

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 745 580**

51 Int. Cl.:

C12Q 1/6876 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **22.03.2017 PCT/EP2017/056878**

87 Fecha y número de publicación internacional: **28.09.2017 WO17162773**

96 Fecha de presentación y número de la solicitud europea: **22.03.2017 E 17713238 (8)**

97 Fecha y número de publicación de la concesión europea: **12.06.2019 EP 3417074**

54 Título: **Procedimientos de análisis y matrices para su uso en los mismos**

30 Prioridad:

**23.03.2016 WO PCT/EP2016/056465
05.01.2017 GB 201700138**

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
02.03.2020

73 Titular/es:

**SENZAGEN AB (100.0%)
Medicon Village, Scheelevägen 2
223 81 Lund , SE**

72 Inventor/es:

**LINDSTEDT, MALIN MARIE;
BORREBAECK, CARL A.K.;
JOHANSSON, HENRIK;
ALBREKT, ANN-SOFIE;
ZELLER, KATHRIN STEPHANIE;
CHAWADE, AAKASH;
LINDBERG, TIM CARL SEWE y
FORRERYD, ANDY ANDREAS SEBASTIAN**

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 745 580 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

Procedimientos de análisis y matrices para su uso en los mismos

Campo de la invención

La presente invención se refiere a un procedimiento para identificar la potencia como sensibilizador de la piel de un agente de ensayo y matrices y kits de análisis para su uso en dichos procedimientos.

Antecedentes

La dermatitis alérgica de contacto (DAC) es una enfermedad inflamatoria de la piel que afecta a una significativa proporción de la población. Es comúnmente causada por respuestas inmunológicas hacia haptenos químicos y conlleva una sustancial carga económica para la sociedad. Los ensayos actuales para productos químicos de sensibilización se basan en la experimentación con animales. La legislación sobre el registro y uso de productos químicos, p.ej., dentro de las industrias farmacéutica y cosmética, ha estimulado importantes esfuerzos de investigación para desarrollar ensayos alternativos basados en células humanas para la predicción de la sensibilización. El objetivo es reemplazar los experimentos con animales por ensayos *in vitro* que desplieguen un mayor poder predictivo.

La DAC se caracteriza por eccema y episodios recurrentes de prurito [Akhavan, A. and S.R. Cohen, The relationship between atopic dermatitis and contact dermatitis. Clin Dermatol, 2003. 21(2): p. 158-62]. La enfermedad afecta a una significativa proporción de la población, con tasas de prevalencia de 7,2 % a 18,6 % en Europa [Mortz, C.G., y col., prevalence of atopic dermatitis, asthma, allergic rhinitis, and hand and contact dermatitis in adolescents. The Odense Adolescence Cohort Study on Atopic Diseases and Dermatitis. Br J Dermatol, 2001. 144(3): p. 523-32; y Nielsen, N.H., y col., Allergic contact sensitization in an adult Danish population: two cross-sectional surveys eight years apart (the Copenhagen Allergy Study). Acta Derm Venereol, 2001. 81(1): p. 31-4], y la incidencia está aumentando debido a la reiterada exposición a productos químicos de sensibilización. DAC es una respuesta de hipersensibilidad de tipo retardado de tipo IV causada principalmente por linfocitos T auxiliares 1 (Th1) y linfocitos T CD8⁺ que producen interferón γ (IFN) en el sitio de contacto con pequeños haptenos químicos en individuos previamente expuestos e inmunológicamente sensibilizados [Fonacier, L.S., S.C. Dreskin, and D.Y. Leung, Allergic skin diseases. J Allergy Clin Immunol. 125(2 Suppl 2): p. S138-49]. Las células dendríticas (DC) en la epidermis inician las reacciones inmunes respondiendo a los haptenos unidos a las auto-moléculas y activando la inmunidad mediada por linfocitos T.

