

De las redes de información celular a la medicina molecular digital

From cellular information networks to digital molecular medicine

*Rafael Rangel-Aldao*¹

Resumen

La transformación que experimenta la biotecnología con la información biológica a escala de sistemas es de tal magnitud y profundidad, que solo a través de redes globales de cooperación es posible realizar investigación competitiva. Las aplicaciones que ahora emergen, aparte de las biofarmacéuticas y agroindustriales que ya entran en su madurez, dan lugar a una nueva medicina molecular, predictiva, preventiva, personalizada, participativa y accesible en forma digital. Una combinación sin precedentes de técnicas genómicas y proteómicas permite seguir el curso de la expresión de todos los genes y proteínas de cada individuo así como la vida social de tales moléculas informacionales, que es donde radican los eventos cruciales para la predicción del riesgo y la detección precoz de enfermedades comunes. Esa sociedad de bioinformación a escala molecular sigue el mismo tipo de patrones de redes complejas propias de los sistemas sociales y ecológicos, que se repiten en forma fractal a toda escala incluyendo las redes neuronales, genómicas, proteómicas y metabolómicas. A cada nivel de complejidad aparece la firma indeleble de la autoorganización, redes entrelazadas cuyos nodos se conectan entre sí de acuerdo con una ley de potencia con una distribución desigual de enlaces entre los nodos, donde muy pocos dominan todas las conexiones y, por ende, la red completa. La importancia práctica de esos súper nodos, genes y proteínas, se aprecia a partir de 2007 con más de una docena de publicaciones que demuestran la asociación de variaciones genómicas con enfermedades comunes. El valor predictivo de estas asociaciones genómicas, no necesariamente deterministas, se incrementa notablemente al combinarlas con otros factores de riesgo de tipo clínico y metabólico haciendo posible una determinación más precisa del riesgo, la prevención y atención personalizada de la enfermedad.

Palabras clave: proteoma, sistemas, redes, biotecnología, medicina.

Abstract

The emerging field of systems biology is transforming biotechnology to an extent that it is making individual investigators or even consolidated groups become obsolete, thereby making it necessary to work in global collaboration networks to be competitive. The new biology (as just another branch of information science) combined with genomics and proteomics is giving rise to personalised medicine at the molecular level and the ability to predict and prevent the risk of contracting major common diseases, as well as facilitating patients'

¹ Departamento de Tecnología de Procesos Biológicos y Bioquímicos, Sección de Biofísica, Universidad Simón Bolívar, Caracas, Venezuela. rrangel@usb.ve

active participation or even that of healthy individuals in their own care. This so-called P4 medicine (predictive, preventative, personalised and participatory) essentially reflects people's social life in informational biological molecules which are arranged in complex networks following a power law by which very few nodes or hubs made of either genes or their transcription and translation products dominate the entire network through unequal distribution of links or edges. Around one dozen publications of genome-wide association studies (GWAS) have shown how the genomic variations of some of these hubs can be applied to predicting the risk of contracting multigenic and common diseases. Moreover, combining GWAS with clinical and metabolic indices of risk significantly improves the power of such techniques for personalised medicine.

Key words: Proteome, systems, networks, biotechnology, medicine.

Recibido: abril 11 de 2008

Aprobado: mayo 21 de 2008

Introducción

La transformación que experimenta la biotecnología con la genómica y la proteómica, sumada al análisis de la información biológica a escala de sistemas, es de tal magnitud y profundidad, que solo a través de redes globales de cooperación es posible realizar investigación competitiva. Las aplicaciones que ahora emergen de la biotecnología, aparte de las biofarmacéuticas y agroindustriales que ya entran en su madurez, han dado lugar a una nueva medicina molecular, predictiva, preventiva, personalizada y participativa (Hood, 2008), y accesible en forma digital (Rangel-Aldao, 2005).

Esta nueva biotecnología basada en medicina molecular ya permite detectar particularidades moleculares en el ADN de cada persona, en apenas una sustitución de una base nucleotídica en una secuencia objetivo, o polimorfismos nucleotídicos sencillos, abreviados en inglés como SNP, para establecer correlaciones entre la aparición de una enfermedad particular y la terapéutica que se debe seguir en una persona determinada (Pearson y Manolio, 2008).

Con la aplicación de tecnologías genómicas y proteómicas también es posible seguir el curso de la expresión de todos los genes y proteínas de cada individuo, así como la vida social de las moléculas informacionales (Han, 2008), que es donde radican los eventos cruciales para la detección precoz de enfermedades como ve-

remos más adelante. Esos estudios moleculares revisten una complejidad tan grande que escapan al médico practicante común, quien, sin embargo, requiere esa información a la mano para prevenir o curar sus pacientes a tiempo. Igual ocurre con las comunidades y el Estado en su misión básica de asegurar un nivel razonable de salud a los ciudadanos.

