



# Análisis de redes en la investigación de las enfermedades respiratorias crónicas

## De los fundamentos a la aplicación

Diego Diez<sup>1</sup>, Alvar Agustí<sup>2</sup> y Craig E. Wheelock<sup>3</sup>

<sup>1</sup>Quantitative Immunology Research Unit, World Premier International Immunology Frontier Research Center, Osaka University, Osaka, Japón; <sup>2</sup>Instituto del Tórax, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Universidad de Barcelona, Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Madrid, España; <sup>3</sup>Department of Medical Biochemistry and Biophysics, Division of Physiological Chemistry II, Karolinska Institutet, Estocolmo, Suecia

### Resumen

Las enfermedades respiratorias crónicas son trastornos multifuncionales complejos cuya patogénesis depende de la interrelación entre factores del huésped y ambientales. Para comprenderlas bien e identificar nuevos tratamientos, se necesita un enfoque holístico que integre múltiples tipos y niveles de datos clínicos y biológicos. Con este fin, la aplicación de estrategias basadas en biología de sistemas, en concreto, el análisis de redes, ofrece grandes posibilidades. Estas estrategias basadas en sistemas dependen en gran medida de métodos computacionales que pueden ser todo un reto para los no especialistas. En consecuencia,

en esta Perspectiva pulmonar: 1) se definen los conceptos básicos de redes en biología y los principios del análisis de redes y 2) se comentan recientes aplicaciones del análisis de redes para comprender las enfermedades respiratorias. La intención de esta Perspectiva es ofrecer a los lectores un mayor conocimiento de los puntos débiles y fuertes de los métodos de análisis de redes, además de su utilidad para abordar preguntas de investigación relacionadas con las enfermedades respiratorias crónicas.

**Palabras clave:** *systems biology; systems medicine; protein-protein interaction network; signaling network; gene regulatory network*

En muchos aspectos, el pulmón es un órgano excepcional. Tiene una compleja estructura multidimensional y un contenido celular que cambia rápidamente y está expuesto directamente a un entorno siempre cambiante. El tejido pulmonar contiene más de 40 tipos de células distintas, de los que cuatro se consideran exclusivos del pulmón: células secretoras bronquiales no ciliadas (club [de Clara]), neumocitos de tipo I (células escamosas), neumocitos de tipo II (células alveolares grandes) y macrófagos alveolares (1). En consecuencia, las enfermedades respiratorias crónicas, tales como asma, enfermedad pulmonar obstructiva crónica (EPOC) y fibrosis pulmonar, son trastornos complejos multifactoriales que dependen de la interrelación entre factores genéticos, moleculares, celulares y ambientales (2-4). Por lo tanto, no

sorprende que siga sin comprenderse del todo la patogénesis de numerosas enfermedades respiratorias crónicas ni que falten tratamientos curativos (4). Las estrategias reduccionistas tradicionales en investigación que se centran en una o en un número pequeño de moléculas (p. ej., genes o proteínas) no pueden abordar esta complejidad y la integración de datos de grandes dimensiones de múltiples niveles clínicos y biológicos es el nuevo paradigma propuesto para elucidar la fisiopatología de las enfermedades pulmonares (2, 3). Una consecuencia importante de este enfoque holístico es la generación de conjuntos de datos sumamente grandes (es decir, conjuntos de datos «ómicos», tales como genómica, transcriptómica, proteómica y metabolómica) que pueden proporcionar decenas de miles de datos puntuales de una única muestra (5). El análisis y la

interpretación estos enormes conjuntos de datos ha generado distintos retos y la acumulación de datos por sí misma no ha aumentado nuestros conocimientos de la etiología de la enfermedad (6).

En respuesta a estas dificultades han surgido los campos de la biología de sistemas y la medicina de redes (6-8). Varias revisiones en medicina respiratoria han resaltado la importancia de estos enfoques (9-13) y esta estrategia se ha utilizado con éxito en múltiples estudios (Tabla 1). El análisis de redes es un componente clave de muchas de estas investigaciones (7), pero puede ser confuso para los no especialistas. En consecuencia, en esta Perspectiva pulmonar:

1) se definen los conceptos básicos de redes y análisis de redes en biología y 2) se comenta su aplicación en las enfermedades respiratorias crónicas.

(Original recibido el 5 de marzo de 2014; aceptado en su forma final el 24 de septiembre de 2014)

Subvencionado por: Osaka University y beca JSPS KAKENHI 25430182 (D.D.), becas FIS 12/01117, SEPAR 192/2012 y 065/2013 y RESERCAIXA 2012 (A.A.) y Swedish Heart Lung Foundation y Center for Allergy Research (C.E.W.).

Correspondencia y solicitud de separatas dirigida a Diego Diez, Ph.D., Quantitative Immunology Research Unit, WPI Immunology Frontier Research Center (IFReC), 8F Integrated Life Science Building, Osaka University, 3-1 Yamada-oka, Suita, Osaka 565-0871, Japón. Correo electrónico: diez@ifrec.osaka-u.ac.jp

Am J Respir Crit Care Med Vol 190, Iss 9, pp 981-988, Nov 1, 2014

Copyright © 2014 por la American Thoracic Society

Publicado originalmente en prensa como DOI: 10.1164/rccm.201403-0421PP el 25 de septiembre de 2014

