

MÉTODOS Y DISPOSITIVO PARA IDENTIFICAR Y SELECCIONAR DONANTES CON BASE EN LA DETERMINACIÓN DE VALORES DE COMPATIBILIDAD DEL HLA Y OTRAS VARIABLES BIOLÓGICAS

CAMPO TÉCNICO

[1] La presente invención se encuentra enmarcada en el campo de las invenciones implementadas por computador especialmente adaptadas al campo de la investigación médica con aplicaciones clínicas. En particular, este documento divulga métodos, dispositivos y algoritmos relacionados, que permiten, de manera rápida y eficiente, identificar y seleccionar un donante adecuado a partir del cálculo de valores de compatibilidad del HLA.

ANTECEDENTES

[2] El trasplante de células hematopoyéticas es un tratamiento comúnmente empleado en pacientes que padecen de algunas clases de cáncer, enfermedades inmunes y enfermedades graves de la sangre, tales como leucemia, linfoma, y mieloma múltiple, así como enfermedades genéticas. En este procedimiento, también llamado reemplazo de médula ósea, se infunden en un paciente previamente preparado, células hematopoyéticas obtenidas de médula ósea, sangre de cordón umbilical y sangre periférica movilizada, que pueden provenir del mismo paciente (trasplante autólogo), un gemelo idéntico (trasplante singénico) o de un donante que puede o no estar relacionado (trasplante alogénico).

[3] Una de las complicaciones de un trasplante alogénico, es la denominada enfermedad injerto contra huésped. Esta enfermedad puede ocurrir en cualquier punto luego de realizar el trasplante y se relaciona con el ataque al receptor por parte de las células trasplantadas. La severidad de los síntomas depende del grado de histocompatibilidad entre el donante y el receptor.

[4] Por lo tanto, el éxito de este procedimiento depende de poder tener una identidad genética máxima entre donante y receptor para los genes que codifican las proteínas del complejo mayor de histocompatibilidad, denominadas antígenos leucocitarios humanos (HLA, por sus siglas en inglés). Ahora bien, más del 60% de los pacientes que requieren

este tratamiento no disponen de un donante familiar idéntico, por lo que, en la práctica, es necesario realizar un proceso de selección para encontrar el mejor donante no familiar en un registro de donantes de células hematopoyéticas (o células formadoras de la sangre).

[5] Las búsquedas de donante para pacientes con enfermedades graves de la sangre normalmente se realizan en las bases de datos que contienen las tipificaciones de donantes y en los sitios genéticos de los distintos HLA de diferentes factores, como la compatibilidad entre donante y paciente, que miden el grado de identidad genética en los sitios de histocompatibilidad (loci del antígeno leucocitario humano) y otras variables que determinan el éxito del trasplante.

[6] Los diferentes algoritmos de búsqueda para identificar y seleccionar el mejor donante están sujetos a los avances en la investigación clínica, que con los años se han venido modificando, haciendo que la búsqueda no sea ya una tarea solucionable por inspección humana, sino que requiera de recursos computacionales, que permitan proporcionar resultados de una búsqueda que combina la compatibilidad HLA y otras variables biológicas de los donantes, de forma rápida y eficiente.

[7] En vista de lo anterior, durante la última década se han desarrollado diversas estrategias implementadas por computador para conseguir este objetivo, las cuales abordan desde distintos ángulos este complejo proceso. Por ejemplo, el documento de patente US2013/0132379 divulga un método o sistema para la identificación y selección de una unidad de sangre de cordón umbilical para trasplante, en donde dicho método se basa en la comparación de un conjunto de códigos serológicos y moleculares del loci de HLA y de los posibles valores de los alelos existentes para donante y receptor. Esta configuración estática determina caso a caso el espacio de búsqueda, incluyendo de forma independiente los alelos de baja, media y alta resolución, lo que lleva a realizar un gran número de operaciones, y, por consiguiente, a requisitos informáticos superiores.

[8] Ahora bien, también se han desarrollado estrategias de selección de donantes alogénicos con base en la combinación específica de genes de importancia clínica y HLA. Por ejemplo, en el documento US2016/0034666 se describen metodologías de puntuación y clasificación de donantes de células hematopoyéticas no relacionados (URD) para pacientes con leucemia mieloide aguda (LMA) o síndrome mielodisplásico, que se basa en las combinaciones de genotipos y alelos KIR/HLA. Este documento se centra en la

aplicación exclusiva a tratamiento de LMA, por lo que emplea ciertas interacciones entre HLA y genotipos KIR que otorgan un grado de compatibilidad según estudios efectuados exclusivamente en pacientes con dicha enfermedad.

