

REIVINDICACIONES

1. Un método *in vitro* para identificar genotípicamente ambos alelos de al menos un locus de un gen HLA en muestras de ADN genómico de un sujeto, que comprende:

(a) amplificar por PCR de largo alcance dichas muestras de ADN genómico utilizando al menos un par de cebadores sentido y antisentido, para cada muestra, donde dicho par de cebadores, sentido y antisentido, está identificado por las secuencias seleccionadas del grupo que consiste en SEQ ID NO: 1-2, SEQ ID NO: 3-4, SEQ ID NO: 5-6, SEQ ID NO: 7-8, SEQ ID NO: 9-10, SEQ ID NO: 11-12, SEQ ID NO: 13-14, SEQ ID NO: 15-16, SEQ ID NO: 17-18, SEQ ID NO: 19-20 y SEQ ID NO: 21-22, donde dicho par de cebadores amplifica en un único amplicón al menos un gen HLA completo junto con regiones 5'-UTR y 3'-UTR de dicho gen HLA, una región corriente arriba y una región corriente abajo de dicho gen HLA, de ambos alelos, de al menos un locus de un gen HLA, formando así amplicones, donde dicho gen HLA está seleccionado del grupo que consiste en HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5, HLA-DPB1, HLA-DQB1, HLA-DPA1 y HLA-DQA1,

(b) combinar los amplicones de cada una de las muestras,

(c) fragmentar los amplicones, obteniendo así fragmentos y etiquetar dichos fragmentos, formando así librerías de fragmentos etiquetados,

(d) combinar dichas librerías de fragmentos etiquetados,

(e) secuenciar dichas librerías de fragmentos etiquetados por secuenciación masiva, generando así lecturas de secuencias que solapan parcialmente,

(f) alinear dichas lecturas de secuencias contra un genoma de referencia,

(g) identificar variantes en base a una llamada de variantes,

(h) determinar bloques haplotípicos en base al alineamiento de las lecturas de secuencias contra un genoma de referencia y añadir regiones flanqueantes de dichos bloques haplotípicos a dichos bloques haplotípicos,

(i) alinear localmente las lecturas de secuencias de los bloques haplotípicos obtenidos en la etapa anterior contra todos los alelos de una base de datos de referencia de secuencias genómicas que codifican genes HLA para cada gen HLA, obteniendo así unos primeros alelos candidatos,

(j) pre-calcular las distancias genéticas y determinar la agrupación filogenética de las familias de HLA a diferentes niveles en la base de datos de referencia de secuencias genómicas que codifican genes HLA, obteniendo así una matriz de distancia que contiene los valores de distancia de cada alelo consigo mismo y con los demás,

(k) llevar a cabo una selección primaria de dichos primeros alelos candidatos en base al resultado de contraste de homología global de las lecturas de secuencias a los haplotipos, obteniendo así unos segundos alelos candidatos,

(l) llevar a cabo un estudio de filogenia evolutiva de dichos segundos alelos candidatos, obteniendo así un árbol filogenético, acceder a la matriz de distancia obtenida en la etapa (j) y seleccionar alelos HLA de la base de datos de referencia de secuencias genómicas que codifican genes HLA más cercanos filogenéticamente, obteniendo así alelos HLA contrastados y

(m) llevar a cabo un test de contraste con alelos HLA contrastados, calcular las distancias entre las lecturas de secuencias y los alelos HLA contrastados en cada nivel y asignar cada lectura de secuencias a una familia HLA conocida a diferentes niveles, donde se asigna una lectura de secuencia a una familia HLA conocida a diferentes niveles cuando se obtiene una significancia estadística en dicho test de contraste por encima de un valor de corte definido y donde se asigna una lectura de secuencia como secuencia *de novo* que codifica para un nuevo locus del gen HLA cuando se obtiene una significancia estadística en dicho test de contraste por debajo de un valor de corte definido.

