hoeveelheid gegevens is dit niet voor iedere onderzoeksgroep weggelegd. Zonder de BiG Grid infrastructuur kunnen we niet deelnemen aan dit soort high-end publicaties.’