El reglamento REACH (Registro, Evaluación y Autorización de Productos Químicos) requiere que todos los productos químicos nuevos y existentes dentro de la Unión Europea, lo cual implica aproximadamente 30000 productos químicos, deberán ser analizados en cuanto a sus efectos peligrosos [CE 1907/2006, en Reglamento (CE) No 1907/2006]. Dado que la identificación de posibles sensibilizadores requiere actualmente ensayos con animales, la legislación REACH tendrá un enorme impacto en la cantidad de animales necesarios para los ensayos. Asimismo, la Séptima Enmienda a la Directiva de Cosméticos (76/768/CEE) planteó una prohibición en los ensayos con animales para la mayoría de los ingredientes cosméticos para uso humano en 2013. Por lo tanto, el desarrollo de alternativas *in vitro* fiables a los animales experimentales para la evaluación de la capacidad de sensibilización de los productos químicos es urgente.

Los autores de la invención han desarrollado un ensayo basado en células humanas para la predicción de productos químicos de sensibilización, llamado Detección rápida de alérgenos genómicos, GARD (Figura 8). Para una descripción detallada, véase la patente internacional WO 2012/056236 y Johansson H y col. The GARD assay for assessment of chemical skin sensitizers. Toxicology in vitro 2013. Al analizar el transcriptoma de células de tipo dendrítico (p. ej., La línea celular MUTZ-3) los autores de la invención han identificado una firma de biomarcadores genómicos con una potente capacidad discriminativa [Johansson H y col. A genomic biomarker signature can predict skin sensitizers using a cell-based in vitro alternative to animal assays. BMC Genomics. 2011; Johansson H y col. The GARD assay for assessment of chemical skin sensitizers. Toxicology in vitro 2013; and Johansson H y col. GARD in-house validation - A proof of concept. Toxicological Sciences 2014]. Asimismo, la forma en que se comprometen las rutas de señalización con diferentes grupos de reactividad química parece estar asociada con la potencia de sensibilización [Albrekt y col. Skin sensitizers differentially regulate signaling pathways in MUTZ-3 cells in relation to their individual potency. BMC Pharmacology and Toxicology 2014].

Sin embargo, hasta la fecha, se requieren sustitutos que no sean animales para la identificación de la potencia de productos químicos de sensibilización de la piel, en lugar del ensayo preferente del ensayo de ganglios linfáticos locales de ratón (LLNA) [Basketter, D.A., y col., Local lymph node assay - validation, conduct and use in practice. Food Chem Toxicol., 2002. 40(5): p. 593-8], seguido de la prueba de maximización con cobayas (GPMT) [Magnusson, B. and A.M. Kligman, The identification of contact allergens by animal assay. The guinea pig maximization assay. J Invest Dermatol, 1969. 52(3): p. 268-76].

Una alternativa *in vitro* a estos modelos animales presentaría preferentemente una mayor fiabilidad, precisión de predicción de potencia y, lo que es más importante, correlación con la reactividad humana.

- La patente internacional WO 2012/68474 A2 describe un procedimiento de evaluación exhaustivo de la hipersensibilidad de contacto; Albrekt y col. BMC Pharmacology & Toxicology 15(5) (2014): 1-12 describe sensibilizadores de la piel que regulan de forma diferenciada las rutas de señalización en células MUTZ-3 en relación con su potencia individual; Lambrechts y col. Toxicological Sciences 116(1) (2010): 122-129 describe la evaluación de la potencia de sensibilización de la piel química mediante un ensayo *in vitro* basado en células dendríticas humanas; Schoeters y col. Molecular Immunology 44 (2007): 3222-3233 describe análisis de micromatriz en células dendríticas que revelan posibles biomarcadores para sensibilización de la piel inducida por productos químicos.

Divulgación de la invención

- 10 Sobre la base de estudios anteriores, se ha demostrado que el ensayo GARD es valioso para la identificación de sensibilizador de la piel. Una evaluación más amplia del transcriptoma humano utilizando modelos multivariados ha revelado ahora sorprendentemente que también son posibles modelos de potencia *in vitro* mejorados.

