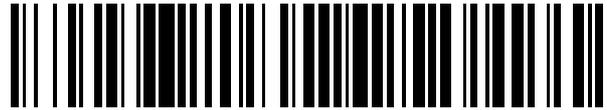


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 573 484**

51 Int. Cl.:

C12Q 1/68 (2006.01)

G01N 33/48 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **02.08.2005 E 05789862 (9)**

97 Fecha y número de publicación de la concesión europea: **02.03.2016 EP 1774323**

54 Título: **Análisis automatizado de patrones de interacción sonda-diana multiplexados: coincidencia de patrones e identificación de alelos**

30 Prioridad:

02.08.2004 US 909638

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

08.06.2016

73 Titular/es:

**BIOARRAY SOLUTIONS LTD. (100.0%)
35 TECHNOLOGY DRIVE, SUITE 100
WARREN, NJ 07059, US**

72 Inventor/es:

**XIONGWU, XIA y
SEUL, MICHAEL**

74 Agente/Representante:

FÚSTER OLAGUIBEL, Gustavo Nicolás

ES 2 573 484 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Análisis automatizado de patrones de interacción sonda-diana multiplexados: coincidencia de patrones e identificación de alelos

5

Antecedentes

Patrones de interacción complejos como marcadores de diagnóstico – Los formatos de ensayo paralelos, que permiten el análisis simultáneo (“multiplexado”) de múltiples loci genéticos en una sola reacción, son muy adecuados para la determinación de configuraciones de diana específicas (“alelos”) encontradas en una muestra dada y para la monitorización de marcadores cuantitativos tales como los niveles de expresión de genes designados o los niveles de biomarcadores de proteína circulantes que se manifiestan por sí mismos en patrones de interacción de receptor-ligando. En lo que sigue, la referencia a interacciones sonda-diana pretende referirse a esta situación más general. Mediante el estudio de la(s) diana(s) con un conjunto seleccionado de sondas de oligonucleótido (véase, por ejemplo, la patente estadounidense n.º 5.837.832, titulada “Arrays of nucleic acid probes on biological chips”) y el análisis de los patrones de interacciones específicas de una o más secuencias diana con ese conjunto de sondas, pueden identificarse rápidamente alelos y combinaciones de alelos.

10

15

20

25

Esta capacidad de diagnóstico probablemente desempeñará un papel cada vez más importante en el estudio de enfermedades complejas tales como artritis, diabetes y cáncer, incluyendo la evaluación de la predisposición a desarrollar una enfermedad que tienen herencia compleja y que requiere la interpretación de todo un conjunto de marcadores moleculares. Sin embargo, el análisis de los resultados (en forma de un patrón de lecturas de intensidad producido en un ensayo multiplexado que refleja la fuerza de la interacción de una o más diana(s) con el conjunto de sondas seleccionado) se enfrenta al reto formidable de interpretar el patrón de interacción mapeándolo para validar combinaciones de alelos o evaluando la predisposición o el riesgo, a la vez que también se determina la fiabilidad y la “exclusividad” de la asignación.

30

35

Un modelo: Tipificación molecular de HLA – El análisis de polimorfismos en el complejo de genes del antígeno leucocitario humano (HLA) proporciona un modelo de la complejidad implicada en analizar la asociación de enfermedades, sirviendo de ese modo para delinear los requisitos que han de abordarse mediante un análisis automatizado rápido y fiable. El complejo de HLA comprende múltiples loci altamente polimórficos que codifican para antígenos variables mediante una respuesta inmunitaria a médula ósea o tejido “extraño”. En la actualidad se han identificado como alelos de clase I 282 HLA-A, 540 HLA-B y 136 HLA-C alelos, y como alelos de clase II 418 HLA-DRB, 24 HLA-DQA1 y 53 HLA-DQB1. Muchas secuencias de alelos conocidos aparecen en bases de datos públicas, por ejemplo, la base de datos IMGT/HLA, www.ebi.ac.uk/imgt/hla/intro.html) para antígenos leucocitarios humanos.

40

45

50

Se han usado ampliamente ensayos de hibridación paralelos (“multiplexados”) de diversos formatos para tipificación molecular de HLA lo que requiere una combinación única de rendimiento y fiabilidad en la identificación de alelos o grupos de alelos asociados con antígenos de clase I y clase II específicos. En el contexto de la tipificación molecular de HLA, las metodologías de ensayo convencionales de la técnica recurren a un formato de “transferencia puntual inversa”. Según este formato, se exponen sondas, situadas en un conjunto de bandas bien separadas, en una tira estrecha de membrana de nailon u otro material de sustrato, a una disolución de diana(s) en condiciones que permiten la captura de la(s) diana(s) para producir, en una etapa de coloración posterior, señales colorimétricas. Otros métodos de la técnica incluyen el uso de sondas presentadas sobre micropartículas codificadas que se suspenden en una disolución diana y se analizan mediante citometría de flujo (véase “Products” <http://www.onelambda.com>). Un método reciente proporciona un entorno de ensayo integrado usando matrices planas de micropartículas codificadas dispuestas sobre chips de silicio (véase, por ejemplo, la solicitud aceptada con n.º de serie 09/690.040, asignada a BioArray Solutions, Ltd.).

55

En la técnica anterior se ha descrito el diseño de formatos de ensayo paralelos para el análisis de loci polimórficos tales como el complejo de HLA, especialmente la selección de conjuntos de pares de cebadores y sondas, así como en varias solicitudes en tramitación junto con la presente (véase, por ejemplo, Concurrent Optimization in Selection of Primer and Capture Probe Sets for Nucleic Acid Analysis”, presentado el 15/7/2004 y cedida a BioArray solutions, Ltd.).

60

Complementariedad de secuencia y representación binaria – La interpretación de patrones de interacción sonda-diana implica la tarea de hacer coincidir una cadena binaria (“patrón de reacción”) derivada de un patrón de intensidad de señales experimental con una (o más) combinaciones de alelos o de establecer la validez de nuevos alelos.

65

Cada alelo tendrá subsecuencias que son perfectamente complementarias, y otras que no son complementarias a un conjunto de sondas construido para estudiar la diana. Esta configuración se representa en la técnica por un código binario que proporciona la base para asignaciones de alelos. Es decir, mediante la asignación a cada sonda con coincidencia perfecta de una puntuación “+” (en el presente documento indicada mediante “8”), y a cada sonda sin coincidencia de una puntuación de “-” (en el presente documento indicada mediante “1”), se construye una

cadena binaria para representar el patrón de interacción del conjunto de sondas elegido con una combinación de alelos específica encontrada. El diccionario que muestra la correspondencia entre alelos y cadenas binarias se conoce en la técnica como la “tabla de resultados positivos”.

5 El patrón de reacción (producido por los conjuntos de sondas seleccionados) puede corresponder a más de una combinación de alelos individual, y el grado de ambigüedad (“degeneración”) determina la precisión (“resolución”) que puede obtenerse en la identificación de combinaciones de alelos. En general, el grado de resolución puede aumentarse mediante la adición de sondas al conjunto.

10 Las intensidades de señal de ensayo reflejan la fuerza de las interacciones sonda-diana. Una sonda ideal produce una señal de ensayo de alta intensidad cuando es perfectamente complementaria (“con coincidencia”) con su subsecuencia diana en una muestra dada y de otro modo produce una intensidad de señal de ensayo de baja intensidad. Es decir, la distribución de intensidades de señal de una sonda de este tipo a lo largo de un gran conjunto de muestras, presentaría de manera ideal dos picos distintos, lo que sugiere una segmentación de intensidades de señal en subpoblaciones que reflejan configuraciones de secuencia de sonda y diana “con coincidencia” o “sin coincidencia”.

15 Sin embargo, en la práctica, la interacción de una o más dianas polimórficas con una multiplicidad de sondas puede producir una amplia variedad de intensidades de señal de ensayo. Por ejemplo, pueden reducirse intensidades de señal de ensayo de otro modo positivas, o pueden potenciarse intensidades de señal de ensayo de otro modo negativas, “extendiendo” de ese modo las distribuciones de intensidades individuales. Por ejemplo, la hibridación sonda-diana se debilita cuando una sonda se encuentra en una subsecuencia diana un alelo que comprende polimorfismos distintos del polimorfismo “designado” de la sonda. A la inversa, una hibridación sonda-diana puede potenciarse inesperadamente cuando una sonda, a la vez que presenta una ausencia de coincidencia significativa con la diana dentro de su subsecuencia designada, coincide con un alelo específico en una subsecuencia no designada.

20 Como con la binarización, generalmente las subpoblaciones se delimitan mediante la selección de un umbral. Particularmente cuando las distribuciones de señales de ensayo no son bimodales, la selección del umbral representa una etapa inicial crítica en el análisis.

25 En el contexto de la tipificación molecular de HLA, el amplio análisis necesario de patrones de interacción y asignación de alelos se basa actualmente en un grado sustancial en la experiencia de los especialistas. Estos especialistas y expertos emplean, habitualmente con un soporte informático mínimo, un proceso que lleva mucho tiempo, difícil y a menudo subjetivo de establecimiento, revisión y refinamiento (“redacción”) de modo interactivo de asignaciones de alelos, a menudo con referencia a compilaciones impresas de alelos conocidos (por ejemplo, la base de datos mantenida por el Programa Nacional de Donantes de Médula Ósea) y “tablas de resultados positivos” correspondientes.

30 Como con la tipificación molecular de antígenos leucocitarios y antígenos eritrocitarios, el análisis y la interpretación fiables y rápidos de patrones de interacción sonda-diana complejos representa un requisito previo para la validación significativa de conjuntos de marcadores genéticos para validar estos “factores de predicción” de predisposición de enfermedad o capacidad de respuesta al tratamiento en poblaciones de pacientes de tamaño suficiente para permitir conclusiones estadísticamente significativas. Surgen retos similares en otras áreas, por ejemplo: en relación con el análisis de polimorfismos genéticos en el análisis de mutación para examen de portador y diagnóstico y evaluación de riesgo asociado; y en relación con la evaluación de la predisposición a adquirir enfermedades genéticas de herencia compleja que puede manifestarse por sí misma en forma de un conjunto completo de marcadores polimórficos o perfiles de expresión génica.

35 Sería deseable un sistema de software conveniente que recurra a algoritmos informáticos y procedimientos robustos para el análisis e interpretación automatizados de patrones y que proporcione un entorno integrado para la revisión interactiva y redacción de asignaciones así como para la gestión y la visualización de datos.

40 Pastinen *et al.*, enseña una detección de alelos de dos colores partiendo de la página n mediante la determinación de una intensidad de señal a una primera longitud de onda (CY5 a 670 nm) y una segunda longitud de onda (CY3 a 570 nm). Posteriormente, la primera intensidad de señal CY5 (muestra de prueba) se divide entre la segunda intensidad de señal CY3 (muestra de control). Estos valores ajustados se usan entonces para calcular la proporción de intensidades de señal entre las muestras de prueba y control.

45 **Sumario**

El objeto anterior se logra mediante un método para asignaciones de alelos según la reivindicación 1. Pueden derivarse realizaciones preferidas de las reivindicaciones dependientes respectivas.

60 Se divulgan adicionalmente métodos y algoritmos (y su implementación) que soportan el análisis automatizado y la revisión y la mejora interactivas (“redacción”) del análisis dentro de un entorno de software integrado, para

asignaciones de alelos automatizadas. La implementación, preferiblemente con un sistema de software y un programa denominado el programa de asignación de alelos automatizada (“AAA”), proporciona una multiplicidad de funcionalidades incluyendo: *Gestión de datos* por medio de una interfaz integrada para una base de datos portátil para permitir la visualización, importación, exportación y creación de informes resumen personalizables; *Configuración del sistema* (“ajuste”) incluyendo autorización de usuario, análisis de conjunto de entrenamiento y enmascaramiento de sonda; *Análisis de patrones* incluyendo coincidencia de cadenas e inversión de sonda; y *Redacción interactiva* que combina cálculos de bases de datos en tiempo real y edición de “corta y pega”, generando declaraciones de “advertencia” y anotación de soporte.

10 *Umbralización* - Métodos de selección y refinamiento de umbrales tal como se divulga, incluyendo una generalización de la representación binaria obtenida segregando distribuciones de intensidad de sonda en tres o más subpoblaciones.

