

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/23709928>

Transcriptome in mexicans. Methodology for analyzing gene expression profile of large scale simultaneous samples of muscle tissue, adipose and lymphocytes obtained from the same in...

Article in *Gaceta medica de Mexico* · January 2008

Source: PubMed

CITATIONS

2

READS

69

13 authors, including:



Raul A Bastarrachea
Texas Biomedical Research Institute

92 PUBLICATIONS 604 CITATIONS

SEE PROFILE



Juan Carlos Lopez-Alvarenga
Hospital General de Mexico

152 PUBLICATIONS 1,589 CITATIONS

SEE PROFILE



Hugo Laviada
Universidad Marista de Mérida

78 PUBLICATIONS 1,311 CITATIONS

SEE PROFILE



Ricardo M Cerda-Flores
Autonomous University of Nuevo León

405 PUBLICATIONS 1,712 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Misceláneos [View project](#)



Poverty, Health and Violence [View project](#)

Transcriptoma en mexicanos: metodología para analizar el perfil de expresión genética de gran escala en muestras simultáneas de tejido muscular, adiposo y linfocitos obtenidas en un mismo individuo

Raúl A. Bastarrachea,^{a,*} Juan Carlos López-Alvarenga,^{a,b,**} Jack W. Kent Jr.,^a Hugo A. Laviada-Molina,^c Ricardo M. Cerda-Flores,^{d,e} Ana Laura Calderón-Garcidueñas,^f Amanda Torres-Salazar,^g Edna J. Nava-González,^h Elizabeth Solis-Pérez,^h Esther C. Gallegos-Cabrales,^e Shelley A. Cole^a y Anthony G. Comuzzie^a

^aDepartment of Genetics, Auxology and Metabolism Working Group, Southwest Foundation for Biomedical Research, San Antonio, Texas, USA; ^bDepartamento de Enseñanza e Investigación, Hospital Regional de Veracruz, Secretaría de Salud, Veracruz, Veracruz, México; ^cDepartamento de Nutrición Humana y Trastornos del Metabolismo, Facultad de Medicina, Universidad Autónoma de Yucatán, Mérida, Yucatán, México; ^dDepartamento de Genética de Poblaciones y Bioinformática, Centro de Investigación Biomédica del Noreste, IMSS, Monterrey, N.L., México; ^eFacultad de Enfermería, Universidad Autónoma de Nuevo León, Monterrey, N.L., México; ^fHospital de Especialidades 25, Centro Médico Nacional Nuevo León, IMSS, Monterrey, N.L., México; ^gUnidad de Investigación Médica en Enfermedades Infecciosas y Parasitarias (UIMEIP), Centro Médico Nacional Siglo XXI, IMSS, México D.F., México; ^hFacultad de Salud Pública y Nutrición, Universidad Autónoma de Nuevo León, Monterrey, N.L., México

Recibido en su versión modificada: 21 de mayo de 2008

Aceptado: 6 de junio de 2008

RESUMEN

Objetivo: Describir la metodología de análisis de múltiples transcritos con técnicas de microarreglo en biopsias simultáneas de tejido muscular, adiposo y sangre en un mismo individuo, como parte de la estandarización del estudio GEMM (Genética de las Enfermedades Metabólicas en México).

Material y métodos: Se incluyó a cuatro sujetos con índice de masa corporal (IMC) entre 20 y 41. Se registró estatura, talla y composición corporal. Se realizó biopsia muscular (vasto lateral), de tejido adiposo subcutáneo y muestra de sangre completa. El ARN total fue extraído de los tejidos y amplificado para análisis de microarreglos.

Resultados: De 48 687 potenciales transcritos, 39.4% fue detectable en al menos uno de los tejidos. La expresión de leptina no fue detectable en linfocitos, débilmente expresada en músculo, alta expresión en el tejido adiposo y correlacionó con el IMC. El GLUT4 también ilustra la especificidad para el músculo sin verse afectado por el IMC. La concordancia en la expresión de transcritos fue 0.70 ($p < 0.001$) para los tres tejidos.

Conclusiones: Fue factible cuantificar simultáneamente la expresión genética de miles de transcritos, hubo concordancia en la expresión entre diferentes tejidos obtenidos en un mismo individuo, y confiabilidad del método al reproducir las relaciones biológicas esperadas. El estudio GEMM podrá analizar las correlaciones de los transcritos expresados dentro de un órgano y luego entre diferentes tejidos, y proveerá endofenotipos cuantitativos novedosos que proporcionarán un amplio panorama de información sobre las enfermedades metabólicas, incluyendo obesidad y diabetes tipo 2.