[9] Otras alternativas para la selección de donantes se apoyan en la comparación de otro tipo de datos, tales como la información del transcriptoma, proteoma o epigenoma, además de la información genética. Por ejemplo, el documento US20110112864A1, describe un sistema automatizado de localización y selección de muestras de sangre de cordón umbilical basado en la compatibilidad de HLA, peso del paciente, número de células nucleadas y número de células hematopoyéticas (CD34+). En particular este documento revela técnicas de predicción cuyo objetivo es establecer etapas y métodos de manipulación, transporte y tiempos requeridos una vez se tiene la compatibilidad entre paciente y donante, siendo dicha propuesta de estadística enfocada a las técnicas posteriores al hallazgo de compatibilidad.

[10] Asimismo, dada la generalizada ausencia de datos de alta resolución de tipificación de HLA, se han planteado estrategias de selección que operan a partir de datos de tipificación de HLA de resolución baja o intermedia. Por ejemplo, el documento US20180225412 proporciona un método basado en la determinación del número de Epítomos HLA predichos indirectamente reconocibles (PIRCHES) para múltiples proteínas específicas de HLA. Este documento se enmarca en la estadística predictiva, pues consiste en el desarrollo y uso de modelos que permiten inferir información faltante, por ejemplo, cuando no hay información de alguno de los genes HLA, se puede generar un valor de compatibilidad estimado usando información de los genes HLA que sí están presentes. Por lo tanto, este documento proporciona medios para determinar un valor de compatibilidad hipotético usado en casos específicos de información faltante.

[11] Además de la literatura patente se encuentran algunas soluciones web, por ejemplo, Stem Cell Software Solutions cuenta con Prometeus, Orpheus y Amadeus, cada uno especializado en la administración de laboratorios en áreas como células hematopoyéticas, inmunogenética o banco de sangre. Estos productos se centran en la promoción y retención de donantes y en otros procesos necesarios relacionados a bancos de donantes y en la búsqueda que se hace de manera exhaustiva.

[12] Otras alternativas, como Optimatch, exigen altos recursos computacionales, puesto que los alelos descubiertos son gestionados usando una base de datos IBM Informix Dynamic Server V9.40, y además requiere de web servers y una interfaz de usuario de .NET framework. Por lo tanto, estas alternativas implican condiciones y requerimientos técnicos que no son fáciles de cumplir.

[13] Match-Me es una herramienta web creada por Daniel Boswell que proporciona la posibilidad de buscar usando dos fuentes de información (bases de datos), no obstante, no se contempla la opción de escalar la búsqueda.

[14] Muchas de las tecnologías actualmente disponibles se basan en la información obtenida de acuerdo con protocolos diseñados y aplicados en el extranjero para la recolección y almacenamiento, y requieren en muchos casos de equipos de alta tecnología, lo cual limita su aplicación en el ámbito nacional.

[15] Así, a pesar de la existencia de diversas metodologías computacionales diseñadas para la selección de donantes para el trasplante de células hematopoyéticas, persiste la necesidad de proporcionar alternativas que permitan realizar la búsqueda de donantes locales de alta compatibilidad que puedan ejecutarse en un ambiente de portabilidad, y sin requerir mayores especificaciones técnicas en el equipo ejecutor.

DESCRIPCIÓN GENERAL DE LA INVENCION

[16] En vista de lo anterior, la presente solicitud proporciona un método para calcular un valor de compatibilidad de HLA, para seleccionar un donante adecuado a partir de un registro de donantes de células, con base en este y otros parámetros biológicos. El método de la invención permite optimizar la cantidad de comparaciones empleando una búsqueda segmentada, es decir, una búsqueda en la que la cantidad de comparaciones se optimiza, y en donde la presencia de homocigotos se emplea como una característica aprovechable de optimización.

[17] En particular, el proceso de la invención permite el manejo de la información de forma que se reduzca la cantidad de operaciones lógicas de igualdad, necesarias para el cálculo

de la compatibilidad, tomando como característica aprovechable la eventual repetición de alelos (homocigosis).

[18] Asimismo, la solicitud divulga un método en el que se tiene en cuenta la igualdad o el emparejamiento en el polimorfismo del péptido líder del HLA-B para la selección de un donante.

[19] En particular, el método de la invención se ve robustecido con estructuras de almacenamiento de información que permiten escalabilidad y adaptabilidad ante el descubrimiento de nuevos valores de alelos.

[20] En segundo lugar, la solicitud divulga un dispositivo legible por computador que almacena instrucciones ejecutables por computador para llevar a cabo el método propuesto.

[21] Convenientemente, la invención proporciona medios que permiten llevar a cabo el proceso de búsqueda y selección en un ambiente de portabilidad, pues su ejecución se enmarca en un medio no transitorio y legible por computador que consigue una solución Stand-Alone, y logra replicabilidad en diversos entornos, sin requerir mayores especificaciones del equipo de cómputo lector.

