



Universiteit  
Leiden  
The Netherlands

## Multi-omics in research: epidemiology, methodology, and advanced data analysis

Faquih, T.O.

### Citation

Faquih, T. O. (2023, March 28). *Multi-omics in research: epidemiology, methodology, and advanced data analysis*. Retrieved from <https://hdl.handle.net/1887/3589838>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3589838>

**Note:** To cite this publication please use the final published version (if applicable).

# *Chapter 9*

## *Summary of Findings*

### *(Dutch, English, Arabic)*



## SAMENVATTING

De recente ontwikkelingen in metabolomics (de studie van metabolieten), proteomics (de studie van eiwitten) en genomics (de studie van genen) geven diepgaande inzichten in de menselijke biologie en de pathofysiologie van ziekten. In dit proefschrift onderzochten we de methodologische uitdagingen waar deze OMICs-technologieën voor staan en pasten we deze toe in epidemiologische studies.

In deel één concentreerden we ons op enkele van de methodologische uitdagingen waar OMICs-onderzoek voor staat. In **Hoofdstuk 2** bestudeerden we de overeenkomst tussen proteomics metingen van biomarkers voor veneuze trombose op basis van aptameren en de gestandaardiseerde klinische metingen. Voor negen onderzochte veneuze trombose biomarkers, vonden we dat er weinig overeenkomst was tussen de twee methoden. De overeenkomst was in het bijzonder slecht voor D-dimeer, een marker die vaak gebruikt wordt in veneuze trombose diagnostiek. In **Hoofdstuk 3** hebben we onderzocht hoe om te gaan met ontbrekende waarden in metabolomics data, vooral bij gebruik van ‘untargeted metabolomics’. We hebben de prestaties van twee eerder gerapporteerde imputatie methoden beoordeeld, door het simuleren van ontbrekende data in metingen van de NEO-studie, namelijk meervoudige imputatie en “*k*-nearest neighbor”. Onze resultaten lieten zien dat de meervoudige imputatie methode in de meeste scenario’s resulteerde in minder bias dan de *k*-nearest neighbor methode, behalve in het scenario met kleine steekproefgrootte en veel ontbrekende waarden. We hebben een openbaar beschikbaar R-script gemaakt om het imputeren van ontbrekende metabolomics data met behulp van deze twee imputatie technieken te versnellen.

Metabolieten kunnen unieke inzichten geven in de pathofysiologie van ziekten, aangezien ze de samengestelde effecten omvatten van genetische, leefstijl- en omgevingsfactoren. Daarom wordt metabolomics ook gebruikt om bijvoorbeeld biologische processen in veroudering te bestuderen. Sommige studies hebben metabolomics data gebruikt, om met behulp van voorspelmodellen de “metabolomische leeftijd” te schatten, een weerspiegeling van veroudering op het niveau van het metabolisme. In **Hoofdstuk 4** bespreken we veelvoorkomende uitdagingen bij het ontwikkelen van voorspelmodellen op basis van metabolomics data. Daarnaast hebben we zelf een model ontwikkeld om de metabolomische leeftijd te voorspellen op basis van metabolieten gemeten met het Metabolon platform. We ontwikkelden het model in de INTERVAL-studie met behulp van ‘ridge regressie’ en ‘bootstrapping’. Deze populatie was uitermate geschikt voor het ontwikkelen van een model omdat het een relatief gezonde, grote populatie is, ( $n = 11.977$ ) met zeer uiteenlopende leeftijden. Uiteindelijk was ons model goed in staat om de metabolomische leeftijd te voorspellen (gecorrigeerde  $R^2 = 0.83$ ).