Palabras clave:

Análisis de microarreglos, ARN mensajero, síndrome metabólico, expresión genética, ARN no traducido

SUMMARY

Objective: We describe the methodology used to analyze multiple transcripts using microarray techniques in simultaneous biopsies of muscle, adipose tissue and lymphocytes obtained from the same individual as part of the standard protocol of the Genetics of Metabolic Diseases in Mexico: GEMM Family Study.

Methods: We recruited 4 healthy male subjects with BMI 20-41, who signed an informed consent letter. Subjects participated in a clinical examination that included anthropometric and body composition measurements, muscle biopsies (vastus lateralis) subcutaneous fat biopsies and a blood draw. All samples provided sufficient amplified RNA for microarray analysis. Total RNA was extracted from the biopsy samples and amplified for analysis.

Results: Of the 48 687 transcript targets queried, 39.4% were detectable in a least one of the studied tissues. Leptin was not detectable in lymphocytes, weakly expressed in muscle, but over-expressed and highly correlated with BMI in subcutaneous fat. Another example was GLUT4, which was detectable only in muscle and not correlated with BMI. Expression level concordance was 0.7 ($p < 0.001$) for the three tissues studied.

Conclusions: We demonstrated the feasibility of carrying out simultaneous analysis of gene expression in multiple tissues, concordance of genetic expression in different tissues, and obtained confidence that this method corroborates the expected biological relationships among LEP and GLUT4. The GEMM study will provide a broad and valuable overview on metabolic diseases, including obesity and type 2 diabetes.

Key words:

Microarray analysis, messenger RNA, metabolic syndrome (syndrome X), genetic expression, untranslated RNA

*Correspondencia y solicitud de sobretiros: Raúl A. Bastarrachea. Department of Genetics, Southwest Foundation for Biomedical Research, 7620 NW Loop 410, San Antonio, Texas, USA. Tel.: 78227 5301. Correo electrónico: raulbs@sfbgenetics.org

**Los dos primeros autores contribuyeron igualmente en la preparación de este artículo.

Introducción

El Proyecto Genoma Humano finalizó hace aproximadamente cuatro años,¹ sin embargo, aún queda mucho por entender sobre las variantes alélicas que pueden proteger, provocar o aumentar la susceptibilidad para desarrollar estados patológicos.

Recientemente se ha profundizado en el conocimiento de la complejidad del genoma humano, y se ha dado importancia no solo a los efectos que genes específicos puedan tener sobre fenotipos de enfermedad, sino también a la frecuencia e intensidad con que cada gen es traducido a diferentes proteínas. El producto inmediato de la traducción de un gen es la formación de ácido ribonucleico mensajero (ARNm) y se conoce como transcrito. El paradigma de "un gen, una proteína" se rompió al descubrirse la edición alternativa del ARNm, situación que aclaró el mecanismo de cómo un mismo gen puede traducirse en diferentes proteínas. Estos cortes alternativos del ARNm tienen una amplia variedad funcional que está siendo revelada por novedosos estudios genómicos.²

El conocimiento sobre el genoma se hizo más complejo al observarse que existe ARN que no codifica proteínas (ARNnc) y que desempeña un papel importante en la regulación del momento propicio en que un gen debe activarse y en la cantidad de proteínas que debe producir.³ Parte de estos ARNnc incluyen los microARN, una familia de pequeñas cadenas de nucleótidos fisiológicamente funcionales que se encuentran integrados dentro de las regiones intrónicas de estos transcritos primarios de ARN, cuyas secuencias pueden realizar una amplia variedad de modificaciones que afectan la expresión de un gen.⁴ Al parecer, solamente existen alrededor de 23 mil genes con un tamaño promedio de 27 kb en el genoma humano. Un paso muy importante ha sido reconocer que la mayoría de los 2.9 Gb del genoma se transcriben.

La actividad de un gen puede ser medida por la cantidad de transcritos que puede producir, aunque estos transcritos no codifiquen proteínas. El conjunto de ARNs de todos los genes es lo que se conoce como transcriptoma. La semejanza entre los genomas de especies de mamíferos diverge en forma muy importante en cuanto a análisis de transcritos.⁵

El estudio Genética de las Enfermedades Metabólicas en México (GEMM)⁶ busca reclutar a 1500 individuos de diversas familias, distribuidos en la República Mexicana. Dicha población es principalmente el resultado de una mezcla de ancestros europeos (hispánicos), amerindios y, en menor grado, africanos.^{7,8} GEMM ha sido diseñado para explotar los avances en la cuantificación del ARNm a través de métodos de alto rendimiento (*high-throughput*), que permiten comparar perfiles de expresión individual o por conglomerados del transcriptoma completo. La adquisición y el análisis de estos fenotipos de expresión transcriptómica podrán ofrecer un sólido panorama en la variación de la herencia a nivel del producto de expresión de genes individuales, así como usar una herramienta de avanzada en población mexicana para la identificación de genes candidatos para enfermedades complejas, comunes y altamente prevalentes como la obesidad, la diabetes tipo 2 y la enfermedad cardiovascular aterosclerosa.⁹

Recientemente se ha especulado sobre la utilización de expresión de ARN en tejidos subrogados de otros de difícil obtención, tal es el caso de la utilización de la expresión de transcritos de células blancas en sangre que puedan concordar con la expresión de transcritos en el sistema nervioso central, sin embargo, dicha relación todavía está en investigación.