Por consiguiente, un primer aspecto de la invención proporciona un procedimiento *in vitro* para determinar la potencia de sensibilización de la piel de un agente que comprende o consiste en las etapas de:

- 15 (a) proporcionar una población de células dendríticas o una población de células de tipo dendrítico;
- (b) exponer las células proporcionadas en la etapa (a) a un agente de ensayo; y
- (c) medir en las células de la etapa (b) la expresión de dos o más biomarcadores enumerados en la Tabla A;

en el que la expresión de los biomarcadores medidos en la etapa (c) es indicativa de la potencia de sensibilización de la piel del agente de ensayo de la etapa (b).

- 20 Se desvela también en el presente documento, que uno o más de los biomarcadores medidos en la etapa (c) se selecciona del grupo definido en la Tabla A (i) y/o la Tabla A (ii).

Se desvela también en el presente documento que la etapa (c) comprende o consiste en la medición de la expresión de uno o más biomarcadores enumerados en la Tabla A (i), por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 de los biomarcadores enumerados en la Tabla A (i). Por ejemplo, la etapa (c) puede comprender o consistir en la medición de la expresión de todos los biomarcadores enumerados en la Tabla A (i).

- El procedimiento puede incluir o excluir la medición de la expresión de HIST1 H2BM. El procedimiento puede incluir o excluir la medición de la expresión de HIST1 H4B. El procedimiento puede incluir o excluir medir la expresión de HIST1 H1D. El procedimiento puede incluir o excluir medir la expresión de PAICS. El procedimiento puede incluir o excluir la medición de la expresión de HIST1 H4D. El procedimiento puede incluir o excluir la medición de la expresión de HIST1 H2AL///HIST1H2BN. El procedimiento puede incluir o excluir la medición de la expresión de PLK1. El procedimiento puede incluir o excluir la medición de la expresión de PHGDH. El procedimiento puede incluir o excluir la medición de la expresión de MCM2. El procedimiento puede incluir o excluir la medición de la expresión de MCM7. El procedimiento puede incluir o excluir la medición de la expresión de CD53. El procedimiento puede incluir o excluir la medición de la expresión de KIFC1. El procedimiento puede incluir o excluir la medición de la expresión de WEE1. El procedimiento puede incluir o excluir la medición de la expresión de TMPO. El procedimiento puede incluir o excluir la medición de la expresión de MCM4. El procedimiento puede incluir o excluir la medición de la expresión de PSRC1. El procedimiento puede incluir o excluir la medición de la expresión de RNF149. El procedimiento puede incluir o excluir la medición de la expresión de MCM6. El procedimiento puede incluir o excluir la medición de la expresión de MCM5. El procedimiento puede incluir o excluir la medición de la expresión de FANCA.

Se desvela también en el presente documento que la etapa (c) comprende o consiste en la medición de la expresión de uno o más biomarcadores enumerados en la Tabla A (ii), por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 o 24 de los biomarcadores enumerados en la Tabla A (ii). Por ejemplo, la etapa (c) puede comprender o consistir en medir la expresión de todos los biomarcadores enumerados en la Tabla A (ii).

- 45 El procedimiento puede incluir o excluir la medición de la expresión de CAD. El procedimiento puede incluir o excluir la medición de la expresión de MRTO4. El procedimiento puede incluir o excluir la medición de la expresión de TOE1. El procedimiento puede incluir o excluir la medición de la expresión de CDC20. El procedimiento puede incluir o excluir la medición de la expresión de PM20D2. El procedimiento puede incluir o excluir la medición de la expresión de LOC344887. El procedimiento puede incluir o excluir la medición de la expresión de UNG. El procedimiento puede incluir o excluir la medición de la expresión de CDKN1A. El procedimiento puede incluir o excluir la medición de la expresión de HIST1H2BF. El procedimiento puede incluir o excluir medir la expresión de HIST1H1B. El procedimiento puede incluir o excluir medir la expresión del transcrito ID 7896697. El procedimiento puede incluir o excluir medir la expresión de JADE1. El procedimiento puede incluir o excluir la medición de la expresión de KIAA0125. El procedimiento puede incluir o excluir la medición de la expresión de CHRNA5. El procedimiento puede incluir o excluir la medición de la expresión de CHAF1 B /// MORC3. El procedimiento puede incluir o excluir la medición de la expresión de SEL1L3. El procedimiento puede incluir o excluir la medición de la

