



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).

Resumen

El cáncer se considera la pandemia silenciosa del siglo XXI y la segunda causa principal de muerte en todo el mundo. La heterogeneidad que lo caracteriza, la cual se manifiesta entre distintos tipos de cáncer, individuos e incluso células tumorales, hace que sea extremadamente difícil de tratar de manera efectiva y segura en todos los pacientes. La oncología personalizada ha surgido como una estrategia eficiente para aprovechar las diferencias presentes en el cáncer y así atacar específicamente las células tumorales. Este enfoque tiene como objetivo reducir los efectos secundarios mientras mantiene o mejora la eficacia terapéutica. Sin embargo, la disponibilidad de terapias personalizadas es actualmente limitada, lo que deja a muchos pacientes con cáncer deseando poder optar a tratamientos más selectivos o específicos. En este contexto, las herramientas computacionales juegan un papel crucial en la exploración de preguntas no resueltas en la investigación del cáncer y en la aceleración del descubrimiento de nuevas proteínas que pueden ser atacadas selectivamente en terapias anticancerígenas como dianas terapéuticas. Una ventaja principal del uso de herramientas computacionales es la capacidad de investigar familias de proteínas prometedoras que han sido pasadas por alto en la investigación del cáncer debido a limitaciones experimentales o de sesgo de publicación, como las proteínas de membrana. Esta tesis profundiza en el potencial de las herramientas computacionales en la priorización de nuevas dianas terapéuticas, mutaciones y candidatos a fármacos para su uso en la oncología personalizada, con un enfoque específico en las proteínas de membrana.

Este concepto se introduce primero en el **Capítulo 1**, donde se vinculan los tres niveles de priorización con la relevancia funcional, la capacidad de ser atacadas con medicamentos, la potencia de la terapia, su selectividad y la posibilidad de generar resistencias. En este capítulo se delinean las principales promesas y desafíos de la oncología personalizada, seguidos de una visión general de los métodos computacionales utilizados en el descubrimiento de fármacos que pueden extrapolarse a la investigación oncológica. En particular, se introduce cómo estos métodos pueden aplicarse al estudio de las proteínas de membrana como dianas terapéuticas anticancerígenas prometedoras pero experimentalmente complejas. Estos conceptos se detallan a lo largo de esta tesis.

El **Capítulo 2** revisa la amplia gama de herramientas computacionales que pueden aplicarse en el descubrimiento de fármacos oncológicos. El enfoque principal de este capítulo está en dos categorías: inteligencia artificial (IA) y métodos basados en la estructura (SB). Estas dos categorías se describen de manera independiente, pero se destaca el potencial sinérgico de su combinación, especialmente en el contexto de la investigación del cáncer, la cual requiere soluciones multidisciplinarias. Al revisar una selección de aplicaciones combinadas en cáncer, el lector obtiene una comprensión del potencial de las metodologías desarrolladas y aplicadas a lo largo de esta tesis.

Las aplicaciones discutidas en el Capítulo 2 se centran principalmente en dianas terapéuticas anticancerígenas ampliamente consolidadas, en particular las proteínas citosolubles. Sin embargo, ampliar el rango de dianas anticancerígenas es crucial para expandir el acceso a los tratamientos de oncología personalizada a una población más

amplia. El **Capítulo 3** enfatiza el uso de proteínas de membrana como posibles nuevas dianas anticancerígenas. Estas pueden ser exploradas usando herramientas computacionales para abordar los desafíos experimentales que las hacen menos atractivas de estudiar en comparación con las proteínas solubles. Este capítulo también identifica los principales desafíos en el descubrimiento computacional de fármacos para proteínas de membrana, que están principalmente relacionados con la disponibilidad de datos y el sesgo de publicación. En este contexto, se destacan y examinan tres familias de proteínas con diferentes niveles de representación en la literatura: receptores tirosina-cinasa (RTKs), receptores acoplados a proteínas G (GPCRs) y transportadores de solutos (SLCs).

El **Capítulo 4** explora en más detalle las diferencias en la disponibilidad de datos entre familias de proteínas y dianas a nivel individual. Este capítulo resalta la fuerte correlación entre el sesgo de publicación y los datos presentes en bases de datos públicas, específicamente los datos de bioactividad relacionados con proteínas mutantes o variantes genéticas en la base de datos ChEMBL. Estos datos, vitales para el descubrimiento de fármacos oncológicos, están mayoritariamente presentes en dianas anticancerígenas y variantes genéticas de relevancia establecida, en particular cinasas y RTKs. El capítulo enfatiza la importancia de estos datos en el descubrimiento computacional de fármacos a través de la evaluación comparativa de modelos de predicción de bioactividad teniendo o no en cuenta la variabilidad genética. Estos modelos pueden ser utilizados en el descubrimiento de fármacos oncológicos si se dan las circunstancias adecuadas. Además, el capítulo ofrece herramientas de datos y recomendaciones para ayudar en la preparación de conjuntos de datos con variantes de alta calidad para la predicción de bioactividad.