Dirección de Internet: www.atsjournals.org

**Tabla 1.** Aplicaciones del análisis de redes en el estudio de las enfermedades respiratorias

Enfermedad	Método utilizado	Principal resultado	Referencia
Asma	Transcriptoma + redes bayesianas	El TGF- $\beta$ 1 es un regulador clave de respuestas del asma dependientes de la IL13	21
Asma	Transcriptoma + red de PPI	Las vías de PPAR e INF regulan la respuesta a los glucocorticoides	29
Asma	SNP + red de PPI	Identificados SNP no sinónimos en las secuencias de codificación de la red de receptores tipo Toll	39
Asma	OMIM + transcriptoma + red de PPI	Sugiere que GNB2L1 tiene un papel como importante mediador de señalización en el asma	40
Asma y tabaquismo	Transcriptoma + IPA	TIMP1 y TSBH1 están relacionados con el estrés oxidativo	41
EPOC	Comorbilidades en la EPOC (experto) + PPI	Las multimorbilidades en la EPOC comparten genes y vías biológicas	42
EPOC	Transcriptoma + información mutua	La EPOC se asocia al fallo de regulación de vías bioenergéticas	19
EPOC	Transcriptoma + IPA	Existe una clara especificidad de género en respuesta al tabaquismo importante y la EPOC	25
Cáncer de pulmón	Red de PPI	Genes hiperregulados tienden a estar muy conectados, a diferencia de los genes hiporregulados	28
Múltiples (diseasoma)	OMIM	Genes compartidos sugieren un origen común para las enfermedades	17
<i>Pseudomonas aeruginosa</i> en la fibrosis quística crónica	Transcriptoma + metaboloma + KEGG	La hiperregulación de genes de un patógeno que intervienen en la adaptación durante la progresión de la enfermedad representa posibles dianas terapéuticas	24
Hipertensión pulmonar	Transcriptoma de miRNA + predicción de dianas	Identificado miRNA-21 como un regulador de vías de enfermedad	22

*Definición de abreviaciones:* EPOC = enfermedad pulmonar obstructiva crónica; IPA = análisis de vías con Ingenuity (QIAGEN), programa informático comercial para análisis de redes/vías; KEGG = *Kyoto Encyclopedia of Genes and Genomes*; OMIM = *Online Mendelian Inheritance in Man*; PPI = interacciones entre proteínas; SNP = polimorfismos de un solo nucleótido; TGF = factor de crecimiento transformante. El diseasoma contiene información relacionada con varias enfermedades respiratorias, tales como asma, EPOC y cáncer de pulmón.

## Las vías celulares pueden representarse como redes

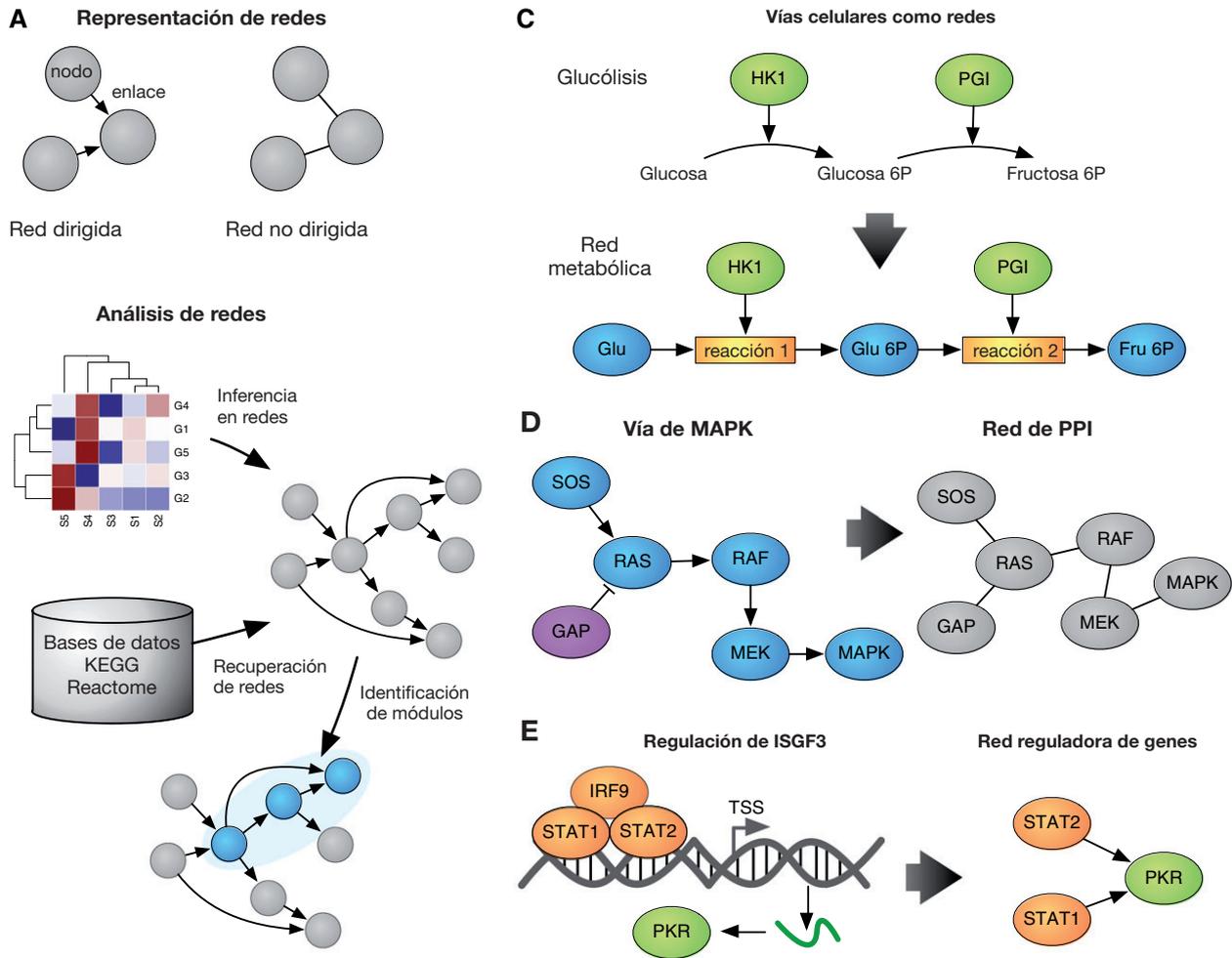
Los procesos bioquímicos y moleculares que se producen en la célula se representan comúnmente como vías (Figuras 1C-1E). Pueden distinguirse tres tipos generales de vías: 1) vías metabólicas, que representan las reacciones que intervienen en la transformación enzimática de sustratos en productos; 2) vías de señalización, que representan las cascadas de proteínas que interaccionan y pequeñas moléculas (ligandos y segundos mensajeros) que permiten la transducción de información desde el exterior de la célula y 3) vías reguladoras de genes, que representan los mecanismos reguladores que controlan los programas transcripcionales que determinan la identidad y los fenotipos celulares (incluida la remodelación

