

Polimorfismos reguladores y su participación en la patogenia de enfermedades complejas en la era posgenómica

Jesús Hernández-Romano, PhD,⁽¹⁾ Jesús Martínez-Barnetche, MD, PhD,⁽²⁾
Verónica Valverde-Garduño, PhD.⁽²⁾

Hernández-Romano J, Martínez-Barnetche J,
Valverde-Garduño V.

Polimorfismos reguladores y su participación en la patogenia de enfermedades complejas en la era posgenómica. *Salud Publica Mex* 2009;51 suppl 3:S455-S462.

Resumen

El estudio de la participación de la variación genética en la predisposición a las enfermedades complejas ha cobrado nuevas dimensiones en la era genómica. Los polimorfismos de un solo nucleótido (SNP) son el tipo de variación más común entre individuos y su vinculación con enfermedades es motivo de investigación intensa. En fecha reciente, el estudio de los SNP que afectan la expresión genética (rSNP) ha suscitado mayor interés, ya que las diferencias de la expresión genética entre un sujeto y otro pueden modificar el fenotipo. El descubrimiento y caracterización funcional de los rSNP y el estudio de su frecuencia alélica representan un nuevo campo en la búsqueda de determinantes genéticos de enfermedades multifactoriales.

Palabras clave: regulación de la expresión génica; polimorfismo genético; transcripción; variación genética; alelormorfos; enfermedades genéticas congénitas

Hernández-Romano J, Martínez-Barnetche J,
Valverde-Garduño V.

Polymorphisms in gene regulatory regions and their role in the physiopathology of complex disease in the post-genomic era. *Salud Publica Mex* 2009;51 suppl 3:S455-S462.

Abstract

The genomic era is imparting a new impulse to the study of the role of genetic variation in susceptibility to disease. The most common type of genetic variation between individuals is single nucleotide polymorphisms (SNP). The association of SNPs with susceptibility to disease is the current focus of intense research. Recently, the study of SNPs that alter the regulatory mechanisms of gene expression (rSNP) has emerged as a promising field for understanding disease, since this type of variation can have a profound effect on human traits related to susceptibility to disease. The finding and functional characterization of biologically significant rSNPs is advancing our knowledge of genetic determinants for multifactorial disease.

Key words: gene expression regulation; genetic polymorphism; transcription; genetic variation; alleles; genetics, disease

(1) Área de Concentración en Enfermedades Infecciosas. Universidad Politécnica del Estado de Morelos. Cuernavaca, Morelos, México.

(2) Centro de Investigación sobre Enfermedades Infecciosas. Instituto Nacional de Salud Pública. Cuernavaca, Morelos, México.

Fecha de recibido: 28 de julio de 2008 • Fecha de aceptado: 3 de junio de 2009

Solicitud de sobretiros: Dra. Verónica Valverde Garduño. Centro de Investigación sobre Enfermedades Infecciosas. Instituto Nacional de Salud Pública. Av. Universidad 655, col. Santa María Ahuacatlán. 62100 Cuernavaca, Morelos, México.
Correo electrónico: vvalverde@insp.mx

La mayoría de las enfermedades comunes tiene un origen multifactorial, es decir, surgen como resultado de la interacción de múltiples variantes genéticas y diversos factores ambientales, razón por la cual no siguen patrones hereditarios mendelianos y se les denomina enfermedades "complejas".¹ Aun en el caso de trastornos infecciosos, se ha documentado la contribución genética del individuo en la evolución natural del padecimiento.²

Una de las promesas de la ciencia genómica es la posibilidad de descifrar las interacciones múltiples entre variantes genéticas relacionadas con un mayor riesgo de desarrollar afecciones complejas. La atención se ha concentrado en los polimorfismos de un solo nucleótido (SNP), cuya frecuencia aproximada es de 1 en 1 000 pb y representan el tipo de variación más abundante en las poblaciones humanas.³ Un SNP en regiones codificantes puede representar un cambio en la secuencia de una proteína (SNP no sinónimo), y por lo tanto alterar su función, o bien puede no cambiar su secuencia (SNP sinónimo). Casi todas las enfermedades de origen monogénico (mendeliano) se deben a mutaciones que modifican la secuencia y la función de una proteína.¹

La proporción del genoma que codifica a proteínas representa sólo 1.5%; empero, se ha calculado que alrededor de 5% del genoma es funcionalmente importante dado que su secuencia está conservada debido a una fuerte selección.⁴ Esto sugiere que una fracción considerable del genoma (3.5 %) está integrada por elementos funcionales no codificantes, muchos de los cuales pueden intervenir en la regulación de la expresión genética.^{4,5}