Breve descripción de las figuras

[22] La figura 1 muestra un esquema que ilustra las etapas del método de acuerdo con la invención.

[23] La figura 2 enseña la estructura matricial empleada en el método de la invención.

[24] La figura 3 enseña los resultados de 127 búsquedas de compatibilidad realizadas entre 1.400 unidades de células de sangre de cordón umbilical mediante el método de la invención, y su clasificación por la compatibilidad HLA receptor-donante.

Definiciones

[25] Como se usa en la presente solicitud, el término “*alogénico*” se refiere al trasplante de células hematopoyéticas en donde se proporcionan células obtenidas de la médula ósea, sangre periférica y sangre de cordón umbilical de un donante que puede o no estar emparentado con el paciente que necesita dicho tratamiento.

[26] En el contexto de la presente invención, las siglas EICH y GVHD se usan indistintamente para referirse a la enfermedad injerto contra huésped o *graft vs host disease*.

[27] En la presente solicitud, el término “*HLA (human leukocyte antigen)*” se refiere al conjunto de genes que codifican para las proteínas del complejo mayor de histocompatibilidad en humanos, también denominado “*MHC, major histocompatibility complex*”, las cuales determinan el reconocimiento y la respuesta inmune humana. Pueden ser de clase I o de clase II dependiente de si presentan péptidos derivados de antígenos endógenos o derivados de péptidos exógenos. Estas proteínas incluyen, pero no se limitan a: HLA-A, HLA-B, HLA-C (clase I), HLA-DR β 1, HLA-DQ β 1, HLA-DPA1 y HLA-DPB1 (clase II).

[28] Como se usa en la presente solicitud, el término “*alelo*” se refiere a las diferentes versiones o secuencia de nucleótidos únicas que existen de un mismo gen o locus genético. Un individuo hereda dos alelos para cada gen, uno proveniente de la madre y otro proveniente del padre. En particular, si bien existe un elevado número de alelos para cada uno de los genes que codifican las moléculas del HLA, un individuo puede tener como máximo un par de alelos diferentes para un mismo gen. La combinación de alelos determina el genotipo.

[29] El término “*heterocigoto*” se refiere a un individuo que presenta alelos diferentes para un gen determinado en ambos cromosomas homólogos. Por su parte, el término “*homocigoto*” se refiere a un individuo que presenta alelos iguales o idénticos para un gen determinado en ambos cromosomas homólogos.

[30] El término “*resolución*” o “*resolución alélica*” se refiere a la cantidad de información de tipificación de los alelos del HLA de un donante particular que se encuentra disponible, la cual depende de las técnicas empleadas (métodos serológicos o métodos moleculares,

como la secuenciación). De acuerdo con la Nomenclatura de Factores del Sistema HLA, a cada alelo del HLA se le asigna un número único compuesto de por lo menos cuatro conjuntos de dígitos separados por dos puntos (:). Los alelos cuyos números difieren en los dos primeros conjuntos de dígitos son aquellos que difieren en una o más sustituciones de nucleótidos que cambian la secuencia de aminoácidos de la proteína codificada. Los alelos que difieren solo por sustituciones de nucleótidos sinónimas (también llamadas sustituciones silenciosas o no codificantes) dentro de la secuencia codificante, se distinguen por el uso del tercer conjunto de dígitos. Los alelos que solo se diferencian por polimorfismos de secuencia en los intrones, o en las regiones no traducidas 5' o 3' que flanquean los exones e intrones, se distinguen por el uso del cuarto conjunto de dígitos. Así, se habla de baja resolución (de dos dígitos) cuando solo se han identificado los dos primeros conjuntos de dígitos y de alta resolución (de 4 dígitos) cuando se tiene información de los cuatro conjuntos de dígitos. De esta forma, cuando se habla de una “*baja resolución*” se entiende que se ha identificado el grupo de alelos HLA en cualquier gen o locus dado, más no se conoce la secuencia de la proteína específica encontrada. Se entiende que la compatibilidad entre receptor y donante para un gen particular debe determinarse a la misma resolución.