epigenética de la estructura de la cromatina y la unión colaboradora de factores de transcripción [TF] a promotores y potenciadores para regular la expresión de genes diana).

Estas vías se describieron originalmente como entidades individuales, pero están interconectadas y sujetas a intercomunicación (*cross-talk*) a múltiples niveles. Por lo tanto, no es adecuado hablar de vías independientes, sino que la atención debe centrarse en el sistema de vías integradas (es decir, la red de vías celulares). En consecuencia, una célula puede considerarse como un sistema de vías metabólicas, de señalización y reguladoras de genes que incluye miles de moléculas, posiblemente implicadas en cientos de miles de interacciones con otras moléculas. Esta complejidad permite que los sistemas celulares muestren una robustez y

adaptabilidad extremas a diferentes condiciones ambientales (14), pero simultáneamente las hace intrínsecamente resilientes a una caracterización detallada (es decir, no puede determinarse fácilmente la contribución de cada componente en el sistema integral). Por suerte, las vías biológicas se definen en términos de moléculas que interaccionan con otras, permitiendo su representación como redes, que pueden investigarse mediante herramientas de análisis de redes (Figuras 1C-1E).

En su forma más simple, una red se define como un grupo de puntos interconectados, que están formados por un conjunto de nodos y enlaces que contienen información sobre el sistema que se está modelando. Los nodos son puntos de unión o conexión que representan las variables de la red y los enlaces reflejan las relaciones entre ellos (Figura 1A).



**Figura 1.** (A) Los nodos representan entidades (p. ej., moléculas, enfermedades, condiciones) y los enlaces relacionan los nodos entre sí. Las relaciones pueden ser dirigidas (*izquierda*) o no dirigidas (*derecha*). (B) Flujo de trabajo del análisis de redes: la estructura de redes puede aprenderse *de novo* a partir de datos experimentales u obtenerse de bases de datos de conocimientos como *Kyoto Encyclopedia of Genes and Genomes* (KEGG) y *Reactome*. La integración de datos experimentales permite identificar subredes relacionadas con enfermedades (módulos). (C-E) Las vías celulares pueden representarse como redes. En las redes metabólicas (C), los nodos representan compuestos metabólicos (glucosa, glucosa-6P, fructosa-6P) y enzimas (HK1 = hexocinasa, PGI = fosfoglucoisomerasa) conectadas a través de una reacción. En las redes de señalización (D), los nodos representan proteínas (p. ej., SOS, RAS, RAF) y los enlaces, los procesos que facilitan ( $\downarrow$ ) o bloquean ( $\perp$ ) el flujo de información (es decir, transducción de señales; p. ej., fosforilación de MEK por RAF1). En las redes reguladoras de genes (E), los nodos representan factores de transcripción (TF) (p. ej., STAT1 y STAT2) y genes diana (p. ej., PKR) y los enlaces indican activación (o represión) de la expresión del gen diana por el TF regulador. ISGF3 = factor génico 3 estimulado por IFN; PPI = interacciones entre proteínas; TSS = sitio de inicio de la transcripción.

Esta estructura (o disposición) se puede usar para representar múltiples tipos de datos desde interacciones sociales (p. ej., Facebook) e historia cultural (15) hasta la señalización celular (16). Esta flexibilidad ha impulsado su incorporación en estrategias de investigación que requieren la capacidad de integrar e interrogar grandes y dispares conjuntos de datos (p. ej., datos basados en la ómica). Por lo tanto, las redes pueden usarse para representar vías celulares, en las que los nodos corresponden a moléculas y los enlaces, a acciones bioquímicas que realizan las moléculas. Los enlaces pueden ser dirigidos (p. ej., la proteína A fosforila la proteína B, pero lo contrario no es cierto) o no dirigidos (p. ej., la proteína A interactúa con la proteína B y lo

contrario también es cierto) (Figura 1A). Los nodos y enlaces de la red transmiten información sobre los procesos biológicos que se producen en la célula y esta información puede ser utilizada por métodos computacionales con el objetivo de identificar el comportamiento dinámico de la red (es decir, cambios en la actividad bioquímica o fisiológica a lo largo del tiempo). En consecuencia, existe interés por identificar la estructura de las redes de enfermedades y comprender los efectos dinámicos de las perturbaciones (p. ej., genes inactivados [*knock out*]/silenciados [*knock down*], intervención farmacológica, etc.).

Las redes también pueden usarse para modelar otros tipos de información. Por ejemplo, el llamado *diseasoma humano* conecta enfermedades (nivel clínico) con

genes (nivel molecular) (17). Esta red bipartita (es decir, una red que contiene dos conjuntos desagrupados de nodos: enfermedades y genes) puede producir dos proyecciones diferentes: 1) la red de genes, cuyos nodos son genes (y dos genes están conectados si se asocian a la misma enfermedad) y 2) la red de enfermedades, cuyos nodos son enfermedades (y dos enfermedades están conectadas si se asocian a los mismos genes) (17). El análisis de la proyección de enfermedades reveló que muchas enfermedades comparten conexiones génicas, lo que sugiere que muchos trastornos tienen causas comunes subyacentes (p. ej., la mayoría de cánceres están conectados a través de TP53 y PTEN, resaltando el papel que desempeñan las mutaciones de estos genes en la patogénesis del cáncer).