La variación fenotípica como resultado de la variación de la regulación de la expresión genética se propuso desde mediados de la década de 1970.⁶ El estudio de la expresión genética a escala genómica ha sido un avance crucial para establecer que la variación de la expresión genética entre una persona y otra es un fenómeno común y que se vincula con un fenotipo.⁷⁻⁹ El propósito de esta revisión es resumir los avances recientes de las medidas ideadas para la identificación de SNP en regiones reguladoras (rSNP), su validación funcional y el estudio de su profundo efecto fisiopatológico consecutivo a la sobreexpresión, subexpresión o expresión aberrante de un gen. Asimismo, se analiza el hecho de que la identificación de SNP reguladores (rSNP) abre un campo promisorio a la búsqueda de determinantes genéticos de afecciones de origen multifactorial. Como preámbulo, se presenta una breve introducción a los conceptos actuales sobre la regulación de la expresión genética. Los aspectos generales de la búsqueda de

determinantes genéticos en enfermedades complejas se han descrito en otras investigaciones.^{1,10-12}

Mecanismos de regulación de la expresión genética

La regulación de la expresión genética se refiere a los mecanismos celulares que controlan el perfil espaciotemporal del producto funcional de un gen. La regulación puede actuar a nivel de la transcripción (mecanismos de regulación transcripcional) o de manera postranscripcional; aquí sólo se alude a los mecanismos transcripcionales, que tienen relevancia directa para esta revisión.

La intensidad de la transcripción de un gen que codifica a una proteína depende, al menos, de la unión de factores de transcripción activados a regiones reguladoras en la molécula de ADN y del reclutamiento del complejo activo de la polimerasa II de ARN, que en conjunto determinan la frecuencia de síntesis del ARNm correspondiente.^{13,14} Los factores de transcripción son proteínas que al unirse a sitios específicos del ADN, los denominados elementos de respuesta (ER), pueden interactuar con otros factores transcripcionales o con cofactores para formar un complejo de proteínas y DNA que permite la iniciación de la transcripción del gen.

Los ER son secuencias de ADN de aproximadamente cuatro a 15 pares de bases (pb) de longitud cuya secuencia casi siempre presenta algunas variaciones. Diferentes familias de factores transcripcionales reconocen distintos ER. Los cambios sutiles en la secuencia de un ER pueden inducir cambios en la afinidad con la que se une el factor transcripcional correspondiente y modificar así los niveles de transcripción (figura 1).

En los vertebrados, los ER pueden encontrarse en tres tipos de regiones reguladoras conocidas. Los que componen al promotor suelen localizarse dentro de las primeras 1 000 pb en la región 5' del inicio de la transcripción; otros pueden situarse a grandes distancias de dicho sitio (hasta 100 kb 5' o 3' del gen) en estructuras conocidas como *enhancers* (o potenciadores), que pueden incrementar la frecuencia de inicio desde el promotor con el que se relacionan, así como en los *silencers* (o silenciadores), que pueden reducir o suprimir la transcripción.¹⁵ Las mutaciones en los ER de cualquiera de estos tres tipos de regiones reguladoras en el genoma (rSNP) pueden modificar la transcripción e influir en el fenotipo del individuo.

El ADN genómico se encuentra unido con proteínas y forma un complejo nucleoproteínico llamado cromatina. Las regiones del genoma que están organizadas de manera compacta (heterocromatina) están apagadas