En el presente documento se describe la metodología de análisis de los transcritos y la cuantificación de ARNm con técnicas de alto rendimiento. Es de destacar que este estudio se realizó con muestras simultáneas de tres tejidos en los mismos individuos. También se analiza la concordancia normalizada de la expresión de los tres tejidos obtenida de las biopsias de tejido muscular y adiposo y linfocitos de cuatro sujetos masculinos mexicanos, como parte de la estandarización de la metodología del estudio GEMM.

Material y métodos

Población, recolección y envío de muestras y fenotipos

En agosto de 2007, colaboradores de la Universidad Autónoma de Nuevo León reclutaron a cuatro sujetos masculinos originarios de Monterrey, dos de ellos con índice de masa corporal (IMC, peso kg/estatura m²) menor de 25 y otros dos con IMC mayor de 30. Estos sujetos firmaron un consentimiento informado por escrito previamente aprobado por el Comité de Ética de la misma universidad.

Cada sujeto fue sometido a examen clínico, se registraron fenotipos antropométricos (estatura, peso, IMC y composición corporal por bioimpedancia eléctrica), presión arterial y frecuencia cardíaca. Además, se pidió una muestra de sangre para medición de glucemia y HbA1c. Otra alícuota de 3 ml de sangre fue colectada en tubos Tempus para estabilización del ARN (Applied Biosystems). Un cirujano plástico obtuvo biopsias de tejido muscular (vasto externo del cuádriceps) y tejido adiposo subcutáneo infraumbilical. Estas biopsias fueron incubadas durante la noche en RNALater (Ambion) en refrigeración y posteriormente en congelación (-20°C). Los participantes fueron nuevamente examinados al segundo y quinto día del procedimiento; dicha evaluación determinó que todos se recuperaron de las heridas quirúrgicas sin complicaciones.

La sangre y los tejidos congelados se enviaron desde Monterrey, México, por mensajería especializada en transporte internacional de muestras biológicas (World Courier Management, Inc.), al Departamento de Genética de la Southwest Foundation for Biomedical Research (SFBR) en San Antonio, Texas, en Estados Unidos. El análisis de las muestras fue aprobado por el Comité de Ética (*Institutional Review Board*) de la *University of Texas Health Science Center*, San Antonio, y por el Comité de Ética de la Universidad Autónoma de Nuevo León.

Amplificación del ARN

La amplificación del ARN por transcripción reversa es el método estándar para adecuar las muestras y realizar un

análisis de microarreglos, que requirió 1.5 mg de ARN amplificado por muestra. Dicha amplificación se realizó de acuerdo a técnicas establecidas.¹⁰

La calidad del ARN obtenido fue el factor más importante que afecta la eficiencia de la amplificación del ARN debido a que las impurezas en la muestra disminuyen la eficiencia de la transcriptasa inversa. La integridad del ARN contenido en las muestras se midió a través de Agilent® 2100 Bioanalyzer. Estos microfluidos viajan en pequeños tubos capilares de vidrio (minilaboratorios) que miden los cambios en gradientes y los tiempos de migración que tienen las moléculas. La interpretación de la calidad, pureza y cantidad apropiada es determinada respecto a que típicamente los ARN ribosomales se observan en las bandas 18S y 28S. Lo anterior estuvo en concordancia con nuestro análisis del ARN (Figura 1), sugiriendo que la aparición de estos dos picos es necesaria e indicativa de la integridad apropiada del ARN. Por el contrario, cuando el ARN ha sido parcialmente degradado, genera pequeñas cadenas de ADNc, lo que a su vez reduce la longitud promedio del ARNc, situación que puede detectarse con la aparición de picos más pequeños antes de las

bandas 18S y 28S. Dichos picos disminuidos indican que la calidad del ARN es inapropiada.

La cantidad de ARN que se obtiene es generalmente menor en las muestras de tejido adiposo, probablemente debido a la menor expresión genética de este tejido, aunada a la menor densidad celular por volumen cuando se compara con otros tejidos como el músculo.

El control del Illumina es 1 mg/ml de ARN proveniente de células HeLa. Este control no puede ser comparado con la cantidad de ARN proveniente de los tejidos experimentales debido a que el control es ARN de gran pureza y calidad.¹¹ Por lo tanto, se utiliza solo para corroborar la calidad del método.