15 *Determinación de umbral inicial*- También se divulga un método de ajuste de umbrales por medio del análisis de un conjunto de referencia (“entrenamiento”) y seleccionando, para cada sonda en un conjunto de sondas seleccionado, un umbral que maximiza el grado de concordancia de los resultados de ensayo y los alelos asignados con los proporcionados para el conjunto de entrenamiento. El método de determinación de los ajustes de umbral iniciales también proporciona una cifra de mérito (“bondad”) como método de base de evaluar la robustez de ese umbral: Un método relacionado de determinación de umbral inicial divulgado en el presente documento aplica un algoritmo de binarización a perfiles de intensidad de sonda individuales. *Refinamiento de umbral: Coincidencia de patrones*- Se divulga un método de refinamiento de patrones haciendo coincidir una cadena binaria experimental (“patrón de reacción”), producida mediante la aplicación de ajustes de umbral iniciales, con un compendio de patrones de reacción correspondientes a combinaciones de alelos válidas. El sistema de software en el presente documento soporta un modo de alterar (“invertir”) bits específicos dentro de la cadena (“palabra”) experimental. El programa identifica sondas, y combinaciones de sondas, como candidatas para “inversión” con el fin de producir la concordancia completa o parcial entre la “palabra” experimental modificada y la palabra o palabras más cercanas, en el diccionario. La inversión de una sonda (para determinadas muestras en el conjunto que está considerándose) corresponde a un refinamiento en el ajuste de umbral para esa sonda.

30 *Enmascaramiento de sonda* – También se divulga una característica de programa que soporta un modo de configuración (“ajuste”) en el que sondas seleccionadas pueden excluirse temporalmente del análisis (“emascarse”). También pueden enmascarse señales de ensayo producidas por sondas que no contribuyen significativamente a discriminar entre alelos (o que puede considerarse que producen patrones de intensidad de baja fiabilidad) cuando se analizan los resultados, y entonces visualizarse sólo si su contribución se considera necesaria.

35 *Datos estadísticos de frecuencia de alelos* - En otro aspecto, el sistema de software proporciona un método para realizar un seguimiento y presentar la frecuencia relativa de aparición de grupos de alelos (y combinaciones de los mismos).

40 *“Redacción” interactiva* - El sistema de software proporciona un entorno integrado para facilitar el acceso simultáneo a los datos que están analizándose y a bases de datos y tablas de resultados positivos que están consultándose, por ejemplo en el transcurso de la redacción. Se proporcionan operaciones de “cortar y pegar” en múltiples pantallas para permitir la edición rápida y conveniente de asignaciones automatizadas (“programa”) incluyendo una función de anotación.

45 *Pruebas de confirmación para resolución de ambigüedad* - El programa también alberga información adicional que ayuda a la resolución de ambigüedades por medio de amplificación específica de grupo o por medio del uso de análisis de polimorfismos mediado por elongación (véase “Multiplexed Analysis of Polymorphic Loci by Concurrent Interrogation and Enzyme-Mediated Detection” presentada el 15/10/2002; n.º de serie 10/271,602).

50 *Análisis distribuido: Procesamiento, análisis, interpretación, archivo* - La arquitectura del sistema de software soporta un modo de análisis distribuido, que permite diferentes funciones tales como registro de imágenes de ensayo, análisis automatizado, redacción interactiva y evaluación y “aprobación final” y generación de informes que va a realizarse por diferentes individuos en diferentes ubicaciones geográficas. Este modo de análisis distribuido amplía las capacidades de laboratorios de prueba individuales para ampliar sus menús de pruebas respectivos sin la necesidad de especialización local relacionada con las muchas áreas de especialización distintas. Por ejemplo, pueden elegirse ubicaciones centrales de pruebas para facilitar la recogida de muestras de pacientes, mientras que médicos acreditados pueden revisar y liberar resultados de prueba finales desde una ubicación diferente, a la vez que sirven a múltiples centros de prueba.

60 También se divulga un método y pseudocódigo para el análisis de alelos completamente automatizado, que se expone a continuación.

Breve descripción de los dibujos

65 La figura 1A ilustra un conjunto de intensidades de señal de ensayo registradas para la sonda HA109 en el análisis de un conjunto de entrenamiento de muestras. Mediante un método independiente, la intensidad de sonda

normalizada se puntuó negativa para muestras marcadas “-” y positiva para muestras marcadas “+”.

5 La figura 1B muestra una determinación de umbral para una sonda en un conjunto entrenamiento de sondas, donde el valor de umbral se representa en el eje X, y la medición de umbral se representa sobre el eje Y. El umbral óptimo produce la medición máxima en Y, que es 1 en este caso.

10 La figura 1C muestra los ajustes del sistema para varias sondas de HLA diferentes. La tolerancia de asignación de alelos (véase la figura 2) se introduce en los recuadros de texto. Para HLA-A se permite un máximo de 6 inversiones; para HLA-B 8 inversiones; y HLA-DR 5 inversiones. A cada sonda puede asignarse según se requiera, alta confianza, baja confianza o no usarse. El conjunto de sondas central (véase la figura 3) consiste solo en las sondas de alta confianza, mientras que el conjunto de sondas expandido incluye las sondas de alta y baja confianza. Mediante el cambio de estos ajustes, puede cambiarse interactivamente el conjunto central y el conjunto expandido. Por ejemplo, HA120 puede ajustarse como alta confianza y HA121 como baja confianza.

15 Las figuras 2A a 2C muestran, respectivamente, la intensidad normalizada (“proporción”) para las sondas HB103, HB123A, HB154, ordenadas en orden de proporción creciente para ilustrar una discontinuidad en el perfil de proporción de sondas. HB103 (figura 2A) tiene la mayor diferencia en el perfil de proporción. HB123A (figura 2B) no tiene un salto obvio en el perfil. HB154 (figura 2C) tiene dos saltos en el perfil. En el patrón de reacción, 8 indica positivo, 1 indica negativo (sin señal) y 0 indica que la sonda no se usa.

20 La figura 3 es un ejemplo de asignación de alelos, donde el patrón de reacción se muestra en la primera fila, oscilando entre 0 y 8, y la cadena de hibridación es el patrón mostrado en las columnas. Las columnas 119, 121, 122, 135A, 142A y 145 son sondas de baja confianza. Puesto que sólo hay una asignación sugerida, el conjunto de sondas expandido está vacío.

25 La figura 4 es el patrón de reacción y la tabla de resultados positivos para una reacción a modo de ejemplo entre sondas y una diana, que también muestra una captura de pantalla del programa para realizar redacción manual, asignación de alelos y un lugar para insertar comentarios.

30 La figura 5 es un gráfico de barras para la distribución de la frecuencia de alelos de una población particular.

La figura 6 es un gráfico de barras que muestra la comparación entre estudios de genotipado notificados de una distribución de alelos en una población “judía normal”, y los resultados experimentales para tal población.

35 La figura 7 es una captura de pantalla que ilustra la información resumen de asignaciones para un panel designado “03250443,” e incluye el nombre del panel, el nombre de la muestra, la posición de la muestra, la asignación de alelos, las sondas de inversión, mensaje de advertencia y comentarios.

40 La figura 8A es un perfil de proporción de sondas.

La figura 8B es la derivada numérica que muestra los puntos de inflexión derivados de la figura 8A.

Descripción detallada

45 Tras el registro de un patrón de intensidad de señales de ensayo para una muestra dada, se realiza una secuencia de etapas analíticas para identificar la combinación de alelos correspondiente.

2.1 Conversión de patrones de intensidad de ensayo en cadenas binarias

50 *Intensidades de señal de ensayo normalizadas: Perfiles de intensidad de sonda* – Determinados métodos para sondar polimorfismos dentro de un ácido nucleico diana tales como el análisis de polimorfismos mediado por elongación (eMAP™, también denominado en el presente documento “elongación mediada por captura”), divulgado en la solicitud de patente estadounidense en tramitación junto con la presente con n.º de serie 10/271.602 (PCT/US02/33012) producen señales de ensayo que se basan en un proceso de reconocimiento molecular cuya alta especificidad produce una distribución de señales de ensayo aproximadamente “binaria”. En cambio, métodos tales como el análisis de polimorfismos multiplexado mediado por hibridación (hMAP™, solicitud de patente estadounidense con n.º de serie 10/847.046) producen intensidades de señal de ensayo que reflejan la afinidad eficaz que rige la interacción de cada sonda en un conjunto de tales sondas múltiples con la diana. Para corregir variaciones en el trasfondo, en la concentración de la diana original o en otras condiciones experimentales, se normalizan las interacciones sonda-diana que registran intensidades de señal experimentales usando señales registradas a partir de sondas de control positivas y negativas (y pares sonda-diana) incluidas en la reacción.

60 A partir de cada intensidad de señal, habitualmente se resta el valor medio, I_k , para el tipo k^o de sonda que incluye la señal de control positiva, I_{PC} , y la señal de control negativa, I_{NC} , y el resultado se divide entre la señal de control positiva corregida para obtener una intensidad normalizada (proporción):

$$r = (I - I_{NC}) / (I_{PC} - I_{NC})$$

Para facilitar una evaluación del rendimiento de cualquier sonda dada en el panel, se construye un perfil de intensidad de sonda, también denominado en el presente documento perfil de proporción, ordenando los valores de r registrados para esa sonda a lo largo de un conjunto de muestras, normalmente 100 muestras, por ejemplo, en orden ascendente. Ejemplos de tales perfiles se muestran en las figuras 2A-2C, donde la figura 2A ilustra un perfil que presenta una transición brusca de gran amplitud desde valores de r inferiores hasta superiores, mientras que la figura 2B ilustra un perfil que presenta una transición gradual de amplitud pequeña.

Para métodos tales como hMAP, las intensidades de señal normalizadas se convierten primero en una representación binaria: si r supera un umbral ajustado previamente, T, la puntuación binaria correspondiente es positiva, s = 1 (también indicada en el presente documento como "8"), de lo contrario es negativa, s = -1 (también indicada en el presente documento como "1"). En las subsecciones siguientes se divulgan métodos de implementación de esta etapa crítica de procesamiento de señales.

2.2 Determinación de umbrales: Binarización – Se divulga un algoritmo para la determinación y el refinamiento iterativo de ajustes de umbral de binarización. Al igual que para la etapa análoga en el análisis de imágenes de convertir intensidades de píxeles en escala de grises en una representación "en blanco y negro", la binarización asigna intensidades de señal de ensayo normalizadas a uno de dos subconjuntos. Esto no resulta problemático siempre que la distribución de intensidades de señal normalizadas para un conjunto de muestras en consideración tenga una forma bimodal que presente picos bien separados: entonces puede situarse un umbral casi en cualquier parte entre los dos picos sin afectar al resultado; la figura 2A corresponde a un histograma bimodal. Sin embargo, en otros casos, cuando los picos separados no están claramente resueltos, la binarización presenta una fuente de incertidumbre o error potencial: la asignación de valores de intensidad específicos a uno u otro subconjunto dependerá de manera sensible de la colocación precisa del umbral; la figura 2B corresponde a un caso de este tipo.

Ajustes de umbral iniciales: Análisis de conjuntos de "entrenamiento" – Los ajustes de umbral iniciales pueden basarse en el análisis de una referencia o "conjunto de entrenamiento". Preferiblemente, las muestras de referencia se eligen para reflejar características del grupo de muestras de interés; por ejemplo la frecuencia predominante de aparición de combinaciones de alelos y haplotipos. Tal información puede proporcionar restricciones adicionales sobre asignaciones de alelos probables. A continuación se elaboran métodos de recogida y análisis estadístico automatizados de los datos estadísticos de la población de muestra.

Se analiza un conjunto de referencia ("entrenamiento") de S muestras, con patrón de reacción de referencia determinado y validado de manera independiente $\{\sigma_k, 1 \leq k \leq P_T\}$, y asignaciones de alelos determinadas y validadas de manera independiente, con un conjunto seleccionado de sondas P, para obtener el patrón de intensidad normalizada (proporción) $\{r_k, 1 \leq k \leq P\}$ y, para cada sonda, k, en el conjunto seleccionado (véase también a continuación), se determina un umbral, T_k , para maximizar la concordancia entre el patrón de reacción real, $s_k = s_k(\mathbf{T})$, y el patrón de referencia $\{\sigma_k, 1 \leq k \leq P_T\}$.

Es decir, para cada sonda en el conjunto real, se determina un umbral para cada sonda mediante el análisis del perfil de intensidad normalizado a lo largo del conjunto de entrenamiento de S muestras para maximizar la correlación cruzada $C = \sum_i ((r_i - T_k) * \sigma_i) / \sum_i |(r_i - T_k)|$, $1 \leq i \leq S$. A su vez, para cada sonda, para encontrar el máximo de la función C, el ajuste de umbral, T_k , se aumenta gradualmente hasta que el signo de la cantidad $r_i - T_k$ coincide con el del bit correspondiente, σ_i , en el patrón de referencia. Para las sondas usadas en el ensayo, pero no en el estudio del conjunto de entrenamiento, se "retrocalcula" un patrón de reacción a partir de la tabla de resultados positivos usando los alelos asignados. La figura 1B ilustra la forma de la función $C = C(\mathbf{T})$, $r_{\min} \leq T \leq r_{\max}$. El ajuste de umbral se elige para maximizar la función C.