[31] Como se usa en la presente solicitud, el término “*péptido líder*” se refiere al péptido codificado por la secuencia del primer exón en los genes de HLA, la cual dirige las proteínas sintetizadas al retículo endoplasmático para modificaciones posteriores a su producción. El término “*péptido líder del HLA-B*” se refiere al péptido codificado por la secuencia del primer exón del gen que codifica para el HLA-B, mientras que la mayoría de los alelos del HLA-B se determinan a partir de la secuencia del exón 2 o 3. Se ha encontrado que el péptido líder cumple un papel fundamental tanto en las vías inmunitarias adaptativas, mediante la estimulación de células T, como en las vías inmunitarias innatas, en donde influye en el reconocimiento de las células natural killer (células NK), por lo que, clínicamente, la compatibilidad del péptido líder del HLA-B puede ser incluso más determinante que la compatibilidad del HLA-B. El péptido líder del HLA-B se distingue de los líderes de otros genes clásicos de clase I debido a un dimorfismo en la posición –21 que conduce a péptidos con metionina (M) o treonina (T) en la segunda posición del péptido líder, los cuales influyen de manera diferencial en las respuestas de células NK y células T. De esta forma, si bien existen aproximadamente 8,756 alelos reportados para el HLA-B, que dan lugar a un

elevado número de combinaciones o genotipos, el dimorfismo del péptido líder da lugar a solo 3 posibles genotipos (TT, MT o MM). La identidad del péptido líder puede determinarse a partir de la información disponible para los alelos del HLA-B en baja resolución.

[32] Como se usa en la presente solicitud, el término “*valores o datos de entrada de los donantes*” hace referencia al conjunto de datos que contiene, entre otros, los datos genéticos o de tipificación del HLA de los donantes o potenciales donantes, y que puede incluir el genotipo del péptido líder del HLA-B. El conjunto comprende la identidad de los alelos de un donante para cada uno de los genes o loci analizados, de manera que los datos de cada gen se expresan en forma de pares de datos genéticos o pares de información genética. Así, para un gen homocigoto se tiene un par de datos iguales o repetidos, mientras que para un gen heterocigoto se tiene un par de datos de entrada diferentes. Los valores de entrada del donante también pueden incluir datos de celularidad y de disponibilidad. Los valores o datos de entrada de los donantes pueden provenir de una base de datos de entrada, por ejemplo, un registro de donantes de células de sangre de cordón umbilical. Cuando el genotipo del péptido líder no está incluido, dicha información puede determinarse a partir de la información disponible para los alelos del HLA-B en baja resolución.

[33] Como se usa en la presente solicitud, el término “*valores o datos de entrada del receptor*” hace referencia al conjunto de datos que contiene, entre otros, los datos genéticos o de tipificación del HLA, y que puede incluir el genotipo del péptido líder del HLA-B del receptor. Al igual que para los valores de entrada del donante, el conjunto comprende la identidad de los alelos identificados en un receptor expresados en forma de pares de datos genético o pares de información genética, para cada uno de los genes o loci analizados, de manera tal que el conjunto comprende tantos pares de datos como genes sean analizados.

[34] Tal como se usa en la presente invención, el término “*celularidad*” se refiere a la cantidad o proporción de células productoras de sangre en una muestra determinada.

[35] Como se usa en la presente solicitud, el término “*valor de compatibilidad de HLA*” hace referencia al resultado obtenido de la comparación de todos los alelos entre el donante y el receptor. El valor se expresa en términos de una fracción, cuyo numerador corresponde al número de alelos compatibles entre donante y receptor y el denominador corresponde al número loci comparados. Así, un donante en donde todos los alelos sean compatibles con el receptor para 3 genes analizados tendrá un valor calculado de compatibilidad de 6/6.

[36] Como se usa en la presente solicitud, el término “*matriz*” o “*estructura matricial*” se refiere a un arreglo de datos en un esquema de filas y columnas.

[37] En el contexto de la presente solicitud, un “*medio no transitorio legible por computador*” debe entenderse como cualquier medio tangible que pueda contener o almacenar un programa informático (algoritmos) para su uso en, o en conexión con, un sistema de cómputo. De este modo, este término se emplea en el presente documento para referirse a todos aquellos aparatos o dispositivos que puedan realizar la ejecución de instrucciones mediante uno o más procesadores.

[38] Como se usa en la presente solicitud, el término “*comparación segmentada*” se refiere a una comparación en donde se analiza una correspondencia en un determinado segmento del espacio de búsqueda entre donantes y receptor. Por ejemplo, los alelos del gen HLA-A del receptor se buscarán en el segmento que corresponde al gen A dentro de la matriz del registro de donantes de células productoras de sangre, posteriormente, la información de los siguientes genes se buscará en los segmentos correspondientes.

[39] Como se usa en la presente solicitud, el término “*comparación simplificada*” se refiere a la operación lógica de igualdad realizada solamente en uno de los dos alelos de un gen cuando hay homocigosis en el receptor o en el donante para un determinado gen, implicando 2 o 3 combinaciones de dicha operación. De esta manera, la comparación simplificada implica 2 combinaciones de dicha operación en caso de que se haya detectado homocigosis tanto en el donante como en el receptor, o 3 combinaciones de dicha operación en caso de que se haya detectado heterocigosis en el donante o en el receptor.