La ciencia que sustenta todo ese progreso también experimenta cambios no menos complejos a los de la biotecnología y la medicina. La biología, por ejemplo, se ha hecho insuficiente para explicar el funcionamiento de organismos vivos a escala sistémica, y hoy vemos inéditas fusiones en bloque de aquella con las otras ciencias, las matemáticas, la física, la química, y con disciplinas netamente tecnológicas como la ingeniería y la computación (Kirchner, 2005), con enormes implicaciones en la propiedad intelectual (Rangel-Aldao, 2007). La biotecnología también ha unido fuerzas con otras tecnologías como la nanotecnología para adquirir la complejidad necesaria que demandan las soluciones de la nueva medicina molecular (Crispin et ál., 2008).

Otro aspecto interesante de la transformación biotecnológica y la emergente medicina molecular es que ya no es una ventaja competitiva dominar todas las técnicas biotecnológicas fundamentales pues es posible, por ejemplo, contratar con terceros la secuenciación y el clo-

namiento de genes y proteínas, células, plantas, y hasta animales enteros a un precio que es cada vez más razonable y asequible (Pennisi, 2005). Por tanto, dejó de ser determinante disponer de laboratorios con personal y técnicas altamente especializadas para producir resultados o productos que ahora son fácilmente contratables en el mercado global de los servicios tecnológicos (Rangel-Aldao, 2005); o, peor aún, que tales técnicas ahora resulten inadecuadas para hallar una solución médica personalizada, o siquiera para entender a escala sistémica la fisiología del organismo entero. Solo hay que poseer las conexiones precisas para realizar los objetivos deseados sin tener que hacerlo todo en casa.

La forma de realizar investigación científica propiamente dicha también ha variado considerablemente, y en tales mutaciones hallamos coincidencias entre la biotecnología y la medicina. En ambas, está en curso un proceso de transición del enfoque reduccionista, hasta hace poco típico de la bioquímica y la biología molecular, por ejemplo, hacia otro del tipo sistémico u holístico. Esa transición ha sido impulsada por las dificultades y limitaciones del pensamiento reduccionista para entender la información del genoma humano, pues el conocimiento de las secuencias de todos los genes y de sus elementos reguladores no dio el resultado esperado según el cual esas secuencias serían suficientes para hallar la cura de enfermedades complejas (Lewontin, 2001), y hasta para explicar el comportamiento de un individuo (Keller, 2000), de allí que todas las empresas que basaron su apuesta de valor en la genómica paradójicamente quedaron fuera del mercado al conocerse el genoma y no hallarse en éste las soluciones ofrecidas (Burrill, 2003).

El genoma se revela más bien como un sistema digital cuya complejidad reside en el rol de las secuencias no codificantes (98,8% del genoma) como mecanismo paralelo de control de la expresión genética (Taft y Mattick, 2003). Más aún, a lo largo de esa cadena de expresión que incluye ADN, ARN y proteínas, se produce una intensa vida social de interacciones macromoleculares en respuesta a señales del ambien-

te, que se auto-organizan en redes complejas de “toma de decisiones” (Helikar et ál., 2008) siguiendo una topología de redes de mundo pequeño (Watts y Strogatz, 1998), o libres de escala (Fernández y Sole, 2005) donde muy pocos nodos se distribuyen en forma desigual la mayoría de los enlaces de toda la red (Albert y Barabasi, 2000).

Este enfoque sistémico de la investigación científica moderna tiene como consecuencia la virtual desaparición del investigador y la del médico individual aislado. Ni siquiera el trabajo de grupo puede compensar la brecha de conocimientos necesarios para entender el sistema, de allí que la tendencia sea hacia la colaboración en redes transdisciplinarias sin fronteras facilitadas por las tecnologías de información y comunicación (Castells, 1989), en casi todas las ramas de la ciencia y del saber (Gimerà et ál., 2005).

El código genético digital y las redes sociales de información molecular

A medida que aumenta la complejidad genética por la evolución de las especies, se evidencia el predominio del ADN no codificante en las especies más evolucionadas hasta llegar a las proporciones de los vertebrados (Taft y Mattick, 2003). Mattick (2004) va un poco más allá y encuentra una casi perfecta correspondencia del aumento de las regiones de ADN no codificantes con secuencias que corresponden a redes de ARN (micro-ARN, y de supresión o de interferencia, ARNi). Basado en este hallazgo, el desarrollo de la complejidad genómica sería el resultado de un sistema digital de expresión y regulación genética constituido por intrones, exones, ADN y ARN no codificantes de los organismos multicelulares:

Se trata de un sistema eferente de control endógeno, un programa que en teoría puede dictar trayectorias de diferenciación y desarrollo celular guiado por señales ambientales que proveen contexto para corregir el ruido estocástico del propio programa.

Esta visión de Mattick (2004) coincide con un conjunto de evidencias experimentales que explican cómo esos ARN de cadena corta, o micro ARN, pueden regular la expresión genética a distintos niveles (Novina y Sharp, 2004). El sistema digital, entonces, se construye con dos tipos de información codificadas en el ADN que se expresan en forma paralela: uno, contenido en los genes traducidos (secuencias codificantes) de los elementos estructurales y funcionales que ya conocemos, las proteínas; y el otro, las redes regulatorias (secuencias no codificantes, ADN basura) que especifican la conducta y la vida social de esos genes y sus productos de expresión (Hood y Gala, 2003).