El pseudocódigo para determinar el ajuste de umbral inicial es el siguiente:

```

/*
** ρ es el patrón de intensidad normalizada ("proporción") para una muestra dada; la binarización convertirá ** cada
patrón de intensidad en un patrón de reacción compuesto por P bits; habrá ** S de tales patrones:
** π es el conjunto de perfiles de sonda; habrá P de tales perfiles, cada uno con un umbral T;
** τ es el conjunto de umbrales óptimos (iniciales), que va a determinarse mediante la maximización de C;
*/
Generar perfiles de sonda (ρ, S, π, P)                                     /* ordenar los patrones de reacción
por sonda */
{
    PARA (cada sonda)
    {
        Extraer Intensidad (de los patrones de reacción)
        Ordenar intensidades de sonda();
    }
}

```

```

}
/*
**maximizar C =  $\sum_i ((r_i - T_k) * \sigma_i) / \sum_i |(r_i - T_k)|$ 
** ri indica los elementos del perfil de proporción R */
5 */
Encontrar umbral (R,  $\Sigma$ , S)
{
    Cmáx = 0;
    T = rmín;
10 DO
    {
        SI ((C =  $\sum_i ((r_i - T) * \sigma_i) / \sum_i |(r_i - T)|$ ) > Cmáx) Cmáx = C;
        T = T +  $\Delta T$ ;
    } MIENTRAS QUE (T <= rmáx);
15 Volver (T);
}

Determinar umbrales iniciales()
{
20 Generar perfiles de sonda ( $\rho$ , S,  $\pi$ , P)
    PARA (cada perfil de proporción P,  $\pi_k$ )
    {
         $\tau_k$  = Encontrar umbral ( $\pi_k$ ,  $\Sigma_k$ , S)
    }
25 Volver ( $\tau$ )
}

```

30 “Bondad” de los ajustes de umbral – Los ajustes de umbral pueden ser robustos para algunas sondas, pero menos robustos para otras sondas en el conjunto. Es decir, cuando la composición de las dos subpoblaciones de muestras, generada por la aplicación de un umbral T al perfil de intensidad de sonda cambia en respuesta a un pequeño cambio en el valor T a T + ΔT o T - ΔT , donde $\Delta T / T \ll 1$, entonces ese umbral no es robusto y la confianza estadística es baja.

35 Para permitir la colocación de intervalos de confianza en ajustes de umbral individuales también se divulga en el presente documento un factor de mérito (“bondad de umbral”). Esto se deriva de la forma del pico de la función C = C(T), producida en el transcurso de maximizar la correlación cruzada con un conjunto de asignaciones para un conjunto de entrenamiento. Cuanto más pronunciado es el pico en la función, más sensible es el valor seleccionado de T para pequeñas variaciones en T, tal como se ilustra en la figura 1B.

40 Una “bondad”, G, de un umbral seleccionado mediante el método de maximización de la correlación cruzada, C, tal como acaba de describirse, se define tal como sigue:

$$G = (C_L + C_R) / 2C_{Máx}$$

45 en la que C_{Máx}, C_L y C_R, respectivamente, indican el valor máximo de C, C_L el valor de C obtenido cuando se disminuye el valor de umbral en un 30% y C_R el valor de C obtenido cuando se aumenta el valor de umbral en un 30%.

50 Pesos – Una vez que se ha determinado un umbral, T, puede reformularse el perfil de intensidad de sonda, {r_i, 1 i S} de forma en escala, w_i = (r_i - T)/T, en la que los pesos, w_i, representan la magnitud relativa de intensidades de señal individuales (normalizadas). El sistema de software descrito en el presente documento (designado “AAA”) realiza un seguimiento de pesos y los presenta en uno de varios formatos, por ejemplo, mediante una simple clasificación en “Cerrado” (“C”, w 0,5) o “Distante” (“D”) en relación con “inversión” de bits, tal como se comenta adicionalmente a continuación.

55 2.3 Coincidencia de cadenas: Correlaciones dentro de patrones de interacción sonda-diana

El análisis de patrones de intensidad experimentales se dirige a identificar, o identificar aproximadamente, el/los alelo(s) diana subyacentes. Para este fin, se binarizan patrones de intensidad mediante la aplicación de un conjunto de umbrales, y la cadena binaria resultante (“patrón de reacción”) se compara con combinaciones de pares de tales cadenas correspondientes a alelos conocidos enumerados en una “tabla de resultados positivos.” Cada entrada en una “tabla de resultados positivos” representa un alelo válido y proporciona una secuencia binaria en la que cada posición contiene una puntuación de “Coincidencia” o “Sin coincidencia” que hace referencia al grado de complementariedad del alelo con la sonda en esa posición (véase la figura 3A). Los alelos se designan mediante

65 códigos de 4 dígitos y se agrupan mediante los 2 dígitos principales en grupos de antígenos.

Corrección de errores mediante coincidencia de cadenas: "Inversiones" – Para identificar los alelos diana, se compara el patrón de reacción binario, $\{S_k, I k P\}$, bit por bit, con todas las cadenas de referencia que representan combinaciones de 2 alelos; estas se generan mediante la aplicación de una operación O en las entradas de la tabla de resultados positivos. La coincidencia de cadenas de bits completas refuerza las correlaciones y permite un método de "corrección de errores" dando la vuelta ("invirtiendo") bits individuales en la cadena según se considere necesario con el fin de producir una coincidencia con una cadena de referencia válida. Este proceso es parecido a comprobar errores topográficos para reducir palabras válidas; cambiando la letra "t" en "valit" por "d", se obtiene una palabra válida en inglés; se obtiene otra palabra válida en inglés cambiando "i" por "u" y "t" por "u", pero "valit" está más "próxima" a "valid" que a "value", y por tanto es más probable que la primera represente la palabra deseada.

De manera análoga, el programa AAA está diseñado para encontrar cadenas de bits válidas más próximas ("palabras") que representen alelos válidos, tal como se identifica, por ejemplo, en una base de datos que puede introducirse en el programa AAA. El programa AAA enumera las cadenas válidas "más próximas", agrupadas mediante códigos de "grupos" de 2 dígitos comunes, en el orden de distancia de Hamming creciente (es decir, el número de bits sin coincidencia) para la cadena experimental. Específicamente, el programa identifica los bits sin coincidencia y sugiere las "inversiones" necesarias, concretamente "de 1 a 8" o "de 8 a 1," lo que produciría una coincidencia completa entre la cadena experimental y las cadenas válidas adicionales dentro de una distancia de Hamming máxima ajustada previamente.

El programa AAA también permite una búsqueda más profunda del espacio de cadenas de referencia con el fin de producir una lista de "casi coincidencias" (véase el menú desplegable ilustrado en la figura 4). Este análisis ampliado frecuentemente revela posibles cadenas alternativas que representan asignaciones de alelos más probables basándose en consideraciones adicionales tales como la frecuencia de aparición de determinados alelos o haplotipos en la población de interés. Esta característica reduce sustancialmente el tiempo y el esfuerzo invertidos en la edición interactiva.

Distancia entre cadenas - Dentro de cada grupo, las cadenas correspondientes a combinaciones de alelos válidas se ordenan en el orden de distancia de Hamming ponderada creciente a partir del patrón de reacción. Esta función de distancia se define en términos de los pesos, $w_i = r_i - T)/T$, asociados con las sondas sin coincidencia. Por ejemplo, suponiendo que hay M sin coincidencia, una posible función de distancia es:

$$X^2 = (1/M) \sum_{\text{sondas sin coincidencia}} W^2$$

2.4 Refinamiento de umbral iterativo

De manera ideal, el procedimiento de coincidencia de cadenas que acaba de describirse producirá una coincidencia inequívoca entre el patrón de reacción y una cadena que representa una combinación de alelos válida. Sin embargo, aun cuando se denomine una coincidencia perfecta, esta denominación puede no ser inequívoca si implica bajos pesos para una o más de las sondas. Es decir, en la práctica, el patrón de reacción puede contener falsos negativos o falsos positivos, dependiendo del ajuste de umbral para sondas individuales y los pesos de intensidades normalizadas. Especialmente la confianza estadística asociada con ajustes de umbral de sondas que tienen perfiles de proporción continuos para el conjunto de muestras que está considerándose será baja, y por tanto resulta beneficioso tener un proceso de ajuste ("ajuste fino") de tales ajustes de umbral.

El procedimiento de coincidencia de cadenas proporciona una base para el refinamiento de los umbrales iniciales. Después de todo, invertir una sonda es equivalente a ajustar el umbral correspondiente para cambiar el signo de la proporción normalizada en relación con el ajuste de umbral. Es decir, si tras un paso inicial de asignación de alelos automatizada para un conjunto de muestras, una determinada sonda se "disminuye" de manera sistemática, o bien como falso positivo o bien como falso negativo, esto es una indicación de que está vigente un refinamiento de umbral para esa sonda. Por consiguiente, la optimización de umbral implica un proceso iterativo de ajustar los ajustes de umbral de una o más sondas "disminuidas" para minimizar el número total de inversiones identificadas por el programa AAA. Dado que este proceso de optimización de umbral se basa en coincidencia de cadenas, en lugar de en la inspección de perfiles de razón de sonda individuales, como en la etapa de ajuste de umbrales iniciales, la optimización de umbral refleja las correlaciones entre múltiples sondas en el conjunto y mejora la confianza estadística. El refinamiento de umbral puede realizarse de manera continua usando muestras analizadas como un conjunto de referencia que se expande de manera continua. En esta aplicación, cada nuevo conjunto de muestras se convierte en un nuevo conjunto de entrenamiento.

En general, el número de inversiones incluso tras la optimización de umbral permanecerá finito. En tales casos, deben tenerse en cuenta los pesos asociados con las inversiones indicadas. El programa AAA designa de manera conveniente proporciones de "inversiones" o bien como "Próxima" (C, r 0,5) al umbral, o bien como "Distante" (D) del umbral. El punto de corte representa un parámetro de rendimiento ajustable que puede ajustarse de manera más o menos conservativa, implicando generalmente un ajuste más conservativo un mayor grado de revisión y edición interactiva, tal como se comenta en mayor detalle a continuación. Una necesidad de invertir sondas "distantes", es decir aquellas que tienen un gran peso, con el fin de obtener una coincidencia representa una indicación de que

puede estar disponible un nuevo alelo.

El pseudocódigo a continuación resume el procedimiento de refinamiento de umbral tal como se implementa en el sistema de software AAA de la invención.

```

5  /* calcular la asignación de alelos para una lista de muestras, entonces
   /* identificar y analizar las sondas invertidas para esa lista

       Calcular asignación (listas de muestras);
10      Obtener inversiones (listas de muestras);
       Analizar inversiones();

   /* Seleccionar sondas que requieren refinamiento de umbral
15      Conjunto de sondas = Seleccionar sondas();

   /* para cada sonda, encontrar el umbral óptimo minimizando el número de inversiones totales
       PARA cada (sonda en el conjunto de sondas)
           {
20                 T0 = Obtener el umbral inicial (sonda);                                /* obtener el
                                                                                          umbral inicial
                                                                                          /* ajustar
                   PARA (T = T0-intervalo; T<=T0+intervalo; T+=delta T)
                   umbral
                   {
25                         Volver a calcular la asignación (listas de muestras);
                                                                                          /* volver a analizar
                                                                                          las sondas invertidas
                                                                                          y obtener el
                                                                                          número de inversiones
                                                                                          totales
30                         Inversiones totales = Analizar inversiones()
                                                                                          /* encontrar inversiones
                                                                                          totales mín.
                   }
                   Encontrar las inversiones totales mín.();

35                 /* el nuevo umbral será el que minimiza las inversiones totales
                   Tnuevo = Obtener nuevo umbral (sonda);
           }

```

A continuación se exponen características adicionales incluidas en el sistema de software AAA.

40 **Modificación de conjuntos de sondas** – Un patrón de reacción binario particular puede coincidir con más de un alelo, y a menudo lo hará si la cadena de bits tiene sólo algunos elementos (“8s” y “1s”) y la diana incluye múltiples regiones polimórficas. El grado de ambigüedad se calcula enumerando simplemente el número de alelos no resueltos (“degenerados”). El alargamiento de la cadena, mediante la inclusión de sondas adicionales, puede proporcionar un medio para lograr la resolución aumentada con el fin de resolver ambigüedades.

50 **Enmascaramiento de sonda: Conjuntos de sondas central y expandido** – A continuación se describe un método para la designación interactiva de conjuntos centrales y conjuntos expandidos de sondas, junto con una característica de “enmascaramiento de sonda”, que puede usarse para corregir señales de aquellas sondas que no funcionan tan bien como otras. En el enmascaramiento de sondas se ignoran los resultados de aquellas sondas que se hibridan con una amplia variedad de muestras, en lugar de sólo con muestras particulares con alelos particulares. Un nivel tan amplio de hibridación puede resultar de hibridación cruzada o de sondas que seleccionan como diana subsecuencias que se expresan ampliamente.

