

Mapeo Genético mediante desequilibrio de ligamiento por mezcla y su aplicación en enfermedades complejas

Genetic mapping by means of ligament unbalance by mix and its application in complex diseases

Candelaria Vergara; Beatriz Mercedes Martínez Alfaro

Laboratorio de Genética Molecular, Instituto de Investigaciones Inmunológicas, Universidad de Cartagena, Cartagena, Colombia.

RESUMEN

La identificación de los genes implicados en enfermedades complejas es uno de los grandes desafíos en el campo de la genética humana. Dentro de las estrategias utilizadas para tal fin están los estudios de ligamiento en familias y los estudios de asociación poblacional usando casos y controles. Sin embargo, una alternativa propuesta en los últimos años es el método de mapeo genético por desequilibrio de ligamiento por mezcla, también denominado "Admixture Mapping" (AM). Es una estrategia que metodológicamente se ubica entre los estudios de ligamiento en familias y los rastreos genómicos para asociación y que es aplicable a poblaciones genéticamente mezcladas.

Palabras clave: AMs, marcador de ancestría, enfermedades complejas, poblaciones mezcladas, Mapeo genómico

ABSTRACT

Characterization of gens involved in complex diseases is one of greatest challenges in field of human genetics. Included in the more used strategies to this objective are studies of ligament in families, and studies of population association, using cases and controls. However, in pas years an alternative proposal is the genetic mapping method by unbalance of ligament by mix, also called "Admixture Mapping" (AM). It is an strategy that methodologically is placed among ligament studies in families and the genomic searching for association, and it is applicable in populations presenting with genetic mixture.

Key words: AMs, ancestral marker, complex diseases, mixed populations, genomic mapping

MEZCLA POBLACIONAL Y MAPEO GENÉTICO POR DESEQUILIBRIO DE LIGAMIENTO POR MEZCLA

La mezcla poblacional es la formación de una población nueva mediante el entrecruzamiento de dos poblaciones genéticamente divergentes (parentales) seguida por el apareamiento de sus descendientes ¹ cuyos genomas están compuestos de fragmentos cromosómicos originados en las distintas poblaciones parentales. Durante la mezcla se genera en forma transitoria el denominado "desequilibrio de ligamiento por mezcla" (ALD, por sus cifras en inglés: *Admixture Linkage Disequilibrium*) en los individuos de la población híbrida, consistente en grandes bloques de haplotipos y asociaciones no aleatorias entre polimorfismos de un solo nucleótido (SNPs por su nombre en inglés *Single Nucleotide Polymorphism*) a lo largo de distancias de cientos de centimorgans (cM). La longitud de los bloques depende en gran medida de las proporciones en la contribución de las poblaciones parentales y el grado de mezcla continua y reciente entre la población nueva y cualquiera de las parentales. ² El ALD, el cual surge principalmente entre los loci que tienen distintas frecuencias alélicas entre las poblaciones parentales, disminuye en función de la tasa de recombinación entre los loci y el tiempo representado por el número de generaciones después del evento de mezcla. ³ La disolución del ALD es rápida entre aquellos loci separados por distancias mayores de 30 cM o entre loci ubicados en diferentes cromosomas pero persiste por más tiempo y es más difícil de disolver entre los loci separados por distancias menores a 10cM.

El ALD es la base genética del AM; este consiste teóricamente en hacer un rastreo genómico, ya sea en casos y controles o en casos solamente, pertenecientes a una población mezclada en la que una de las poblaciones parentales tiene una mayor incidencia de la enfermedad identificando inicialmente los segmentos genómicos provenientes de cada población parental mediante el uso de marcadores informativos de ancestría (AIMs, por *Ancestry Informative Markers*). De manera que si un segmento específico que proviene de la población parental con mayor incidencia de la enfermedad se encuentra en una frecuencia significativamente mayor en casos que en controles, es sugestivo de que en él se localiza el gen o alelo asociado con la enfermedad. ¹ A este análisis inicial le sigue el mapeo más denso de la región o regiones identificadas y la caracterización de los genes candidatos contenidos en esa región. Este método tiene el mismo poder que los rastreos genómicos para asociación en poblaciones homogéneas y tiene las ventajas de que no es influenciado por la heterogeneidad alélica y que además puede realizarse a un menor costo debido a que se usan entre 1,200 y 3,000 marcadores, es decir, unas 200 a 500 veces menos que los rastreos genómicos.