De esta forma, a partir del genoma viajan juntos como en un sistema de comunicación digital (voz, sonido, imágenes, datos, texto) tanto la información necesaria para construir la maquinaria biosintética, como toda aquella requerida para regular los distintos niveles de complejidad que demandan el desarrollo y los patrones básicos de conducta de los seres vivos.

La organización y dinámica de ese andamiaje digital que regula la expresión genética y la subsiguiente diferenciación celular de los organismos superiores, es el objeto de estudio de la nueva biología como fuente de innovación y sustento de una biotecnología emergente basada en la medicina personalizada a escala molecular. A cada nivel de complejidad aparece la firma indeleble de la auto-organización, redes entrelazadas cuyos nodos se conectan de acuerdo con una ley de potencia mediante la cual, como hemos visto, hay una distribución desigual de enlaces entre los nodos, donde muy pocos dominan todas las conexiones y, por ende, la red completa (Barabási y Oltvai, 2004).

Esa sociedad de bioinformación a escala molecular sigue el mismo tipo de patrones de redes característicos de los sistemas ecológicos, neuronales, y hasta de la Internet, entre otros (Milo et ál., 2002). Así vemos que tanto el genoma (Luscombe et ál., 2002), el transcriptoma

(Van Noort et ál., 2004), los dominios de proteínas (Koonin et ál., 2002), los sustratos del metaboloma (Jeong et ál., 2000), y el fisioma (Buchman, 2002), siguen todas las leyes de potencia en su organización y evolución natural (Han, 2008). En el genoma también se hallan familias génicas, pseudogenes, y secuencias codificantes de dominios y pliegues de proteínas, que también siguen una ley de potencia, y en este sentido es elocuente la conclusión de Luscombe et ál. (2002):

“La ley de potencia provee una descripción matemática concisa de una importante propiedad biológica: el dominio completo de unos pocos sobre toda la población”.

Con el transcriptoma, es decir, el conjunto de todos los transcritos del genoma, ocurre algo similar al genoma si consideramos a los genes como nodos y a la coexpresión genética como enlaces de una red en un contexto fisiológico determinado (Van Noort et ál., 2004). De esta forma los nodos adquieren la topología de una red libre de escala y/o del tipo de mundo pequeño (Watts y Strogatz, 1998). Igual ocurre cuando los pliegues y dominios de proteínas son vistos como nodos, y sus conexiones como las relaciones proteína-proteína (Koonin et ál., 2002). Es obvio que tal distribución solo puede significar una cosa, de nuevo, que muy pocos dominios-nodos pueden interactuar con casi toda la red, pero la gran mayoría de los nodos solo puede conectarse a muy pocos, igual que ocurre a macro escala con las personas y sus amistades en las redes sociales (Kosinets y Watts, 2006).

Los sustratos del metaboloma, por su parte, también están organizados en redes de conectividad cuya topología refleja la existencia subyacente de una ley de potencia como lo demuestra un estudio del grupo de Barabási (Jeong et ál., 2000). Este modo de auto-organización en redes libres de escala se puede visualizar en su conjunto como una serie de redes estratificadas con niveles de creciente complejidad, relacionadas todas entre sí, y con un alto grado de coordinación y sincronización desde

el genoma al fisioma. Cada estrato estaría localmente gobernado por súper nodos de genes, RNA, proteínas, metabolitos, u otras moléculas señalizadoras del entorno al cual está sujeto el organismo entero como lo señala la figura 1 (Rangel-Aldao, 2005):

El resultado final de esta red de redes, ilustrada arriba, sería el sistema digital referido por Mattick (2004), mecanismo último responsable del funcionamiento del fisioma. Sobre este andamiaje es que debe ocurrir la sociedad molecular de la información biológica.

La vida social del reguloma

Llamaremos reguloma al conjunto de redes que regulan a escala intracelular el funcionamiento y la respuesta celular ante las señales del entorno, es decir, las del genoma, transcriptoma, proteoma y metaboloma, dejando al fisioma como sistema último para explicar

el comportamiento global de los organismos multicelulares. La enorme complejidad del genoma ha dado lugar a que en paralelo se estudien todos los componentes del reguloma y sus interacciones, es decir, no solo las de los genes entre sí –quién regula a quién (pocos a muchos, sabemos ya)– sino también los mecanismos mediante los cuales ocurren esas relaciones corriente abajo en la expresión genética, es decir, a nivel del proteoma y del metaboloma, y sus respectivas moléculas informacionales capaces de alterar la expresión genética.