2. El método según la reivindicación 1, caracterizado por que dicha significancia estadística se calcula en base al valor estadístico p.

3. El método según la reivindicación 1 ó 2, caracterizado por que dicho valor de corte definido se calcula en base al valor estadístico p y tiene un valor de 10^{-6} .

4. El método según cualquiera de las reivindicaciones 1 a 3, caracterizado por que la longitud de los fragmentos generados en la etapa (c) es entre 1200 y 1500 pb.

5. El método según cualquiera de las reivindicaciones 1 a 4, caracterizado por que las librerías se secuencian en la etapa (e) por secuenciación masiva con una longitud de lectura entre 200 y 300 pb.

6. Un par de cebadores identificado por las secuencias seleccionadas del grupo que consiste en SEQ ID NO: 1-2, SEQ ID NO: 3-4, SEQ ID NO: 5-6, SEQ ID NO: 7-8, SEQ ID NO: 9-10, SEQ ID NO: 11-12, SEQ ID NO: 13-14, SEQ ID NO: 15-16, SEQ ID NO: 17-18, SEQ ID NO: 19-20 y SEQ ID NO: 21-22, para uso en la identificación genotípica *in vitro*

de ambos alelos de al menos un locus de un gen HLA de un sujeto, donde dicho par de cebadores amplifica en un único amplicón un gen HLA completo junto con regiones 5'-UTR y 3'-UTR de dicho gen HLA, una región corriente arriba y una región corriente abajo de dicho gen HLA, de ambos alelos de al menos un locus de un gen HLA, donde dicho gen HLA está seleccionado, respectivamente, del grupo que consiste en HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5, HLA-DPB1, HLA-DQB1, HLA-DPA1 y HLA-DQA1.

7. Un kit para identificar genotípicamente *in vitro* ambos alelos de al menos un locus de un gen HLA en muestras de ADN genómico de un sujeto, que comprende:

(a) al menos un par de cebadores, sentido y antisentido, identificado por las secuencias seleccionadas del grupo que consiste en SEQ ID NO: 1-2, SEQ ID NO: 3-4, SEQ ID NO: 5-6, SEQ ID NO: 7-8, SEQ ID NO: 9-10, SEQ ID NO: 11-12, SEQ ID NO: 13-14, SEQ ID NO: 15-16, SEQ ID NO: 17-18, SEQ ID NO: 19-20 y SEQ ID NO: 21-22, donde dicho par de cebadores amplifica en un único amplicón un gen HLA completo junto con regiones 5'-UTR y 3'-UTR de dicho gen HLA, una región corriente arriba y una región corriente abajo de dicho gen HLA, de ambos alelos de al menos un locus del gen HLA, donde dicho gen HLA está seleccionado del grupo que consiste en HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5, HLA-DPB1, HLA-DQB1, HLA-DPA1 y HLA-DQA1,

(b) una ADN polimerasa e

(c) instrucciones de uso.

8. El kit según la reivindicación 7, caracterizado por que comprende adicionalmente reactivos para amplificación y secuenciación.

9. El kit según la reivindicación 7 u 8 para uso en la identificación genotípica de ambos alelos de al menos un locus de un gen HLA, donde dicho gen HLA está seleccionado del grupo que consiste en HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5, HLA-DPB1, HLA-DQB1, HLA-DPA1 y HLA-DQA1.

