

Fuentes de información para modelados en multi-ómicos y metabólicos

Sources for multi-omics and metabolics modelling pipelines

Julio César Osorio

Las aproximaciones “ómicas” generan una cantidad inconmensurable de datos que son producto de intentar conocer a un organismo y su relación con el ambiente. Este conocimiento resulta necesario ya que existen intereses biotecnológicos y clínicos. Pero aunque existe un gran interés, el vacío en el conocimiento es gigante y se requiere de un gran esfuerzo computacional para intentar conocer las fluctuaciones que pueden existir entre un genotipo determinado y un fenotipo para un momento y lugar dado. Tenemos que partir del hecho de que la vida de cualquier organismo es un evento complejo y resumir esa información en variables medibles y con capacidades predictivas requiere de la integración de mucha información en distintas escalas.

La “ómica” múltiple e integrada permite la evaluación de procesos biológicos complejos como la producción de combustible y su relación con microorganismos, procesos biofarmacéuticos, o la investigación médica que trata de conocer las relaciones patógeno hospedero, entre otras. Este tipo de integración requiere la unificación de los formatos de los datos, la comparabilidad de los datos desde distintas escalas y la normalización de los mismos. Pero hay ciertas desventajas al tratar de hacer derivaciones de potencial metabólico de un microorganismo teniendo en cuenta solo su genoma debido a que este solo nos da un instante de la real capacidad fenotípica del organismo. Esta capacidad fenotípica puede estar mediada por respuestas al ambiente, perturbaciones genéticas, fluctuaciones en el tiempo, etc. Dada esta limitación el análisis “ómico” múltiple trata de obtener información desde la transcriptómica y ver como varían los niveles de RNA en la célula. La metabólica aporta sobre el conocimiento de los metabolitos y, las interacciones proteicas nos ayudan a ampliar el panorama sobre los tipos de regulación post traducionales. Además, la metagenómica y la metatranscriptómica nos dan un panorama de la diversidad de metabolismos en las comunidades microbianas.

Este tipo de integración se puede hacer uniendo de a dos capas. Se pueden hacer flujos “ómicos” que unan las capas transcriptómica con la capa metabolómica, proteómica con metabolómica y proteómica con transcriptómica. Este tipo de integraciones pueden ayudar a entender como la regulación de la expresión génica esta desacoplada de las dinámicas de la proteína. Además, cómo la eficiencia de la traducción puede tener un impacto regulatorio mayor sobre la abundancia de la proteína que el recambio proteico. Otro tipo de integraciones como metatranscriptómica y la metagenómica generan indicios sobre el rol funcional de cada microorganismo en relación con la comunidad microbiana. Los modelamientos de este tipo pueden arrojar diferentes resultados y algunos de estos nos pueden sorprender como es el caso del modelamiento para *Mycoplasma pneumoniae* que muestra rigidez de las vías metabólicas ya que están siendo reguladas como unidades funcionales.

Al trabajar en distintas escalas y en diferentes capas comenzamos a ver como las ciencias “ómicas” se especializan en diferentes límites. De esta forma cuando hablamos de los límites del gen tratamos de representar la información presente en el DNA aunque existe algún tipo de información adicional que no podrá ser completamente explorada. Este tipo de información tiene mucho valor pues nos dan guías sobre las interacciones que un

Grupo de Investigación - Ciencias Odontológicas UNICOC

1. Biólogo. Magíster en Biología énfasis en Genética Humana

Autor responsable de correspondencia: Julio César Osorio.

Correo electrónico: jcosorio@unicoc.edu.co

Citar como: Osorio JC. Fuentes de información para modelados en multi-ómicos y metabólicos. Journal Odont Col. 2015;8(15):60-65

Recibido: Abril 2015, aceptado: Mayo 2015

conjunto de genes puede tener y como éstas relaciones afectan la proteínas y el fenotipo de los organismos ante un ambiente en particular. Se usan varias estrategias para tratar de llenar los huecos en la información entre ellos la “asociación más culposa”, la cual permite inferir funciones de un gen desde los genes con funciones relacionadas o aquellos que comparten interacciones de tipo genético o físico.

Los límites genómicos estudian como los dominios GC pueden estar relacionados con la transferencia horizontal de genes, la patogenicidad y resistencia a antibióticos en los microorganismos. Además, puede ser utilizado sobre la estructura 3D del cromosoma y cómo esta tiene implicaciones sobre la reparación, replicación y transcripción del DNA.

También podemos llegar a los límites de la información metabólica y estudiar las redes bioquímicas que nos permitan la reconstrucción de redes metabólicas a escala genómica o, situarnos en los límites filogenéticos y utilizar el amplio rango de secuencias de genomas disponibles en las bases de datos públicas con el fin de llenar los vacíos en la información.

Podemos sobrepasar los límites determinísticos y tratar de explicar el fenotipo que no está siendo codificado por el DNA. El contexto epigenómico nos comienza a dar explicaciones sobre las diferencias fenotípicas que surgen en células con idéntica información genética. En general, queremos poder cuantificar como ésta memoria epigenética afecta el metabolismo, las relaciones ecológicas y los cambios en el fenotipo.