Análisis de la expresión genética en tejido adiposo, muscular y linfocitos

Los perfiles transcripcionales generados en las plataformas del Illumina Human-6 v2 BeadChip contienen 48 687 oligonucleótidos específicos de 50-mer en aproximadamente 30 diferentes pozos en promedio. De ellos, 47% se considera secuencias de referencia (RefSeq) que consisten en trans-

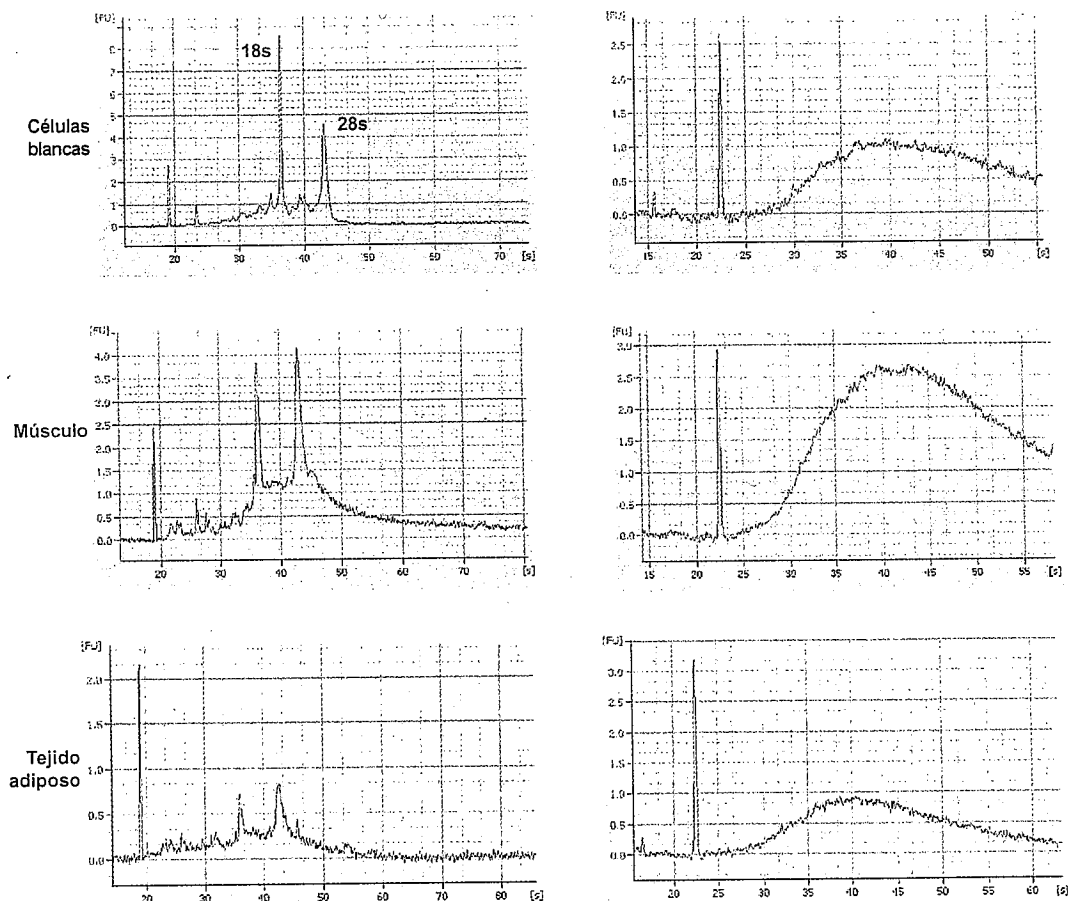


Figura 1. Electroferogramas del ARN total de células sanguíneas (luego de eliminar restos de hemoglobina), músculo y grasa. El tamaño de los ARN es estimado como una función del tiempo de migración calculado por el software del bioanizador. En el lado izquierdo pueden observarse los ARN 5S, 18S y 28S. En el derecho, una vez que se ha amplificado el ARN se modifican completamente las curvas.

critos bien caracterizados; en cambio, 53% corresponde a transcritos predichos y poco caracterizados (no-RefSeq).¹²