expresión de SSRP1. El procedimiento puede incluir o excluir la medición de la expresión de FOXM1. El procedimiento puede incluir o excluir la medición de la expresión de LMNB1. El procedimiento puede incluir o excluir la medición de la expresión de HIST1H2AK. El procedimiento puede incluir o excluir la medición de la expresión de CENPA. El procedimiento puede incluir o excluir la medición de la expresión de NCAPH. El procedimiento puede incluir o excluir la medición de la expresión de HIST1H2BB. El procedimiento puede incluir o excluir la medición de la expresión de ATP6V1 B2. El procedimiento puede incluir o excluir la medición de la expresión de TRAP1.

Se desvela también en el presente documento que la etapa (c) puede comprender o consistir en la medición de la expresión de uno o más biomarcadores enumerados en la Tabla A(iii), por ejemplo, 1, 2, 3, 4, 5, 6 o 7 de los biomarcadores enumerados en la Tabla A(iii). Por ejemplo, la etapa (c) comprende o consiste en la medición de la expresión de todos los biomarcadores enumerados en la Tabla A (iii).

El procedimiento puede incluir o excluir la medición de la expresión de PFAS. El procedimiento puede incluir o excluir la medición de la expresión de TMEM97. El procedimiento puede incluir o excluir la medición de la expresión de DHCR24. El procedimiento puede incluir o excluir la medición de la expresión de FTH1. El procedimiento puede incluir o excluir la medición de la expresión de HIST1H2AE. El procedimiento puede incluir o excluir la medición de la expresión de NQO1. El procedimiento puede incluir o excluir la medición de la expresión de CD44.

Se desvela también en el presente documento que la etapa (c) comprende o consiste en la medición de la expresión de tres o más de los biomarcadores enumerados en la Tabla A, por ejemplo, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 o 51 de los biomarcadores enumerados en la Tabla A. Por ejemplo, la etapa (c) puede comprender o consistir en la medición de la expresión de todos los biomarcadores enumerados en la Tabla A.

El procedimiento puede incluir o excluir la medición de la expresión de HIST1H2BM. El procedimiento puede incluir o excluir la medición de la expresión de HIST1H4B. El procedimiento puede incluir o excluir la medición de la expresión de HIST1H1D. El procedimiento puede incluir o excluir la medición de la expresión de PAICS. El procedimiento puede incluir o excluir la medición de la expresión de HIST1H4D. El procedimiento puede incluir o excluir la medición de la expresión de HIST1H2AL /// HIST1H2BN. El procedimiento puede incluir o excluir la medición de la expresión de PLK1. El procedimiento puede incluir o excluir la medición de la expresión de PHGDH. El procedimiento puede incluir o excluir la medición de la expresión de MCM2. El procedimiento puede incluir o excluir la medición de la expresión de CM7. El procedimiento puede incluir o excluir la medición de la expresión de CD53. El procedimiento puede incluir o excluir la medición de la expresión de KIFC1. El procedimiento puede incluir o excluir la medición de la expresión de WEE1. El procedimiento puede incluir o excluir medir la expresión de TMPC. El procedimiento puede incluir o excluir medir la expresión de MCM4. El procedimiento puede incluir o excluir la medición de la expresión de PSRC1. El procedimiento puede incluir o excluir la medición de la expresión de RNF149. El procedimiento puede incluir o excluir la medición de la expresión de MCM6. El procedimiento puede incluir o excluir la medición de la expresión de MCM5. El procedimiento puede incluir o excluir la medición de la expresión de FANCA.