55 El sistema de software AAA proporciona una pantalla de configuración (“ajuste”) que permite al usuario designar sondas dentro de un panel para formar parte de un conjunto central o un conjunto expandido. La función de enmascaramiento de sonda induce a los usuarios a introducir una lista de sondas que han de ignorarse (“enmascaradas”) en el primer paso de la asignación de alelos automatizada (es decir, el programa calcula primero las asignaciones basándose en un conjunto de sondas central que se hibridan más estrechamente. El objetivo de usar el conjunto central es obtener una asignación de nivel de grupo para alelos (es decir, un grupo de varios alelos posibles) que usan sondas que proporcionan discriminación de nivel de grupo con un alto nivel de confianza.

65 En el modo de enmascaramiento de sonda, el programa AAA realiza en primer lugar asignaciones de nivel de grupo usando sólo el conjunto de sondas central. En un segundo paso (opcional), la asignación puede refinarse repitiendo el cálculo con el conjunto extendido que contiene todas las sondas en el conjunto central, así como las sondas menos fiables restantes. El segundo paso producirá asignaciones adicionales que siguen siendo compatibles con las

asignaciones realizadas en el primer paso. El programa también realiza este segundo paso siempre que el primer paso no produzca una asignación de nivel de grupo única.

5 El conjunto extendido es útil para guiar la “redacción” y permite que el usuario seleccione la asignación de alelos más probable. En algunos casos, puede ser necesario generar y usar la versión complementaria (por ejemplo, antisentido) de una o más sondas (y los transcritos y amplicones correspondientes) para evitar hibridación cruzada excesiva. En tales casos, las sondas no complementarias se excluyen entonces de los pasos primero y/o segundo.

10 *Datos estadísticos de la población: Análisis de distribución de la frecuencia de alelos* – La tasa de recurrencia de cada alelo depende de la población sobre la que está realizándose la tipificación de HLA. Para un panel que contiene un gran número de muestras, la aparición de un alelo particular es representativa de su abundancia o excepcionalidad en toda la población de interés. Puede calcularse la distribución de alelos en una población de etnia conocida para un panel.

15 Se usaron los resultados para un conjunto de paneles del programa de base de datos AAA para calcular la frecuencia de alelos. El programa asigna a cada muestra un conjunto de dos alelos basándose en su patrón de reacción a lo largo de todo el conjunto de sondas. Los cálculos de frecuencia se basan en asignaciones de alelos de dos dígitos. Los dos primeros dígitos de las asignaciones de alelos para todas muestras se extraen y se compilan en un único vector. Se calcula un histograma que enumera cada alelo con el número de veces que se encuentra en el panel (su recuento). Se calcula la frecuencia cuando el recuento se normaliza por el número total de asignaciones.

20 Un objetivo de realizar un seguimiento de los datos estadísticos de la frecuencia de alelos, tal como se implementa en el programa, es proporcionar ayudas de edición, tales como indicaciones de advertencia siempre que se identifica un alelo poco frecuente. Esto ayudará en el caso de asignaciones degeneradas, donde es posible más de una única asignación. En tales casos, pueden eliminarse manualmente aquellas asignaciones que implican alelos poco frecuentes. En el caso de asignaciones individuales, una indicación de este tipo de alelos poco frecuentes garantiza que las asignaciones se comprueban manualmente, y o bien se verifican o bien se invierten.

25 *“Redacción” interactiva* – Tras las asignaciones de alelos automatizadas usando el programa y los métodos descritos anteriormente, las supuestas asignaciones pueden “editarse” frente a o bien una base de datos de alelos que indica alelos conocidos, o bien mediante una combinación de datos experimentales para alelos (que forma un conjunto de referencia que se expande continuamente) y una base de datos de alelos. Una ilustración de las etapas clave se facilita en un ejemplo.

30 Los pesos representan una medida de confianza en el bit asignado a cualquier intensidad de sonda dada: es menos probable que los bits se asignen incorrectamente y por tanto es menos probable que se ejecuten las inversiones, cuanto mayor sea el peso de una señal de intensidad de sonda específica (o una función adecuada basándose en esos pesos). Por tanto, los pesos pueden ayudar a guiar la redacción interactiva.

35 Se dispone de otra guía en forma de frecuencias de alelos en toda la población, o en una subpoblación que está monitorizándose. De nuevo, las muestras analizadas forman una base de datos de referencia en expansión continua, que se añaden al conjunto de entrenamiento cuyas frecuencias de alelos (y haplotipo) se adaptan en tiempo real.

40 *Detección de nuevos alelos* – Pueden indicarse nuevos alelos mediante dianas que producen patrones de reacción binarios que pueden hacerse coincidir con cadenas de referencia existentes que representan combinaciones de alelos conocidos solo invirtiendo sondas que tienen pesos significativos. Esto se comenta en mayor detalle en el ejemplo I (Asignación de alelos) a continuación.

45 *Generalización de representación binaria* - El programa AAA también alberga representaciones de patrones de intensidad distintos de la representación binaria y las cadenas binarias correspondientes (“palabras”).

50 *Alfabeto de tres letras y tablas de resultados positivos* – Como generalización inmediata, deben considerarse representaciones que recurren a un alfabeto de tres o más letras. Una representación de tres letras de este tipo surge de manera natural cuando se proporciona un par de sondas degeneradas para uno o más de los sitios diana polimórficos designados. Por ejemplo, en un enfoque nuevo que recurre al formato de análisis de polimorfismos multiplexados mediado por elongación (eMAP) para analizar mutaciones en un conjunto de genes que codifican para antígenos de grupo sanguíneo humano, se proporciona un par de sondas de elongación degeneradas para cada uno de los sitios variables designados. Los miembros del par difieren en o cerca de sus extremos 3’ terminales, estando designado un miembro para coincidir con el alelo diana normal esperado, estando designado el otro miembro para coincidir con el alelo variante esperado. Sólo se elonga la sonda de elongación que coincide con la diana de manera que se produce una señal de ensayo correspondiente asociada con los productos de elongación (véase la solicitud estadounidense con n.º de serie 10/271.602). Es decir, eMAP produce uno de tres valores posibles en cada sitio polimórfico designado, concretamente normal, variante (mutante “homocigoto”), o heterocigoto.

55 La representación refleja los tres posibles desenlaces de la determinación de eMAP en cada sitio designado,

concretamente:

- sonda normal con coincidencia, sonda variante sin coincidencia: normal - indicado mediante 1

5 - sonda normal sin coincidencia, sonda variante con coincidencia: variante - indicado mediante -1

- sonda normal con coincidencia, sonda variante con coincidencia: heterocigoto – indicado mediante 0

10 Esto refleja las posibles combinaciones de los alelos subyacentes, concretamente AA (normal o “de tipo natural”), BB (variante, homocigoto) y AB o BA (heterocigoto). Una tabla de resultados positivos para los sitios de interés estará compuesta por códigos de letras que se combinan mediante normas ya expuestas.

15 El ejemplo III ilustra el uso de un alfabeto de 3 letras (1, 0, -1) para representar combinaciones bialélicas observadas.

20 *Umbrales superior e inferior* – Una representación de tres letras también surge, de manera análoga, en relación con la introducción de un umbral superior y uno inferior. Para cada sonda que está considerándose, un ensayo intensidad de señal por debajo del umbral inferior corresponde a una ausencia de coincidencia con ambos alelos diana, un ensayo intensidad de señal por encima del umbral inferior, pero por debajo del umbral superior, corresponde a una coincidencia con un alelo pero no con el otro, y un ensayo intensidad de señal por encima del umbral superior corresponde a una coincidencia con ambos alelos.

25 Las designación de bits negativos y positivos puede realizarse con confianza aumentada si se definen dos umbrales, que segregan intensidades de ensayo normalizadas registradas para cualquier sonda dada en tres subpoblaciones. Las tres subpoblaciones serían aquellas para las que: (i) una sonda dada presenta ausencia de coincidencia con ambos alelos asignados (1,1), (ii) una sonda coincide con un alelo (1, 8; 8, 1) y (iii) una sonda coincide con ambos alelos (8, 8).

30 Debido a la posible existencia del segundo (8, 8) umbral, sería posible establecer un umbral para una sonda particular de manera incorrecta; es decir, el umbral para una sonda (8, 8) diferente de una sonda (8, 1), podría identificarse de manera incorrecta como el umbral para una sonda (8, 1) diferente de una sonda (1, 1). Las designaciones de umbral incorrectas de este tipo pueden detectarse y corregirse mediante el refinamiento y la expansión continuados del conjunto de entrenamiento, y/o comprobando de manera doble las asignaciones de alelos frente a referencias de la base de datos de alelos conocidos y garantizando la coherencia.

35 Otra situación en la que han de considerarse las necesidades, es que las intensidades de ensayo normalizadas registradas para cualquier sonda dada que están por encima del segundo umbral (superior), podrían deberse a la reacción con la subsecuencia diana designada en ambos alelos (indicando un homocigoto) o podrían deberse a la reacción con dos alelos independientes, que de manera coincidente, son reactivos con esa sonda. De nuevo, esta situación puede detectarse y corregirse mediante el refinamiento y la expansión continuados del conjunto de entrenamiento, y/o comprobando de manera doble las asignaciones de alelos frente a referencias de la base de datos de alelos conocidos.

45 En la determinación de la ubicación de umbrales, cuando hay más de un umbral para una sonda particular, puede examinarse la proporción de perfiles de intensidad (tal como se muestra en las figuras 2A a 2C; véanse también las figuras 8A y 8B). Pero si no hay inflexión pronunciada en el perfil, tal como se ilustra claramente en la figura 2B, pueden localizarse los puntos de inflexión, y por tanto los umbrales, adoptando la derivada numérica usando un filtro de desconvolución. Los resultados de adoptar la derivada numérica de esta manera se muestran en la figura 8A, que es un perfil de proporción, y en la figura 8B, que es la derivada numérica que muestra los puntos de inflexión derivados de la figura 8A.

50 *Digitalización de patrones análogos* – También pueden digitalizarse intensidades normalizadas, en lugar de binarizarse, con cualquier grado deseable de precisión superior a la permitida por la binarización. Por ejemplo, en lugar de dos subpoblaciones, podría escogerse segregar intensidades en ocho subpoblaciones o 16 subpoblaciones. Inherente a esta representación es la información representada en forma de pesos en la representación binaria comentada anteriormente en el presente documento. Cada intensidad normalizada digitalizada representa de hecho una medida de la coafinidad de una interacción sonda-diana particular (véase la solicitud estadounidense n.º 10/204.799 “Multianalyte Molecular Analysis”; documento WO 01/98765). Se comparan patrones de reacción digitalizados experimentales y patrones digitalizados de referencia por medio del cálculo de correlaciones cruzadas usando métodos convencionales.

60 *Acceso a distancia para múltiples usuarios, servidor de aplicaciones* – el uso de un programa también permite el establecimiento de una red para permitir el análisis, la redacción y la notificación a distancia de los resultados de asignación de alelos. Por ejemplo, puede accederse a una base de datos que forma parte del entorno de software AAA a través de una conexión de red segura. El programa AAA también soporta un modo de servicio de aplicación que permite la edición interactiva desde una ubicación distinta de la ubicación del laboratorio experimental.

65

Realización preferida de análisis multiplexado: Detección de matrices codificadas al azar - En un formato de análisis multiplexado, las sondas de detección se presentan en micropartículas codificadas (“perlas”). Se asocian marcadores con las dianas. Las perlas codificadas unidas a las sondas en la matriz son preferiblemente fluorescentes y pueden distinguirse usando filtros que permiten la discriminación entre diferentes tonos. Preferiblemente, se disponen conjuntos de perlas codificadas en forma de una matriz plana al azar sobre un sustrato plano, permitiendo de ese modo el examen y el análisis mediante microscopía. Se monitoriza la intensidad de los marcadores diana para indicar la cantidad de diana unida por perla. Este formato de ensayo se explica en detalle adicional en la publicación internacional n.º WO 01/98765 titulada: “Multianalyte molecular analysis”.

Se dispone de varios métodos de producción de firmas ópticas, por ejemplo mediante captura de dianas marcadas o mediante elongación de sonda mediada por diana (eMAP), realizándose esta última preferiblemente usando oligonucleótidos específicos de alelos inmovilizados que pueden actuar como cebadores de una reacción de elongación catalizada por polimerasa. (véase, por ejemplo, la publicación internacional n.º WO 03/034029). Se producen una o más dianas adecuadas, por ejemplo, mediante transcripción inversa de ARN y/o amplificación de ADN genómico, seguido opcionalmente por etapas adicionales tales como fragmentación (véase la solicitud provisional estadounidense 60/515.413), desnaturalización o selección de cadenas (solicitud estadounidense con n.º de serie 10/847.046).