Se pudo obtener cantidad abundante de ARN de todas las muestras de músculo, grasa y linfocitos, sin embargo, la cantidad de ARN obtenida del tejido adiposo fue menor, situación que se anticipaba. Se realizó una prueba de χ^2 para cada uno de los transcritos con el objeto de determinar si hubo exceso significativo de las muestras con valores por arriba del percentil 95, de acuerdo a una distribución basada en muestras negativas que el proveedor proporciona. Esta prueba de χ^2 es efectuada por el programa del Illumina BeadStudio para cada transcrito en cada muestra. Considerando este criterio, de los 48 687 transcritos diana en el Illumina Human-6 v2 BeadChip fueron detectados 19 188 (39.4%) en al menos un tejido (linfocitos 30.9%, músculo 28.4%, tejido adiposo 30.8%). Este criterio fue ajustado para esta muestra pequeña: si el valor de p fue menor de 0.05 para un transcrito en las cuatro muestras, se consideró este transcrito como "detectable" en el tejido estudiado. Este criterio es muy conservador y resulta apropiado solo para este estudio piloto. Cuando se obtengan las muestras completas del estudio GEMM se utilizará el criterio de razón de falso descubrimiento (*false-discovery rate*), usado en el *San Antonio Family Heart Study* (SAFHS).¹³ A manera de comparación, en el SAFHS se ha informado una detección de 20 413 (43.2%) oligonucleótidos de 1240 muestras de linfocitos. Cada oligonucleótido detectado es un fenotipo de expresión transcricional específico.¹³

En nuestra muestra piloto se calculó una calificación de Z (*Z-score*) por cada transcrito obtenido, basada en el promedio de expresión de cada muestra. Estas calificaciones de Z se usaron para calcular los análisis de coeficiente de correlación intraclase de dos factores aleatorios y para estimar los componentes principales basados en matrices de correlación. También se utilizaron para calcular la regresión producto-momento de Pearson entre el índice de masa corporal y la expresión de leptina.

Resultados

Calidad de la amplificación del ARN

El ARN extraído de los tejidos, posteriormente amplificado, se sometió a cuantificación. En la figura 1 puede observarse cómo aparecen picos altos de ARN 18S y 28S en linfocitos

Cuadro I. Características de los participantes

Sujeto	Edad	Peso (kg)	IMC	TA (mmHg)	HbA1c (%)	Grasa corporal (%)
A	34	63.7	20.3	100/70	4.8	12.2
B	33	65.5	24.4	110/86	5.2	21.7
C	33	105.2	35.6	136/90	5.0	38.3
D	37	132.2	41.3	126/90	5.4	36.3

TA=Presión arterial.

y músculo; en cambio, en el tejido adiposo las señales mostraron menor amplitud.

Expresión de LEP y GLUT4, ejemplos de la especificidad de la expresión de los tejidos

Una vez obtenidos los transcritos, se seleccionaron dos genes como ejemplo del comportamiento específico de acuerdo con el tejido donde se expresan. Ambos genes de leptina (*LEP*) y Glut4 (*GLUT*) ayudaron a determinar que la expresión de los genes que obtuvimos tiene el sentido biológico esperado. Se observó la expresión de leptina (*LEP*) en tejido adiposo (~2-7 desviaciones estándares por arriba del promedio de expresión en este tejido) con una correlación directa con el IMC del sujeto. El transportador de glucosa en respuesta a la insulina *SLC2A4* (*GLUT4*) se detectó solo en músculo. Muchos otros genes también mostraron especificidad de tejido (observaciones aún no publicadas).

En el cuadro I pueden observarse las características de los sujetos estudiados. La figura 2 muestra la correlación entre la cantidad de transcrito e IMC en tejido adiposo y tejido muscular.

Concordancia entre la expresión normalizada de tejidos

El análisis de concordancia en la expresión normalizada nos da una aproximación en cuanto a la cantidad del transcrito producido en cada uno de los tejidos respecto al otro, considerados como factores aleatorios provenientes de una muestra mayor de tejidos. El coeficiente de correlación intraclase para factores aleatorios de los tres tejidos fue de 0.704 (IC95%=0.695-0.712). El coeficiente de células en sangre y músculo fue de 0.601 (IC95%=0.589-0.613), en sangre y tejido adiposo de 0.749 (IC95%=0.741- 0.756) y, finalmente,

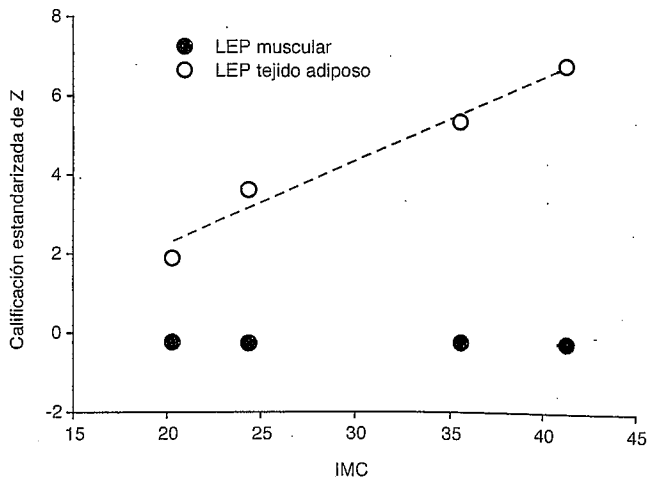


Figura 2. Correlación entre el número de transcritos del gen de leptina (*LEP*) y el índice de masa corporal de los sujetos estudiados. La correlación entre la expresión del transcrito de leptina e IMC fue $r=0.98$ (IC95%=0.43-1.0) para el tejido adiposo y $r=-0.67$ (IC95%=-0.99-0.82) para el tejido muscular.