El resultado inicial de tales estudios es un trazado cartográfico estático sobre la vida social de todo el reguloma, es decir, mapas globales de interacciones moleculares a escala de organismos completos, desde unicelulares como las levaduras, hasta llegar al fisioma de multicelulares como *Drosophila*, metazoarios como *C. elegans*, mamíferos como el ratón, y el más evolucionado de todos, el humano. Vista la

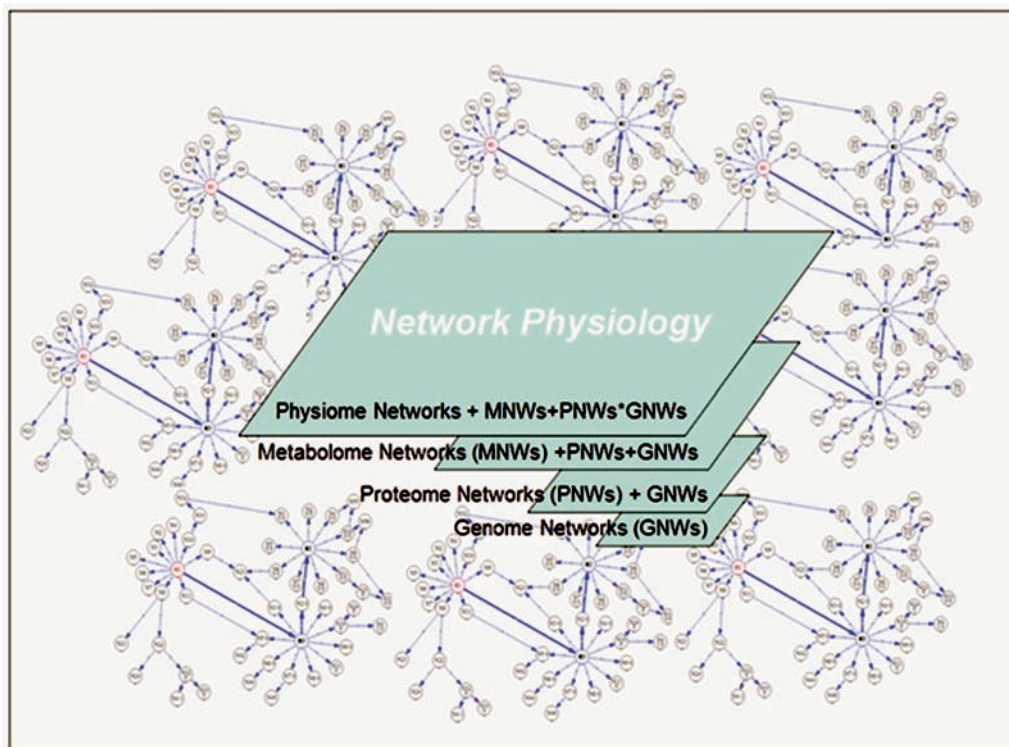


Figura 1. (Niveles de redes interconectadas Redes Genómicas (GNWs), Redes Proteómicas (PNWs), Redes Metabolómicas (MNWs), Redes del Fisioma (Physiome Networks))

existencia de redes libres de escala o de mundo pequeño, desde el genoma al metaboloma de todos estos organismos, no sorprende la topología similar de cada uno de ellos.

En primer lugar, están las relaciones de genes de levaduras, según estudios de Tong y colaboradores de la Universidad de Toronto (2004). La arquitectura de la red se organiza en módulos o racimos de nodos (genes) conectados a otros con funciones relacionadas (movimiento, división celular, replicación de ADN). Para producir tal observación, los autores cruzaron colonias con deleciones viables en 132 genes objetivos (relacionados con las funciones citadas arriba) para hallar 4.000 interacciones en 4.700 mutantes dobles de fenotipo similar (que comparten defectos de adaptación celular). Esta topología es típica de redes de mundo pequeño, vale decir, muchos nodos agrupados por escasos grados de proximidad, e igualmente, libre de escala por la distribución desigual de las conexiones entre los nodos.

En el siguiente nivel de la expresión genética, la transcripción, también en levaduras, ocurre un fenómeno similar al del genoma, con la distribución en redes libres de escala de los genes (nodos) reguladores de otros (conexiones) que son responsables de importantes funciones celulares como el ciclo celular, la diferenciación celular, síntesis de ADN, ARN, y proteínas, la respuesta al ambiente y el metabolismo (Lee et ál., 2002). En ese trabajo del grupo de Richard Young, del MIT, los autores desarrollaron un algoritmo para computar y visualizar todas las conexiones entre genes conocidos y los factores de transcripción que regulan esas funciones celulares listadas arriba. Antes de ello, sin embargo, hicieron sus propias construcciones genéticas con inserciones de secuencias de epitopos (*c-myc*) en los factores de transcripción para la detección inmunológica conjunta con el respectivo gene objetivo.

Si los factores de transcripción son los nodos, y el número de regiones promotoras los enlaces o conexiones, se observa en el trabajo citado que muy pocos reguladores reconocen a

múltiples regiones promotoras mientras que la mayoría de éstas solo son reconocidas por muy pocos de aquéllos. Concluye el grupo de Young con la siguiente afirmación:

Estas observaciones contienen varias implicaciones importantes. El control de la mayoría de los procesos celulares, está caracterizado por la existencia de redes de reguladores de la transcripción que a su vez regulan a otros reguladores. Es también evidente que los efectos de las mutaciones de los reguladores transcripcionales sobre la expresión genética global, medidos estos por patrones de expresión [en micro matrices de ADN/ARN], reflejarán también el efecto de las redes reguladoras en la identificación de los genes objetivo de un regulador en particular.

A escala de proteínas también ocurre un mecanismo similar de regulación, y es revelador el número exponencial de trabajos sobre redes de interacciones proteína-proteína que se descubren cada año (Aloy y Russell, 2004). Al igual que en la transcripción, donde un factor en particular regula muchos genes mediante varias redes, también una proteína en particular puede conectarse como súper nodo a varios procesos celulares a la vez, como la proteína quinasa CDK, uno de los reguladores clave del ciclo celular.