El procedimiento puede incluir o excluir la medición de la expresión de CAD. El procedimiento puede incluir o excluir la medición de la expresión de MRT04. El procedimiento puede incluir o excluir la medición de la expresión de TOE1. El procedimiento puede incluir o excluir la medición de la expresión de CDC20. El procedimiento puede incluir o excluir la medición de la expresión de PM20D2. El procedimiento puede incluir o excluir la medición de la expresión de LOC344887. El procedimiento puede incluir o excluir la medición de la expresión de UNG. El procedimiento puede incluir o excluir la medición de la expresión de CDKN1A. El procedimiento puede incluir o excluir la medición de la expresión de HIST1H2BF. El procedimiento puede incluir o excluir medir la expresión de HIST1H1B. El procedimiento puede incluir o excluir medir la expresión del transcrito ID 7896697. El procedimiento puede incluir o excluir medir la expresión de JADE1. El procedimiento puede incluir o excluir medir la expresión de KIAA0125. El procedimiento puede incluir o excluir medir la expresión de CHRNA5. El procedimiento puede incluir o excluir la medición de la expresión de CHAF1B /// MORC3. El procedimiento puede incluir o excluir la medición de la expresión de SEL1L3. El procedimiento puede incluir o excluir la medición de la expresión de SSRP1. El procedimiento puede incluir o excluir la medición de la expresión de FOXM1. El procedimiento puede incluir o excluir la medición de la expresión de LMNB1. El procedimiento puede incluir o excluir la medición de la expresión de HIST1H2AK. El procedimiento puede incluir o excluir la medición de la expresión de CENPA. El procedimiento puede incluir o excluir la medición de la expresión de NCAPH. El procedimiento puede incluir o excluir la medición de la expresión de HIST1H2BB. El procedimiento puede incluir o excluir la medición de la expresión de ATP6V1B2. El procedimiento puede incluir o excluir la medición de la expresión de TRAP1.

El procedimiento puede incluir o excluir la medición de la expresión de PFAS. El procedimiento puede incluir o excluir la medición de la expresión de TMEM97. El procedimiento puede incluir o excluir la medición de la expresión de DHCR24. El procedimiento puede incluir o excluir la medición de la expresión de FTH1. El procedimiento puede incluir o excluir la medición de la expresión de HIST1H2AE. El procedimiento puede incluir o excluir la medición de la expresión de NQO1. El procedimiento puede incluir o excluir la medición de la expresión de CD44.

El procedimiento puede incluir o excluir la medición de la expresión de uno o más entre MCM7, PFAS, HIST1H2AE.

Se desvela también en el presente documento que la etapa (c) puede comprender o consistir en la medición de la expresión de uno o más biomarcadores enumerados en la Tabla B. Por ejemplo, la etapa (c) puede comprender o consistir en la medición de la expresión de 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 de los biomarcadores enumerados en la Tabla B. Por ejemplo, la etapa (c) puede comprender o consistir en medir la expresión de todos los biomarcadores enumerados en la Tabla B.

5

Se desvela también en el presente documento que la etapa (c) comprende o consiste en la medición de la expresión de al menos 2 de los biomarcadores enumerados en la Tabla B, por ejemplo, ≥ 3 , ≥ 4 , ≥ 5 , ≥ 6 , ≥ 7 , ≥ 8 , ≥ 9 , ≥ 10 , ≥ 11 , ≥ 12 , ≥ 13 , ≥ 14 , ≥ 15 , ≥ 16 , ≥ 17 o 18 de los biomarcadores enumerados en la Tabla B.

Se desvela también en el presente documento que la etapa (c) comprende o consiste en medir la expresión de uno o más biomarcadores enumerados en la Tabla B (i), por ejemplo, ≥ 2 , ≥ 3 , ≥ 4 , ≥ 5 , ≥ 6 , ≥ 7 , ≥ 8 , ≥ 9 , ≥ 10 , ≥ 11 , ≥ 12 , ≥ 13 , ≥ 14 , ≥ 15 o 16 de los biomarcadores enumerados en la Tabla B (i).

10

Se desvela también en el presente documento que la etapa (c) comprende o consiste en la medición de la expresión de uno o más biomarcadores enumerados en la Tabla B (ii), por ejemplo, 2 de los biomarcadores enumerados en la Tabla B (ii).