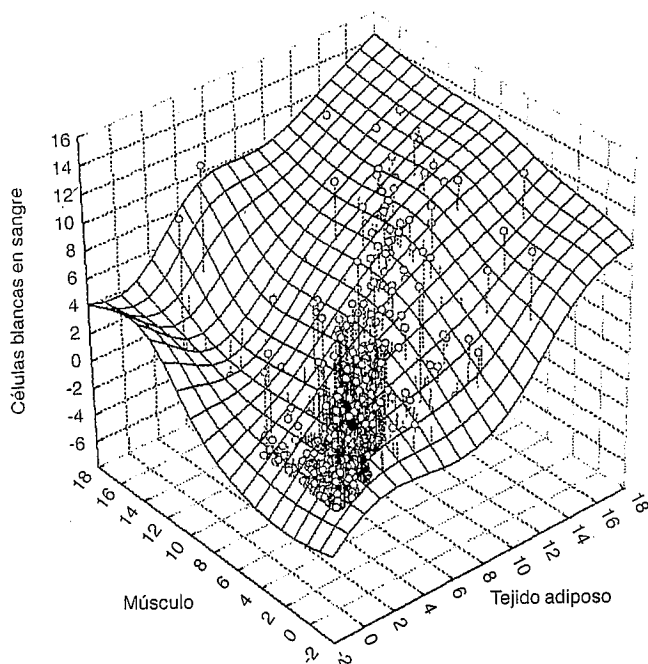


Figura 3. Superficie de respuesta de la correlación de transcritos de tejido adiposo y muscular con la expresión génica de células blancas en sangre. La concordancia fue de 0.74 (IC95%=0.70-0.71). La escala de los ejes son unidades normalizadas en Z. Ver texto para detalles.

en músculo y tejido adiposo de 0.69 (IC95%=0.68-0.70) (Figura 3).

El análisis de componentes principales muestra que los tres tejidos pueden combinarse linealmente para formar un factor único, que explicaría 80.56% de la variabilidad de los tres tejidos; la extracción de la varianza fue de 94% para el tejido adiposo y de 88% para músculo y sangre (Figura 4). Uno de los principales objetivos del estudio GEMM es caracterizar los patrones de correlación en la expresión génica entre los diferentes tejidos de un mismo individuo. Este enfoque comparativo desde el transcriptoma representa la adquisición de información genómica altamente novedosa.

Discusión

Prácticamente no existen estudios en la literatura donde se haya podido coleccionar muestras de diferentes tejidos en un mismo individuo y se haya analizado la expresión de los diferentes transcritos conocidos del genoma; la metodología y el análisis de este tipo de abordaje apenas está conociéndose. Este estudio piloto nos da una pauta de seguimiento para lograr los objetivos propuestos por el estudio GEMM. La recolección de tejidos en un gran número de sujetos es un componente nuevo e importante para la investigación genómica de enfermedades metabólicas.

Los métodos utilizados en este estudio piloto mostraron eficiencia, factibilidad y correlación clínico-molecular entre los genes seleccionados (*LEP* y *GLUT4*), remarcando en

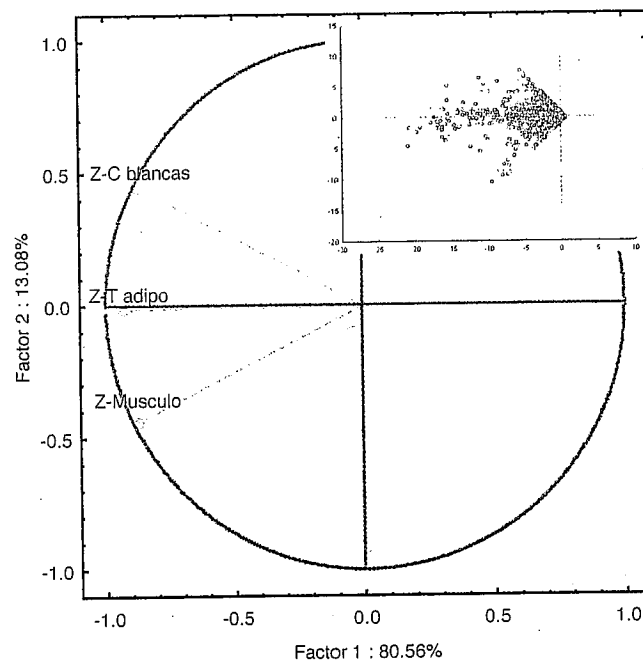


Figura 4. Análisis de componentes principales. Los tres tejidos pueden combinarse linealmente en un solo factor que explica 80% de la variabilidad. El recuadro muestra la correlación de las calificaciones de los factores. Ver texto para detalles.