A escala del organismo completo, la vida social de las proteínas, del proteoma, en el ciclo celular de levaduras se puede visualizar cómo la sucesión de complejos moleculares se forman y disocian en sincronía con cada etapa de la división celular tal como lo demuestra el trabajo de Lichtenberg et ál. (2005). En cada uno de esos complejos macro moleculares organizados en redes de mundo pequeño hay súper nodos de proteínas, como los de la proteína quinasa dependiente de ciclinas (CDK) que regulan a las otras quinasas y proteínas objetivo mediante interacciones multimoleculares.

En las levaduras, estos módulos dinámicos se organizan de forma tal que por lo general encontramos en ellos dos o más proteínas que usualmente andan juntas (como en una red social, donde siempre hay “sospechosos

habituales”), que se reúnen en los mismos sitios subcelulares con múltiples complejos macromoleculares de hasta 23 proteínas cada uno (Gavin et ál., 2006).

Esas proteínas de tan intensa vida social se sintetizan en sincronía con las otras relacionadas a ellas para luego encontrarse en espacios subcelulares específicos como el núcleo, endosomas, o mitocondria, entre otros, donde cumplen sus funciones en forma coordinada como se pudo demostrar a partir del uso de la técnica de híbridos apareados en levaduras para detectar los vecinos de cada proteína (Gavin et ál., 2006).

A escala del proteoma humano, las redes de unas 40.000 interacciones proteína-proteína trazadas por el grupo de Rhodes y colaboradores (2005), siguen un modelo probabilístico agrupado como redes de mundo pequeño donde cada módulo representa una función macro molecular. La interconexión de cada uno de esos módulos o subredes podría visualizarse a través de interruptores de un tablero virtual cuya activación/desactivación sería controlada por supernodos (genes y proteínas) para coordinar la toma de decisiones de la red que controla la maquinaria celular (Brandman et ál., 2005).

En el siguiente nivel del reguloma, el metaboloma, al menos en organismos procariontes ocurre también la distribución desigual de nodos de las redes libres de escala tal como lo demuestra el grupo de Barabasi (Almaas et ál., 2004). Mediante un mapa metabólico, similar al de las interacciones proteína-proteína, se agrupan en racimos las respectivas vías metabólicas que usa *E. coli* para adaptarse de un medio rico en glutamato a otro con succinato. Los autores resumen sus hallazgos de la siguiente manera:

Notamos que la red metabólica fluye de manera altamente desigual. Mientras la mayoría de las reacciones ocurren a bajos flujos, la mayor parte de la actividad está dominada por muy pocas reacciones de alto flujo. *E. coli* responde a los cambios de los nutrientes regulando la velocidad de ciertos flujos seleccionados, predo-

minantemente aquellos de la red de alto flujo. Esta conducta probablemente representa una propiedad universal de la actividad metabólica de todas las células, con implicaciones potenciales para la ingeniería metabólica.

De comprobarse esta teoría del grupo de Barabasi, podría concluirse en forma análoga a lo que podría ocurrir con los interruptores del proteoma, que la red de tránsito del metaboloma también tiene su tablero de control para la toma de decisiones (Helikar et al, 2008). Al detectar el organismo los cambios del ambiente se activarían ciertos conmutadores del “tablero central” para conectar eficazmente las rutas metabólicas de flujo lento con las “super autopistas” de las vías rápidas, y producir de este modo la respuesta adaptativa requerida. Es así como, en otras palabras, describe Almaas et ál. (2004) el caso específico de *E. coli*:

Por ejemplo, el cambio de glutamato a succinato como sustrato del crecimiento celular apaga la biosíntesis de vitamina B6, y reconecta hacia otros destinos varias rutas de sustratos metabólicos como las de la coenzima A y la biosíntesis de NAD así como las vías de la respiración, y modifica también el uso del ciclo del ácido cítrico. Aparte de muy pocos otros cambios, las demás rutas metabólicas [lentas] permanecen inalterables. Estos reacomodos resultan en grandes pero específicos cambios de flujos metabólicos.

La sociedad molecular de la información biológica y la medicina personalizada

El enfoque sistémico de la nueva biología permite al menos distinguir tres grandes grupos de aplicaciones genómicas a la medicina, aparte de otras proteómicas que también veremos a continuación. Esos tres grupos corresponden a:

1. El uso de ADN y ARN inmovilizado en micromatrices para detectar redes de expresión de genes relacionados con enfermedades de alto impacto

económico y social como, por ejemplo, el cáncer.

2. La detección de secuencias genéticas o de proteínas como marcadores moleculares de enfermedades crónicas.
3. La detección de polimorfismos genéticos basados en la sustitución de bases nucleotídicas sencillas (SNP) para entender y curar enfermedades.

