

Los resultados, publicados el pasado marzo en *Nature*, revelaron que tales diferencias no solo se debían a la expresión de genes presentes en el cromosoma adicional, sino que también aparecían cambios en determinadas regiones distribuidas a lo largo de todo el genoma. En concreto, el estudio identificó 337 regiones alteradas en las células con trisomía que no lo estaban en las células normales.

Hasta la fecha, la mayoría de las investigaciones sobre la expresión génica en esta patología se han centrado en genes del cromosoma 21. Algunos de ellos se han vinculado claramente con los impedimentos cognitivos de los enfermos. Sin embargo, no se han podido identificar los genes responsables de todas las anomalías. Ya se sospechaba que algunos de los síntomas estaban relacionados con genes de otros lugares y cuya expresión estaría alterada debido a cambios generales en el equilibrio celular. Pero estos genes podrían estar situados en cualquier parte del genoma, lo que hacía difícil su identificación.

Aunque se desconoce el motivo por el que estas zonas presentan una expresión

anómala, se ha visto que están asociadas en gran medida a regiones de ADN compactado situadas en la periferia del núcleo, las cuales suelen presentar genes con una expresión reprimida. No obstante, parece que en la trisomía del cromosoma 21 esta represión resulta menor. Basándose en esta observación, se han planteado dos hipótesis no excluyentes para explicar los síntomas del síndrome.

Por un lado, la sobreexpresión de genes del cromosoma 21 implicados en el mantenimiento de la estructura de la cromatina (forma condensada que adoptan los cromosomas en el núcleo en condiciones normales, cuando la célula no se divide) provocaría alteraciones en la arquitectura de esta, lo que a su vez modificaría la expresión génica en las regiones afectadas. Por otro lado, la presencia de un cromosoma adicional añade un porcentaje nada despreciable de ADN al núcleo, lo que perturbaría el equilibrio celular.

Este estudio corrobora las sospechas que se tenían de que en el síndrome de Down intervienen genes que se hallan

repartidos por todo el genoma, los cuales se ubican en 337 regiones distintas. El hallazgo reduce extraordinariamente la proporción del genoma que debe examinarse para identificar los genes en cuestión, lo que permitirá avanzar en el entendimiento de los cambios genéticos que dan lugar a los síntomas de esta enfermedad.

—David González Knowles
Integromics S.L.
Parque Científico de Madrid

PARA SABER MÁS

Domains of genome-wide gene expression dysregulation in Down's syndrome.

A. Letourneau et al. en *Nature*, n.º 508, págs. 345-350, abril de 2014.

EN NUESTRO ARCHIVO

La vida interior del genoma. T. Misteli en *lyC*, abril de 2011.

Evolución de la cromatina. G. A. Babbitt en *lyC*, mayo de 2011.

COMUNICACIÓN

Avances y retos en la comprensión de datos biológicos

Crónica del V congreso internacional sobre visualización

MICHAEL P. SCHROEDER Y NÚRIA LÓPEZ BIGAS

La conferencia internacional sobre visualización de datos biológicos (www.vizbi.org) reúne cada año a científicos de todos los ámbitos para debatir sobre las formas de representar la multitud de datos biológicos que se están generando en la actualidad. Durante cuatro días se presentan numerosas herramientas y enfoques centrados en la representación visual orientada a facilitar la interpretación de datos complejos.

Una de las particularidades más interesantes de ese encuentro es que a él asisten, además de biólogos, especialistas en informática y visualización de datos. Así, el inicio de la última y quinta edición de la conferencia (celebrada en Heidelberg a principios del pasado mes de marzo), fue protagonizado por el informático experto en interpretación de grandes conjuntos de datos Jeffrey Heer, de la Universidad de Washington, quien introdujo a los parti-

cipantes en los principios de la visualización de datos.