## Aplicación del análisis de redes en la investigación de las enfermedades respiratorias

### Métodos de inferencia en redes

Una de las aplicaciones más comunes del análisis de redes es la inferencia en redes, en la que se desconoce la estructura de la red (es decir, las relaciones entre nodos de conexión) y el objetivo es reconstruirla a partir de datos experimentales. Dicho de otro modo, el objetivo es identificar las relaciones reguladoras o las interacciones entre componentes biológicos. Por ejemplo, en redes de señalización, el objetivo es identificar qué proteína regula a qué otra proteína y, en las redes reguladoras de genes, qué TF regulan a qué genes diana.

La determinación experimental de vías celulares es costosa y larga; por lo tanto, los métodos que reconstruyen automáticamente vías celulares a partir de datos ómicos experimentales han despertado un interés considerable. Se han desarrollado numerosas estrategias para resolver esta cuestión y, en algunos casos, han aportado conocimientos mecanísticos. Por ejemplo, la información mutua, una medida de la interdependencia estadística entre dos variables aleatorias, puede usarse para revertir la ingeniería de redes celulares (18). Turan y colaboradores aplicaron esta estrategia para integrar la expresión génica, los niveles sanguíneos de citocinas, el intercambio de gases pulmonares y la función del músculo esquelético en la EPOC, e identificaron la hipoxia como un factor potencialmente crítico que contribuye en la disfunción del músculo esquelético en estos pacientes (19).

Otra estrategia de inferencia en redes comporta el uso de redes bayesianas, que son un tipo de modelo gráfico probabilístico en el que los nodos representan densidades de probabilidades condicionales y la red proporciona la distribución de probabilidad conjunta sobre el conjunto de variables. Las redes bayesianas se han aplicado para reconstruir satisfactoriamente vías de señalización y reguladoras conocidas a partir de la expresión génica y otros tipos de datos ómicos (16) y también para construir sistemas de decisión para el diagnóstico y el tratamiento de la neumonía asociada al respirador (20). Novershtern y colaboradores utilizaron redes bayesianas para integrar perfiles de expresión génica en un modelo murino de asma, que resaltó el papel del factor de crecimiento transformante beta 1 (TGF- $\beta$ 1) como regulador fundamental de respuestas dependientes de la IL13 (21).

Una limitación general que impide la aplicación satisfactoria de las estrategias de inferencia en redes a las enfermedades humanas es la necesidad de grandes conjuntos de datos generados bajo diferentes condiciones experimentales (es decir, perturbaciones del sistema) para predecir de manera fiable interacciones entre elementos del sistema. Este es un obstáculo concreto de muchos estudios clínicos en los que las cifras de las muestras y la capacidad para realizar perturbaciones a menudo están limitadas. Una estrategia algo tangencial para reconstruir redes es integrar pruebas experimentales en combinación con predicciones computacionales. Por ejemplo, Parikh y colaboradores crearon una red de genes diana/miRNA asociados a la hipertensión pulmonar utilizando una predicción generada por ordenador de dianas de miRNA a partir de datos de expresión de miRNA (22). Demostraron que miRNA-21 reprime la expresión de RhoB y Rho-cinasa, que está relacionada con un descenso de angiogénesis y vasodilatación. Esta estrategia también se aplica comúnmente para reconstruir redes reguladoras de genes utilizando datos de ChIP-seq de TF.

### Métodos de identificación de módulos

La segunda aplicación más común del análisis de redes es la identificación de módulos. Un módulo, o subred, es un subgrupo distinto de nodos y enlaces relacionados con la funcionalidad que se espera interaccionen físicamente con mayor frecuencia y muestren dependencias funcionales más fuertes entre sí que otras moléculas en la red (23). En la identificación de módulos, la red global se interroga con un rango de métodos de análisis topológico (revisados en la referencia 23) para identificar regiones densamente interconectadas asociadas a la misma característica de interés (p. ej., tratamiento del asma, Figura 1B). Estos métodos se basan en el principio de que estructuras modulares como cascadas bioquímicas/de señalización o complejos proteicos muestran patrones de interacción característicos. Los módulos pueden investigarse además por estrategias de enriquecimiento funcional (p. ej., análisis del enriquecimiento en ontología génica) para evaluar si las moléculas implicadas en un análisis son funcionalmente relevantes, además de descubrir funciones compartidas imprevistas. La identificación de módulos puede ser una potente estrategia para identificar subfenotipos de enfermedad o respondedores terapéuticos únicos y se está incorporando en estudios

basados en medicina de sistemas en la enfermedad respiratoria (5).

### Bases de datos de conocimientos

A partir de los datos podría deducirse la estructura de la red, pero también podría obtenerse de bases de datos que acumulan los conocimientos actuales sobre vías celulares. Habitualmente, estas bases de datos organizan la información distinguiendo entre vías metabólicas y de señalización. Las redes metabólicas son, con diferencia, la red biológica mejor caracterizada, de la que se han determinado la mayoría de vías y relaciones (reacciones) más importantes. Diversas bases de datos dan acceso a información conocida de redes metabólicas en seres humanos y otros organismos, tales como la *Kyoto Encyclopedia of Genes and Genomes* (KEGG; [www.kegg.jp](http://www.kegg.jp)) y el proyecto Reactome ([www.reactome.org](http://www.reactome.org)). Por ejemplo, Oberhardt y colaboradores integraron datos de expresión génica con la red metabólica para el patógeno *Pseudomonas aeruginosa* en el contexto de la infección pulmonar por fibrosis quística e identificaron adaptaciones del patógeno a la progresión de la enfermedad (24). La red metabólica se determinó experimentalmente por análisis de flujo y la asignación de genes a reacciones se consiguió con información de vías metabólicas de KEGG y otras bases de datos (24).