El ADN inmovilizado de múltiples genes en micro matrices ha permitido dilucidar el origen de una las características más interesantes de los tumores cancerosos, la heterogeneidad celular, aparentemente responsable de las diferencias en patrones clínicos, respuesta terapéutica, y sobrevida de los pacientes. El grupo de Patrick Brown de Stanford estudió el fenómeno en linfoma humano derivado de células B (DLBCL) mediante el análisis de la expresión de múltiples genes de tumores de pacientes con cuadros clínicos claramente diferenciables (Alizadeh et ál., 2000). Los autores encontraron dos formas de DLBCL con distintos patrones de expresión genética indicativos de estadios diferentes de diferenciación a partir de linfocitos B. Un tipo de tumores expresa los genes característicos de linfocitos B del centro germinal; el otro, expresa los genes que son normalmente inducidos *in vitro* en linfocitos B de la circulación periférica. El patrón de expresión genética se visualiza en una micro matriz que emite fluorescencia de distintos colores de acuerdo con el grado de expresión de los genes estudiados, es decir, activados, inactivados, o indiferentes. Lo más interesante y revelador de este estudio de expresión genética de tumores es que sirve de base para el pronóstico terapéutico y sobrevida de los pacientes. No hay duda, pues, que si bien la genómica no ha resuelto el mecanismo por el cual esos genes diferencialmente activados pueden producir distintos tipos de linfomas, sí ayuda en mucho a predecir el curso de la enfermedad y su respuesta al tratamiento médico.

Algo similar ocurre con el segundo grupo de aplicaciones genómicas listado arriba, es de

cir, el uso de secuencias génicas (Srinivas et ál., 2001) y sus productos de expresión como marcadores moleculares para el diagnóstico de muchas enfermedades (Sander, 2000). Entre tales técnicas se destaca la del laboratorio de Charles Lieber, de Harvard, para la detección eléctrica y simultánea de varios marcadores de cáncer de próstata, por ejemplo, con transistores de efecto de campo en matrices de nanotubos de silicio con anticuerpos inmovilizados altamente específicos (Zheng et ál., 2005). Los autores logran una sensibilidad a escala femtomolar por la señal eléctrica producida en los nanotubos ante la unión del antígeno con anticuerpos monoclonales contra: a) el antígeno prostático específico (PSA, ver Brawer, 2001); b) el complejo de α_1 -antiquimiotripsina-PSA (Stenman et ál., 1991); c) el antígeno carcino-embriónico CEA, y d) mucina-1. Con esta técnica también fue posible medir ciertas actividades enzimáticas que son reveladoras de la presencia de cáncer como es el caso de la telomerasa (Kim et ál., 1994). Para ello se sustituyeron los anticuerpos por otro tipo de biosensores como oligonucleótidos cebadores de la telomerasa.

En cuanto al tercer grupo de aplicaciones genómicas, los polimorfismos genéticos, es quizás el campo más activo de la medicina molecular pues trata del tipo más abundante de variaciones genéticas a escala individual y poblacional (Melton, 2003):

El genotipaje con SNP promete revelar por qué ciertas personas son más susceptibles que otras a enfermedades como el cáncer o la diabetes, y qué predispone a otros a sufrir los efectos adversos colaterales de los medicamentos. Las implicaciones para el descubrimiento de nuevos fármacos, investigaciones biomédicas, diagnóstico, y hasta para la medicina forense son considerables.

Cada individuo difiere de otro en una base nucleotídica de cada mil doscientos de los tres billardos de bases que tiene el genoma, es decir, que cada persona tiene al menos unos tres millones de SNP como base para obtener su "firma" individual. Cada SNP tiene un número específico asignado en la base de datos del

Centro Nacional de Información sobre Biotecnología de Estados Unidos, y hasta febrero de 2008 habían sido registrados 12 millones de SNP debidamente caracterizados en alelos específicos con sus respectivas frecuencias e información genómica complementaria (NCBI, 2008).

El genotipaje masivo con SNP comienza a revelar a partir del año 2007 (Christensen y Murray, 2007) por qué ciertas personas son más susceptibles que otras a enfermedades tales como el cáncer o la diabetes, y qué predispone a otros a sufrir los efectos adversos colaterales de los fármacos. A partir de ese mismo año 2007 más de una docena de publicaciones demuestran que mediante los estudios masivos de asociación de variaciones genómicas con enfermedades comunes (abreviados con el acrónimo en inglés de *GWAS*), sí es posible predecir cierto nivel de riesgo genómico ante las enfermedades que más nos afectan: las cardiovasculares, el cáncer, la diabetes y la artritis, entre otras (*The Wellcome Trust Case Control Consortium*, 2007). Hasta el presente se han acumulado casi 100 sitios o *loci* cromosómicos de variaciones genómicas asociadas a unas 40 enfermedades comunes (Pearson y Manolio, 2008). Sin embargo, esas dolencias, dicho sea de paso, son bien complejas en cuanto a los genes que allí intervienen de manera no determinista, es decir, que cuando funcionan mal no necesariamente ocurrirá la enfermedad pues hay factores desencadenantes como hábitos de vida y cambios de ambiente, o intervenciones clínicas que pueden prevenir o retardar la aparición y severidad de la dolencia. En consecuencia, el valor predictivo de estas asociaciones genómicas mejora notablemente cuando se combina la determinación de SNP con otros factores de riesgo de tipo clínico y metabólico como tensión arterial, el índice de masa corporal, y marcadores como colesterolemia, lipoproteínas y proteína C reactiva (PCR), por ejemplo (Kathiresan et ál., 2008).