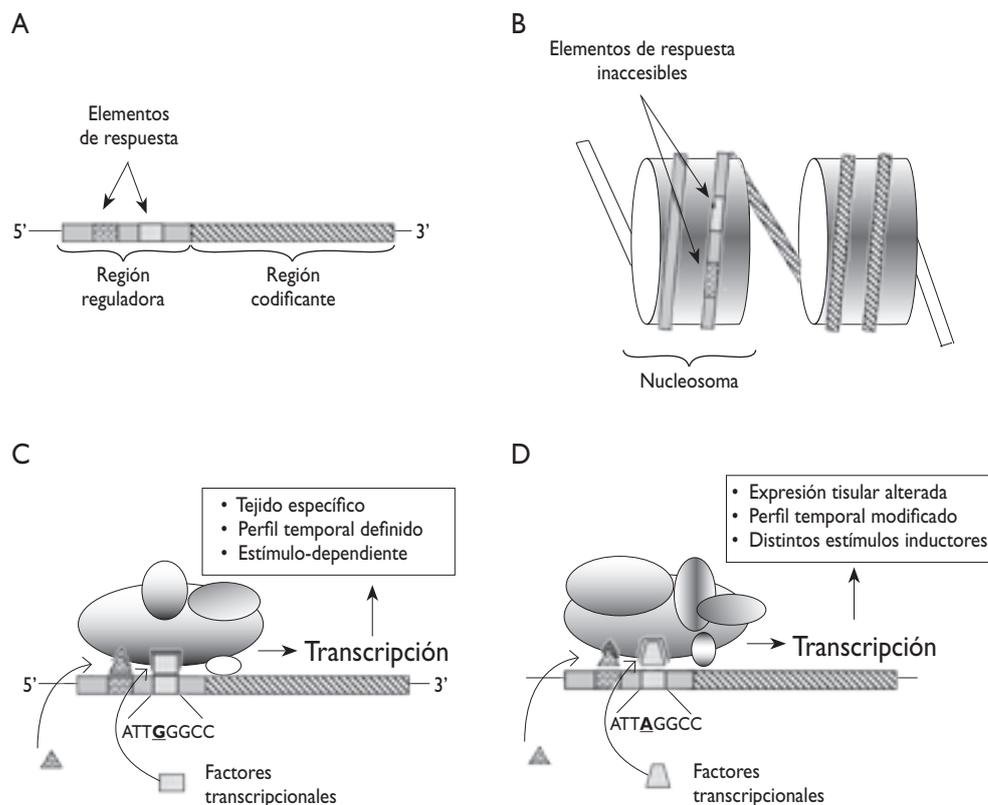


FIGURA 1. IMPORTANCIA DE LOS POLIMORFISMOS DE UN SOLO NUCLEÓTIDO EN REGIONES REGULADORAS (rSNPs). **A. ESTRUCTURA BÁSICA DE UN GEN A NIVEL DE SECUENCIA DE ADN;** UN GEN POSEE UNA REGIÓN REGULADORA QUE CONTIENE ELEMENTOS DE RESPUESTA RECONOCIDOS POR FACTORES DE TRANSCRIPCIÓN Y UNA REGIÓN CODIFICANTE QUE TIENE LA INFORMACIÓN NECESARIA PARA SINTETIZAR UNA PROTEÍNA. PARA QUE UN GEN SE EXPRESA EN UNA CÉLULA DETERMINADA, ÉSTA DEBE POSEER LOS FACTORES DE TRANSCRIPCIÓN ADECUADOS; ADEMÁS, LOS ELEMENTOS DE RESPUESTA DEBEN ESTAR EXPUESTOS PARA QUE PUEDAN RECONOCERLOS LOS FACTORES TRANSCRIPCIONALES.¹⁶ **B. LA ESTRUCTURA DE LA CROMATINA PUEDE IMPEDIR EL ACCESO DE FACTORES DE TRANSCRIPCIÓN A SUS ELEMENTOS DE RESPUESTA.** **C. EL DESPLAZAMIENTO DE LOS NUCLEOSOMAS EXPONE LOS ELEMENTOS DE RESPUESTA Y PERMITE LA AGRUPACIÓN DE LA MAQUINARIA TRANSCRIPCIONAL Y LA EXPRESIÓN DEL GEN.**¹⁶ **D. LAS MUTACIONES EN LOS ELEMENTOS DE RESPUESTA PUEDEN MODIFICAR LA EXPRESIÓN DEL GEN Y AFECTAR EL FENOTIPO DEL ORGANISMO**¹⁷

desde el punto de vista transcripcional. En contraste, las regiones transcripcionalmente activas del genoma presentan una organización más relajada (eucromatina).

La organización del ADN genómico en cromatina es fundamental en la regulación de la expresión genética. La modulación diferencial de la compactación del ADN genera diferentes fenotipos heredables, sin cambio alguno en la secuencia del gen blanco de esta modulación, lo cual se conoce como epigenética. Ejemplos comunes de la regulación epigenética y expresión diferencial son la inactivación de un cromosoma X en la mujer (cuerpo

de Barr) o la supresión alélica de la expresión de algunos genes por impronta (*imprinting*).

Los factores que afectan la transcripción también pueden catalogarse como factores en *cis*, como aquellos que ejercen su efecto en la misma cadena de ADN y cerca del gen en cuestión. Los elementos de respuesta (y sus variantes polimórficas) afectan la transcripción en *cis*. Por otro lado, los efectos en *trans* surgen de un gen o factor adicional. Los efectos en *trans* pueden originarse por factores no genéticos, como estímulos ambientales (drogas, patógenos, enfermedades) o bien variaciones

genéticas en otros genes (p. ej., factores transcripcionales) que afectan la expresión del gen en cuestión.^{18,19}

Análisis de los determinantes genéticos de la variación de la expresión genética

La identificación de variantes genéticas relacionadas con la variación de la expresión genética supone retos muy particulares inherentes a la complejidad de los mecanismos de regulación transcripcional. La identificación de una diferencia en la expresión genética entre individuos es un reto complejo, pero asequible. Incluso al definir que el efecto opera en *cis*, se debe intentar esclarecer la posible influencia de factores epigenéticos. A continuación se describen los enfoques desarrollados en fecha reciente para poder inferir, detectar y validar la presencia de los rSNP que participan en la fisiopatología de enfermedades complejas específicas.