Las redes de señalización no están totalmente caracterizadas y constantemente se están descubriendo nuevas vías. Los conocimientos actuales se catalogan en bases de datos web como KEGG y Reactome. Una característica especialmente útil de la base de datos KEGG es que contiene «vías de enfermedad», que son grupos de vías que contienen genes asociados a una enfermedad concreta (p. ej., asma). Esta información puede combinarse con alteraciones en la expresión génica para centrarse en alteraciones específicas de la enfermedad. Un ejemplo es nuestro estudio en el que examinamos el transcriptoma de los leucocitos después de la exposición al humo de cigarrillo (25). Los resultados identificaron un número de genes (además de sus ontologías y redes de interacción) que se expresaron de manera diferencial y exclusiva en pacientes con enfermedad pulmonar obstructiva crónica y revelaron diferencias significativas entre sexos (25). Por desgracia, la representación de las vías de enfermedad respiratoria en la KEGG es bastante limitada, solo incluye información sobre «cáncer de pulmón microcítico», «cáncer de pulmón no microcítico» y «asma», siendo esta última

una representación muy sencilla de los complejos mecanismos que conducen a la enfermedad. Esto pone de relieve una limitación importante en el campo respiratorio y es necesario dedicar significativos esfuerzos para aumentar el contenido y la funcionalidad de las vías específicas de la enfermedad respiratoria para aprovechar al máximo las estrategias de la biología de sistemas.

**Redes de interacciones entre proteínas**

Una característica clave de las redes de señalización es que las proteínas deben interactuar con otras proteínas para realizar sus funciones reguladoras. Por lo tanto, una forma de descifrar sistemáticamente las redes de señalización celular es identificar la red de interacciones entre proteínas (PPI) o interactoma (26). A pesar de algunas limitaciones (26), las redes de PPI se han aplicado con éxito para elucidar mecanismos biológicos por integración de datos de PPI con cambios transcripcionales (27). En estas estrategias, la significación estadística del cambio en la expresión puede usarse para puntuar cada proteína en la red de PPI: puntuaciones más altas se corresponden con genes con valores de *p* menores. Esta red puntuada puede usarse para identificar módulos que contienen proteínas con puntuaciones altas, que se asocian a las respuestas celulares (Figura 2A) (23). Wachi y colaboradores utilizaron una estrategia interactoma/transcriptoma combinada para estudiar las características

topológicas de genes expresados de manera diferente en tejido de pulmón con cáncer de células escamosas (28). Publicaron que genes expresados de manera diferente en cáncer de pulmón tienden a estar conectados en la red de PPI con otros genes expresados de manera diferente, lo que sugiere la existencia de módulos funcionales de genes coregulados. De forma similar, combinamos datos de PPI y expresión génica para estudiar el efecto de corticosteroides inhalados en pacientes con asma tratados con fluticasona (29). El módulo de red de PPI identificado sugirió una relación entre la regulación de la expresión génica por los glucocorticoides, la activación de la vía del PPAR (por hiperregulación de PPARGC1A y RXRA) y la represión de la vía del interferón de tipo 1 (por hiporregulación de STAT1 e IRF9).

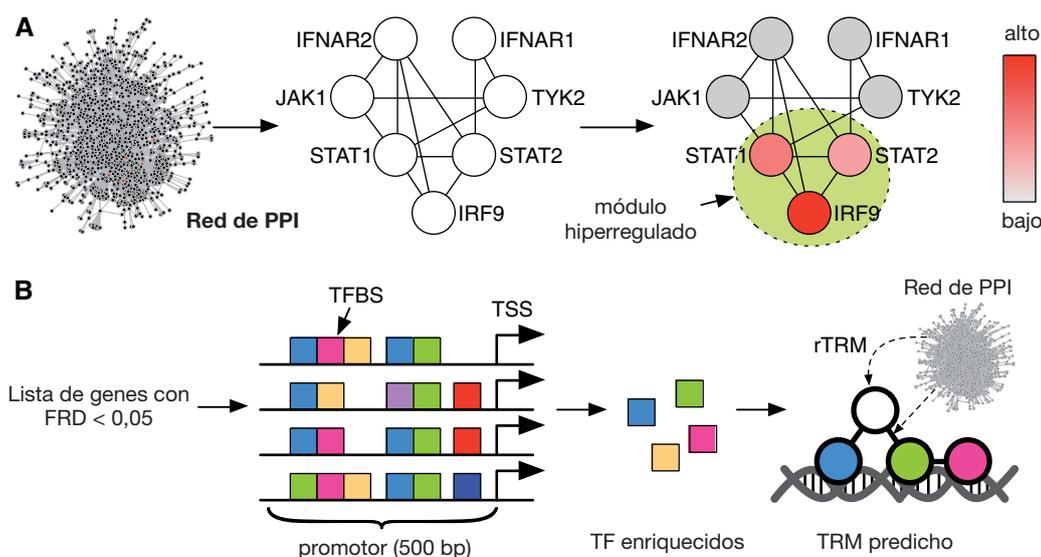
**Redes reguladoras de genes**

A pesar del papel fundamental que desempeñan los TF en la regulación de los procesos celulares y su asociación con las enfermedades, las redes reguladoras de genes están, con diferencia, menos caracterizadas (30). Estudios pioneros en bacterias y levaduras permitieron descubrir motivos de redes esenciales (es decir, un tipo recurrente de subred, igual que un bucle de prealimentación) que regulan la dinámica de la regulación transcripcional (31). Sin embargo, la regulación transcripcional es más compleja en eucariotas superiores. Los potenciadores pueden localizarse a miles de nucleótidos de distancia de sus dianas, desafiando la