El procedimiento puede incluir o excluir la medición de la expresión de CCNA2. El procedimiento puede incluir o excluir la medición de la expresión del transcrito ID 8151252. El procedimiento puede incluir o excluir la medición de la expresión de PIGW. El procedimiento puede incluir o excluir la medición de la expresión de SNRPD1. El procedimiento puede incluir o excluir la medición de la expresión de ARHGAP19 /// ARHGAP19-SLIT1. El procedimiento puede incluir o excluir la medición de la expresión de HISTH2AB. El procedimiento puede incluir o excluir la medición de la expresión de LRCH2. El procedimiento puede incluir o excluir la medición de la expresión de PKP4 /// PKP4. El procedimiento puede incluir o excluir la medición de la expresión de RAVR2 /// RAVR2. El procedimiento puede incluir o excluir la medición de la expresión de DUT /// DUT. El procedimiento puede incluir o excluir la medición de la expresión de AURKA. El procedimiento puede incluir o excluir medir la expresión del transcrito ID 8055309. El procedimiento puede incluir o excluir la medición de la expresión de NDC1. El procedimiento puede incluir o excluir la medición de la expresión de KIF22 /// KIF22. El procedimiento puede incluir o excluir la medición de la expresión de OK/SW-CL.58. El procedimiento puede incluir o excluir medir la expresión del transcrito ID 7994343. El procedimiento puede incluir o excluir medir la expresión de HIST1H2AB /// HIST1H2AE. El procedimiento puede incluir o excluir la medición de la expresión de HMGB3.

15

20

25

"Expresión" significa la presencia, nivel y/o cantidad del biomarcador.

Dentro de "biomarcador" se incluye cualquier molécula biológica, o componente o fragmento de la misma, cuya medición puede proporcionar información útil para determinar la potencia del sensibilizador. Por lo tanto, en el contexto de la Tabla A, el biomarcador puede ser una molécula de ácido nucleico, como un ARNm o ADNc. Alternativamente, el biomarcador puede ser una proteína codificada por la molécula de ácido nucleico o resto de carbohidrato, o un componente antigénico o fragmento del mismo.

30

En una realización adicional o alternativa, se sabe ya o se sospecha que el agente de ensayo es, capaz de inducir sensibilización de la piel.

35

Por ejemplo, se puede saber ya que el agente de ensayo es capaz de inducir sensibilización de la piel aplicando un procedimiento ya conocido en la técnica, por ejemplo, un procedimiento descrito en la patente internacional WO 2012/056236 y/o Johansson H y col. The GARD assay for assessment of chemical skin sensitizers. Toxicology in vitro 2013.

40

En una realización alternativa o adicional, el procedimiento sirve para identificar la potencia del sensibilizador de la piel y el estatus del sensibilizador/no sensibilizador de la piel del agente de ensayo (es decir, identificar si el agente de ensayo es un sensibilizador o no e identificar su potencia como sensibilizador de la piel). En una realización alternativa o adicional, el procedimiento comprende identificar si el agente de ensayo es un sensibilizador aplicando el procedimiento descrito en la patente internacional WO 2012/056236 y/o Johansson H y col.

45

En "potencia de sensibilización de la piel" se incluye o significa la fuerza de la capacidad de sensibilizar la piel de un agente. Por ejemplo, la potencia relativa o la fuerza de la capacidad de sensibilización de un agente podrían conllevar ordenar un grupo de agentes de ensayo desde el más potente al menos potente o *viceversa*, y/o podría conllevar su categorización de acuerdo con uno o más reglamentos o sistemas conocidos.

En una realización adicional o alternativa, la potencia de sensibilización de la piel determinada según el procedimiento se clasifica de acuerdo con el Reglamento Europeo de Clasificación, Etiquetado y Envasado (CLP) (CE) 1272/2008 (<http://echa.europa.eu/clp-2015>). Este sistema se basa en el Sistema Globalmente Armonizado (SGA) de Naciones Unidas y, a partir de junio de 2015, la única legislación que se aplique a la clasificación y el etiquetado de sustancias y mezclas. Requiere que las empresas clasifiquen, etiqueten y envasen sus productos de manera apropiada antes de comercializarlos. Proporciona las categorías: 1A (fuerte), 1B (débil) o sin categoría (no es sensibilizador).