El siguiente paso en el desarrollo de la medicina molecular personalizada va mucho más allá de estas tres grandes aplicaciones ge-

nómicas y proteómicas que hemos revisado, y corresponde al estudio de la vida social del proteoma y del reguloma. En efecto, ya se dio el primer intento de producir mapas sobre la red de interacciones proteína-proteína de humanos (Rual et ál., 2005):

Aunque distante de ser completos, este y otros mapas similares revelan propiedades topológicas y dinámicas de las redes del interactoma que se relacionan con funciones biológicas específicas, sugiriendo que un mapa del interactoma humano proveerá conocimientos sobre el desarrollo celular y los mecanismos de la enfermedad a escala de sistemas.

Lo más interesante de este mapa con relación a la medicina molecular en desarrollo es que revela las interacciones de 100 proteínas relacionadas con distintas enfermedades en forma de redes libres de escala. De esas interacciones, el trabajo revela la existencia de 424 pares interactivos de proteínas de las cuales al menos una de cada dúo ha sido previamente asociada con alguna enfermedad humana. Sin embargo, el mapa también revela su enorme potencial diagnóstico al hallar los autores que de esos 424 pares de proteínas interactuantes, 352 dúos habían pasado desapercibidos como tales por estudios anteriores no sistémicos. Como ejemplo de esas asociaciones recién descubiertas citan los autores la interacción entre RTN4, un inhibidor del crecimiento neurítico, y SPG21, una proteína asociada a la paraplejia espática. Las mutaciones de SPG21 causan un trastorno autosómico recesivo que produce demencia, el síndrome de Mast. La interacción entre las dos proteínas arroja luz sobre su posible participación en esa enfermedad por cuanto SPG21 se localiza en el sistema de transporte de vesículas del trans-Golgi endosómico y se cree que participa en el transporte y la distribución intracelular de proteínas. Aunque se desconoce la función de la recién descubierta "socia" de SPG21, RNT4 pertenece a una familia de proteínas que también se localiza en el retículo endoplásmico y que son marcadoras de la diferenciación neuroendocrina. Aparte de ello, la estructura de RNT4 comparte con SPG21

dos motivos regulatorios conservados a través de todo el genoma, y que pueden desempeñar un rol importante en el desarrollo del síndrome de Mast. Éste y otros ejemplos de proteínas “sospechosas habituales” que aparecen juntas en lugares y contextos determinados, ilustran cómo se puede conectar la red con procesos biológicos y entender mejor su relación con la aparición de enfermedades. Estos mapas, sin embargo, aún son estáticos y requieren de una siguiente etapa que revele el dinamismo de esas relaciones proteína-proteína.

En resumen, el impacto del enfoque sistémico de la nueva biología y biotecnología sobre la emergente medicina molecular ha hecho que de ahora en adelante sea necesario el análisis en cada paciente de una serie de redes interconectadas en cada uno de los estratos del reguloma. De esta forma será posible abordar las distintas fuerzas innovadoras que moldean tanto a la investigación biomédica como a la práctica clínica del futuro.

Referencias bibliográficas

- Albert, R.; Barabasi, A. L. (2000). Topology of evolving networks: local events and universality. *Phys. Rev. Lett.* 85: 5234-5237.
- Alizadeh, A. A. (2000). Distinct types of diffuse large B-cells lymphoma identified by gene expression profiling. *NEJM* 403: 503-511.
- Alloy, P.; Russell, R. B. (2004). Ten thousand interactions for the molecular biologist. *Nature Biotechnology* 22: 1317, 1321.
- Almaas, E. et ál. (2004). Global organization of metabolic fluxes. *Nature* 427: 839-843.
- Barabasi, AL.; Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5: 101-113.
- Brandman, O. et ál. (2005). Interlinked Fast and Slow Positive Feed Back Loops Drive Reliable Cell decisions. *Science* 310: 496-498.
- Brawer, M. K. (2001). Prostate specific antigen. New York: Marcel Dekker.
- Buchman, T. G. (2002). The community of the self. *Nature* 420: 246-251.
- Burrill, S. (2003). Biotech 2003, Revaluation and Restructuring, March 2003. Burrill and Company.
- Castells, M. (1989). The Informational City. Information Technology, Economic Restructuring and Urban-Regional Process. Oxford: Blackwell Publishers.
- Crispin, R.; Choon, P. F. M.; Khachigian, L. M. (2008). DNzyme technology and cancer therapy: cleave and let die. *Mol. Cancer Ther.* 7: 243-251.
- Christensen, K.; Murray, J. C. (2007). What genome-wide association studies can do for medicine. *N Engl J Med.* 356 (11): 1094-1097.
- Fernandez, P.; Sole, R. V. (2005). Power Laws. Scale-Free Networks and Genome Biology. Koonin, E.V.; Wolf, Y.I.; Karev, G. P. (eds.). New York: Springer. p. 206-224.
- Gavin, A. C. et ál. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature*, published online doi:10.1038. *Nature* 04532.
- Guimerá, R.; Uzzi, B.; Spiro, J.; Nuñez Amaral, L. A. (2005). Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance. *Science* 308: 697-702.
- Han, J. D. (2008). Understanding biological functions through molecular networks. *Cell Res.* 18: 224-237.
- Helikar, T. et ál. (2008). Emergent decision-making in biological signal transduction networks. *PNAS* 105: 1913-1918.
- Hood, L.; Galas, D. (2003). The Digital Code of DNA. *Nature* 421: 444-448.
- Hood, L. (2008). Institute of Systems Biology. Seattle, Washington. Disponible en: http://www.systemsbio.org/Intro_to_ISB_and_Systems_Biology/Predictive_Preventive_Personalized_and_Participatory
- Jeong, H. et ál. (2000). The large-scale organization of the metabolic networks. *Nature* 407: 651-654.
- Karev, G. P. et ál. (2002). The structure of the protein universe and genome evolution. *Nature* 420: 14, 223.
- Katiresan, E. R. et ál. (2008). Polymorphisms Associated with Cholesterol and Risk of Cardiovascular Events. *NEJM* 358: 1240-1249.
- Kirchner, M. (2005). The meaning of systems biology. *Cell* 2005; 121: 503-504.
- Kim, N. W. et ál. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* 266: 2011-2015.