Cuantificación de ARNm específico de alelo

En este caso se analiza el nivel de transcripción de cada alelo. La cuantificación de ARNm específico de alelo requiere el conocimiento previo de la presencia de al menos un SNP en el transcrito que permita identificar el ARNm producido a partir de cada cromosoma. El ARNm purificado del tejido se utiliza como molde para una amplificación de ADN con oligonucleótidos específicos para cada marcador. Un cociente de expresión entre ambos alelos diferente a 1 sugiere la presencia de un rSNP en desequilibrio de ligamiento con el SNP del transcrito, pero no excluye la posibilidad de que las diferencias sean de origen epigenético.²⁰ En estudios de páncreas dirigidos a analizar la susceptibilidad genética a la diabetes tipo I se ha encontrado que los alelos clase I de la región de las secuencias repetitivas de número variable en tándem (VNTR) de la insulina/IGF-2 se asocian con incrementos del nivel de expresión cuando se comparan con los alelos de la clase III.²¹ El método de cuantificación de ARNm específico de alelo se ha aplicado en la evaluación de 129 genes, de los cuales 18% mostró diferencias alélicas de expresión en líneas linfoblastoides humanas.²² Una desventaja importante en este caso es que el SNP marcador debe estar presente cuando menos en el transcrito primario y es posible que una gran parte de los rSNP no se acompañe de un marcador de este tipo.

Enfoques bioinformáticos para la identificación de rSNP

La explosión del número de bases de datos públicas del genoma humano, así como de sus variaciones, representa una fuente abundante de información disponible

para la búsqueda *in silico* de candidatos a determinantes genéticos de enfermedades. Sin embargo, para la eventual confirmación de relación causal es inevitable la validación funcional de las variantes candidatas.

Las principales bases de datos que catalogan SNP en general son la HGMD, dbSNP, HGBASE, ALFRED y OMIM.²³⁻²⁷ Para rSNP existe una base de datos especializada denominada *rSNP_Guide*, la cual contiene versiones curadas de rSNP validados de manera experimental, así como mutaciones en elementos reguladores generados también de forma experimental.²⁸

Una forma de identificar *in silico* rSNP requiere en principio el reconocimiento de un elemento de respuesta con alta probabilidad de ser funcionalmente relevante y la subsiguiente búsqueda de variaciones que puedan repercutir en la transcripción del gen en cuestión. La identificación de secuencias reguladoras *cis* por métodos informáticos ha progresado en grado considerable. Sin embargo, debido a que la secuencia de los elementos de respuesta son altamente frecuentes, dado que tienden a ser cortas y altamente degeneradas y su funcionalidad depende del contexto, dichos algoritmos se caracterizan por una especificidad limitada. Es decir, muchas de las secuencias o motivos potenciales encontrados son biológicamente irrelevantes²⁹ y se debe hacer uso de información adicional para su validación.

Mediante esta forma de búsqueda de los rSNP incluidos en la red reguladora del factor de transcripción p53 se han identificado a nivel genómico más de 600 ER a p53 putativos polimórficos. Con criterios cada vez más estrictos, y con base en el conocimiento de los mecanismos de transactivación de p53, 40 de ellos se seleccionaron como ER altamente probables y ocho se evaluaron funcionalmente mediante transactivación de genes reporteros en *S. cerevisiae* (levadura) y en líneas celulares humanas; esto confirmó que las variantes resultan en una capacidad de transactivación diferencial.³⁰ Con un enfoque similar se identificaron variantes polimórficas dentro de límites de 5 kb en la región 5' al inicio de la transcripción de 176 genes de receptores acoplados a proteínas G (GPCR) conservados entre el ser humano y el ratón y mediante la búsqueda indiscriminada de ER en la secuencia de 110 pb alrededor del SNP mediante matrices de peso disponibles en bases de datos públicas. De un total de 800 SNP encontrados, sólo 200 correspondían a segmentos conservados entre ser humano y ratón y 36 se situaban dentro de un ER putativo. Siete de estos 36 SNP se sometieron a validación funcional mediante EMSA y transactivación de reporteros en líneas celulares humanas; se encontró que cuatro de ellos afectan la transcripción en grado notable.³¹

En virtud del gran número de proyectos de transcriptoma y la abundancia de secuencias cortas

expresadas (*expressed sequence tags*, EST), ha sido posible inferir rSNP por medio de un ensayo de cuantificación de ARNm específico de alelo *in silico*.³² Al igual que *in vivo*, se requieren secuencias de EST provenientes de una genoteca de un solo individuo (usualmente accesibles al público en internet) para las cuales se seleccionan grupos de EST polimórficas y se cuantifica y evalúa estadísticamente la sobrerepresentación de cada variante, o bien se calcula si la representación de alguna variante difiere en términos estadísticos de la frecuencia alélica esperada para ese gen en particular, según el HapMap.³³ Al igual que las aproximaciones *in vitro*, este método no permite discernir si la variación en la cantidad del ARNm de cada alelo se debe a un rSNP o a factores epigenéticos.

