



Universiteit
Leiden
The Netherlands

Getting personal: advancing personalized oncology through computational analysis of membrane proteins

Gorostiola González, M.

Citation

Gorostiola González, M. (2025, January 24). *Getting personal: advancing personalized oncology through computational analysis of membrane proteins*. Retrieved from <https://hdl.handle.net/1887/4093962>

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/1887/4093962>

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

Samenvatting

Kanker wordt beschouwd als de stille pandemie van de 21e eeuw en is de tweede belangrijkste doodsoorzaak wereldwijd. De aanzienlijke heterogeniteit van de ziekte kanker, gezien in verschillende kankersoorten, individuen en zelfs tumorcellen, maakt het buitengewoon uitdagend om het effectief en veilig bij alle patiënten te behandelen. Gepersonaliseerde oncologie is naar voren gekomen als een efficiënte strategie om gebruik te maken van de verschillen die aanwezig zijn in kanker voor de selectieve targeting van tumorcellen. Deze aanpak streeft ernaar de bijwerkingen te verminderen terwijl de therapeutische effectiviteit behouden blijft of verbetert. Echter, de beschikbaarheid van gepersonaliseerde therapieën is momenteel beperkt, waardoor veel kankerpatiënten verlangen naar meer selectieve behandelingen. In deze context spelen computationele hulpmiddelen een cruciale rol bij het verkennen van onopgeloste vragen in kankeronderzoek en het versnellen van de ontdekking van nieuwe eiwitten die selectief getarget kunnen worden in antikankertherapieën. Een belangrijk voordeel van het gebruik van computationele hulpmiddelen is het vermogen om veelbelovende eiwitfamilies te onderzoeken die over het hoofd zijn gezien in kankeronderzoek door experimentele beperkingen of publicatiebias, zoals membraaneiwitten. Dit proefschrift onderzoekt het potentieel van computationele hulpmiddelen bij het voorrang geven aan nieuwe doeleiwitten, mutaties en medicijnen voor gebruik in gepersonaliseerde oncologie, met een specifieke focus op membraaneiwitten.

Dit concept wordt eerst geïntroduceerd in **Hoofdstuk 1**, waar de drie prioriteitsniveaus worden gekoppeld aan functionele relevantie, kans op genezing, therapie-effectiviteit, selectiviteit en resistentie. De belangrijkste beloften en uitdagingen in gepersonaliseerde oncologie worden in dit hoofdstuk uiteengezet, gevolgd door een overzicht van computationele methoden die worden gebruikt in geneesmiddelenonderzoek en te extrapoleren zijn naar oncologisch onderzoek. In het bijzonder wordt belicht hoe deze methoden kunnen worden toegepast op de studie van membraaneiwitten als veelbelovende maar experimenteel uitdagende antikankertargets. Deze concepten worden verder uitgewerkt in dit proefschrift.

Hoofdstuk 2 bespreekt de brede reeks computationele hulpmiddelen die kunnen worden toegepast in oncologische medicijnontdekking. In het bijzonder komen twee hoofdcategorieën aan bod: kunstmatige intelligentie (AI) en structuur-gebaseerde (SB) methoden. Deze twee categorieën worden onafhankelijk uiteengezet, maar het verhoogde potentieel van hun combinatie wordt benadrukt, vooral in de context van kankeronderzoek dat multidisciplinaire oplossingen vereist. Door een selectie van gecombineerde toepassingen in kankergerelateerde doelen te beoordelen, krijgt de lezer inzicht in het potentieel van de methodologieën die zijn ontwikkeld en toegepast in dit proefschrift.

De toepassingen die in Hoofdstuk 2 worden besproken, richten zich voornamelijk op gevestigde antikankertargets, met name oplosbare proteïne-kinasen. Het verbreden van het scala aan antikankertargets is echter cruciaal voor het uitbreiden van de toegang tot gepersonaliseerde oncologiebehandelingen voor een grotere populatie. **Hoofdstuk 3** benadrukt membraaneiwitten als potentiële nieuwe antikankertargets die kunnen

worden onderzocht met behulp van computationele hulpmiddelen om de experimentele uitdagingen aan te pakken die ze minder aantrekkelijk maken om te bestuderen in vergelijking met oplosbare eiwitten. Dit hoofdstuk identificeert ook de belangrijkste uitdagingen in computationele medicijnondekking voor membraaneiwitten, die voornamelijk te maken hebben met de beschikbaarheid van gegevens en publicatiebias. In deze context worden drie eiwitfamilies met verschillende niveaus van vertegenwoordiging in de literatuur uitgelicht en onderzocht: receptor tyrosine kinasen (RTKs), G-eiwit gekoppelde receptoren (GPCRs) en solute carriers (SLCs).

Hoofdstuk 4 onderzoekt verder de verschillen in de beschikbaarheid van gegevens tussen eiwitfamilies en individuele doelen. Dit hoofdstuk benadrukt de sterke correlatie tussen publicatiebias en gegevens in openbaar beschikbare databases, specifiek bioactiviteitsgegevens gerelateerd aan gemuteerde eiwitten of genetische varianten in de ChEMBL database. Deze gegevens, die van vitaal belang zijn voor oncologische medicijnondekking, zijn aanzienlijk verrijkt op bekende antikankertargets en genetische varianten, met name kinasen en RTKs. Het hoofdstuk benadrukt het belang van deze gegevens in computationele medicijnondekking door benchmarking van variant-agnostische en variant-bewuste bioactiviteitsmodellen die kunnen worden gebruikt in oncologische medicijnondekking onder geschikte omstandigheden. Daarnaast biedt het hoofdstuk gegevenshulpmiddelen en aanbevelingen om te helpen bij het samenstellen van hoogwaardige variant-geannoteerde datasets voor bioactiviteitsmodellering.