Heer citó dos factores que definen una buena visualización: la expresividad y la eficiencia. La opción más expresiva será la que incluya todos los datos que pretendemos representar, y solo esos. La más eficiente será la más rápida de interpretar. Para ilustrar estos aspectos se presentó un estudio que comparaba la interpretación de los mismos datos representados mediante diferentes formas: gráficos circulares y diagramas de barras. Se destacó en este trabajo el cambio de magnitud percibido por las personas entrevistadas respecto a las diferentes versiones. Conclusión: los gráficos circulares resultan engañosos y las barras son más adecuadas. De ahí que Manolis Savva, de la Universidad Stanford, y otros colaboradores desarrollaran en 2011 ReVision, un programa informático que permite «re-

visualizar» gráficos circulares en forma de diagramas de barras.

Coordenadas, matrices y redes

Pero los datos biológicos poseen ciertas características que entrañan dificultades adicionales a su visualización. Nos referimos a la densidad y la multidimensionalidad.

En los estudios sobre genómica del cáncer, para determinar el perfil molecular de un tumor con frecuencia se realizan múltiples ensayos (secuenciación del genoma, secuenciación del transcriptoma y nivel de metilación del ADN). En los proyectos actuales del Consorcio Internacional del Genoma del Cáncer —en el que España participa—, se aplican los ensayos mencionados a miles de muestras de pacientes. La meta es analizar, como mínimo, el tumor de unas 500 personas para los 50 tipos de cáncer más

frecuentes. Además de solventar los retos puramente computacionales del análisis de esta gran cantidad de datos, se necesitan estrategias para su exploración y visualización. ¿Cómo pueden representarse tantos datos multidimensionales de varios centenares de pacientes y casi 20.000 genes?

En enero de 2013 publicamos un artículo de revisión en *Genome Medicine* en el que mostrábamos que las herramientas de representación disponibles pueden clasificarse en tres categorías: ligadas a la posición genómica, de matrices de colores y de redes. Aparte de la visualización, la interacción con el usuario constituye un componente clave; es decir, la visualización debe ser adaptable a las necesidades e intereses de cada investigador. Por ello revisten importancia aspectos como la navegación por la representación visual, la opción *zoom*, la visualización a diferentes niveles de resolución y la posibilidad de filtrar y ordenar los datos.

En las visualizaciones ligadas a la posición genómica se representa cada medición en horizontal, alineada con el genoma de referencia. Al reducir cada medición a uno o pocos píxeles pueden visualizarse a la vez muchos datos, procedentes de diferentes tipos de ensayos. A ojo pueden observarse regiones genómicas que tienden a presentar valores extremos. En el caso de que sea necesario ver los datos con mayor precisión, puede ampliarse el detalle de la medición y reducir la extensión de la región genómica.

La representación de matrices de colores (*heatmaps*) viene utilizándose desde hace varias décadas en biología, sobre todo para representar los resultados de micromatrices de expresión. A menudo, su imagen estática resulta insuficiente para

extraer información. Para solventarlo, en nuestro laboratorio hemos desarrollado *Gitools* (www.gitools.org), un visualizador interactivo de matrices multidimensionales de colores. Estos mapas ofrecen una vía de exploración muy interesante, ya que permiten decidir libremente el orden de las columnas y las filas, por lo que no están ligados a la posición genómica; ello facilita el estudio de asociaciones entre entidades alojadas en regiones distantes del genoma, como sería el caso de un grupo de genes que pertenecen a la misma ruta molecular. Asimismo, permiten superponer visualmente las distintas mediciones, lo que facilita el análisis de las asociaciones.

Imaginemos, por ejemplo, que tenemos un mapa de colores en forma de matriz, cuyas columnas representan pacientes y cuyas filas corresponden a genes, y en la que cada celda tiene asociado el valor de expresión y el número de copias del gen en el genoma. El hecho de tener varios valores asociados a las celdas se puede aprovechar para estratificar los datos según un criterio y observar el comportamiento del otro. Con *Gitools* podemos ordenar las columnas (pacientes) según el número de copias del gen y observar rápidamente el efecto que ejerce un cambio de copias en la expresión del gen en las muestras respectivas. Con un mapa de colores se pueden representar una gran cantidad de muestras a la vez.