Posteriormente al registro de una imagen de decodificación de la matriz de perlas, se expone la matriz a las dianas en condiciones que permiten la captura de sondas presentadas por partículas. Tras un tiempo de reacción adecuado, se lava la matriz de partículas codificadas para retirar dianas restantes libres y débilmente apareadas. Entonces se obtiene una imagen del ensayo para registrar la señal óptica de los complejos sonda-diana en la matriz (o para registrar la señal de las sondas alargadas, en el caso de elongación mediada por captura en el formato de ensayo que está usándose). Puesto que cada tipo de partícula se asocia de manera única con una sonda específica de secuencia, la etapa de descodificación permite la identificación de moléculas diana apareadas determinada a partir de la fluorescencia de cada tipo particular de partícula.

Se usa un microscopio de fluorescencia para la descodificación. Los conjuntos de filtro de fluorescencia en el descodificador están diseñados para distinguir la fluorescencia producida por colorantes de codificación usados para teñir partículas, mientras que otros conjuntos de filtro están diseñados para distinguir señales de ensayo producidas por los colorantes asociados con las dianas. Puede incorporarse una cámara CCD en el sistema para el registro de imágenes de descodificación y ensayo. La imagen del ensayo se analiza para determinar la identidad de cada una de las dianas capturadas correlacionando la distribución espacial de las señales en la imagen del ensayo con la distribución espacial de las partículas codificadas correspondientes en la matriz.

En este formato de análisis multiplexado, existe una limitación en el número de tipos de sonda, porque el número total de tipos de perla en la matriz está limitado por el método de codificación usado (por ejemplo, el número de colores distinguibles disponibles) y por los límites de la instrumentación usada para la interpretación, por ejemplo, el tamaño del campo en el microscopio usado para leer la matriz. También debe considerarse, al seleccionar las sondas, que determinados tipos de sonda se hibridan más eficazmente a su diana que otros, en las mismas condiciones. La eficacia de la hibridación puede resultar afectada por varios factores incluyendo la interferencia entre sondas vecinas, la longitud de la sonda y la secuencia de la sonda, y, significativamente, por la temperatura a la que se realiza el apareamiento. Una eficacia de hibridación baja puede dar como resultado una señal falsa negativa. Por consiguiente, un diseño de ensayo debe intentar corregir un apareamiento sonda/diana de poca eficacia de este tipo.

Una vez realizado un ensayo real, puede usarse el sistema de obtención de imágenes de matriz (tal como se describe en la solicitud estadounidense con n.º de serie 10/714.203) para generar una imagen del ensayo, que puede usarse para determinar la intensidad de las señales de hibridación de diversas perlas (sondas). La imagen del ensayo puede aplicarse entonces mediante un sistema para asignación de alelos automática, tal como se describe en el presente documento.

Ejemplos

I. Asignación de alelos – A modo de ilustración (véase también la “ilustración de captura de pantalla” en la figura 3), AAA, usando el conjunto de sondas central de sondas, enumera dos asignaciones de alelos en grupo sugeridas, concretamente A*03 + A*29 y A*29 + A*74. Los dos grupos se ordenan en el orden de los pesos de las ondas invertidas. Cuanto menor es el peso, mayor es el orden de los grupos. Si el conjunto de sonda central produce asignaciones sugeridas degeneradas, como en este caso, el análisis se repite automáticamente usando el conjunto de sondas expandido. Este segundo paso produce una asignación sugerida de A*03 + A*29 lo que requeriría la inversión de HA120+, lo que indica que HA120 puede representar un falso positivo. El patrón de reacción y la tabla de resultados positivos (para alelos de HLA-A) también se presentan en la captura de pantalla. En el patrón de reacción, 8 indica que la sonda es positiva, 1 negativa y 0 significa que la sonda no se usa.

En el modo de redacción manual, el usuario puede editar las asignaciones de alelos iniciales comprobando las referencias conocidas para los alelos identificados en los conjuntos central y expandido, y adaptando entonces las cadenas de bits a las esperadas a partir de los alelos conocidos. La verificación de los resultados experimentales

frente a alelos conocidos de esta manera proporciona una validación de los resultados de ensayo, y un medio de editar la cadena de bits. El usuario escoge A*03011 y A*2901101 como asignación, HA120 como sonda de inversión en un modo de redacción manual en la figura 4.

5 EJEMPLO II: Datos estadísticos de la frecuencia de alelos

Se examinaron 1155 usando un panel de HLA-A, y se analizaron los patrones de intensidad de muestra usando el programa AAA para obtener asignaciones de alelos en grupos de dos dígitos. En la tabla 1 se muestran el recuento y la frecuencia relativa de aparición de denominaciones de grupo calculadas por el programa AAA, inmediatamente a continuación.

10

N.º de alelo	Recuento	Frecuencia
01	332	0,144
02	448	0,194
03	251	0,109
11	139	0,060
23	86	0,037
24	260	0,113
25	30	0,013
26	203	0,088
29	87	0,038
30	105	0,045
31	46	0,020
32	66	0,029
33	86	0,037
34	6	0,003
36	4	0,002
43	0	0,000
66	25	0,011
68	100	0,043
69	28	0,012
74	5	0,002
80	3	0,001
	2310	

El gráfico de barras para la siguiente distribución se muestra en la figura 1. Resulta evidente que los alelos 01, 02 y 03 son mucho más abundantes en esta población mientras que los alelos 36, 43 y 80 son comparativamente poco frecuentes.

15

EJEMPLO III. Captura de pantalla de información resumen de asignaciones

La captura de pantalla en la figura 3 ilustra la información resumen de asignaciones para el panel 03250443. Incluye el nombre del panel, el nombre de la muestra, la posición de la muestra, la asignación de alelos, las sondas de inversión, mensaje de advertencia y comentarios. La asignación de alelos enumera la asignación de nivel de alelos mediante un algoritmo informático. Las inversiones y los mensajes de advertencia también se presentarán según una asignación informática. Si hay redacción manual, la asignación de alelos se realizará según la elección del redactor manual. Véase la figura 4 para una ilustración de la redacción manual. También se presentan el comentario y las inversiones introducidas durante cualquier redacción manual. Las sondas de inversión se insertarán en una cadena (manual) en el extremo, lo que indica que se genera mediante redacción manual.

25

El software enumera los dos primeros dígitos en la asignación de alelos y los siguientes dígitos, si es aplicable. Por ejemplo, A*24(020101) indica que la denominación de dos dígitos es A*24. La ventana de información resumen presenta toda la información vital en una ventana, lo que facilita examinar y navegar a través de diferentes muestras.

30

EJEMPLO IV: Alfabeto de tres letras. Tipificación molecular de antígenos de grupo sanguíneo

En un conjunto de aproximadamente 500 muestras clínicas y controles, se identificaron varias combinaciones de alelos mediante un diseño de ensayo eMAP, diseñado para sondear antígenos de grupo sanguíneo minoritarios incluyendo Duffy (FYA/FYB), GATA, Landsteiner-Weiner (LWA/LWB), Colton (CoA/CoB), Scianna (SC1/SC2), Diego (DIA/DIB) y Dombrock (DoA/DoB), comprendiendo este último tres mutaciones. Véase la tabla 2 a continuación.

35

TABLA 2 - Combinaciones de alelos observadas de antígenos de grupo sanguíneo humano minoritarios

40

ID de la muestra	FYA/FYB	GATA	LWA/LWB	CoA/CoB	SC1/SC2	DIB/DIA	DO-793	DO-624	DO-378
------------------	---------	------	---------	---------	---------	---------	--------	--------	--------

N21 BAS	-1	-1	1	1	1	1	-1	-1	-1
N40 BAS	-1	0	1	1	1	1	-1	-1	-1
M17	0	0	1	1	1	1	-1	-1	-1
A10	-1	1	1	1	1	1	-1	-1	-1
A4	0	1	1	1	1	1	-1	-1	-1
A1	1	1	1	1	1	1	-1	-1	-1
N39 BAS	-1	-1	1	1	1	1	-1	-1	0
N71	-1	0	1	1	1	1	-1	-1	0
N62-BAS	0	0	1	1	1	1	-1	-1	0
N66-BAS	-1	-1	1	1	1	1	-1	-1	1
1	-1	0	1	1	1	1	-1	-1	1
N72	-1	1	1	1	1	1	-1	-1	1
M12	-1	1	1	1	1	1	-1	-1	1
16	-1	-1	1	1	1	1	0	0	-1
34	-1	0	1	1	1	1	0	0	-1
A21	-1	1	1	1	1	1	0	0	-1
N34 BAS	0	0	1	1	1	1	0	0	-1
A28	0	1	1	1	1	1	0	0	-1
A14	1	1	1	1	1	1	0	0	-1
N70	-1	-1	1	1	1	1	0	0	0
A6	0	1	1	1	1	1	0	0	0
A7	1	1	1	1	1	1	0	0	0
N78-BAS	-1	-1	1	1	1	1	0	0	1
2	-1	1	1	1	1	1	1	1	0
M23	0	1	1	1	1	1	1	1	0
U79	1	1	1	1	1	1	1	1	0
N35 BAS	-1	-1	1	1	1	1	1	1	1
N51	-1	0	1	1	1	1	1	1	1
A9	-1	1	1	1	1	1	1	1	1
26	-1	1	1	1	0	1	1	1	1
62	-1	1	1	1	1	1	1	1	1
N51-BAS	0	0	1	1	1	1	1	1	1
A25	0	1	1	1	1	1	1	1	1
N7 BAS	1	1	1	1	1	1	1	1	1

EJEMPLO V: Establecimiento de relaciones con enfermedades y estados

5 También pueden usarse asignaciones de alelos determinadas por los métodos anteriores para establecer riesgo o presencia de enfermedades o estados. Se conoce bien que determinados trastornos inmunitarios están asociados con el locus de HLA. Si se conocen los alelos asociados pueden escribirse y si se desconocen pueden usarse los métodos descritos en el presente documento para establecer una base de datos de alelos para indicar el riesgo o la presencia de enfermedades o estados. La base de datos puede actualizarse de manera continua basándose en la monitorización de pacientes cuyas muestras se usaron en la base de datos; es decir, cuando se produce un determinado desarrollo de la enfermedad, sus alelos pueden analizarse para determinar coincidencias con aquellos con una enfermedad o estado particular.

15 Debe entenderse que los términos, expresiones y ejemplos en el presente documento son a modo de ejemplo y no limitativos, y que la invención se define únicamente en las reivindicaciones que siguen, e incluye todos los equivalentes del contenido de las reivindicaciones. Las etapas del método en las reivindicaciones no están necesariamente en orden, y a menos que se especifique en la reivindicación, pueden llevarse a cabo en cualquier orden, incluyendo el especificado en las reivindicaciones.

REIVINDICACIONES

- 5 1. Método para asignaciones de alelos donde las asignaciones se realizan basándose en someter una muestra que incluye ácido nucleico diana a un ensayo de hibridación o ensayo de elongación mediado por captura, o ambos, y donde se analizan varios marcadores polimórficos, presentes en alelos del ácido nucleico diana, comprendiendo el método:
- 10 determinar un patrón de intensidad de señales de ensayo generado por interacciones sonda-diana para generar un patrón de reacción designando como positivas aquellas intensidades de señal que superan un umbral predeterminado y designando como negativas aquellas intensidades de señal que no superan el umbral predeterminado, binarizando de ese modo dicho patrón de intensidad de señales de ensayo generado por interacciones sonda-diana donde cada una de dicha intensidad corresponde al número de acontecimientos de hibridación o elongación correspondientes, según sea aplicable;
- 15 realizar asignaciones de alelos basándose en comparar y hacer coincidir dicho patrón de reacción con un patrón de reacción de referencia similar que representa combinaciones de dos o más alelos de referencia, tal como se enumera en una colección de referencia de alelos para el ácido nucleico diana en construcción;
- 20 y designar, como nuevos alelos, los que tienen patrones de reacción que difieren de patrones de reacción de referencia que representan combinaciones de dos o más alelos de referencia mediante una disparidad mínima predeterminada.
- 25 2. Método según la reivindicación 1, en el que dicho patrón de reacción se edita basándose en la comparación con el patrón de reacción de referencia.
3. Método según la reivindicación 1, que incluye además la etapa de realizar asignaciones de alelos iniciales que luego se confirman o cambian tras dichas etapas de comparación y coincidencia.
- 30 4. Método según la reivindicación 1, en el que el patrón de reacción de referencia similar es el patrón de reacción de referencia más similar.
5. Método según la reivindicación 1 ó 4, realizado mediante un sistema informático de software.
- 35 6. Método según la reivindicación 1, en el que el ensayo de hibridación genera una señal de ensayo positiva cuando tiene lugar la hibridación.
7. Método según la reivindicación 1, en el que el ensayo de elongación mediado por captura genera una señal de ensayo positiva cuando tiene lugar la elongación.
- 40 8. Método según la reivindicación 6, en el que el ensayo de elongación mediado por captura genera una señal de ensayo positiva cuando tiene lugar la elongación.
- 45 9. Método según la reivindicación 1, que incluye además la etapa de restar de cada intensidad de señal, I, la señal de control negativa, I_{NC}, y dividir el resultado entre la señal de control positiva corregida, I_{PC} - I_{NC}, para obtener una intensidad normalizada (proporción):
- $$r = (I - I_{NC}) / (I_{PC} - I_{NC})$$
- 50 10. Método según la reivindicación 1, que incluye además la etapa de seleccionar dicho umbral predeterminado para maximizar la similitud entre dicho patrón de reacción y un patrón de reacción de referencia.