Diseño de mapeos genéticos mediante desequilibrio de ligamiento por mezcla

Aunque las bases teóricas del AM fueron descritas originalmente hace más de una década, ³ sólo hasta ahora ha sido factible su realización gracias al desarrollo de paneles de AIMS cada vez más específicos para cada población y la disponibilidad de genotipificación a gran escala. A pesar de que la información sobre SNPs y haplotipos derivada de los proyectos Genoma Humano y HapMap es muy útil en el desarrollo de paneles de AIMs que diferencian ancestría Africana y Europea, estos

proyectos no incluyen individuos de poblaciones Amerindias, lo que limita así el diseño de mapas de marcadores para las poblaciones Latinas. Sin embargo, en un esfuerzo colectivo, un grupo de investigadores analizaron varias bases de datos personales y públicas y seleccionaron 1.646 marcadores distribuidos a lo largo del genoma para conformar un panel discriminativo de las poblaciones Europea, Amerindia y Africana, las tres principales poblaciones fundadoras de las poblaciones Latinas actuales; de manera que en la actualidad ya es posible aplicar este método a poblaciones como la Colombiana. De hecho, la efectividad de este grupo de marcadores fue validada experimentalmente, con excelentes resultados, en 232 muestras provenientes de una población latina habitante en Norteamérica, mejicanos, brasileros y colombianos.⁴ Por su parte, en otro estudio publicado simultáneamente, se describió otro grupo de 2.120 AIMS que diferencia específicamente ancestría Europea/Africana de la Amerindia (discriminando para Meso América y Suramérica) con una distancia genética y física promedio entre marcadores de 1.74 cM y 1.22 Mb (SD \pm 832kb), respectivamente. Según los autores, el diseño de este panel, en el que se excluyeron AIMS con frecuencias alélicas significativamente diferentes entre poblaciones Amerindias de Norte y Suramérica, permite usarlo para hacer AM en muchas poblaciones mezcladas a lo largo del continente Americano.⁵

En términos prácticos, el AM es un proceso que consta de varios pasos: el primero es seleccionar la población mezclada y estimar el tamaño muestral de la población a utilizar en el AM que tenga un poder aceptable (> 0.8). El tamaño muestral depende de varios factores como las proporciones de ancestría en la población mezclada, los cuales se pueden determinar analizando inicialmente varias docenas de AIMS. Las proporciones a su vez dependen de las contribuciones de las poblaciones parentales y de la dinámica de la mezcla, es decir si hubo un único evento de mezcla inicial entre las poblaciones parentales o en forma alternativa, si además del evento inicial, las poblaciones parentales siguieron mezclándose con la nueva población híbrida en forma continua a través del tiempo (flujo genético continuo). Otro factor determinante en el tamaño muestral es el riesgo relativo atribuible a la etnicidad, definido por algunos autores como la relación entre el riesgo de una ancestría sobre la otra (s)⁶ y que es igual a la relación de riesgos genotípicos (GRR por *Genotypic Risk Ratio*) en el caso de que la variante causante tenga frecuencias extremas en las poblaciones fundadoras (0 y 100%). El mínimo GRR detectable en un AM es 1.5 en estudios que analizan varios miles de individuos afectados.⁶⁻⁸