[40] Como se usa en la presente solicitud, el término “*comparación exhaustiva*” se refiere a la operación lógica de igualdad ejecutada para cada alelo cuando hay heterocigosis para

un determinado gen, tanto en el receptor como en el donante, e, implicando 4 combinaciones de dicha operación.

[41] Como se usa en la presente solicitud, el término “inicializar el arreglo de compatibilidad con un valor en cero, hace referencia a iniciar con un arreglo en donde todos los valores son iguales a cero.

[42] Como se usa en la presente solicitud, el término “*arreglo de compatibilidad*” se refiere a una estructura computacional tipo vector de valores enteros cuyos campos almacenan el valor de compatibilidad entre el paciente y un posible donante.

[43] Como se usa en la presente solicitud, el término “*arreglo de resultados*” se refiere a una estructura computacional tipo vector de valores *string* o conjunto de caracteres cuyos campos almacenan la información de genes, celularidad y demás valores del donante concatenados con el valor de compatibilidad obtenido del arreglo de compatibilidad.

[44] Como se usa en la presente solicitud, el término “*vectorización*”, se refiere la utilización de funciones y operaciones equivalentes para eliminar bucles y mejorar el rendimiento.

[45] Como se usa en la presente solicitud, el término “FOR” se refiere a un ciclo que permite repetir una serie de pasos, un determinado número de ocasiones.

DESCRIPCIÓN DETALLADA DE LA INVENCION

Métodos de acuerdo con la invención

[46] En una primera realización, ilustrada en la Figura 1, la invención se refiere a un método para seleccionar un donante a partir de un valor calculado de compatibilidad de HLA. En donde el método de la invención comprende las etapas de:

- a. recibir los valores de entrada de donantes en una estructura matricial;
- b. seleccionar un modo para realizar el análisis;
- c. ingresar los datos del receptor, en forma de pares de información genética para cada loci a analizar;

- d. determinar la presencia de homocigotos y heterocigotos para cada loci;
- e. inicializar el arreglo de compatibilidad con valor cero;
- f. iniciar el ciclo FOR, comparando la información encontrada en las filas de la estructura matricial (donantes) y el arreglo de compatibilidad;
- g. ejecutar las comparaciones de acuerdo con la presencia de homocigotos y heterocigotos:
 - g1.realizar una comparación simplificada en caso de homocigosis, en caso de igualdad incrementar en uno (1) el campo actual del arreglo de compatibilidad o
 - g2.realizar una comparación exhaustiva en caso de heterocigosis, en caso de igualdad incrementar en uno (1) el valor entero en el campo correspondiente en el arreglo de compatibilidad;
- h. finalizar el ciclo FOR;
- i. ordenar los resultados del arreglo de compatibilidad de acuerdo con los valores de celularidad asociados a cada fila del registro de donantes;
- j. proporcionar los resultados al usuario.

[47] Como se describe en el presente documento, el método para identificar y seleccionar un donante comprende la comparación entre el valor del receptor y cada fila de los valores de los donantes para cada uno de los pares genéticos. Dicha comparación consiste en la operación lógica de igualdad, cuyo resultado verdadero incrementará en uno el valor entero de la fila evaluada en el arreglo de compatibilidad.

[48] De esta manera, una vez se hayan ejecutado todas las operaciones lógicas de igualdad en todas las filas, se genera un nuevo arreglo de resultados consecuencia de la operación “es mayor a” entre los valores del arreglo de compatibilidad. Para determinar que el donante es adecuado, el valor deberá ser superior a 4, teniendo en cuenta que se evalúan como mínimo 3 genes, que corresponden a 6 posibles alelos. De esta forma, este parámetro establece la visualización de las compatibilidades superiores a 4 alelos entre la cantidad de genes escogidos en el modo, por ejemplo, en el caso de que se evalúen 3 genes, se muestran todos los donantes que tengan compatibilidad 6/6, 5/6 y 4/6.

[49] De acuerdo con el método de la invención, las comparaciones en la dimensión de columnas se hacen de modo segmentado según los pares de datos genéticos mientras que

en la dimensión de filas se hace de manera exhaustiva mediante un FOR o con posibilidad de optimización empleando vectorización cuando hay homocigosis.

[50] Por último, se genera un documento de texto que contiene los datos de los donantes ordenados en función del parámetro de celularidad, empleando el arreglo de resultados. En caso de no encontrar donantes compatibles, el documento obtenido lo indica. Mientras que, en caso de existir compatibilidad adecuada, el archivo incluye la información pertinente, es decir, el identificador (ID) del donante, los alelos que posee, el valor de celularidad, la disponibilidad y el valor de compatibilidad, por ejemplo 6/6 si selecciona como modo la evaluación de 3 genes. El archivo también contiene las unidades que tienen resultados negativos o con compatibilidades menores a la óptima pero dentro del rango aceptado, tales como 5/5 y 4/6.