Análisis genético de la variación de la expresión genética a escala genómica

Además de representar un avance tecnológico clave para determinar que la variación de la expresión genética entre individuos es un fenómeno común, la tecnología de microarreglos de oligonucleótidos se ha adaptado para la identificación de un vasto número de SNP en el mismo ensayo. De esta forma, la determinación de diferencias alélicas de expresión (véase la sección anterior) puede realizarse en microarreglos, siempre y cuando los individuos analizados sean heterocigotos para un SNP exónico. De esta forma ha sido posible determinar para cientos de genes que al menos la mitad presenta diferencias alélicas de expresión, lo cual sugieren la presencia de un rSNP en desequilibrio de ligamiento con el SNP exónico.^{34,35} Una limitación de estos estudios es que no permiten dilucidar si la causa de las diferencias alélicas de expresión es atribuible a factores epigenéticos; y, en caso contrario, no indica cual es el rSNP candidato.

El nivel de expresión de genes puede considerarse en sí mismo un fenotipo cuantificable que en muchos casos es heredable (*expression Quantitative Trait Loci*, eQTL) y, por lo tanto, es susceptible de análisis por estudio genético de segregación o relación en organismos como levadura, ratón, rata y seres humanos.^{7-9,36} Con la finalidad de identificar candidatos a determinantes genotípicos de perfiles de expresión (fenotipos de expresión), Morley y colaboradores³⁷ analizaron por microarreglos los perfiles de expresión de 8 500 genes de líneas linfoblastoides derivadas de 195 individuos pertenecientes a 14 familias. En forma paralela, tipificaron 2 756 SNP autosómicos distribuidos a lo largo del genoma y se realizó un análisis de segregación, de tal modo que se determinó que 984 fenotipos de expresión que mostraron mayor varianza entre individuos no relacionados estaban ligados claramente a una región cromosómica y los consideraron como locus regulado-

res en *cis*, si se encontraban a menos de 5 mb del gen en cuestión, o en *trans* si se hallaban a más de 5 mb o en otro cromosoma. Sin embargo, mediante un sistema similar, pero con diferente plataforma de microarreglos y a partir de un método diferente de análisis estadístico de la segregación, Monks y colegas³⁸ encontraron resultados discordantes respecto de los de Morley y colaboradores.³⁷

Se ha sugerido que los QTL comunes con un impacto moderado en el fenotipo pueden detectarse con mayor facilidad mediante estudios de asociación.³⁹ Como complemento del estudio de Morley,³⁷ Cheung y colegas⁴⁰ seleccionaron 27 fenotipos de expresión con mayor evidencia de ligamiento en *cis* para realizar un estudio de asociación en 57 individuos no relacionados con más de 770 000 SNP obtenidos del proyecto HapMap³ y consiguieron posicionar rSNP putativos para 14 de 27 fenotipos en una ventana menor de 100 kb del gen en cuestión y mostraron una validación funcional de un rSNP.

Validación por genes reporteros

Una vez que los estudios de asociación u otros métodos reconocen la presencia de uno o más rSNP en regiones reguladoras vinculadas con susceptibilidad a enfermedades comunes, es necesario validar su función reguladora. En este caso se construye una molécula de ADN con la región codificadora de un gen reportero fácil de detectar (luciferasa, proteína fluorescente verde, CAT, etc.) bajo el control transcripcional del promotor que contiene el polimorfismo; como control se utiliza la secuencia del promotor con la variante de individuos no susceptibles. Cada una de estas construcciones se introduce en células relevantes en cultivo y se analiza el nivel de expresión del reportero en cada una de las variantes alélicas del promotor. El cambio de la expresión del gen reportero a partir de las moléculas que contienen el o los polimorfismos sugiere que efectivamente su mecanismo de acción afecta de forma directa la regulación de la expresión del gen. El estudio del gen que codifica al factor de coagulación VII mediante ensayos de expresión de genes reporteros en una línea celular de hepatocitos ha revelado que un SNP (una transversión del nucleótido T a G) en la posición - 61 de la región reguladora reduce la expresión de un gen reportero a 6.7% en comparación con el alelo silvestre.⁴¹ Este método también se ha aplicado en el plano genómico: después de detectar la presencia de SNP en los promotores putativos de genes relevantes para enfermedades mentales, se clonaron 247 promotores y se analizó la expresión de gen reportero en dos líneas celulares relevantes: 20% de rSNP presentó alteración mayor de 50% en el nivel de transcripción.⁴²

Haplochip

Uno de los avances más recientes en el estudio funcional de rSNP es el análisis por haplochip.⁴³ Éste consiste en cuantificar la carga específica de polimerasa II de ARN fosforilada para cada alelo, que es una medida relativa de la actividad transcripcional. Entre las ventajas de este método figura el hecho de que pueden usarse como marcadores SNP en cualquier región del ADN sin restricción hasta una distancia de 2 kb del sitio de iniciación de la transcripción, incluidos los propios rSNP causantes de la actividad transcripcional diferencial. El análisis funcional por haplochip del SNP -308 del gen para el factor de necrosis tumoral (TNF) (que se ha relacionado con susceptibilidad a la malaria, leishmaniasis, lepra, asma y otras enfermedades) reveló que los haplotipos en el locus extendido son complejos y que este rSNP modifica los niveles de unión de la polimerasa II de ARN al gen de la linfotóxina alfa (LTA) que se localiza en la región 5' a TNF en este locus.⁴³ El establecimiento de un ensayo de discriminación alélica por amplificación de ADN en tiempo real pone el haplochip al alcance de muchos investigadores.⁴⁴ Mediante este haplochip modificado lograron identificar un rSNP (alelo -63C) en el gen para la S-transferasa de glutatión M3 (GSTM3). Esta enzima cataliza la conjugación de compuestos tóxicos y carcinogénicos al glutatión para su procesamiento. En el alelo -63C se detecta nueve veces menos capacidad de unión a la polimerasa II de ARN fosforilada con respecto al alelo -63A; se ha conjeturado que la disminución consecuente de GSTM3 podría incrementar el daño al ADN celular.⁴⁴