identificación de potenciadores/promotores funcionales y el emparejamiento de TF y genes diana. Además, los TF colaboran con otros TF para formar un código combinado que regula la expresión de genes diana en los llamados módulos de regulación transcripcional (TRM). Elucidar estos TRM es fundamental para comprender cómo los TF regulan los programas transcripcionales, pero requiere la caracterización sistemática de muchos perfiles de unión de TF (32). Recientemente explotamos esta idea y desarrollamos un nuevo método de análisis de redes (llamado rTRM) que usa información acerca de las PPI y la unión de TF para identificar TRM (Figura 2B) (32, 33). La aplicación de rTRM a la enfermedad respiratoria debería servir para elucidar el papel de los TF en el desarrollo de fenotipos de enfermedad.

**Análisis de redes como una plataforma integradora**

En las últimas dos décadas se ha producido una proliferación de tecnologías capaces de medir diversas moléculas en las células y sus interacciones asociadas (5), tales como tecnologías de hibridación con sondas (p. ej., micromatrices [microarrays], ChIP-on-chip), tecnologías de secuenciación (p. ej., RNA-seq, ChIP-seq) y espectrometría de masas (p. ej., proteoma, fosfoproteoma, metaboloma, glicoma, lipidoma). Estas estrategias se han traducido en un aumento considerable de datos disponibles para investigar enfermedades respiratorias (5). Debido a que los avances en estas tecnologías sin duda seguirán aumentando,



**Figura 2.** (A) Uso de redes de interacciones entre proteínas (PPI) para tener conocimientos de las redes de señalización. Los *nodos rojos* representan proteínas con una puntuación alta (valor de *p* bajo). (B) Uso de redes de PPI para investigar redes de regulación transcripcional. FDR = tasa de descubrimientos falsos; TF = factor de transcripción; TFBS = sitios de unión de TF; TSS = sitio de inicio de la transcripción; TRM = módulos de regulación transcripcional.

**Tabla 2.** Herramientas informáticas seleccionadas para el análisis de redes

Nombre	Plataforma	Propósito	Ref
bayelviraApp	Cytoscape	Inferencia en redes bayesianas	43
jActiveModules	Cytoscape	Identificación de módulos de redes	27
CyTargetLinker	Cytoscape	Información de relación reguladora (factor de transcripción)	44
DisGeNET	Cytoscape	Consultar redes de genes-enfermedades	45
minet	Bioconductor	Inferencia en redes utilizando información mutua	46
birta	Bioconductor	Inferencia en redes bayesianas	47
BioNet	Bioconductor	Identificación de módulos de redes	48
rTRM	Bioconductor	Identificación de módulos de regulación transcripcional	32
Cyrface	Cytoscape	Interfaz entre Cytoscape y R	49
RCytoscape	Bioconductor	Interfaz entre Cytoscape y R	50

Herramientas seleccionadas de los paquetes Cytoscape (<http://apps.cytoscape.org>) y Bioconductor (<http://www.bioconductor.org/packages/release/>) para el análisis de redes biológicas. Para información adicional de otras herramientas informáticas para el análisis de redes en R, consultar CRAN (<http://www.r-project.org>).

al mismo tiempo también crecerán la cantidad, el tipo y la estructura de los datos disponibles para los estudios basados en la medicina de sistemas. En consecuencia, un reto importante es cómo integrar estos conjuntos de datos en modelos de enfermedad coherentes. El análisis de redes proporciona un marco para incorporar esta información. Por ejemplo, los nodos pueden integrar información en diferentes mediciones realizadas por diferentes tecnologías en la misma molécula. Asimismo, los nodos que representan enfermedades o fenotipos también pueden integrarse en la red por medio de su conexión con otros nodos (p. ej., enfermedades conectadas a genes en la red del diseasoma [17, 34]). Sin embargo, un problema de esta estrategia es que la creciente complejidad del modelo de redes comporta un aumento proporcional de la complejidad computacional (es decir, un aumento del número de parámetros). Además, existe una limitación en la disponibilidad de datos experimentales que miden el(los) efecto(s) de las perturbaciones (p. ej., estado de enfermedad, inactivación, silenciamiento, etc.) a lo largo del tiempo, necesarios para estimar de manera fiable el comportamiento dinámico de la red. Sin embargo, con la mayor disponibilidad de datos experimentales acerca de interacciones bioquímicas y fisiopatológicas en el pulmón, se ampliará nuestra capacidad de aplicar metodologías de redes.

## Herramientas informáticas para el análisis de redes

Existen numerosas herramientas informáticas para el análisis de redes,

que no pueden revisarse aquí de manera sucinta, y se remite al lector interesado a revisiones recientes (35, 36). Aquí destacaremos dos proyectos populares, Cytoscape y Bioconductor, que ofrecen varias herramientas gratuitas para el análisis de redes. Ambos recursos son proyectos bien establecidos, que probablemente garantizan su continuidad a largo plazo y la adaptación a nuevas tecnologías. Se espera que estos proyectos sean las principales herramientas utilizadas en las aplicaciones del análisis de redes para investigar enfermedades pulmonares.

Cytoscape (<http://cytoscape.org>) proporciona una interfaz sencilla de señalar y activar (37). La funcionalidad básica incluye leer, anotar y manipular redes. El entorno informático puede aceptar múltiples tipos de datos en diferentes formatos que después pueden incorporarse en la generación de redes. También es posible importar información de redes de bases de datos biológicas, tales como KEGG, Reactome y BioGRID. Disponen de capacidades adicionales como accesorios aportados por la comunidad investigadora (Tabla 2).