este caso el ejemplo de *LEP* correlacionado con el IMC de los sujetos y su falta de expresión muscular, situación que indica patrones de especificidad. Este nuevo tipo de análisis permitirá diseccionar la complejidad de la comunicación celular, que al parecer no necesita relacionarse necesariamente con los cortes alternativos del ARNm.¹⁴ Todo parece indicar que la ciencia médica se dirige hacia la medición de endofenotipos cuantitativos biológico-moleculares que servirán indudablemente de apoyo categórico a los fenotipos clínico-observacionales convencionales en el diagnóstico causal de diversas patologías metabólicas. Los análisis efectuados en este número pequeño de sujetos permiten comprobar que el método de análisis de transcritos es confiable y los hallazgos se comportan conforme a lo esperado.

Los tejidos estudiados tienen muchos tipos de células, cuya interacción seguramente producirá variaciones entre los diferentes tipos de transcritos, los que también se verán afectados por el medio ambiente. El problema de la heterogeneidad celular en estudios a nivel transcriptómico ha sido confrontado en investigaciones muy recientes, y ha sido abordado con análisis de conglomerados de microarreglos, con los que se extrae información simultánea en grupos de genes preseleccionados.¹⁵ Se ha informado también el desarrollo de nuevas técnicas sobre métodos para obtener muestras para estudiar el transcriptoma celular. La microdissección de captura con láser es un claro adelanto en cómo solventar esta clase de dilemas metodológicos, que seguramente en el futuro será de uso cotidiano.^{16,17} Sin embargo, más allá de la interpretación metodológica señalada secundaria a medicio-

nes de escasa magnitud, la cuantificación de la concentración total de transcritos debe ser entendida como endofenotipos cuantitativos que pueden ser utilizados para efectuar mediciones estadísticas con base en análisis oligogénicos secuenciales de ligamiento capaces de identificar la arquitectura genética subyacente a nivel poblacional.¹⁸

La concordancia entre los tres tejidos es mayor que la que se podría encontrar por azar. Algunos tejidos mostraron mayor concordancia en la expresión de los transcritos, como el tejido adiposo y las células en sangre. Esto da pie a realizar análisis entre la expresión de ARN que puede encontrarse en células circulantes y otros tejidos como el sistema nervioso central. La concordancia entre estos dos tejidos es de suma importancia en un futuro cercano, ya que aportará la posibilidad de determinar la validez de subrogar una muestra sencilla para recolectar, como el tejido sanguíneo, contra una muestra difícil de obtener, como el tejido proveniente del sistema nervioso central.

El abordaje estadístico de este tipo de microarreglos de expresión debe considerar la identificación de regulación *cis* y regulación *trans*, que ejercen fuerte influencia en la expresión de oligonucleótidos. La designación *cis* es para elementos que regulan la actividad y perfil de expresión de genes que se encuentran dentro o cerca de su propio *locus*. El término *trans* se utiliza para señalar los factores reguladores que se localizan en algún cromosoma diferente del que reside el gen en cuestión. Estudios muy recientes han sugerido que la regulación *cis* es una característica que favorece la estabilidad de los transcritos individuales, que parece protegerlos ante las influencias del medio ambiente y también parece ser consistente entre tipos celulares diferentes, incluyendo tejidos.^{19,20}

Ésta es la primera experiencia con muestras mexicanas en las que se efectúa un análisis trascricional de alto rendimiento en tres tejidos diferentes tomados en un mismo paciente. Este tipo de análisis orientará los próximos pasos en la investigación de enfermedades metabólicas a nivel genómico-transcriptómico. Aunque el análisis de una muestra de cuatro sujetos es escasa para obtener datos significativos, la obtención de información conjunta en amplia escala de perfiles de expresión genética global en linfocitos, músculo y tejido adiposo facilita el camino para efectuar el proyecto GEMM en una gran cantidad de individuos, tanto desde el punto de vista metodológico como desde la seguridad de los participantes. Esta estrategia nos permitirá elegir el fenotipo trascricional o endofenotipo más apropiado (secundario al perfil de expresión total de oligonucleótidos), para posteriormente utilizarlo y efectuar análisis de ligamientos (*linkage*) a través de escaneo genómico amplio (*genome-wide scan*).

En el estudio GEMM efectuaremos escaneo genómico con endofenotipos obtenidos de la expresión de linfocitos, ya que las biopsias musculares y de tejido adiposo nos servirán para comparación cruzada en el perfil de expresión de estos tres tejidos.