En el caso de las redes, si bien no ofrecen tanta resolución como los dos formatos de visualización ya presentados, aportan una información crucial: las interacciones entre los diferentes componentes o proteínas. Un nodo puede representar un gen o una proteína, y

la información de la cohorte de pacientes se puede alinear circularmente en capas, asignando una capa a cada ensayo. Así, pueden representarse relaciones de interacción como la siguiente: un gen (o nodo) que resulta crítico en la regulación de la expresión de otros genes y se encuentra frecuentemente mutado puede enlazarse con los genes que regula y dibujar el estado de expresión. En casos de rutas moleculares, la visualización reticular resulta muy informativa.

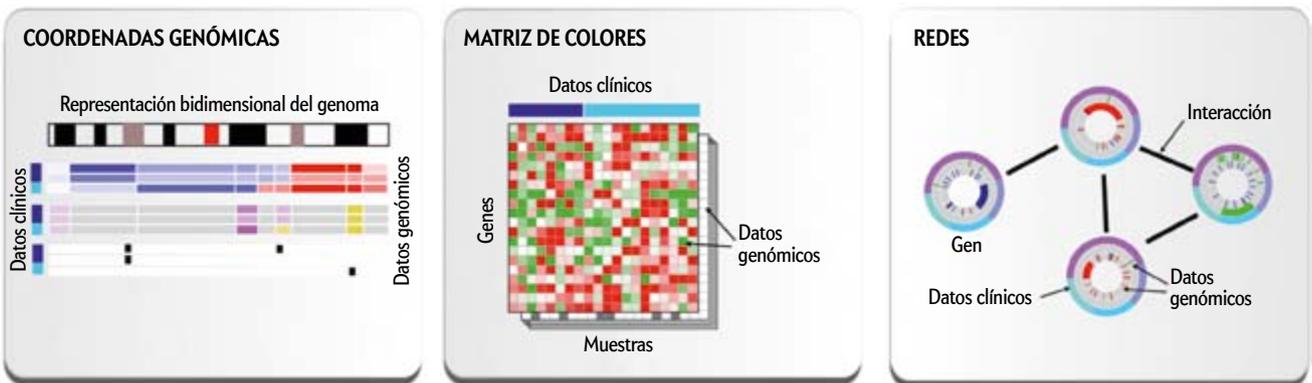
Cada esquema presenta ventajas y desventajas, por lo que se recomienda utilizar múltiples herramientas para investigar los mismos datos.

Genómica y proteómica en 3D

Aparte de discutir sobre la representación de datos genómicos, en la conferencia de este año hubo sesiones enfocadas a la visualización de datos del transcriptoma, estructuras tridimensionales de proteínas, sistemas celulares, tejidos y organismos, y también metagenómica y poblaciones.

Eduardo Eyra, de la Universidad Pompeu Fabra, y Joseph Barash, de la Universidad de Pensilvania, debatieron sobre los retos de la visualización del proceso de corte y empalme del ARN. Eyra mostró que este se halla frecuentemente alterado en el cáncer, ya que, en las células tumorales, las variantes génicas se presentan en proporciones distintas de las del tejido normal.

Marc A. Martí-Renom, del Centro Nacional de Análisis Genómico y el Centro de Regulación Genómica, se centró en las soluciones para la visualización y exploración de los modelos tridimensionales de la cromatina humana. Estos ayudan a entender qué partes del genoma están cercanas



TRES FORMAS DE REPRESENTAR DATOS GENÓMICOS de un conjunto de unas 500 muestras (o pacientes). Las tres soluciones tienen en común que representan información de múltiples genes, para muchas muestras de pacientes, y que esta información se asocia a un contexto de datos clínicos para facilitar la interpretación. La visualización permite al investigador navegar por los abundantes datos. Según el tipo de indagación, resultará más útil un formato u otro.