Perspectivas y conclusiones

Las ciencias genómicas y el desarrollo tecnológico subyacente han hecho posible el estudio sistemático del efecto de la variación de la expresión genética en la variación fenotípica entre individuos; se ha acuñado el término "genética genómica".⁴⁵ Debido a la complejidad de los mecanismos de la regulación transcripcional, los enfoques actuales empleados para identificar eQTL a escala genómica y validar su relación con rSNP no han sido concluyentes en todos los casos. Esto se debe a que casi todos ellos dependen del uso de líneas linfoblastoides que no representan la diversidad de la expresión genética en diferentes tejidos, utilizan diferentes plataformas de microarreglos y diversos métodos estadísticos para analizar los datos, pero sobre todo se debe a que emplean un número reducido de individuos en virtud de los costos elevados, lo cual limita el poder estadístico del análisis.^{17,46,47}

Pese a las limitaciones actuales, este tipo de estudios ha recibido gran apoyo, ya que diversas agencias financian grandes consorcios multidisciplinarios e internacionales para la realización de este tipo de análisis. De esta manera se reduce el costo para mejorar el uso del nuevo conocimiento del genoma humano y con ello revelar las bases fisiopatológicas de las enfermedades complejas. Asimismo, se ha impulsado la inclusión de un número mayor de individuos en los estudios de asociación y se tiene claro que deben repetirse en diversas poblaciones para establecer la validez de las asociaciones y poder continuar con la elucidación de los mecanismos fisiopatológicos subyacentes.^{48,49} Las herramientas generadas hasta ahora por los estudios en salud pública son indispensables; no existe sustituto para la realización de análisis estadísticos rigurosos en los estudios de asociación y su validación en diversas poblaciones.

De manera adicional, en la búsqueda de rSNP es indispensable el estudio detallado del haplotipo,^{50,51} debido a la abundancia de polimorfismos en regiones no codificantes y a que los ER se organizan de modo modular y funcionan en forma dependiente del contexto. La evidencia obtenida hasta la fecha tanto en modelos experimentales como en humanos^{52,53} (cuadro I) sugiere que la variación de la expresión de múltiples genes funcionalmente relacionados puede determinar efectos fenotípicos profundos, algunos de los cuales pueden ser determinantes en la fisiopatología de enfermedades complejas, como la esquizofrenia, diabetes mellitus y trastornos autoinmunitarios. Es indispensable la estandarización de los protocolos, plataformas y fuentes de tejido humano, así como la depuración de métodos estadísticos para analizar la cantidad masiva de información de expresión y variación genética que se genera a nivel mundial. Como lo sugieren los estudios listados en el cuadro I, existe un gran potencial para la comprensión de la fisiopatología de enfermedades complejas si se concede más atención en la búsqueda y análisis de polimorfismos reguladores, especialmente debido a que los rSNP constituyen probablemente un alto porcentaje de la variación en el genoma humano.

Referencias

1. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet* 2003;33(Suppl):228-237.
2. Frodsham AJ, Hill AV. Genetics of infectious diseases. *Hum Mol Genet* 2004;13:R187-194.
3. International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;437:1299-1320.

Cuadro I
EJEMPLOS DE rSNP RELACIONADOS CON ENFERMEDADES HUMANAS

Gen	Posición	Elemento de respuesta afectado	Fenotipo	Enfermedad relacionada	Población estudiada	Referencia
Antígeno Duffy (DRAC)	-46C	GATA-1	Pérdida de la expresión de DRAC en eritrocitos	Resistencia a la malaria por <i>P. vivax</i>	Individuos de raza negra, negativos al antígeno Duffy	54
Globina α	149 709*	GATA-1	Ganancia de función de un sitio GATA-1 y pérdida de la expresión de la globina α	Talasemia alfa	Melanesios afectados por talasemia alfa	55
LTA (linfotóxina α)	+80A	ABF-1	Reducción de la producción de LT α	Infarto agudo del miocardio, asma, sepsis, cáncer, enfermedades autoinmunitarias, malaria cerebral	Europeos con diferentes haplotipos en el locus de LTA	56
IFNG (interferón γ)	-179T	Receptor de estrógeno	Producción de IFN γ en linfocitos T de lámina propia insensible a la supresión por estrógeno	Rápida progresión a sida	Mujeres afectadas por carcinoma de colon, colitis ulcerativa o enfermedad de Crohn	57
GSTM3 (S-transferasa de glutatión)	-63C	ND	Baja producción de GST M3	Sensibilidad a agentes xenobióticos	Líneas celulares de gliomas humanos	44
CCR5 (receptor de quimiocina 5)	208G/T	NF κ B	Sobreproducción de CCR5	Progresión acelerada a sida	Linfocitos T humanos	58
IL-8	Varios ‡	C/EBP	Sobreproducción de IL-8	Patogenia de VSR	Europeos con diferentes haplotipos para el locus de la IL-8	59