- Keller, E. F. (2000). *The Century of the Gene*. Harvard University Press.
- Koonin, E. V. et al. (2002). The structure of the protein universe and genome evolution. *Nature* 420: 218-223.
- Kossinets, G.; Watts, D. (2006). Empirical Analysis of an Evolving Social Network. *Science* 311: 88-90.
- Lee, T. I. et al. (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298: 799-804.
- Lewontin, R. (2001). It ain't necessarily so: the dream of the human genome and other illusions. New York: Review of Books.
- Lichtenberg, U. et al. (2005). Dynamic complex formation during the yeast cell cycle. *Science* 307: 724-727.
- Luscombe, N. M. et al. (2002). The dominance of the population by a selected few: power-law behaviour applies to a wide variety of genomic properties. Disponible en: <http://genomebiology.com/2002/3/8/research/0040.1>
- Mattick, J.S. (2004) RNA Regulation: a new genetics? *Nature Reviews Genetics*, 5, 316-323.
- McAdams, H.H. and Arkin, A. (1998) Stochastic mechanisms in gene expression. *PNAS*, 94, 814.
- Melton, L. (2003). On the Trail of SNPs. *Nature* 422: 917-923.
- Milo, R. et al. (2002). Networks Motifs: Simple Building blocks of complex networks. *Science* 298: 824-827.
- National Center for Biotechnology Information, National Library of Medicine. Database of Single Nucleotide Polymorphisms. Disponible en: <http://www.ncbi.nlm.nih.gov/SNP/>.
- Novina, C. D.; Sharp, P. A. (2004). The RNAi Revolution. *Nature* 403: 161-164.
- Pearson, T. A.; Manolio, T. A. (2008). How to Interpret a Genome-wide Association Study. *J Amer Med Ass.* 299: 1335-1344.
- Pennisi, E. (2005). Cut-rate genomes on the horizon? *Science* 309: 862.
- Rangel-Aldao, R. (2005). Innovation, Complexity, Networks and Health. *Innovation Strategy Today*, 1, 46-67. Disponible en: <http://www.biodevelopments.org/innovation/index.htm>
- Rangel-Aldao, R. (2007). Patenting the Gene-Hubs of Endoplasmic Reticulum Stress: The Systems Biology Approach. *Recent Patents on Biotechnology* 1: 243-251.
- Rhodes, D. R. et al. (2005). Probabilistic model of the human protein-protein interaction network. *Nature Biotechnology* 23 (8): 951-9.
- Rual, J. F. et al. (2005). Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, doi:10.1038/Nature 04209.
- Sander, C. (2000). Genomic medicine and the future of health care. *Science* 287: 1977-1978.
- Srinivas, P. R. et al. (2001). Trends in biomarker research for cancer detection. *Lancet Oncology* 2: 698-704.
- Stenman, U. H. et al. (1991). A complex between prostate specific antigen and α 1-antitrypsin is the major form of prostate specific antigen in serum of patients with prostatic cancer: assay of the complex improves clinical sensitivity for cancer. *Cancer Research* 51: 222-226.
- Taft, R. J.; Mattick, J. S. (2003). Increasing biological complexity is positively correlated with the relative genome-wide expression of non-protein-coding DNA sequences. Disponible en: <http://www.arxiv.org/abs/q-bio.GN/0401020>
- The Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447: 661-678.
- Tong, A. H. Y. et al. (2004). Global Mapping of the Yeast Genetic Interaction Network. *Science* 303: 808-813.
- Van Noort, V. et al. (2004). The yeast co-expression network has a small-world, scale-free architecture and can be explained by a simple model. *EMBO Reports* 5: 280-284.
- Watts, D. J.; Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature* 393: 440-442.
- Zheng, G. et al. (2005). Multiplexed electrical detection of cancer markers with nanowire sensor arrays. *Nature Biotechnology* 23: 1924-1301.