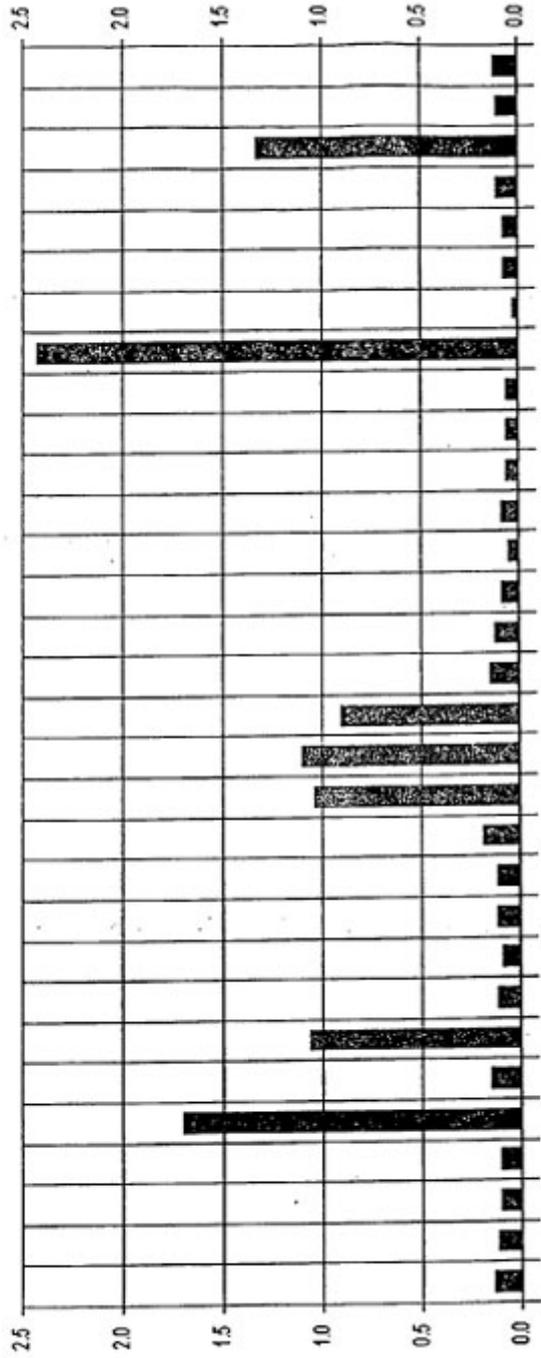


Fig. 1A

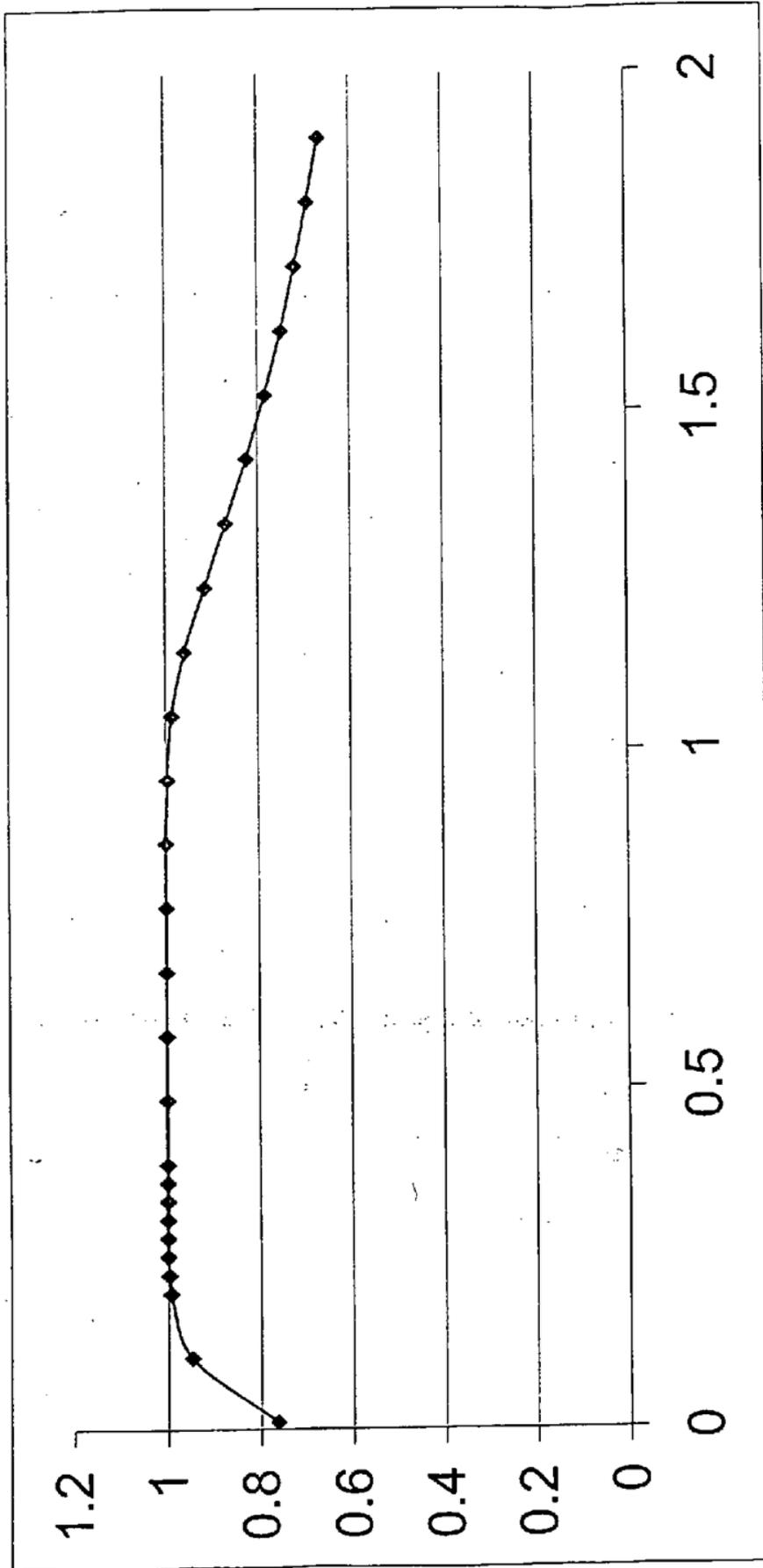


Fig. 1B: Determinación de umbral empírica a modo de ejemplo para una sonda

Ajuste de asignación

Tolerancia de asignación

HLA-A: HLA-B: HLA-DR:

Lista de sondas

Nb. de sonda	Requerida	Alta confianza	Baja confianza	Sin usar
HA114		X		
HA114AT				X
HA115		X		
HA116		X		
HA119			X	
HA120		X		
HA121			X	
HA122		X		
HA122AT				X
HA122B				X
HA123		X		

Barra de desplazamiento

Fig. 1C: Ajustes del sistema

[Barra de herramientas]

Molde: HLA-B [menú desplegable]

Sonda: HB103 [menú desplegable]

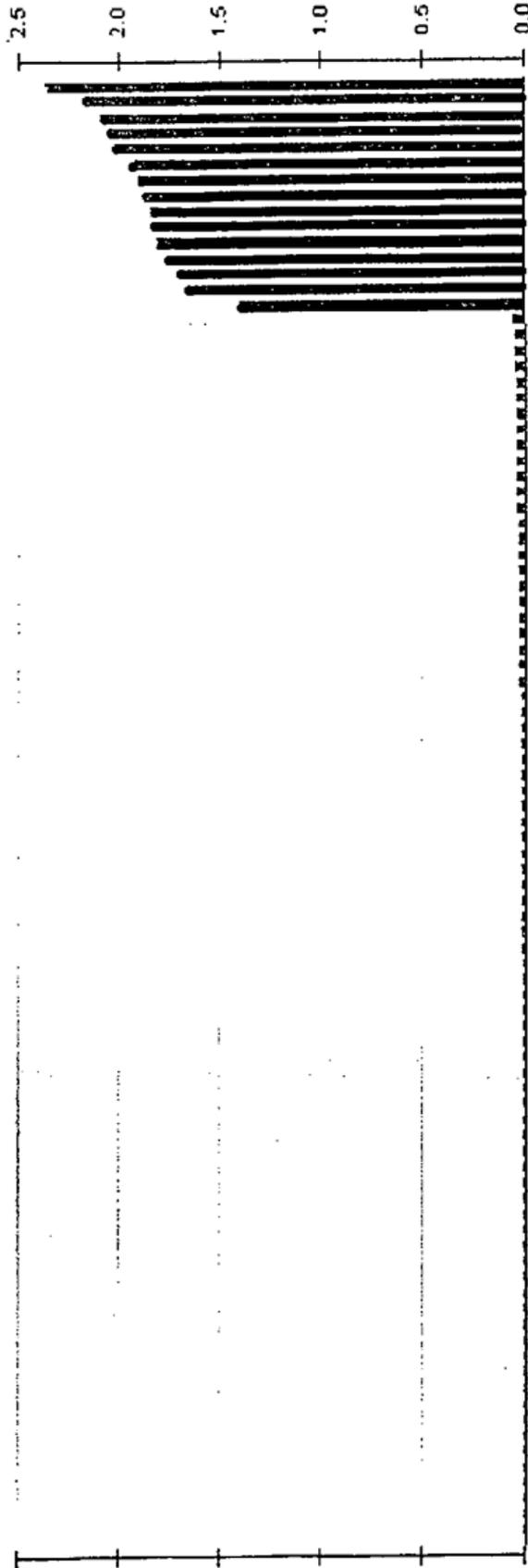
Umbral: 0.700

Panel:

03250444
 03250445
 03250446
 03250447
 03250448

Barra de despl.

Diagrama de proporciones de HB103



Nb. de panel: 03250444 Umbral: 0.700 Recuento de muestras: 93

FIG. 2A Perfil de proporción de sonda de HB103

[Barra de herramientas]

Molde: HLA-B [menú desplegable]

Sonda: HB123A [menú desplegable]

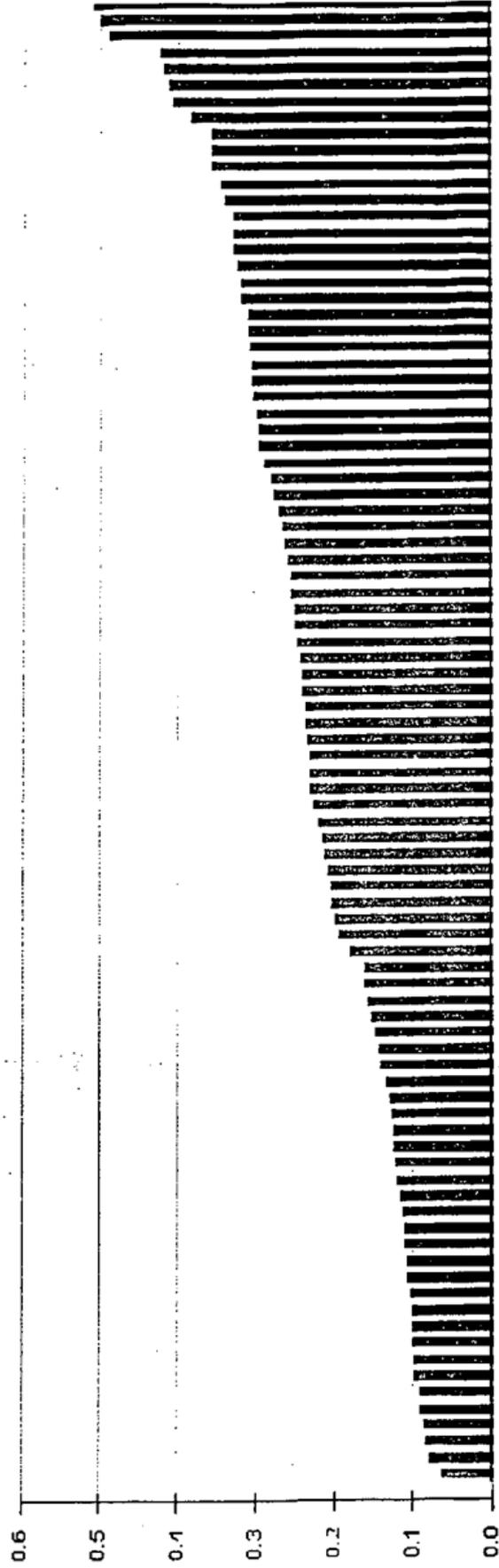
Umbral: 0.300

Panel:

Barra de despl.