* Coordenada en BAC donde se identificó
 \ddagger Dependiente de un haplotipo complejo

4. Cooper GM, Sidow A. Genomic regulatory regions: insights from comparative sequence analysis. *Curr Opin Genet Dev* 2003;13:604-610.

5. Drake JA, Bird C, Nemes J, Thomas DJ, Newton-Cheh C, Reymond A, et al. Conserved noncoding sequences are selectively constrained and not mutation cold spots. *Nat Genet* 2006;38:223-227.

6. King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. *Science* 1975;188:107-116.

7. Sandberg R, Yasuda R, Pankratz DG, Carter TA, Del Rio JA, Wodicka L, et al. Regional and strain-specific gene expression mapping in the adult mouse brain. *Proc Natl Acad Sci U S A* 2000;97:11038-11043.

8. Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G. The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat Genet* 2001;29:389-395.

9. Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ, McCusker JH, et al. Dissecting the architecture of a quantitative trait locus in yeast. *Nature* 2002;416:326-330.

10. Glazier AM, Nadeau JH, Aitman TJ. Finding genes that underlie complex traits. *Science* 2002;298:2345-2349.

11. Carlson CS, Eberle MA, Kruglyak L, Nickerson DA. Mapping complex disease loci in whole-genome association studies. *Nature* 2004;429:446-452.

12. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005;6:95-108.

13. Woychik NA, Hampsey M. The RNA polymerase II machinery: structure illuminates function. *Cell* 2002;108:453-463.

14. Kadonaga JT. Regulation of DNA polymerase II transcription by sequence-specific DNA binding factors. *Cell* 2004;116:247-257.

15. Levine M, Tjian R. Transcription regulation and animal diversity. *Nature* 2003;424:147-151.

16. Khorasanizadeh S. The nucleosome: from genomic organization to genomic regulation. *Cell* 2004;116:259-272.

17. Leung TH, Hoffmann A, Baltimore D. One nucleotide in a kB site can determine cofactor specificity for NF- κ B dimmers. *Cell* 2004;118:453-464.

18. Buckland PR. The importance and identification of regulatory polymorphisms and their mechanisms of action. *Biochim Biophys Acta* 2006;1762:17-28.

19. Pastinen T, Ge B, Hudson TJ. Influence of human genome polymorphism on gene expression. *Hum Mol Genet* 2006;15:R9-16.

20. Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW. Allelic variation in human gene expression. *Science* 2002;297:1143.

21. Bennett ST, Lucassen AM, Gough SC, Powell EE, Undlien DE, Pritchard LE, et al. Susceptibility to human type 1 diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nat Genet* 1995;9:284-292.

22. Pastinen T, Sladek R, Gurd S, Sammak A, Ge B, Lepage P, et al. A survey of genetic and epigenetic variation affecting human gene expression. *Physiol Genomics* 2003;16:184-193.

23. Krawczak M, Ball EV, Fenton I, Stenson PD, Abeyasinghe S, Thomas N, et al. Human gene mutation database—a biomedical information and research resource. *Hum Mutat* 2000;15:45-51.

24. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29:308-311.

25. Brookes AJ, Lehtvaslaihio H, Siegfried M, Boehm JG, Yuan YP, Sarkar CM, et al. HGBASE: a database of SNPs and other variations in and around human genes. *Nucleic Acids Res* 2000;28:356-360.