El segundo paso es establecer el número, densidad y tipo de marcadores a utilizar. El número y la densidad están relacionadas con el número de generaciones desde el evento de mezcla dado que de esto depende el tamaño y distribución del ALD; a mayor número de generaciones, menor es el tamaño de los bloques de ALD y mayor el número de marcadores a utilizar, asumiendo sólo un evento de mezcla inicial sin contribución posterior de las poblaciones parentales. Los marcadores a utilizar, ya sean microsatélites (STRs, por sus siglas en inglés *Short Tandem Repeats*) o SNPs se escogen dependiendo de la informatividad de los mismos; aunque los STRs son más informativos, la tendencia es a usar "mapas" de SNPs dada la mayor facilidad para tipificarlos a gran escala. Existen varias medidas de informatividad, la más sencilla desde el punto de vista estadístico es r , que es la diferencia absoluta en las frecuencias alélicas de cada AIM entre las poblaciones parentales; se considera que un buen "mapa" está compuesto de marcadores con SNPs cuyos rs sean >0.6 entre las poblaciones parentales.¹ La informatividad individual de los marcadores es indicativa de la habilidad del "mapa" para extraer la información de ancestría en la población mezclada, es decir, para asignar la ancestría de los distintos segmentos cromosómicos. Por ejemplo un "mapa" de 4.222 AIMS diseñado para Afro americanos fue capaz de extraer más del 70% de información de ancestría para el 90% del genoma asumiendo un modelo con 6

generaciones después del evento de mezcla, flujo genético continuo y una relación de ancestría Africana: Europea de 80:20.⁹ En el caso de las poblaciones mejicanas que viven en Norteamérica, las cuales tuvieron una mezcla más antigua (15 generaciones) y proporciones de ancestría europea: amerindia de 50:50, se necesitaría un "mapa" con 5000 AIMs para tener un poder aceptable. Uno de los "mapas" mencionados antes para poblaciones latinas tienen un 50% de informatividad para AM en estas poblaciones,⁴ un porcentaje muy similar al primer "mapa" descrito para Afroamericanos.¹⁰

Toda la información obtenida de los AIMs debe ser analizada en conjunto para obtener las proporciones de ancestría a lo largo del genoma y hacer las comparaciones ya sea entre casos y controles o entre los distintos segmentos del genoma en casos. Los programas de computador disponibles para el público actualmente son ANCESTRYMAP,⁸ ADMIXMAP⁷ y STRUCTURE/MALD;¹¹ además de un nuevo algoritmo descrito por Tang y colaboradores el cual tiene en cuenta el desequilibrio de ligamiento en las poblaciones parentales.¹² ANCESTRYMAP utiliza la prueba Bayesiana de razón de probabilidades para calcular la ancestría genómica individual y obtiene el promedio de toda la población analizada; con ese valor como base, identifica las regiones genómicas con la mayor ancestría de la población parental con mayor incidencia de la enfermedad en las cuales puede estar el gen asociado. Aunque no necesita de controles para hacer las asociaciones entre ancestría y la enfermedad, es recomendado incluirlos para estimar en forma robusta las frecuencias alélicas de las poblaciones ancestrales y así incrementar el poder del análisis. ADMIXMAP utiliza fundamentos estadísticos tanto Bayesianos como clásicos; permite inferir un modelo de historia de la mezcla dado que estima la distribución de la ancestría individual y la variación estocástica de la ancestría de los cromosomas híbridos mediante el método de simulación de cadenas de Markov vía Monte Carlo. Este método puede analizar poblaciones mezcladas con más de dos poblaciones ancestrales y es capaz de manejar marcadores bialélicos y multialélicos para hacer análisis de asociación de la ancestría tanto con rasgos dicotómicos (enfermedad si/no) como con rasgos cuantitativos (Niveles séricos de IgE total).¹³ ADMIXMAP tiene la ventaja de que reconstruye la estructura genética de poblaciones ancestrales aunque no haya disponibilidad de muestras de esos individuos a diferencia del ANCESTRYMAP en el que el usuario debe aportar la información de las poblaciones parentales para hacer el análisis. Los programas STRUCTURE y MALD, por su parte, son programas complementarios que usan la información de los genotipos de AIMs en casos y controles provenientes de la población mezclada y de individuos representativos de las parentales; no asumen un modelo particular de la enfermedad y analizan qué tanto la ancestría de un locus se desvía del promedio genómico. De acuerdo con los autores, estos dos programas tienen la ventaja de que no se requiere que los marcadores sean necesariamente informativos de ancestría sino que se puede hacer el cálculo con SNP distribuidos al azar en el genoma.¹¹