03250444
03250445
03250446
03250447
03250448

Diagrama de proporciones de HB123A



Nb. de panel: 03250444 Umbral: 0.300 Recuento de muestras: 93

FIG. 2B Perfil de proporción de sonda de HB123A

[Barra de herramientas]

Molde: HLA-B [menú desplegable]

Sonda: HB123A [menú desplegable]

Umbral: 0.100

Panel: 03250444
03250445
03250446
03250447
03250448

Barra de despl.

Diagrama de proporciones de HB154

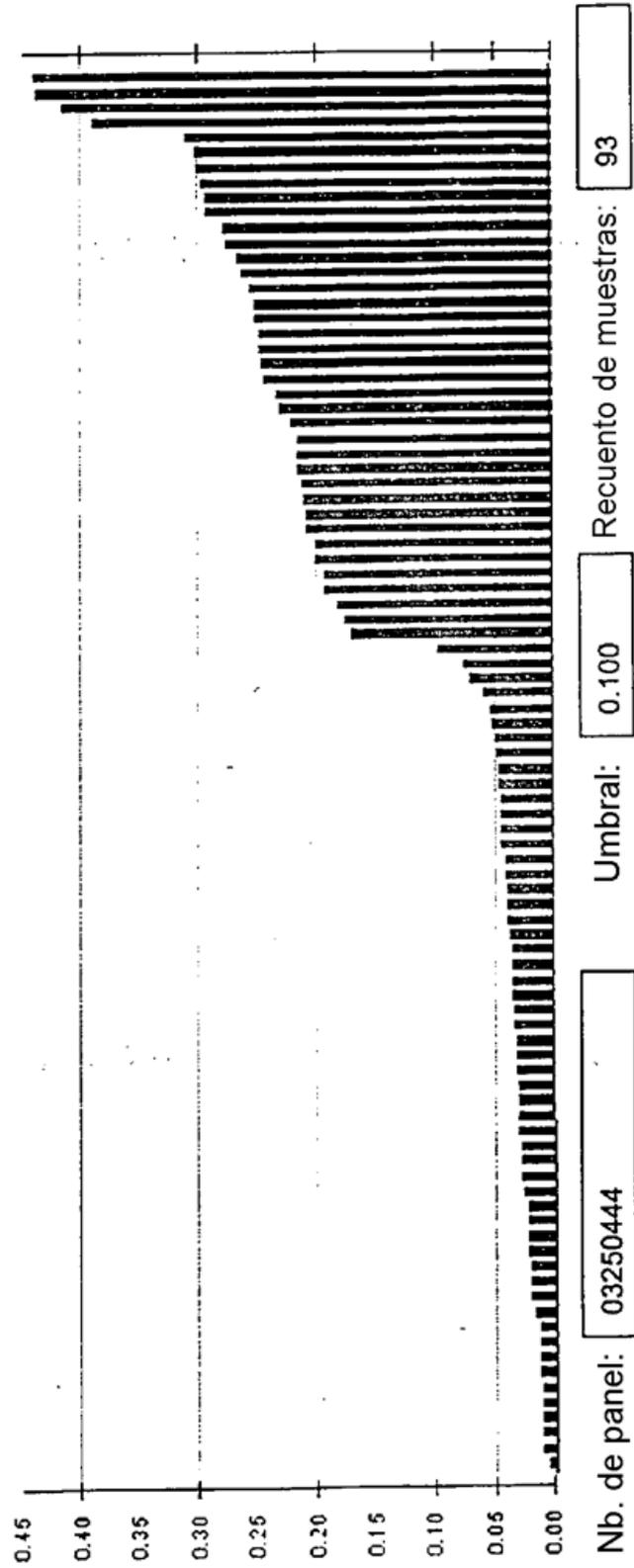
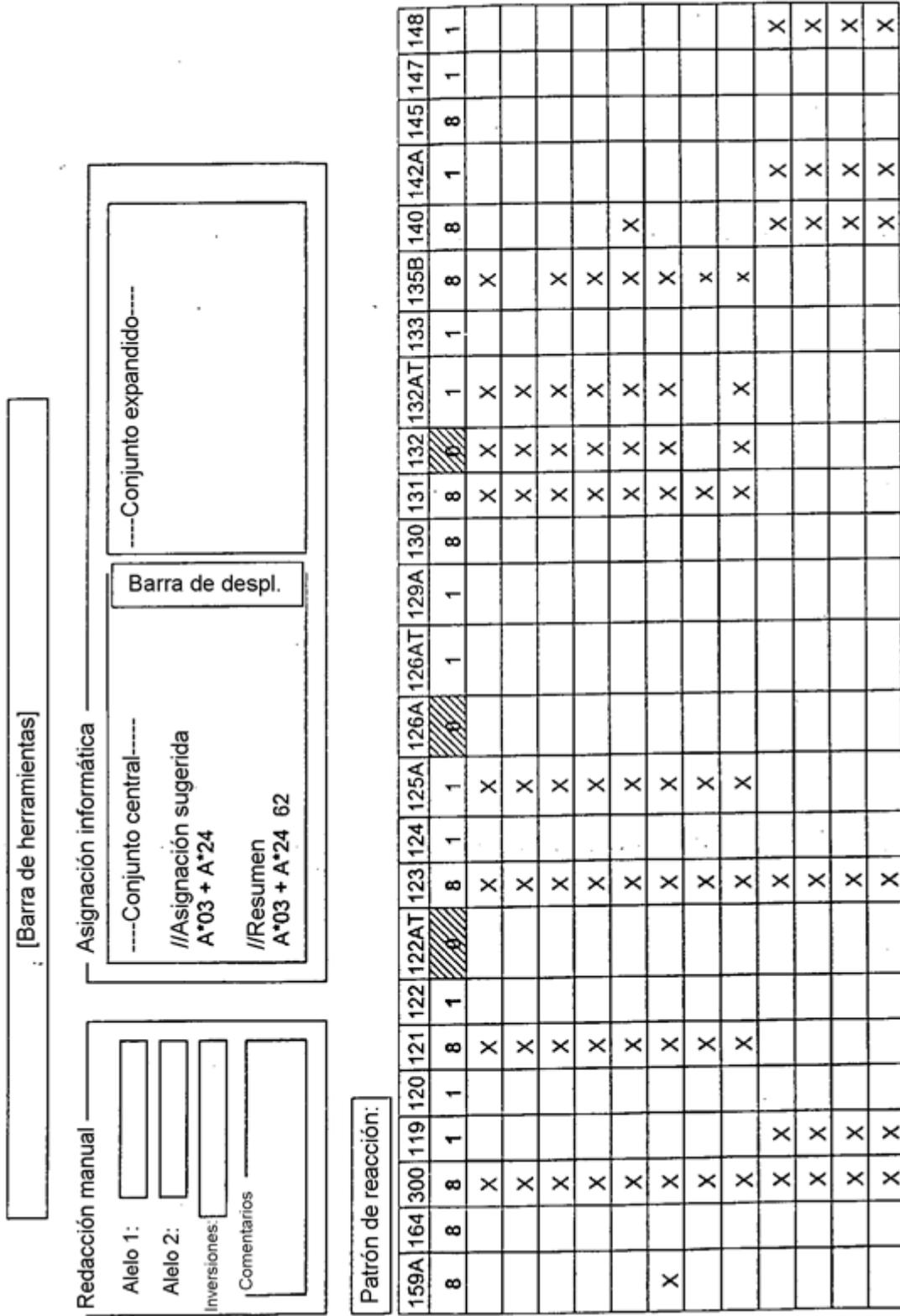