El proyecto Bioconductor (<http://bioconductor.org> [38]) implementa métodos para analizar datos biológicos de la plataforma R (<http://r-project.org>). Aunque trabajar en R puede ser un reto para el usuario no iniciado, su flexibilidad y potentes capacidades estadísticas, junto con la gran colección de paquetes aportados, hacen de R uno de los paquetes informáticos estadísticos más populares. Existen muchos paquetes de R centrados en aplicaciones y redes biológicas. En su mayoría, estos paquetes pueden encontrarse en el depósito de CRAN (<http://cran.r-project.org>) y en Bioconductor (Tabla 2). Por último, diversos

paquetes permiten la comunicación entre Cytoscape y R, aportando lo mejor de cada plataforma (Tabla 2).

## Conclusiones

La creciente acumulación de conjuntos de datos ómicos que describen mecanismos celulares y su relación con fenotipos clínicos tiene potencial para elucidar mecanismos de enfermedades pulmonares complejas. Sin embargo, no es fácil entender ni analizar esta información utilizando estrategias de investigación clásicas y se necesitan nuevos métodos integradores basados en sistemas. El análisis de redes proporciona un punto de partida para estas nuevas estrategias analíticas, que deben combinarse con un fenotipado clínico riguroso para proporcionar perfiles cuantitativos de pacientes. Aunque prometedores, existen numerosos obstáculos para la amplia adopción de estas estrategias de investigación, tales como: 1) métodos insuficientes para la integración de datos y la generación/interrogación de redes, especialmente para los no especialistas; 2) pocas bases de datos con suficiente contenido biológico, especialmente en relación con las enfermedades pulmonares y 3) escasez de interacciones entre investigadores clínicos e informáticos. Creemos que estas limitaciones se abordarán en un futuro próximo y los análisis de redes acabarán siendo una estrategia común en estudios de enfermedades respiratorias humanas basados en la medicina de sistemas.

**Declaraciones de autoría** disponibles con el texto de este artículo en [www.atsjournals.org](http://www.atsjournals.org).