10. Un producto informático que comprende un medio legible por ordenador, en el que hay codificadas instrucciones para controlar un sistema informático que comprende los medios siguientes para llevar a cabo una identificación genotípica *in vitro* de ambos alelos de al menos un locus de un gen HLA en muestras de ADN genómico de un sujeto, donde dicha identificación genotípica comprende amplificar por PCR de largo alcance dichas muestras de ADN genómico utilizando al menos un par de cebadores sentido y antisentido, para cada muestra, donde dicho par de cebadores, sentido y antisentido, está

identificado por las secuencias seleccionadas del grupo que consiste en SEQ ID NO: 1-2, SEQ ID NO: 3-4, SEQ ID NO: 5-6, SEQ ID NO: 7-8, SEQ ID NO: 9-10, SEQ ID NO: 11-12, SEQ ID NO: 13-14, SEQ ID NO: 15-16, SEQ ID NO: 17-18, SEQ ID NO: 19-20 y SEQ ID NO: 21-22; donde dicho par de cebadores amplifica en un único amplicón al menos un gen HLA completo junto con regiones 5'-UTR y 3'-UTR de dicho gen HLA, una región corriente arriba y una región corriente abajo de dicho gen HLA, de ambos alelos, de al menos un locus de un gen HLA, formando así amplicones, donde dicho gen HLA está seleccionado del grupo que consiste en HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5, HLA-DPB1, HLA-DQB1, HLA-DPA1 y HLA-DQA1, combinar los amplicones de cada una de las muestras, fragmentar los amplicones, obteniendo así fragmentos y etiquetar dichos fragmentos, formando así librerías de fragmentos etiquetados, combinar dichas librerías de fragmentos etiquetados, secuenciar dichas librerías de fragmentos etiquetados por secuenciación masiva, generando así lecturas de secuencias que solapan parcialmente y donde dicho sistema comprende los medios siguientes:

- (a) medios para llevar a cabo un alineamiento de dichas lecturas de secuencias contra un genoma de referencia,
- (c) medios para identificar variantes en base a una llamada de variantes,
- (d) medios para determinar bloques haplotípicos en base al alineamiento de las lecturas de secuencias contra un genoma de referencia y medios para añadir regiones flanqueantes de dichos bloques haplotípicos a dichos bloques haplotípicos,
- (e) medios para alinear localmente las lecturas de secuencias de los bloques haplotípicos obtenidos en la etapa anterior contra todos los alelos de una base de datos de referencia de secuencias genómicas que codifican genes HLA para cada gen HLA, obteniendo así unos primeros alelos candidatos,
- (f) medios para pre-calcular las distancias genéticas y determinar la agrupación filogenética de las familias de HLA a diferentes niveles en la base de datos de referencia de secuencias genómicas que codifican genes HLA, obteniendo así una matriz de distancia que contiene los valores de distancia de cada alelo consigo mismo y con los demás,
- (g) medios para llevar a cabo una selección primaria de dichos primeros alelos candidatos en base al resultado de contraste de homología global de las lecturas de secuencias a los haplotipos, obteniendo así unos segundos alelos candidatos,

(h) medios para llevar a cabo un estudio de filogenia evolutiva de dichos segundos alelos candidatos, obteniendo así un árbol filogenético, medios para acceder a la matriz de distancia obtenida en la etapa (f) y medios para seleccionar alelos HLA de la base de datos de referencia de secuencias genómicas que codifican genes HLA más cercanos filogenéticamente, obteniendo así alelos HLA contrastados,

(i) medios para llevar a cabo un test de contraste con alelos HLA contrastados y

(j) medios para calcular las distancias entre las lecturas de secuencias y los alelos HLA contrastados en cada nivel y medios para asignar cada lectura de secuencias a una familia HLA conocida a diferentes niveles, donde se asigna una lectura de secuencia a una familia HLA conocida a diferentes niveles cuando se obtiene una significancia estadística en dicho test de contraste por encima de un valor de corte definido y donde se asigna una lectura de secuencia como secuencia *de novo* que codifica para un nuevo locus del gen HLA cuando se obtiene una significancia estadística en dicho test de contraste por debajo de un valor de corte definido.

11. El producto informático según la reivindicación 10, caracterizado por que dicha significancia estadística se calcula en base al valor estadístico p.

12. El producto informático según la reivindicación 10 u 11, caracterizado por que dicho valor de corte definido se calcula en base al valor estadístico p y tiene un valor de 10^{-6} .