Hoofdstukken 5-7 richten zich op het ontwikkelen van verschillende computationele toepassingen om de drie niveaus van prioritering te behandelen die in Hoofdstuk 1 zijn geïntroduceerd. Deze toepassingen maken gebruik van de methoden die zijn vermeld in Hoofdstuk 2 en worden toegepast op de minder onderzochte membraaneiwitfamilies die zijn geïntroduceerd in Hoofdstuk 3.

Target-prioritering voor GPCRs wordt besproken in **Hoofdstuk 5**. Dit hoofdstuk evalueert de functionele relevantie van individuele GPCRs in kanker door pan-kanker gerelateerde mutaties te analyseren in vergelijking met natuurlijke variatie. De resultaten in dit hoofdstuk en de daaropvolgende zijn gebaseerd op een kankerpatiëntendataset die is gemaakt voor dit proefschrift uit de Genomic Data Commons (GDC) database, die beschikbaar is gesteld voor openbaar gebruik en computationeel vriendelijk en versie-stabiel is. Mutaties die verrijkt zijn in kanker en zich bevinden in functioneel bewaarde motieven worden beschouwd als prioriteiten bij het identificeren van 52 GPCRs als potentiële antikankertargets met behulp van een multi-objectieve optimalisatie aanpak. Deze aanpak maakt het ook mogelijk om praktische doelen, zoals aanvullende computationele en experimentele middelen, op te nemen en kan verder worden geïntegreerd met SB-analyses voor specifieke receptoren van interesse, inclusief de methoden die worden beschreven in Hoofdstuk 6.

Hoofdstuk 6 behandelt mutant-prioritering voor de glutamaat transporter EAAT1, een lid van de SLC-familie. Kankergerelateerde mutaties uit de GDC-dataset, gevonden in de buurt van de orthostere en allosterie bindingsplaats, worden computationeel getest om hun impact op eiwitconformatie en functie te beoordelen. Moleculaire dynamica (MD) simulaties en docking experimenten suggereren dat bepaalde kankergerelateerde

mutaties, met name R479W, een conformationele verandering veroorzaken die kan worden benut in gepersonaliseerde oncologie. Daarnaast demonstreert dit hoofdstuk de overdraagbaarheid van computationele bevindingen naar de echte wereld door middel van *in vitro* experimentele validatie van de effecten van de mutaties op transportfunctie en respons op farmacologische interventie.

Medicijnprioritering wordt besproken in **Hoofdstuk 7**, waarin een methode wordt geïntroduceerd om virtuele screening van grote bibliotheken van kandidaat-medicijnen voor mutante GPCRs te verbeteren. Dit hoofdstuk presenteert de creatie van nieuwe 3D dynamische eiwitdescriptoren (3DDPDs) op basis van MD simulaties om de representatie van gemuteerde eiwitten voor proteochemometrics bioactiviteitsmodellering te verbeteren. Resultaten tonen aan dat deze nieuwe descriptoren beter presteren dan sequentie-gebaseerde descriptoren in wild-type GPCR bioactiviteitsmodellering. Echter, evaluatie van hun toepasbaarheid in gemuteerde GPCRs wacht nog op verbetering van de beperkingen in gegevens beschikbaarheid die worden besproken in Hoofdstuk 4.

De lessen uit Hoofdstukken 4-7 culminereren in **Hoofdstuk 8**, waar een holistische benadering wordt gevolgd om alle eerder besproken gegevenssoorten te integreren. In dit hoofdstuk wordt een patiëntgerichte kennisgrafiek ontwikkeld met als doel het prioriteren van gemuteerde eiwitten voor gerichte therapie bij kanker. Deze benadering combineert de structurele en bioactiviteitsgegevens analyses uit Hoofdstuk 4 evenals de kanker- en natuurlijke variatiegegevens uit Hoofdstuk 5. Daarnaast bouwt het voort op de concepten die zijn verkend in Hoofdstukken 4-7 om doelen en mutaties te prioriteren die functioneel, structureel en klinisch relevant zijn. Vanwege beperkingen in gegevensbeschikbaarheid ligt de focus van dit hoofdstuk voornamelijk op kinasen, specifiek RTKs. Echter, net als de voorgaande hoofdstukken, is het ontworpen om toepasbaar te zijn in elk eiwittype als de noodzakelijke gegevens in de toekomst beschikbaar komen. Geavanceerde modellering algoritmen zouden ook kunnen worden gebruikt om de kennisgrafiek te verbeteren naarmate meer gegevens beschikbaar komen.

Ten slotte biedt **Hoofdstuk 9** een samenvatting van de conclusies die zijn getrokken uit de voorgaande hoofdstukken in de bredere context van computationele oncologische medicijnondekking. Over het algemeen breiden de methoden die in dit proefschrift zijn ontwikkeld het scala aan beschikbare hulpmiddelen uit voor het selecteren van nieuwe doeleiwitten, mutanten en medicijn kandidaten voor gepersonaliseerde oncologische toepassingen. Om echter klinische significantie te bereiken, moeten er verder samengewerkt worden binnen de wetenschappelijke gemeenschap om te focussen op kankeronderzoeksinitiatieven. Computationele hulpmiddelen zoals die in dit proefschrift ontwikkeld zijn kunnen uiterst nuttig zijn voor taken zoals het ontwerpen en implementeren van machine-leesbare open-source kankerdatabases, het identificeren van belangrijke biomarkers voor diagnose en gepersonaliseerde behandeling, het voorspellen van optimale behandelingsstrategieën en het prioriteren van belangrijke onderzoeksgebieden. Zo kunnen door samenwerking met andere wetenschappers effectieve en veilige behandelingen worden ontwikkeld voor alle patiënten die tegen kanker vechten.

