



Universiteit
Leiden
The Netherlands

Data-driven knowledge discovery in polycystic kidney disease

Malas, T.

Citation

Malas, T. (2021, March 30). *Data-driven knowledge discovery in polycystic kidney disease*. Retrieved from <https://hdl.handle.net/1887/3158169>

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/1887/3158169>

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

Cover Page



Universiteit Leiden



The handle <https://hdl.handle.net/1887/3158169> holds various files of this Leiden University dissertation.

Author: Malas, T.

Title: Data-driven knowledge discovery in polycystic kidney disease

Issue Date: 2021-03-30

APPENDIX

Nederlandse Samenvatting

In de afgelopen 20 jaar zijn de medische wetenschappen getuige geweest van belangrijke resultaten die werden gekenmerkt door de voltooiing van het menselijk genoomproject in 2003. Vanaf dat moment daalden de kosten van DNA-sequentiebepaling aanzienlijk en dit markeerde vervolgens de komst van het data-tijdperk in de medische wetenschappen. In dit proefschrift hebben we ons gericht op het onderzoeken van de voordelen van big data bij de behandeling van ziekten bij de mens, en specifiek op het onderzoeken van nieuwe behandelingen voor autosomaal dominante polycysteuze nierziekte (ADPKD). ADPKD is een ziekte die de nieren van de patiënt aantast en kenmerkt zich door met vloeistof gevulde cysten die zich in beide nieren vormen. ADPKD heeft momenteel zeer beperkte behandelingsopties; patiënten kunnen nierfalen ontwikkelen en hebben uiteindelijk vaak een niertransplantatie nodig. In hoofdstuk 2 van dit proefschrift zijn we begonnen met het combineren van de verschillende openbaar beschikbare expressieprofielen van ADPKD om te komen tot een uniform ADPKD-genexpressie profiel. Expressieprofielen worden gemaakt door de volledige set RNA-transcripten (tot expressie gebrachte genen) onder specifieke omstandigheden of in een specifieke cel vast te leggen - met behulp van high-throughput sequencing-methoden. Het hebben van een uniforme expressieprofiel vergroot ons begrip van de genen die bij de ziekte betrokken zijn en minimaliseert fouten die zouden kunnen voortkomen uit het gebruik van verschillende technologieën of verschillende diersystemen. Enkele experimenten zijn beperkt tot een specifieke technologie, diersysteem en / of ziekte-toestand. Het is waardevol om verschillende onderzoeken te combineren om deze beperkingen te overwinnen. In ons werk hebben we aangetoond dat het combineren van experimenten met verschillende ziektemodellen en sequentie technologieën een sterker expressieprofiel creëert met een grotere percentage van genen die betrokken zijn bij de ziekte. Verder hebben we in hoofdstuk 3 deze expressie-signatuur gebruikt om een klasse van genen genaamd transcriptiefactoren (TF's) en hun rol in ADPKD te bestuderen. Transcriptiefactoren zijn bijzonder interessant omdat ze in staat zijn de expressieniveaus van andere genen te reguleren. Het zijn daarmee belangrijke organisatoren van moleculaire gebeurtenissen die aanleiding kunnen geven tot de ontwikkeling van cysten bij ADPKD-patiënten. In hoofdstuk 3 ontdekten we twee TF's, STAT3 en RUNX1, die verder onderzocht zouden kunnen worden op hun rol in cystenvorming. We hebben ook een paar andere genen belicht die door deze twee TF's worden gereguleerd en als een mogelijk aangrijpingspunt kunnen dienen voor medicijnen voor ADPKD-patiënten. In hoofdstukken 2 en 3 combineerden we grote hoeveelheden transcriptomics-gegevens, mogelijk gemaakt door aanzienlijke vooruitgang in DNA-sequentie technologieën.

Verder hebben we in hoofdstuk 4 de sequentie-analyses uitgevoerd voor 24 nierweefsels uit ons ADPKD muismodel (Pkd1cko) in verschillende ziektestadia om de genen te bestuderen die betrokken zijn bij cystevorming en -progressie. We hebben de nieren op basis van de ernst van de ziekte in drie groepen ingedeeld; Vroege fase met nog bijna geen vorming van cysten, intermediaire fase met zichtbare cysten en eindfase met grote cysten die zich in de nieren vormen. We combineerden de gegevens van deze drie ziektefasen om te onderzoeken hoe veranderingen in genexpressie verband houden met de progressie van ADPKD, en kwam tot een reeks genen die betrokken zouden kunnen zijn bij de progressie van ADPKD, genaamd "ADPKD progressie profiel". Verder hebben we de chemo-informatica-database (ChEMBL) gebruikt om te identificeren welke medicijnen kunnen

worden gebruikt om belangrijke genen in het “ADPKD progressie profiel” te targeten en zo de vorming van cysten en ziekteprogressie te vertragen of te stoppen. Dit werk nomineerde verschillende nieuwe kandidaat-geneesmiddelen om ADPKD te behandelen, zoals Birinapant en Gamma-linoleenzuur. In dit hoofdstuk hebben we specifiek gekeken naar geneesmiddelen die al in de klinische praktijk zijn goedgekeurd om andere ziekten te behandelen. Het onderzoeken van het gebruik van een medicijn dat voor een bepaalde ziekte wordt gebruikt voor een andere indicatie, is een opkomend gebied van medicijnontdekking, genaamd drug repurposing. Drug repurposing zou het ontwikkel- en toelatingstraject van medicijnen aanzienlijk kunnen versnellen. Er wordt gebruik gemaakt van onze klinische en toxicologische kennis van huidige medicijnen, waardoor ze gemakkelijker toe te passen zijn voor andere indicaties dan onbekende medicijnen. Daarnaast hebben we in dit hoofdstuk bio-informatica gecombineerd die zich bezighoudt met gegevens van genen en eiwitten, met chemo-informatica die zich bezighoudt met gegevens met betrekking tot kleine moleculen en medicijnen.

Tenslotte richtte ik me in hoofdstuk 5 op de behoefte aan methoden die de enorme hoeveelheden biologische gegevens die dagelijks worden geproduceerd, kunnen combineren tot een zinvol formaat voor machines om wetenschappers te helpen nieuwe waardevolle relaties in de biologische wetenschappen te vinden. Kennis netwerken zijn een steeds populairder formaat om biologische gegevens op te slaan, waarbij gegevens uit verschillende biologische bronnen worden gecombineerd en geannoteerd tot betekenisvolle relaties die snel door machines kunnen worden doorzocht. Door al deze databronnen op één plek te hebben, kunnen wetenschappers verbanden leggen tussen verschillende vormen van data en komen veel biologische vragen aan de orde, zoals "welke medicijnen kunnen worden gebruikt om kanker te behandelen?", Of "welke genen veroorzaken diabetes?". In hoofdstuk-5 van dit proefschrift hebben we onderzocht hoe we dergelijke kennis netwerken kunnen exploreren met behulp van machine learning om te voorspellen welke biologische relaties “waar” en van hoge kwaliteit zouden kunnen zijn, zodat ze verder kunnen worden getest door wetenschappers in het laboratorium. Dergelijke methoden zijn erg belangrijk om biologische ontdekkingen te versnellen en fouten te minimaliseren. Het belang hiervan zal toenemen naarmate we doorgaan met het produceren van enorme hoeveelheden biologische gegevens die ons vermogen om te interpreteren en om te zetten in zinvolle biologische toepassingen te boven gaan.