FIG. 2C Perfil de proporción de sonda de HB154



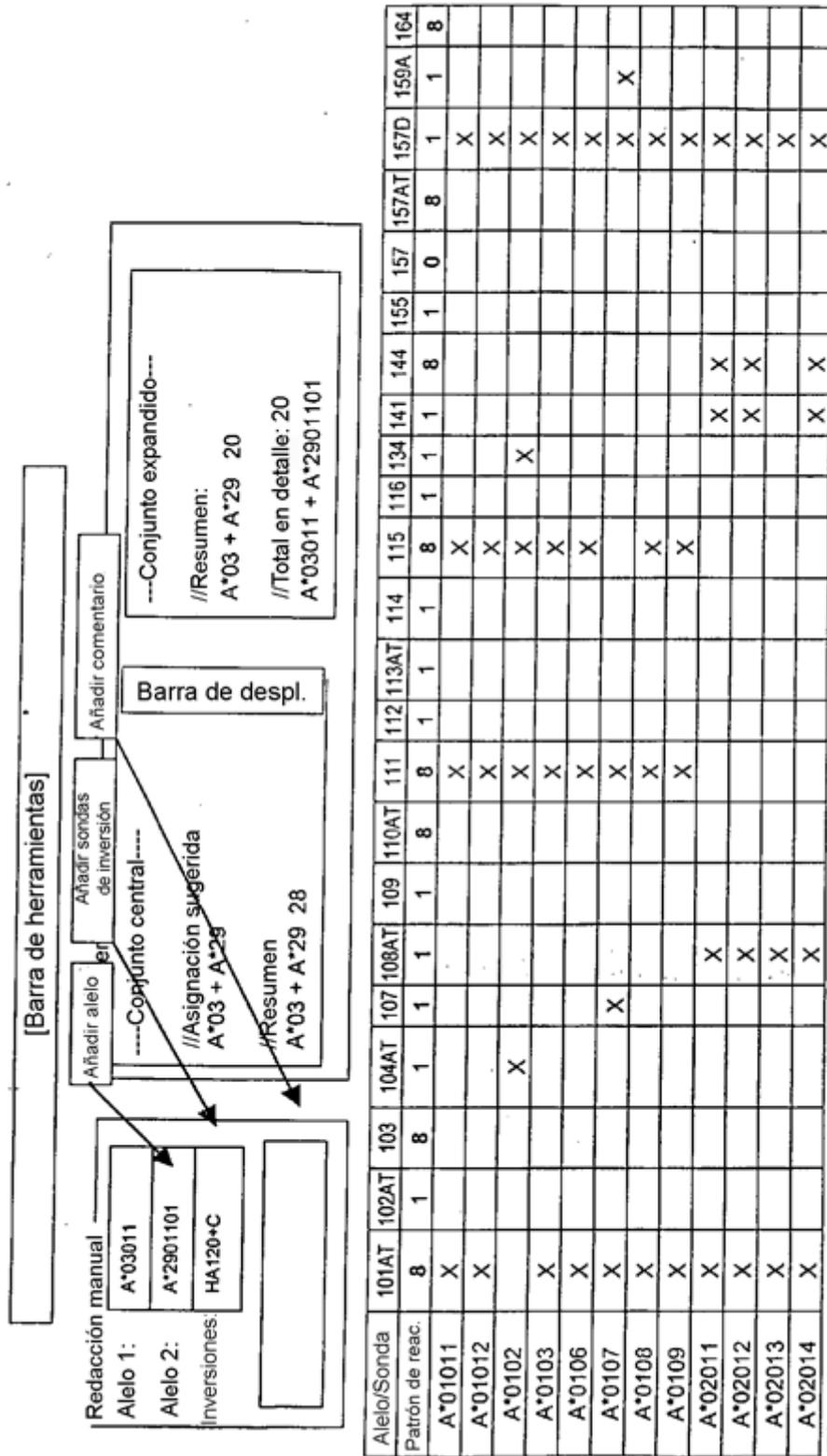


Fig. 4: Asignación de alelos con redacción manual

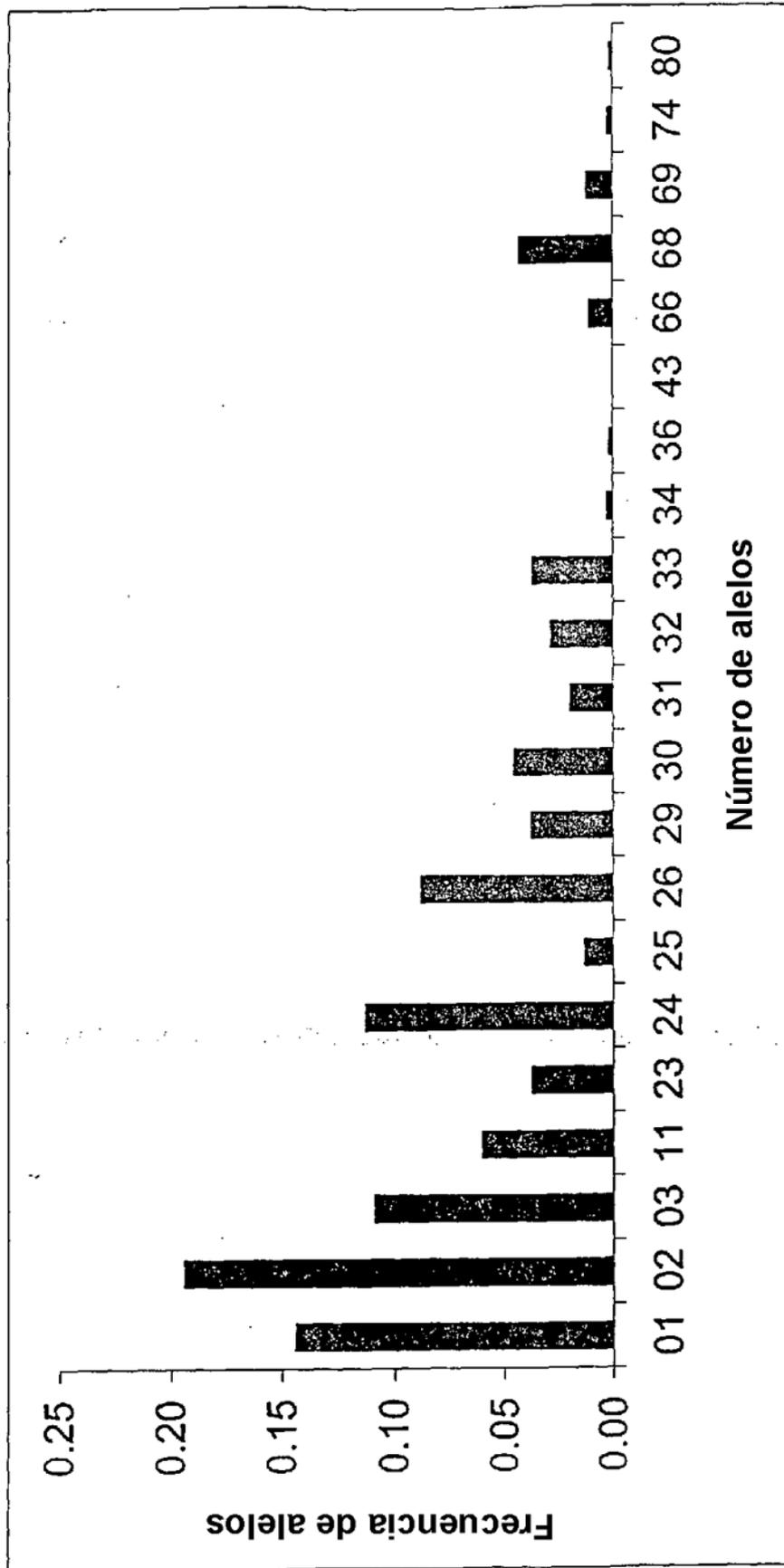


Fig. 5

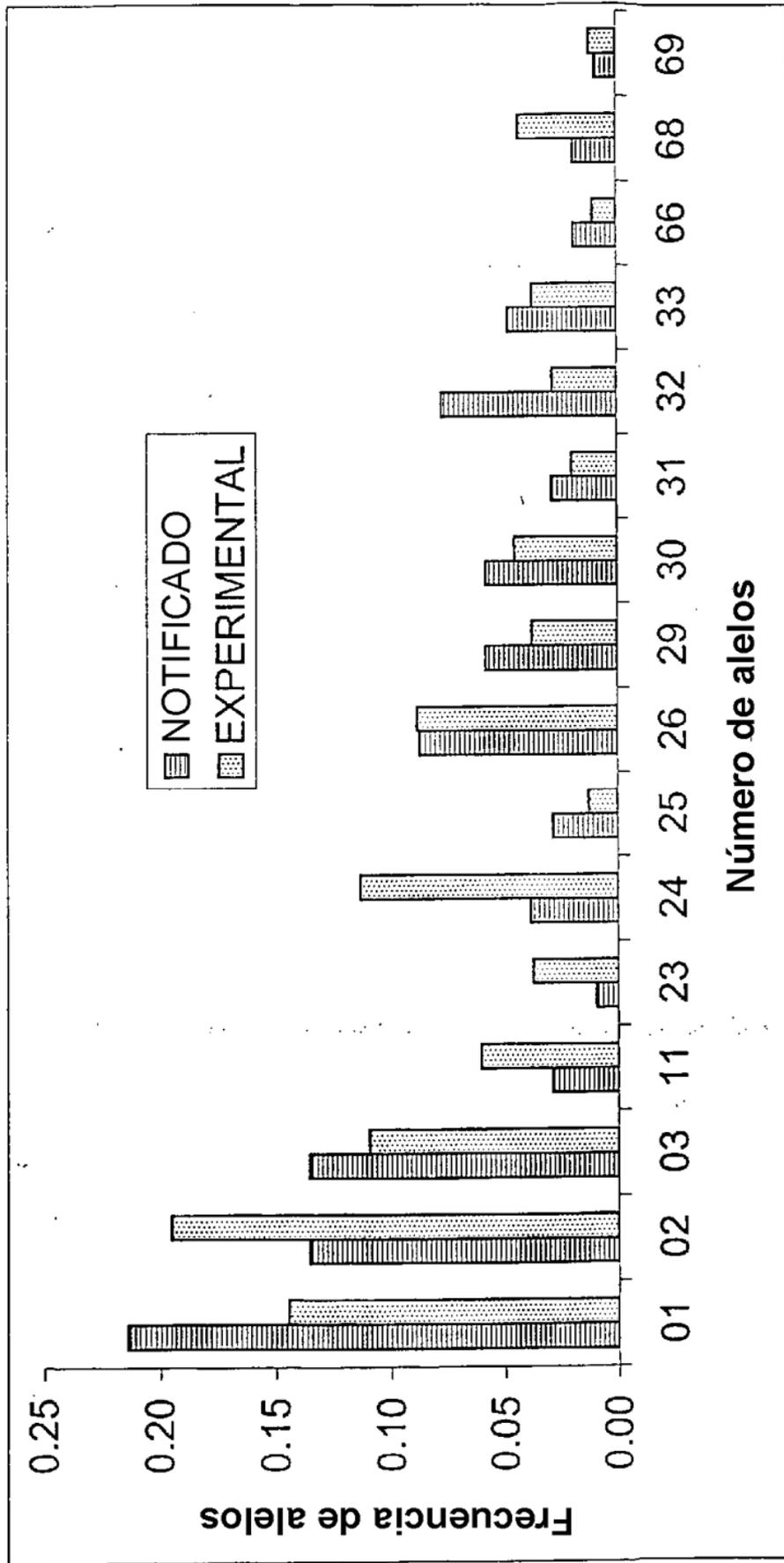


Fig. 6

HLA-A		[MENÚ]									
Criterios de búsqueda		Nombre de muestra:									
Nb. de panel	Nb. de muestra	Posición	Alelo 1	Alelo 2	N.º grupo	N.º alelo	Inv.	Sonda de inversión	Mensaje de advert.	Comentarios	
3250443	D020157	A2	A*24(02101)	A*26(01)	1	45	0				
3250443	D020169	B2	A*02011		0	0	0	(HA163+D)(Manual)		OK	
3250443	D020181	C2	A*01(011)	A*29(01101)	1	25	0				
3250443	D020193	D2	A*3001		2	2	0			OK	
3250443	D020205	E2	A*26(01)	A*30(01)	1	4	0				
3250443	D020217	F2	A*01(011)	A*03(011)	1	31	0				
3250443	D020229	G2	A*01(011)	A*29(01101)	1	25	0				
3250443	D020241	H2	A*01(011)	A*32(01)	1	10	0				
3250443	D020158	A3	A*0302	A*58(011)	1	2	0				
3250443	D020170	B3	A*01(011)	A*02(011)	1	91	0				
3250443	D020182	C3	A*02(02)	A*02(10)	2	71	1	HA157AT+C		OK	
3250443	D020194	D3	A*03(011)	A*24(02101)	1	62	0				
3250443	D020206	E3	A*11(011)	A*3001	1	7	0				
3250443	D020218	F3	A*01(011)	A*03(011)	1	31	0				
3250443	D020230	G3	A*01(011)	A*02(02)	1	8	0				
3250443	D020242	H3	A*01(011)	A*24(02101)	1	61	0				
3250443	D020158	A4	A*02011	A*58011	0	0	0	(HA129A-C)(Manual)		podría ser OK	
3250443	D020171	B4	A*03(011)	A*3001	1	7	0				
3250443	D020183	C4	A*02(011)	A*24(02101)	1	209	0				
3250443	D020195	D4	A*01(011)		3	26	0			conjunto expandido, OK	
3250443	D020207	E4	A*01(011)	A*02(011)	1	91	0				
3250443	D020219	F4	A*01(011)	A*03(011)	1	31	0				
3250443	D020231	G4	A*02(011)	A*26(01)	1	71	0				
3250443	D020243	H4	A*0302	A*6901	1	1	0				
3250443	D020160	A5	A*01(011)	A*68(012)	1	32	0				
3250443	D020172	B5	A*0205	A*2402101	0	0	0	(HA157AT+C)(Manual)		OK	

Fig. 7

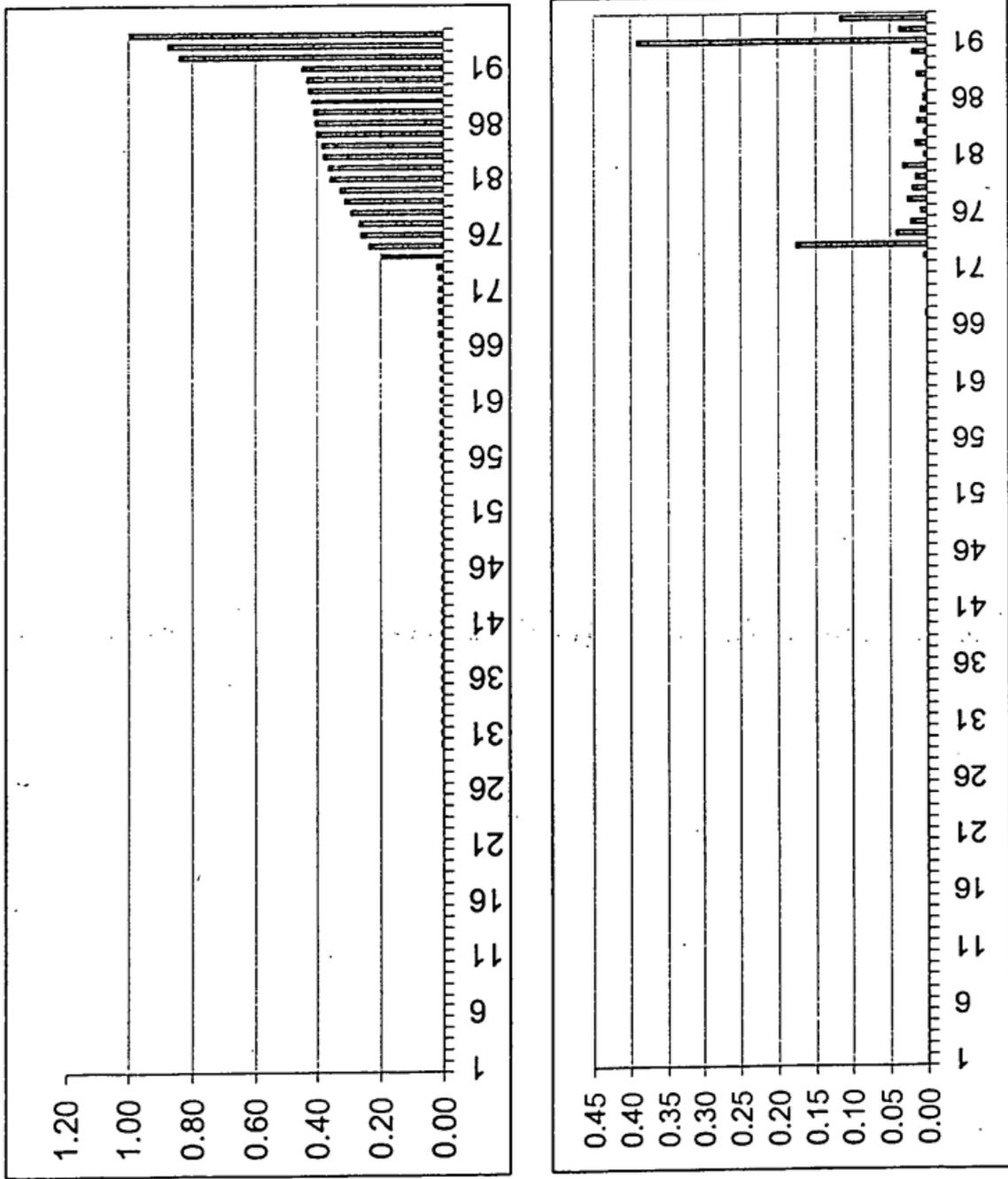


Fig. 8A (panel superior); Fig. 8B (panel inferior)