Retomando el tema de la complejidad y la necesaria ayuda de análisis computacionales se logran identificar tres tareas que resultan desafiantes. El primero recae sobre los diferentes componentes moleculares (genes, moléculas, enzimas), el segundo sobre los procedimientos de normalización de los datos y su respectivo análisis y el tercero sobre la visualización efectiva de los datos.

La primera tarea trata de unir los nombres de los genes a códigos que son compartidos a través de múltiples bases de datos. Herramientas como GOP, KAAS, GO permiten realizar anotaciones derivadas de redes genómicas.

La segunda tarea utiliza un sinfín de herramientas cada una con diferentes funciones y escalas. Por ejemplo MONA es una herramienta basada en un modelo bayesiano que permite inferir asociaciones significantes a través de las múltiples capas de la información. Esta además puede ser utilizada en análisis de la metilación del DNA y la regulación de los microRNA. MADMAX data base permite comparar múltiples experimentos “ómicos” mientras que CONFERO guarda una lista de genes derivados de conocimientos a priori biológicos y los integra con resultados obtenidos desde datos “ómicos”. OMICKRING computa las similaridades entre muestras usando una combinación lineal de matrices ómicas lo que le ayuda a predecir rasgos complejos emergentes a nivel celular a través del peso promedio de los fenotipos de los individuos. VANTED provee un marco para la cartografía y la integración de los datos experimentales a través de redes bioquímicas que pueden ser descargadas desde KEGG. NETGESTALT permite presentar los datos de experimentos multi-escala facilitando su integración, visualización y análisis. INCROMAP soporta el análisis simultáneo de mRNA, miRNA, metilación y proteínas. Además puede representar un conjunto de genes predefinidos en múltiples plataformas experimentales.

Además de estos paquetes muchos de los programas que permiten la normalización y análisis de datos son soportados por la plataforma “R”. Librerías como DANTE permiten ejecutar análisis estadísticos de conjuntos de datos mientras INTEGROMICS realiza análisis entre variables “ómicas” para una misma muestra. Esto lo hace utilizando matrices de datos desde los bloques. MCIA identifica las correlaciones entre conjunto de datos de altas dimensiones a través de un análisis de datos exploratorios. Cuando los conjuntos de datos son heterogéneos se unen las dimensiones proteómica y transcriptómica. En cambio con datos homogéneos las dimensiones que se unen son micro arreglos y RNAseq.

La tercera tarea tiene dos objetivos. El primero se enfoca en herramientas para la automatización, representaciones y el análisis de conjuntos de datos biológicos. El segundo al ensamblaje y limpieza de las vías metabólicas que están siendo construidas. Varias herramientas son entonces utilizadas. CYTOSCAPE permite la visualiza-

ción de los datos biológicos. PATHWAY TOOLS genera diagramas de las redes bioquímicas de un individuo sobre su representación de la expresión génica. PROMETRA visualiza e integra conjuntos de datos desde base de datos “ómicas” para una ruta metabólica definida. PAINTOMICS es capaz de generar una lista de genes significativos o una lista de cambios metabólicos y graficar la información sobre mapas metabólicos derivados de KEGG. 30MICS puede ejecutar análisis de correlación “ómicos” y hacer visualizaciones de las relaciones de los datos con respecto al tiempo, además de las condiciones experimentales y los tipos de datos. METSCAPE puede agrupar datos experimentales, genes y vías metabólicas para luego hacer la integración. MAYDAY realiza análisis y visualización de los datos “ómicos” generales.

Después de cumplir con estas tareas se debería poder proponer un esquema que reconstruya una vía metabólica. De forma resumida la reconstrucción parte desde un genoma que proviene de una técnica de secuenciación de última generación. Esta secuencia es acompañada de múltiples experimentos donde se obtienen datos que pueden ser tasas de crecimiento o micro arreglos. Con esta información se puede construir o proponer un modelo que se somete a múltiples herramientas computacionales con el fin de corregir las lagunas o los vacíos en la información. Con esto se realiza la primera comparación del fenotipo con el modelo y los datos experimentales. El modelo se expone a los diferentes datos producto de la integración “ómica” lo que termina de ajustar el modelo. Finalmente se valida el modelo experimentalmente y se da por terminado la reconstrucción completa.

El objetivo de estas reconstrucciones completas es construir experimentos bioinformáticos que permitan discutir y predecir como cierta ruta metabólica en determinado organismo varía ante cambios en determinados ambientes o ante cambios regulatorios en cualquier capa “ómica”. Los flujos computacionales que permiten esta integración son fundamentales y sin ellos el análisis simultáneo de diferentes tipos de datos sería imposible. Además, hay un hecho importante ya que las anotaciones producto de estas simulaciones enriquecen el conocimiento derivado de estos estudios.

REFERENCIA

Fondi M, Liò M. Multi-omics and metabolic modelling pipelines: Challenges and tools for systems microbiology. *Microbiological Research*. 2015; 171:52–64