En el caso de que se quiera diseñar un estudio de AM en la población Colombiana, es necesario tener en cuenta que los estimados de ancestría poblacional varían de una región a otra, dependiendo de la historia demográfica de cada población. Por ejemplo, para el departamento de Antioquia, la ancestría es 79% europea, 15% amerindia y 6% africana.¹⁴ En esta población considerada como aislada, la dinámica de mezcla se describió como una inmigración inicial antes del siglo XVIII y una subsiguiente expansión de la población por crecimiento interno de la misma con bloques de LD más largos que otras poblaciones no aisladas¹⁵ por lo que teóricamente requeriría un gran número de marcadores para hacer AM. Para otras poblaciones colombianas, es necesario hacer un estudio inicial que analice las características étnicas y genéticas y de esa manera establecer las bases metodológicas para el AM.

Mapeos de desequilibrio de ligamiento por mezcla y enfermedades complejas

EL AM ha sido utilizado con éxito para la búsqueda de genes de susceptibilidad para rasgos y enfermedades complejas. Zhu y colaboradores llevaron a cabo un AM para hipertensión arterial, la cual tiene una mayor prevalencia y severidad en poblaciones de origen africano. Estos investigadores tipificaron 269 STRs distribuidos en el genoma en 737 casos y 573 controles Afro americanos tomando como poblaciones parentales 236 individuos de Nigeria y 293 Euro americanos.¹⁶ Utilizando STRUCTURE, los autores encontraron una mayor ancestría africana en los casos y una asociación entre la enfermedad y marcadores ubicados en 6q24 y 21q21. En esta población también se hizo un AM para cáncer de próstata, el cual tiene una mayor incidencia en poblaciones de ascendencia africana. Se incluyeron 1597 individuos no relacionados y se encontró asociación con un segmento de 3.2 Mb en el cromosoma 8p24.¹⁷ Por su parte, en la esclerosis múltiple se ha determinado que existe un mayor riesgo para poblaciones de origen Caucásico en comparación con las africanas, lo que la convierte en otro candidato para analizar mediante AM en poblaciones mezcladas. En un estudio con un alto poder estadístico realizado en Afroamericanos, se tipificaron 1558 SNPs en 605 casos y 1043 controles, encontrándose una asociación de la enfermedad con un locus ubicado en el cromosoma 1.¹⁸ La mayoría de las asociaciones encontradas en estos estudios están en concordancia con resultados anteriores obtenidos en estudios de ligamiento lo que aporta confiabilidad a los resultados.

El AM es aplicable a las poblaciones mezcladas para identificar los factores genéticos de susceptibilidad para las enfermedades previamente anotadas arriba y también para asma, una enfermedad compleja y multifactorial que tiene una mayor prevalencia y severidad, cursa con niveles más altos IgE total y con una mayor dependencia de esteroides en poblaciones africanas en comparación con poblaciones blancas¹⁹ y amerindias, aún después de descartar las variables socioeconómicas como causales, lo que indica que los factores genéticos podrían ser los responsables de estas diferencias entre los grupos étnicos.

CONCLUSIÓN

El mapeo por desequilibrio de ligamiento por mezcla es una estrategia novedosa de epidemiología genética susceptible de realizar en poblaciones con una mezcla reciente como la colombiana y las del Continente Americano en general, que ha demostrado ser útil en la investigación de genes de susceptibilidad asociados con rasgos y enfermedades complejas con prevalencias variables de acuerdo a la etnicidad. Su implementación se ve facilitada cada día más debido a la mayor disponibilidad de bases de datos sobre las variantes genéticas en los distintos grupos étnicos, el desarrollo de mejores y más eficaces métodos de genotipificación a gran escala y el diseño de algoritmos bioestadísticos y robustos programas de computación para el manejo de grandes bases de datos. Esta estrategia puede ser de mucha utilidad para la identificación de factores de riesgo genético para varias enfermedades complejas como el asma.