Datos de entrada de los donantes

[51] De acuerdo con método de la invención, la estructura matricial de la etapa a), tiene un tamaño que depende de la cantidad de donantes. De esta forma, en una realización preferida mostrada en la Figura 2, la estructura matricial posee entre 10 y 20 columnas, preferiblemente 14 columnas. En particular, las columnas de la estructura matricial empleada en el método de la invención se relacionan con la información para las variables de cada uno de los donantes, obtenida a partir de la base de datos de entrada. En modalidades preferidas, las columnas comprenden datos de identificación (ID) de los donantes; datos relacionados con la información genética, que consiste en la información de cada uno de los alelos de los pares genéticos o el genotipo del péptido líder del HLA-B; Además, las columnas comprenden datos relacionados con otras variables seleccionadas del grupo que incluye, pero no se limita a, celularidad, otras secuencias de péptidos de genes de histocompatibilidad, disponibilidad de la unidad, entre otras. La estructura matricial empleada en el método de la invención cuenta con tantas filas como donantes existan en la base de datos de entrada.

[52] En realizaciones particulares del método de la invención, la información de los donantes proviene del registro de donantes de células del banco público colombiano de sangre de cordón umbilical. En otras modalidades, los datos provienen de otros registros.

[53] Los datos del donante se organizan en forma de pares de información genética. En particular, el método de la invención requiere entre 3 y 8 pares de información genética. En modalidades preferidas, la información genética de entrada incluye, pero no se limita a: HLA-A, HLA-B, HLA-C, HLA-DR β 1, HLA-DQ β 1, HLA-DPA1 y HLA-DPB1, genotipo del péptido líder del HLA-B y celularidad. En donde los valores para cada uno de dichos pares son de tipo alfanumérico de longitud entre 2 y 10. No obstante, la longitud de los datos puede incrementarse, por ejemplo, cuando aumente la resolución de la tipificación del HLA.

[54] De acuerdo con la invención, la estructura de almacenamiento que contiene los datos del donante se implementa en un lenguaje de programación apropiado que cuente con estructuras matriciales y de arreglos. En particular, el lenguaje de programación empleado para la implementación de la invención se podrá seleccionar del grupo que consiste en phyton, java o similares.

[55] En modalidades particulares, el dispositivo informático de la etapa a) se selecciona de un grupo que incluye, pero no se limita a, un equipo de cómputo o una memoria de almacenamiento externo. En aún otras modalidades de la invención, los datos de entrada de los donantes se obtienen mediante la transmisión de archivos con la información fuente del registro de donantes de células. En modalidades preferidas de la invención la estructura matricial puede estar almacenada en archivos tipo csv, texto plano, ORC, Avro, Parquet, Pckle, Feather, o similares. De manera particularmente preferida, la estructura matricial de la invención se almacena en un archivo tipo csv. En modalidades preferidas de la invención, la información de los donantes corresponde a datos tipo char.

[56] En realizaciones particulares de la invención, los datos de los receptores pueden ser ingresados mediante una operación de lectura de teclado de un segundo equipo de cómputo. En otras realizaciones, los datos se ingresan mediante un archivo de texto plano. En particular, el método de la invención contempla el ingreso de los datos del receptor por medio del envío de mensajes de texto desde y hacia un dispositivo periférico, tal como un teléfono móvil. En modalidades preferidas, el usuario recibe las instrucciones necesarias para ingresar, de manera secuencial, la información de cada uno de los pares genéticos del receptor mediante mensajes de texto enviados a dicho dispositivo periférico. En aún otras modalidades, la información del receptor está precargada en un archivo, por ejemplo, un archivo de texto plano.

[57] Tal como se describió previamente, las estructuras de almacenamiento de información empleadas en la presente invención permiten escalabilidad y adaptabilidad ante el descubrimiento de nuevos valores de alelos, por ejemplo, cuando aumenta la resolución de la tipificación del HLA. De esta forma, en una modalidad particularmente preferida, el método de la invención se caracteriza por tomar información organizada en un archivo csv (*comma separated value*) o similares, y construir una matriz que cuenta con segmentos verticales de almacenamiento que contribuyen a reconocer la semántica de los datos.

[58] Además, en el método de acuerdo con la presente invención, la asignación de la longitud de cada segmento vertical es dinámica y depende de la longitud presente en las columnas del archivo csv o cualquier otro tipo de archivo que incluya la información fuente del registro de donantes de células. Si bien los segmentos verticales tienen longitud dinámica en cualquier caso, el método aquí descrito identifica la posición inicial de cada segmento haciendo posible su adaptación en caso de variación en la información del registro de donantes de células, por ejemplo, cuando se descubren y agregan alelos de mayor resolución, o nuevas secuencias peptídicas de interés.