CORTESÍA DE LOS AUTORES



LA INTERACCIÓN DEL USUARIO con la información constituye un componente clave de la visualización de datos biológicos complejos. El programa Kartes (desarrollado por el Grupo de Genómica Estructural del Centro Nacional de Análisis Genómico y el Centro de Regulación Genómica) permite al investigador visualizar de forma simultánea las versiones 2D (*secuencias lineales en la pantalla pequeña*) y 3D (*estructura verde en la pantalla grande*) de una región determinada del genoma.

en el espacio y, por tanto, podrían interactuar. El genoma suele representarse de forma lineal (en dos dimensiones). Sin embargo, en el núcleo de las células este adopta una conformación tridimensional que se ha revelado de gran importancia para la regulación de la replicación y la transcripción del ADN. En particular, es importante poder visualizar las regiones del genoma que, si bien alejadas en la representación bidimensional, se hallan en realidad muy próximas. ¿Cómo puede representarse esta información de manera intuitiva? Martí-Renom presentó Kartes (3dgenomes.org), un programa informático que permite cartografiar la tercera dimensión en regiones seleccionadas. Así, el investigador puede orientarse a la vez en el modelo tridimensional y en el bidimensional.

Por otro lado, Roberto Mosca expuso el trabajo que están llevando a cabo en el grupo que Patrick Aloy dirige en el Instituto de Investigación Biomédica. En concreto, presentó Interactome3D (interactome3d.irbbarcelona.org), un servicio web creado para visualizar interacciones entre proteínas, basado en una red que cuenta con la estructura tridimensional de cada proteína (o nodo). Con esta herramienta esperan llenar los huecos existentes entre

diferentes resoluciones del interactoma de proteínas. Cuando se visualiza a baja resolución se muestra la red de interacciones de proteínas; en alta resolución, se visualiza la estructura de la proteína, incluidos los detalles de la interacción entre dos proteínas.

En resumen, VizBi constituye una conferencia interdisciplinaria e inspiradora: una solución de visualización puede servir a varios campos científicos de diferentes maneras.

—Michael P. Schroeder

—Núria López Bigas

ICREA y Universidad Pompeu Fabra

PARA SABER MÁS

ReVision: Automated classification, analysis and redesign of chart images. Manolis Savva et al. en *Proceedings of the 24th annual ACM symposium on user interface software and technology*, págs. 393-402, 2011.

Gitools: Analysis and visualisation of genomic data using interactive heat-maps. Christian Pérez-Llamas y Núria López-Bigas en *PLoS One*, vol. 6, n.º 5, mayo de 2011.

Visualizing multidimensional cancer genomics data. Michael P. Schroeder, Abel González-Pérez y Núria López-Bigas en *Genome Medicine*, vol. 5, n.º 1, enero de 2013.

SciLogs

Ciencia en primera persona

La comunidad de bloggers científicos de habla hispana.

investigacionciencia.es/blogs

Las mariposas del alma

Nuevas ideas en psicología

Antonio Crego
Universidad a Distancia de Madrid

Química, aire y ambiente

La química del mundo que nos rodea

Xavier Giménez
Universidad de Barcelona

De ratones y hombres

Neurociencia imperfecta

Carmen Agustín Pavón
Colegio Imperial de Londres

Meteoritos y ciencias planetarias

Historias sobre meteoritos

J. M. Trigo-López
Instituto de Ciencias del Espacio - CSIC

Mundo vegetal

Los secretos de las plantas

Marta Renato
Universidad de Barcelona

Regeneración animal

Avances en biología regenerativa

Francesc Cebrià
Universidad de Barcelona



Esté al día de las últimas entradas de nuestros bloggers a través de las redes sociales.