26. Cheung KH, Osier MV, Kidd JR, Pakstis AJ, Miller PL, Kidd KK. ALFRED: an allele frequency database for diverse populations and DNA polymorphisms. *Nuc Acid Res* 2000;28:361-363.
27. McKusick VA. Mendelian inheritance in man. A catalog of human genes and genetic disorders. 12th ed. Baltimore: Johns Hopkins University Press, 1998.
28. Ponomarenko JV, Merkulova TI, Vasiliev GV, Levashova ZB, Orlova GV, Lavryushev SV, et al. rSNP_Guide, a database system for analysis of transcription factor binding to target sequences: application to SNPs and site-directed mutations. *Nucleic Acids Res* 2001;29:312-316.
29. Wasserman WW, Sandelin A. Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet* 2004; 5:276-287.
30. Tomso DJ, Inga A, Menendez D, Pittman GS, Campbell MR, Storici F, et al. Functionally distinct polymorphic sequences in the human genome that are targets for p53 transactivation. *Proc Natl Acad Sci U S A* 2005;102:6431-6436.
31. Mottagui-Tabar S, Faghihi MA, Mizuno Y, Engstrom PG, Lenhard B, Wasserman WW, et al. Identification of functional SNPs in the 5-prime flanking sequences of human genes. *BMC Genomics* 2005;6:18.
32. Lin W, Yang HH, Lee MP. Allelic variation in gene expression identified through computational analysis of the dbEST database. *Genomics* 2005;86:518-527.
33. Ge B, Gurd S, Gaudin T, Dore C, Lepage P, Harmsen E, et al. Survey of allelic expression using EST mining. *Genome Res* 2005;15:1584-1591.
34. Lo HS, Wang Z, Hu Y, Yang HH, Gere S, Buetow KH, et al. Allelic variation in gene expression is common in the human genome. *Genome Res* 2003;13:1855-1862.
35. Pant PV, Tao H, Beilharz EJ, Ballinger DG, Cox DR, Frazer KA. Analysis of allelic differential expression in human white blood cells. *Genome Res* 2006;16:331-339.
36. Cheung VG, Conlin LK, Weber TM, Arcaro M, Jen KY, Morley M, et al. Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat Genet* 2003;33:422-425.
37. Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, et al. Genetic analysis of genome-wide variation in human gene expression. *Nature* 2004;430:743-747.
38. Monks SA, Leonardson A, Zhu H, Cundiff P, Pietrusiak P, Edwards S, et al. Genetic inheritance of gene expression in human cell lines. *Am J Hum Genet* 2004;75:1094-1105.
39. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996;273:1516-1517.
40. Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT. Mapping determinants of human gene expression by regional and genome-wide association. *Nature* 2005;437:1365-1369.
41. Arbini AA, Pollak ES, Bayleran JK, High KA, Bauer KA. Severe factor VII deficiency due to a mutation disrupting a hepatocyte nuclear factor 4 binding site in the factor VII promoter. *Blood* 1997;89:176-182.
42. Buckland PR, Hoogendoorn B, Coleman SL, Guy CA, Smith SK, O'Donovan MC. Strong bias in the location of functional promoter polymorphisms. *Hum Mutat* 2005;26:214-223.
43. Knight JC, Keating BJ, Rockett KA, Kwiatkowski DP. In vivo characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading. *Nat Genet* 2003;33:469-475.
44. Liu X, Campbell MR, Pittman GS, Faulkner EC, Watson MA, Bell DA. Expression-based discovery of variation in the human glutathione S-transferase M3 promoter and functional analysis in a glioma cell line using allele specific chromatin immunoprecipitation. *Cancer Res* 2005;65:99-104.
45. Jansen RC, Nap JP. Genetical genomics: the added value from segregation. *Trends Genet* 2001;17:388-391.
46. de Koning DJ, Haley CS. Genetical genomics in humans and model organisms. *Trends Genet* 2005;21:377-381.
47. Gibson G, Weir B. The quantitative genetics of transcription. *Trends Genet* 2005;21:616-623.
48. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, et al. Genome Wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Gen* 2008;9:356-369.
49. Ioannidis JPA, Thomas G, Daly MJ. Validating, augmenting and refining genome-wide association signals. *Nat Rev Gen* 2009;10:318-329.
50. Knight JC. Regulatory polymorphisms underlying complex disease traits. *J Mol Med* 2005; 83:97-109.
51. Pastinen T, Ge B, Gurd S, Gaudin T, Dore C, Lemire M, et al. Mapping common regulatory variants to human haplotypes. *Hum Mol Genet* 2005;14:3963-3971.
52. Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, et al. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet* 2005;37:710-717.
53. Hubner N, Wallace CA, Zimdahl H, Petretto E, Schulz H, Maciver F, et al. Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat Genet* 2005;37:243-253.
54. Tournamille C, Colin Y, Cartron JP, Le Van Kim C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat Genet* 1995;10:224-228.
55. De Gobbi M, Viprakasit V, Hughes JR, Fisher C, Buckle VJ, Ayyub H, et al. A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. *Science* 2006;312:1215-1217.
56. Knight JC, Keating BJ, Kwiatkowski DP. Allele-specific repression of lymphotoxin-alpha by activated B cell factor-1. *Nat Genet* 2004;36:394-399.
57. Gonsky R, Deem RL, Bream JH, Young HA, Targan SR. An IFNG SNP with an estrogen-like response element selectively enhances promoter expression in peripheral but not lamina propria T cells. *Genes Immun* 2006; 1-10.
58. Bream JH, Young HA, Rice N, Martin MP, Smith MW, Carrington M, et al. CCR5 promoter alleles and specific DNA binding factors. *Science* 1999;284:223
59. Hacking D, Knight JC, Rockett K, Brown H, Frampton J, Kwiatkowski DP, et al. Increased in vivo transcription of an IL-8 haplotype associated with respiratory syncytial virus disease-susceptibility. *Genes Immun* 2004;5:274-282.