## Bibliografía

1. Sorokin SP. The cells of the lungs. In: Nettesheim P, Hanna MG Jr, Deatherage JW Jr, editors. Morphology of experimental respiratory carcinogenesis: proceedings of a Biology Division, Oak Ridge National Laboratory Conference. Gatlinburg, TN: National Technical Information Service; 1970. pp. 3–43.
2. Agustí A, Sobradillo P, Celli B. Addressing the complexity of chronic obstructive pulmonary disease: from phenotypes and biomarkers to scale-free networks, systems biology, and P4 medicine. *Am J Respir Crit Care Med* 2011;183:1129–1137.
3. Faner R, Rojas M, Macnee W, Agustí A. Abnormal lung aging in chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2012;186:306–313.
4. Wu W, Kaminski N. Chronic lung diseases. *Wiley Interdiscip Rev Syst Biol Med* 2009;1:298–308.
5. Wheelock CE, Goss VM, Balgoma D, Nicholas B, Brandsma J, Skipp PJ, Snowden S, Burg D, D'Amico A, Horvath I, et al. Application of 'omics technologies to biomarker discovery in inflammatory lung diseases. *Eur Respir J* 2013;42:802–825.
6. Hood L, Balling R, Auffray C. Revolutionizing medicine in the 21st century through systems approaches. *Biotechnol J* 2012;7: 992–1001.
7. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* 2011;12:56–68.
8. Bousquet J, Anto JM, Sterk PJ, Adcock IM, Chung KF, Roca J, Agustí A, Brightling C, Cambon-Thomsen A, Cesario A, et al. Systems medicine and integrated care to combat chronic noncommunicable diseases. *Genome Med* 2011;3:43.
9. Studer SM, Kaminski N. Towards systems biology of human pulmonary fibrosis. *Proc Am Thorac Soc* 2007;4:85–91.
10. Ahmad F, Champion HC, Kaminski N. Toward systems biology of pulmonary hypertension. *Circulation* 2012;125:1477–1479.
11. Maher TM. Beyond the diagnosis of idiopathic pulmonary fibrosis; the growing role of systems biology and stratified medicine. *Curr Opin Pulm Med* 2013;19:460–465.
12. Dahlin A, Tantisira KG. Integrative systems biology approaches in asthma pharmacogenomics. *Pharmacogenomics* 2012;13: 1387–1404.
13. Faner R, Cruz T, López-Giraldo A, Agustí A. Network medicine, morbidity and the lung in the elderly. *Eur Respir J* 2014;44: 775–788.
14. Kitano H. Biological robustness. *Nat Rev Genet* 2004;5:826–837.
15. Schich M, Song C, Ahn YY, Mirsky A, Martino M, Barabási AL, Helbing D. Quantitative social science. A network framework of cultural history. *Science* 2014;345:558–562.
16. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. Causal protein-signaling networks derived from multiparameter single-cell data. *Science* 2005;308:523–529.
17. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. *Proc Natl Acad Sci USA* 2007;104: 8685–8690.
18. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. *Nat Genet* 2005;37:382–390.
19. Turan N, Kalko S, Stincone A, Clarke K, Sabah A, Howlett K, Curnow SJ, Rodriguez DA, Cascante M, O'Neill L, et al. A systems biology approach identifies molecular networks defining skeletal muscle abnormalities in chronic obstructive pulmonary disease. *PLoS Comput Biol* 2011;7:e1002129.
20. Visscher S, Lucas PJ, Schurink CA, Bonten MJ. Modelling treatment effects in a clinical Bayesian network using Boolean threshold functions. *Artif Intell Med* 2009;46:251–266.
21. Novershtern N, Itzhaki Z, Manor O, Friedman N, Kaminski N. A functional and regulatory map of asthma. *Am J Respir Cell Mol Biol* 2008;38:324–336.
22. Parikh VN, Jin RC, Rabello S, Gulbahce N, White K, Hale A, Cottrill KA, Shaik RS, Waxman AB, Zhang YY, et al. MicroRNA-21 integrates pathogenic signaling to control pulmonary hypertension: results of a network bioinformatics approach. *Circulation* 2012;125: 1520–1532.
23. Mitra K, Carvunis AR, Ramesh SK, Ideker T. Integrative approaches for finding modular structure in biological networks. *Nat Rev Genet* 2013;14:719–732.
24. Oberhardt MA, Goldberg JB, Hogardt M, Papin JA. Metabolic network analysis of *Pseudomonas aeruginosa* during chronic cystic fibrosis lung infection. *J Bacteriol* 2010;192:5534–5548.
25. Faner R, Gonzalez N, Cruz T, Kalko SG, Agustí A. Systemic inflammatory response to smoking in chronic obstructive pulmonary disease: evidence of a gender effect. *PLoS ONE* 2014;9:e97491.
26. Vidal M, Cusick ME, Barabási AL. Interactome networks and human disease. *Cell* 2011;144:986–998.
27. Ideker T, Ozier O, Schwikowski B, Siegel AF. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* 2002;18:S233–S240.
28. Wachi S, Yoneda K, Wu R. Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics* 2005;21:4205–4208.
29. Diez D, Goto S, Fahy JV, Erle DJ, Woodruff PG, Wheelock AM, Wheelock CE. Network analysis identifies a putative role for the PPAR and type 1 interferon pathways in glucocorticoid actions in asthmatics. *BMC Med Genomics* 2012;5:27.
30. Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* 2012;13:613–626.
31. Alon U. Network motifs: theory and experimental approaches. *Nat Rev Genet* 2007;8:450–461.
32. Diez D, Hutchins AP, Miranda-Saavedra D. Systematic identification of transcriptional regulatory modules from protein-protein interaction networks. *Nucleic Acids Res* 2014;42:e6.
33. Hutchins AP, Diez D, Takahashi Y, Ahmad S, Jauch R, Tremblay ML, Miranda-Saavedra D. Distinct transcriptional regulatory modules underlie STAT3's cell type-independent and cell type-specific functions. *Nucleic Acids Res* 2013;41:2155–2170.
34. Menche J, Sharma A, Cho MH, Mayer RJ, Rennard SI, Celli B, Miller BE, Locantore N, Tal-Singer R, Ghosh S, et al. A dVIsive Shuffling Approach (ViStA) for gene expression analysis to identify subtypes in chronic obstructive pulmonary disease. *BMC Syst Biol* 2014;8:S8.
35. Thomas S, Bonchev D. A survey of current software for network analysis in molecular biology. *Hum Genomics* 2010;4:353–360.
36. Diez D, Wheelock AM, Goto S, Haeggström JZ, Paulsson-Berne G, Hansson GK, Hedin U, Gabrielsen A, Wheelock CE. The use of network analyses for elucidating mechanisms in cardiovascular disease. *Mol Biosyst* 2010;6:289–304.
37. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–2504.
38. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;5:R80.
39. Renkonen J, Joenvaara S, Parviainen V, Mattila P, Renkonen R. Network analysis of single nucleotide polymorphisms in asthma. *J Asthma Allergy* 2010;3:177–186.
40. Hwang S, Son SW, Kim SC, Kim YJ, Jeong H, Lee D. A protein interaction network associated with asthma. *J Theor Biol* 2008;252:722–731.
41. Freisztat RJ, Benton AS, Watson AM, Wang Z, Rose MC, Hoffman EP. Delineation of a gene network underlying the pulmonary response to oxidative stress in asthma. *J Invest Med* 2009;57:756–764.
42. Grosdidier S, Ferrer A, Faner R, Pinerio J, Roca J, Cosio B, Agustí A, Gea J, Sanz F, Furlong LI. Network medicine analysis of COPD multimorbidities. *Respir Res* 2014;15:111.
43. Elvira Consortium. Elvira: an environment for creating and using probabilistic graphical models. In: Ga'mez JA, Salmerón A, editors. Proceedings of the First European Workshop on Probabilistic Graphical Models (PGM'02). Cuenca, Spain: Electronic Proceedings; 2002. pp. 222–230.
44. Kutmon M, Kelder T, Mandaviya P, Evelo CT, Coort SL. CyTargetLinker: a cytoscape app to integrate regulatory interactions in network analysis. *PLoS ONE* 2013;8:e82160.
45. Bauer-Mehren A, Rautschka M, Sanz F, Furlong LI. DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. *Bioinformatics* 2010;26:2924–2926.
46. Meyer PE, Lafitte F, Bontempi G. minet: a R/Bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinformatics* 2008;9:461.
47. Zacher B, Abnaof K, Gade S, Younesi E, Tresch A, Fröhlich H. Joint Bayesian inference of condition-specific miRNA and transcription factor activities from combined gene and microRNA expression data. *Bioinformatics* 2012;28:1714–1720.
48. Beisser D, Klau GW, Dandekar T, Müller T, Dittrich MT. BioNet: an R-Package for the functional analysis of biological networks. *Bioinformatics* 2010;26:1129–1130.
49. Gonçalves E, Saez-Rodriguez J. Cyrface: an interface from Cytoscape to R that provides a user interface to R packages. *F1000 Res* 2013;2:192.
50. Shannon PT, Grimes M, Kutlu B, Bot JJ, Galas DJ. RCytoscape: tools for exploratory network analysis. *BMC Bioinformatics* 2013;14:217.