Los Capítulos 5-7 se centran en el desarrollo de varias aplicaciones computacionales para abordar los tres niveles de priorización introducidos en el Capítulo 1. Estas aplicaciones utilizan los métodos introducidos en el Capítulo 2 y se aplican a las familias de proteínas de membrana menos exploradas introducidas en el Capítulo 3.

La priorización de dianas terapéuticas se trata en el **Capítulo 5** con enfoque en la familia de GPCRs. Este capítulo evalúa la relevancia funcional de GPCRs en el cáncer mediante el análisis de mutaciones relacionadas con el cáncer en comparación con la variación natural. Los resultados en este capítulo y los siguientes se basan en un conjunto de datos de pacientes con cáncer creado para esta tesis a partir de la base de datos Genomic Data Commons (GDC), que está disponible para su uso público. Las mutaciones más comunes en cáncer y localizadas en motivos conservados funcionalmente se consideran objetivos prioritarios para identificar 52 GPCRs como posibles dianas anticancerígenas utilizando un enfoque de optimización multiobjetivo. Este enfoque también permite la inclusión de objetivos prácticos, como la disponibilidad de recursos computacionales y experimentales adicionales, y puede integrarse con análisis SB para receptores específicos de interés, como por ejemplo los métodos descritos en el Capítulo 6.

El **Capítulo 6** cubre la priorización de mutantes para el transportador de glutamato EAAT1, un miembro de la familia SLC. Las mutaciones relacionadas con el cáncer del conjunto de datos GDC encontradas cerca de los lugares de unión ortostéricos (sitio activo) y alostéricos se evalúan computacionalmente para valorar su impacto en la conformación y función de la proteína. Las simulaciones de dinámica molecular (MD) y los

experimentos de acoplamiento (docking) sugieren que ciertas mutaciones relacionadas con el cáncer, específicamente R479W, inducen un cambio conformacional que puede ser aprovechado en oncología personalizada. Además, este capítulo demuestra la transferibilidad de los hallazgos computacionales a aplicaciones del mundo real a través de la validación experimental *in vitro* de los efectos de estas mutaciones en la función del transportador y su respuesta a intervención farmacológica.

La priorización de moléculas como fármacos se desarrolla en el **Capítulo 7**, que introduce un método desarrollado para mejorar el cribado virtual de grandes bibliotecas de candidatos a medicamentos que ataquen variantes genéticas de GPCRs. Este capítulo presenta la creación de nuevos descriptores de proteínas que son 3D y dinámicos (3DDPDs), que se basan en simulaciones MD para mejorar la representación de proteínas mutantes en la modelización proteoquimométrica de bioactividad. Los resultados muestran que estos nuevos descriptores superan a los descriptores basados en secuencias proteicas en la modelización de bioactividad de GPCRs no mutados. Sin embargo, la evaluación de su aplicabilidad en GPCRs mutantes está pendiente debido a las limitaciones de disponibilidad de datos discutidas en el Capítulo 4.

Las lecciones de los Capítulos 4-7 encuentran su culminación en el **Capítulo 8**, donde se adopta un enfoque holístico para integrar todos los tipos de datos discutidos previamente. En este capítulo se desarrolla un grafo de conocimiento (knowledge graph) centrado en el paciente con el objetivo de priorizar proteínas mutadas para la terapia dirigida en el cáncer. Este enfoque combina los análisis de datos estructurales y de bioactividad del Capítulo 4, así como los datos de cáncer y variación natural del Capítulo 5. Además, se basa en los conceptos explorados en los Capítulos 4-7 para priorizar dianas terapéuticas y mutaciones que sean funcional, estructural y clínicamente relevantes. Debido a las limitaciones en la disponibilidad de datos, el enfoque de este capítulo se centra principalmente en cinasas, específicamente RTKs. Sin embargo, al igual que los capítulos anteriores, está diseñado para amoldarse a cualquier tipo de proteína si los datos necesarios están disponibles en el futuro. A medida que haya más datos disponibles, también se podrían utilizar algoritmos de modelización avanzados.

Finalmente, el **Capítulo 9** ofrece un resumen de las conclusiones extraídas de los capítulos anteriores en el contexto más amplio del descubrimiento computacional de fármacos oncológicos. En general, los métodos desarrollados en esta tesis amplían la gama de herramientas disponibles para seleccionar nuevas dianas terapéuticas, mutantes y candidatos a medicamentos para aplicaciones de oncología personalizada. Sin embargo, para lograr una significancia clínica, se deben mantener los esfuerzos de colaboración dentro de la comunidad científica para enfocar las iniciativas de investigación del cáncer. Las herramientas computacionales similares a las desarrolladas en esta tesis pueden ser extremadamente beneficiosas para tareas como diseñar e implementar bases de datos de cáncer de código abierto, identificar biomarcadores clave para el diagnóstico y tratamiento personalizado, predecir estrategias de tratamiento óptimas y priorizar áreas clave de investigación. En última instancia, solo a través de esfuerzos colaborativos y enfocados se pueden desarrollar tratamientos efectivos y seguros para todos los pacientes que luchan contra el cáncer.

