

Resumen

Los avances en las técnicas de secuenciación de alto rendimiento y el posterior abaratamiento tecnológico han favorecido el desarrollo y la popularización de una nueva gama de disciplinas de investigación genómica, conocidas colectivamente como "ómicas". Estas tecnologías son capaces de realizar mediciones simultáneas de miles de moléculas esenciales para la vida, tales como el ADN, el ARN, las proteínas y los metabolitos. Históricamente, la investigación genómica clásica ha seguido un enfoque reduccionista al estudiar la estructura, regulación y función de estas unidades biológicas de manera independiente. Sin embargo, pese a ser una poderosa herramienta analítica, el método reduccionista es incapaz de explicar muchos de los fenómenos biológicos que tienen lugar en un sistema vivo, sugiriendo que la esencia del sistema no puede explicarse simplemente mediante la enumeración de elementos que lo componen, sino que radica en la dinámica de los procesos biológicos que entre ellos acontecen.

La Biología de Sistemas se ha establecido en los últimos años como el área de investigación multidisciplinaria que trata de modelar el comportamiento dinámico de los sistemas biológicos a través del estudio holístico de las interacciones entre sus partes, combinando mediciones simultáneas de diferentes tipos de moléculas e integrando múltiples fuentes de información para identificar aquellos componentes que cambian de manera coordinada en las condiciones bajo estudio. La Biología de Sistemas es un área interdisciplinar que requiere que biólogos, matemáticos, bioquímicos y otros investigadores trabajen en estrecha colaboración, y en la que la informática tiene un papel fundamental dado el volumen y la complejidad de los datos manejados.

Esta tesis aborda el problema de la gestión, integración y análisis de los datos en estudios *multi-ómicos*. Más específicamente, la investigación realizada se ha centrado en dos de los retos computacionales más característicos de la Biología de Sistemas: el desarrollo de bases de datos integrativas y el problema de la visualización integrativa. Así, la primera parte de este trabajo se ha dedicado al diseño y creación de un recurso bioinformático para la gestión de experimentos *multi-ómicos*. La plataforma desarrollada, conocida como STATegra EMS, ofrece un conjunto de herramientas que facilitan el almacenamiento y la organización de los grandes conjuntos de datos que son generados durante estos experimentos, así como la anotación de las etapas posteriores de procesamiento y análisis de la información. La heterogeneidad, el volumen y la alta variabilidad de los datos *ómicos* son algunos de los obstáculos que han sido abordados durante el desarrollo del STATegra EMS, con el fin de alcanzar un registro detallado de la meta-información que permita discriminar cada conjunto de datos y lograr así una integración exitosa de la información. Para ello, la plataforma desarrollada ofrece una interfaz web colaborativa y de fácil manejo en la que se combinan modernas tecnologías web y conocidos estándares comunitarios para la representación de los diferentes componentes del experimento.

En la segunda parte de esta tesis se discuten la situación actual y las dificultades de la visualización integrativa de datos en experimentos *multi-ómicos*, y se presenta la herramienta web desarrollada, PaintOmics 3. Dado que la capacidad del cerebro humano para el procesamiento visual está altamente evolucionada, la visualización integrativa en combinación con técnicas de análisis de datos es probablemente una de las herramientas más poderosas para la interpretación y validación de los resultados en Biología de Sistemas. PaintOmics 3 proporciona un completo marco de trabajo para realizar análisis de enriquecimiento de funciones biológicas en experimentos con múltiples condiciones y tipos de datos *ómicos*, en el que se combinan potentes herramientas de visualización integrativa de datos sobre diagramas de interacción molecular y redes de reacción KEGG, redes de interacción de procesos biológicos, y estudios estadísticos de los datos. Además, a diferencia de otras herramientas desarrolladas, PaintOmics 3 destaca por su facilidad de uso y su gran interactividad, así como por su flexibilidad y variedad de los datos *ómicos* aceptados, incluyendo datos de

epigenómica basados en regiones genómicas, datos de proteómica o estudios de miRNA.