Het tweede deel van dit proefschrift richtte zich op verschillende epidemiologische onderzoeksvragen waarbij gebruik gemaakt wordt van genomics and metabolomics data (Metabolon en Nightingale platforms), en het toepassen van geavanceerde data-analyse methoden. In **Hoofdstuk 5** onderzochten we de associaties tussen meer dan 1300 metabolieten gemeten met het Metabolon platform en levervet. We vonden associaties voor een reeks endogene metabolieten (vooral aminozuren en lipiden), evenals voor nieuwe en ongedefinieerde metabolieten. Daarnaast hebben we twee complementaire holistische netwerken van metabolieten geconstrueerd, één op biologische basis en één op statistische basis, om de geassocieerde metabolieten toe te wijzen aan biologische ‘pathways’. Deze interactieve netwerken gaven inzicht in potentiële nieuwe pathways die een relatie hebben met levervet en mogelijk met leverziekten. Deze interactieve netwerken hebben we op een online, publiek beschikbare webpagina, beschikbaar gesteld voor verdere verkenning door de onderzoeksgemeenschap.

Een type variatie in het genoom is variatie in het aantal herhaalde cytosine-adenine-guanine-nucleotidesequenties (CAG-herhalingen). Wanneer het aantal CAG-herhalingen boven een specifieke pathogene drempelwaarde ligt, en deze in of in de buurt van specifieke genen liggen, kunnen ze tot ernstige neurodegeneratieve ziekten leiden. Een voorbeeld hiervan is het aantal CAG-herhalingen in het Huntingtine-gen (*HTT*-gen) dat leidt tot het optreden van de ziekte van Huntington. Uit recent onderzoek is gebleken dat het aantal CAG-herhalingen, zelfs onder de pathogene drempelwaarde, gerelateerd is aan gezondheidsuitkomsten, zoals cognitieve functie, depressie en gewicht. In **Hoofdstuk 6** hebben we de mogelijke associaties verkend tussen variatie in het aantal CAG-herhalingen onder de pathogene drempel in het *HTT*-gen en metabool concentraties. Hiervoor hebben we gebruik gemaakt van de genetische data van 3 grote Europese cohort studies ( $n = 10,275$ ) met metabool data van het Nightingale platform. We hebben gevonden dat het aantal niet-pathogene CAG-herhalingen in het *HTT*-gen geassocieerd was met een ongezond metabool profiel dat vergelijkbaar is met dat van patiënten met cardiometabole ziekten.

Eén van de belangrijkste voordelen van metabolomics is dat het ook mogelijk is om metabool te meten die oorspronkelijk uit de omgeving komen, ook wel xenobiotica genoemd, zoals uit medicatie, voedingsstoffen, en vervuilde stoffen uit het milieu. Een bekende groep xenobiotica zijn de per- en polyfluoralkylstoffen (PFAS), die door de mens zijn gemaakt. Deze stoffen worden ook wel ‘forever chemicals’ genoemd, omdat ze heel lang in het milieu en in het lichaam blijven en nauwelijks op natuurlijke wijze worden afgebroken. Hoge concentraties van PFAS kunnen tot ziekte leiden. Ondanks maatregelen om de productie te verminderen, blijven PFAS meetbaar in de grond en het drinkwater, alsmede ook in het bloed van inwoners van Nederland en Duitsland. In **Hoofdstuk 7**, hebben we de associaties bestudeerd tussen PFAS concentraties in het bloed, gemeten met het Metabolon platform, en metabool/lipoproteïnen, gemeten met het Nightingale platform in twee cohorten, één uit Nederland (NEO Studie) en één uit Duitsland (Rhineland Studie). PFAS was meetbaar in nagenoeg alle deelnemers van deze studies. Uit de meta-analyse bleek daarnaast dat PFAS concentraties geassocieerd waren met een ongezond metabolisch profiel, met verhoogde zuren, low-density lipoproteïne en apoproteïne-B concentraties. Deze associatie was het sterkst in jongere mensen. Dit profiel is zeer vergelijkbaar met het profiel dat we kennen van individuen met een verhoogd cardiometabool risico. Dit geeft aan dat PFAS een gezondheidsrisico zijn, zelfs in mensen die schijnbaar niet aan hoge concentraties PFAS blootgesteld zijn.