También es pertinente comentar de manera breve sobre el análisis genético cuantitativo para individuos emparentados entre sí utilizando el método de componentes de la

varianza,²¹ como apropiado y práctico para abordar estudios genéticos en poblaciones de las que se conoce de manera parcial la interacción de genes y el ambiente.²²

El análisis de la expresión génica,¹⁸ la tecnología de microarreglos, la estadística genética cuantitativa, el enfoque bayesiano de análisis bioinformático, los análisis de ligamiento, los escaneos genómicos, los endofenotipos cuantitativos y otros novedosos avances en genómica integrativa, nos ofrecen un amplio y halagador panorama para la identificación de la asociación causal entre los niveles de expresión y enfermedades metabólicas tales como la obesidad y la diabetes tipo 2.

Agradecimientos

Los autores agradecen el apoyo financiero de HEB, San Antonio, Texas, para el traslado de muestras y la compra de reactivos para la realización de este estudio. Esta investigación fue conducida en las instalaciones cuya construcción fue patrocinada por un donativo (Grant) denominado Research Facilities Improvement Program Grant Number C06 RR (017515) del National Center for Research Resources, National Institutes of Health, USA.

Referencias

1. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 2004;431:931-945.
2. Forrest AR, Taylor DF, Crowe ML, Chalk AM, Waddell NJ, Kolle G, et al. Genome-wide review of transcriptional complexity in mouse protein kinases and phosphatases. *Genome Biol* 2006;7:R5.
3. Perkins DO, Jeffries C, Sullivan P. Expanding the "central dogma": the regulatory role of nonprotein coding genes and implications for the genetic liability to schizophrenia. *Molecular Psychiatry* 2005;10:69-78.
4. Kim VN, Nam JW. Genomics of microRNA. *Trends Genet* 2006;22:165-173.
5. Odom DT, Dowell RD, Jacobsen ES, Gordon W, Danford TW, McCluskey KD, et al. Tissue-specific transcriptional regulation has diverged significantly between human and mouse. *Nature Genetics* 2007;39:730-732.
6. Bastarrachea RA, Kent JA, Rozada G, Cole SA, López-Alvarenga JC, Aradillas C, et al. Heritability and genetic correlations of metabolic disease-related phenotypes in Mexico: Preliminary report from the GEMM Family Study. *Human Biology* 2007;78:121-130.
7. Martínez-Marignac VL, Valladares A, Cameron E, Chan A, Perera A, Globus-Goldberg R, et al. Admixture in Mexico City: Implications for admixture mapping of type 2 diabetes genetic risk factors. *Hum Genet* 2007;120:807-819.
8. Cerda-Flores RM, Budowie B, Jin L, Barton SA, Deka R, Chakraborty R. Maximum likelihood estimates of admixture in Northeastern Mexico using 13 short tandem repeat loci. *Am J Hum Biol* 2002;14:429-439.
9. Curran JE, Johnson MP, Dyer TD, Göring HH, Jack WK, Charlesworth JC, et al. Genetic determinants of mitochondrial content. *Hum Mol Genet* 2007;16:1504-1514.
10. Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci USA* 1990;87:1663-1667.
11. Illumina TotalPrep RNA Amplification Kit. Instruction manual. 2006.
12. Pruitt KD, Maglott T. NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 2005;33:D501-D504.
13. Göring HH, Curran JE, Jonson MP, Dyer T, Charlesworth J, Cole SA, et al. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nature Genetics* 2007; advance online publication; doi: 10.1038/ng2119.
14. Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, et al. RIKEN Genome Exploration Research Group; Genome Science Group (Genome Network Project Core Group); FANTOM Consortium. Antisense transcription in the mammalian transcriptome. *Science* 2005;309:1564-1566.
15. Ponnampalam AP, Weston GC, Trajstman AC, Susil B, Rogers PA. Molecular classification of human endometrial cycle stages by transcriptional profiling. *Mol Hum Reprod* 2004;10:879-893.

16. Holter JL, Humphries A, Crunelli V, Carter DA. Optimization of methods for selecting candidate genes from cDNA array screens: application to rat brain punches and pineal. *J Neurosci Methods* 2001;112:173-184.
17. Segal JP, Stallings NR, Lee CE, Zhao L, Socci N, Viale A, et al. Use of laser-capture microdissection for the identification of marker genes for the ventromedial hypothalamic nucleus. *J Neurosci* 2005;25:4181-4188.
18. Jansen RC, Nap JP. Genetical genomics: the added value from segregation. *Trends Genet* 2001;17:388-391.
19. 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.
20. 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.
21. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998;62:1198-1211.
22. Comuzzie GA, Williams JT, Marin LJ, Blangero J. Searching for genes underlying normal variation in human adiposity. *J Mol Med* 2001;79:57-70.