REFERENCIAS BIBLIOGRÁFICAS

1. Smith MW, O'Brien SJ. Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. *Nat Rev Genet* 2005;6(8):623-32.

2. Pfaff CL, Parra EJ, Bonilla C, Hiester K, McKeigue PM, Kamboh MI, et al. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *Am J Hum Genet* 2001; 68(1):198-207.
3. Chakraborty R, Weiss KM. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc Natl Acad Sci U S A* 1988; 85(23):9119-23.
4. Price AL, Patterson N, Yu F, Cox DR, Waliszewska A, McDonald GJ, et al. A genome-wide admixture map for Latino populations. *Am J Hum Genet* 2007; 80(6):1024-36.
5. Mao X, Bigham AW, Mei R, Gutierrez G, Weiss KM, Brutsaert TD, et al. genome-wide admixture mapping panel for Hispanic/Latino populations. *Am J Hum Genet* 2007; 80(6):1171-8.
6. Seldin M. Admixture mapping as a tool in gene discovery. *Curr Opin Gen Dev* 2007; 17:177-181.
7. Hoggart CJ, Shriver MD, Kittles RA, Clayton DG, McKeigue PM. Design and analysis of admixture mapping studies. *Am J Hum Genet* 2004; 74(5):965-78.
8. Patterson N, Hattangadi N, Lane B, Lohmueller KE, Hafler DA, Oksenberg JR, et al. Methods for high-density admixture mapping of disease genes. *Am J Hum Genet* 2004; 74(5):979-1000.
9. Tian C, Hinds DA, Shigeta R, Kittles R, Ballinger DG, Seldin MF. A genome-wide single-nucleotide-polymorphism panel with high ancestry information for African American admixture mapping. *Am J Hum Genet* 2006; 79(4):640-9.
10. Smith MW, Lautenberger JA, Shin HD, Chretien JP, Shrestha S, Gilbert DA. Markers for mapping by admixture linkage disequilibrium in African American and Hispanic populations. *Am J Hum Genet* 2001; 69(5):1080-94.
11. Montana G, Pritchard JK. Statistical tests for admixture mapping with case-control and cases-only data. *Am J Hum Genet* 2004; 75(5):771-89.
12. Tang H, Coram M, Wang P, Zhu X, Risch N. Reconstructing genetic ancestry blocks in admixed individuals. *Am J Hum Genet* 2006; 79(1):1-12.
13. McKeigue PM, Carpenter JR, Parra EJ, Shriver MD. Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. *Ann Hum Genet* 2000; 64(Pt 2):171-86.
14. Bedoya G, Montoya P, Garcia J, Soto I, Bourgeois S, Carvajal L, et al. Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. *Proc Natl Acad Sci U S A* 2006; 103(19):7234-9.
15. Service Sea. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nat Genet* 2006; 38:556-560.
16. Zhu X, Luke A, Cooper RS, Quertermous T, Hanis C, Mosley T, et al. Admixture mapping for hypertension loci with genome-scan markers. *Nat Genet* 2005; 37(2):177-81.

17. Freedman ML, Haiman CA, Patterson N, McDonald GJ, Tandon A, Waliszewska A, et al. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc Natl Acad Sci U S A* 2006; 103(38): 14068-73.

18. Reich D, Patterson N, De Jager PL, McDonald GJ, Waliszewska A, Tandon A, et al. A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. *Nat Genet* 2005; 37(10): 1113-8.

19. Grant EN, Lyttle CS, Weiss KB. The relation of socioeconomic factors and racial/ethnic differences in US asthma mortality. *Am J Public Health* 2000; 90(12): 1923-5.

Recibido: Abril 2008

Aprobado: Mayo 2008

Candelaria Vergara, MSc, Profesor Auxiliar, Actualmente Estudiante de doctorado en Genética de Poblaciones, Pie de la Popa, Edificio Parque Real apto 2^a3, Cartagena, Bolívar

Beatriz Mercedes Martínez Alfaro, MSc, Profesor Asociado, Jefe del Laboratorio de genética Molecular del Instituto de Investigaciones inmunológicas, San diego Calle de la Tablada 7-57, Cartagena, Bolívar Colombia