Modos del método de la invención

[59] Ventajosamente, el método de la invención puede desarrollarse empleando distintos modos. De manera preferida, dichos modos establecen la cantidad de genes involucrados en el cálculo de compatibilidad. En realizaciones particularmente preferidas, el método de la invención permite seleccionar modos en los que se compararan 3, 4, 5, 6 o 7 genes del HLA, preferiblemente priorizados según sea su relevancia clínica para la compatibilidad. De esta forma, en una realización particular en la que el modo seleccionado se relacione con tres genes, estos serán: los genes HLA-A, HLA-B y HLA-DR β 1. De otro lado, cuando el modo seleccionado requiera cuatro genes, se tratará de: HLA-A, HLA-B, HLA-C y HLA-DR β 1. Asimismo, en las realizaciones en las que el modo requiera de 5 genes, estos serán: HLA-A, HLA-B, HLA-C, HLA-DR β 1 y HLA-DQ β 1. Realizaciones adicionales en las que el modo requiera de 7 genes, estos serán: HLA-A, HLA-B, HLA-C, HLA-DR β 1, HLA-DQ β 1, HLA-DPA1 y HLA-DPB1.

[60] La invención también proporciona un modo en el que la comparación relativa al gen HLA-B entre donante y receptor se hace también con respecto al genotipo del péptido líder.

En particular, este modo se caracteriza por buscar la compatibilidad de acuerdo con la igualdad en el genotipo del péptido líder del gen HLA-B. La comparación relativa al genotipo del péptido líder del HLA-B es especialmente ventajosa cuando no se detecta compatibilidad para el HLA-B.

[61] Adicionalmente, el modo también se relaciona con la selección de la resolución (alta o baja) en la que se lleva a cabo la búsqueda, es decir, si esta se realiza teniendo en cuenta 2 o 4 dígitos por cada uno de los alelos. Por lo tanto, el modo se relaciona con la combinación de varios parámetros: número y tipo de pares genéticos y resolución. De esta forma, la invención contempla distintos modos relativos a las diferentes combinaciones entre número de genes y resoluciones disponibles. Por ejemplo, un modo preferido del método de la invención implica la búsqueda de tres genes en baja resolución. En otra realización, el modo implica la búsqueda de tres genes en alta resolución. De esta forma, la invención contempla distintos modos relativos a las diferentes combinaciones entre número de genes y resoluciones disponibles.

[62] En modalidades preferidas, el modo seleccionado es tal que la búsqueda de los genes de HLA de clase I se realizan en baja resolución (2 dígitos), mientras que las búsquedas de los genes de HLA de clase II se realiza en alta resolución (4 dígitos). Otro modo de la invención implica el análisis de la compatibilidad del HLA-A en baja resolución (2 dígitos), el análisis de la compatibilidad del péptido líder del HLA-B se realiza con base en la variante (M o T) del péptido líder, y la búsqueda de un gen de clase II en alta resolución (4 dígitos).

[63] La selección del modo la realiza preferencialmente el usuario, mediante mensajes de texto enviado a un dispositivo móvil o cualquier otro tipo de herramienta que permita la interacción entre el usuario y el dispositivo que almacena y ejecuta las instrucciones del método de la invención. Alternativamente, la selección del modo se hace automáticamente por el sistema, según la resolución de la información disponible en la base de datos de entrada.

[64] Tal como se describe, el método de la presente invención puede emplearse para seleccionar unidades de un registro de donantes de células de un banco de sangre de cordón umbilical o de donantes adultos de médula ósea o sangre periférica. No obstante,

la selección del modo particular dependerá de la información clínica disponible para cada una de estas opciones.

[65] En general, la selección del modo depende de la información de entrada del registro de donantes de células, y de la de los pacientes, que por motivos técnicos y, en cierta medida, disponibilidad de recursos económicos, puede encontrarse por completo en baja resolución o por completo en alta resolución.

Ventajas del método de la invención

[66] Convenientemente, el método de la invención permite optimizar la cantidad de comparaciones en una búsqueda segmentada o en la que los pares de genes deben corresponder según su tipo y en donde la homocigosis se emplea como una característica aprovechable de optimización, de manera que una disminución en la cantidad de comparaciones disminuye el tiempo necesario para obtener un valor de compatibilidad. Así, la búsqueda segmentada empleada en la presente invención implica que hay una correspondencia en el espacio de búsqueda entre donantes y receptor. Por ejemplo, los alelos del gen HLA-A del receptor se buscarán en el espacio que corresponde al gen A en la matriz del registro de donantes de células, posteriormente, la información de los siguientes genes se buscará en los espacios segmentados correspondientes, lo que evita hacer comparaciones innecesarias.

[67] De esta manera, el método de la invención permite el manejo de la información de forma que se reducen la cantidad de operaciones lógicas de igualdad, ya que toma como característica aprovechable la homocigosis en donantes y/o receptores, al resumir la información de los dos progenitores a un solo valor repetido, en lugar de dos valores distintos por progenitor.

[68] Particularmente, debido a la integración del modo péptido líder, en el cual los datos relacionados con la compatibilidad de acuerdo con la presencia de M o T en el péptido líder del HLA-B, el método de la invención optimiza las búsquedas en el loci del HLA-B con un sistema que mejora las opciones de conseguir un donante no relacionado y disminuye la posibilidad de que se presenten complicaciones relacionadas con la enfermedad de injerto contra huésped. De esta forma, al disminuir la complejidad de la búsqueda para este loci,

el modo de péptido líder del método de la invención permite incrementar las posibilidades de éxito del proceso.

[69] Asimismo, el método de la invención optimiza las búsquedas de acuerdo con las frecuencias alélicas de la población donante, en este caso adaptado a la población colombiana.

[70] Esto conlleva a una disminución en los tiempos de procesamiento, pues, en general los tiempos requeridos para realizar las búsquedas son inferiores a un segundo, lo cual permite realizar las acciones encaminadas a asegurar la disponibilidad de la unidad de células del donante de manera inmediata, disminuyendo el tiempo para el trasplante. Asimismo, la invención presenta menores requerimientos informáticos, lo cual puede facilitar su implementación en distintos centros prestadores de salud que lo necesiten.

Dispositivos de acuerdo con la invención

[71] En otra modalidad, la invención se refiere a un dispositivo, tal como un medio de almacenamiento legible por computador no transitorio, que almacena las instrucciones ejecutables por computador para llevar a cabo el método de la invención.

[72] En realizaciones preferidas, el medio legible por computador no transitorio puede ser una computadora, una CPU (procesador), un dispositivo móvil, un PDA, una tableta, un portátil o similares. En particular, el dispositivo de la invención lleva a cabo el método reclamado en un ambiente de portabilidad, ya que su ejecución se enmarca en un medio no transitorio y legible por computador logrando así una solución autónoma que logra reproducibilidad en diversos entornos sin requerir mayores especificaciones del equipo de cómputo lector.

[73] En donde la identificación, la selección y la determinación del método se puede realizar en un computador que puede ejecutar un programa informático (algoritmos) almacenado en el medio legible por computador no transitorio y los resultados pueden introducirse en el dispositivo o computador que realiza otras etapas del presente método.

EJEMPLOS

[74] Partiendo de la información encontrada para aproximadamente 1.400 unidades de células de sangre de cordón umbilical asociadas al inventario del banco público colombiano de sangre de cordón umbilical, y con el fin de encontrar donantes adecuados para 127 pacientes, se realizaron búsquedas siguiendo el método de la invención.

[75] De esta manera, se llevó a cabo la búsqueda para los alelos HLA-A y HLA-B en baja resolución (dos dígitos), mientras que en HLA- DRβ1 se realizó en alta resolución (considerando 4 dígitos).

[76] En general, cada una de las búsquedas realizadas tomó entre 324,2 y 350 milisegundos. Esto representa una mejora significativa en el tiempo empleado convencionalmente para encontrar un donante, pues anteriormente se requerían cerca de 5 horas de trabajo de personal especializado para la selección de un donante.

[77] El método permitió encontrar donantes compatibles para un 88,2% de los pacientes. Los resultados de compatibilidad positivos de dichas búsquedas se muestran en la tabla 1 a continuación, y se ilustran en la Figura 3:

Compatibilidad	Porcentaje
4/6	55,1
5/6	31,5
6/6	1,6

[78] De esta forma, es posible observar que la invención proporciona medios que permiten identificar y seleccionar, de manera rápida y eficiente donantes de células hematopoyéticas adecuados, partiendo de datos disponibles para poblaciones colombianas.

[79] De otro lado, las búsquedas empleando el modo de péptido líder implican una simplificación del proceso y además, un incremento en las posibilidades de éxito, dado que, por ejemplo, en el caso de la población colombiana se presentan alrededor de 107 alelos del HLA-B (baja resolución), los cuales dan lugar a 5.778 posibles combinaciones de alelos para el HLA-B. Sin embargo, los inventores han encontrado que el 76% de estos 107 alelos del HLA-B presentan un péptido líder con T (treonina) en la posición -21, mientras que 23%

presentan un péptido líder con M (metionina), mientras que solo se desconoce el dimorfismo del péptido líder para un alelo (HLA-B*35:117).

Se deberá entender que las variaciones o modificaciones secundarias hechas al método o dispositivo particular que ha sido descrito en esta solicitud se encuentran dentro del ámbito de aplicación de la presente invención.