50

55

Por ejemplo, el procedimiento puede proporcionar:

- (i) uno o más agentes de categoría de potencia 1A;
- (ii) uno o más agentes de categoría de potencia 1B; y/o
- (iii) uno o más agentes de categoría de potencia sin categoría

- 5 En una realización adicional o alternativa, la potencia de sensibilización de la piel determinada según el procedimiento se clasifica de acuerdo con el sistema descrito en Basketter y col., 2014, Categorization of chemicals according to their relative human skin sensitizing potency, *Dermatitis*, 25(1):11-21, es decir, categorías 1 (sensibilizador más fuerte), 2, 3, 4, 5 o 6 (verdadero no sensibilizador) (p.ej., Tabla 4, Figura 4).

Por ejemplo, el procedimiento puede proporcionar:

- 10 (i) uno o más agentes de la categoría de potencia 1;
 (ii) uno o más agentes de la categoría de potencia 2;
 (iii) uno o más agentes de la categoría de potencia 3;
 (iv) uno o más agentes de la categoría de potencia 4;
 (v) uno o más agentes de la categoría de potencia 5; y/o
 15 (vi) uno o más agentes de la categoría de potencia 6 (p. ej., véase la Tabla 8 y/o Basketter y col., 2014 supra.).

En una realización adicional o alternativa, la potencia de sensibilización de la piel se clasifica de acuerdo con la clasificación del ensayo con ganglios linfáticos locales (LLNA), la prueba de maximización con cobayas (GPMT) o prueba de niveles sin efecto observable (NOEL).

- 20 Para una descripción detallada de LLNA, véase Basketter, DA, y col., Local lymph node assay - validation, conduct and use in practice. *Food Chem Toxicol.*, 2002. 40(5): p. 593-8. Para una descripción detallada de la prueba de maximización con cobayas véase Magnusson, B. and A.M. Kligman, The identification of contact allergens by animal assay. The guinea pig maximization assay. *J Invest Dermatol*, 1969. 52(3): p. 268-76. Para una descripción detallada de la prueba de niveles sin efecto observable (NOEL) en relación con la potencia de sensibilización de la piel véase Basketter y col., 2005, "Evaluation of the skin sensitizing potency of chemicals by using the existing methods and considerations of relevance for elicitation" *Contact Dermatitis*, 52(1):39-43; y Griem, P., y col., 2003, "Proposal for a risk assessment methodology for skin sensitization based on sensitization potency data." *Regul. Toxicol. Pharmacol.*, 38:269-290. Para una correlación entre NOEL y los niveles de potencia, véase también Los datos de catalogación de la biblioteca de la OMS. Sensibilización cutánea en la evaluación del riesgo químico documento de proyecto de armonización de IPCS: no. 5), ISBN 978 92 4 156360 4 (en particular, la Tabla 1 en las páginas 26-28) Para una descripción detallada de CLP, véase (<http://echa.europa.eu/clp-2015>).
- 25 30

En una realización adicional o alternativa, la expresión de los biomarcadores medidos en la etapa (c) se mide en las células proporcionadas en la etapa (a) antes y después de la exposición al agente sensibilizador de la piel de potencia predeterminada, y en la que la diferencia en la expresión entre los biomarcadores antes y después de la exposición al agente de ensayo es indicativa de la potencia del agente sensibilizador de la piel de la etapa (b).

- 35 En una realización adicional o alternativa, la expresión de los biomarcadores medidos en la etapa (c) se mide en las células proporcionadas en la etapa (a) antes y después de la exposición al agente sensibilizador de la piel de potencia predeterminada, y en la que la diferencia en la expresión entre los biomarcadores antes y después de la etapa (c) es indicativa de la potencia del agente sensibilizador de la piel de la etapa (b).

- 40 En "diferencia en la expresión" los autores de la invención incluyen que la presencia y/o cantidad en una primera muestra (p.ej., una muestra de agente de ensayo) difiere de la de una segunda muestra (p.ej., una muestra de agente de control). Preferentemente, la presencia y/o cantidad no es superior a 40 % de la de la muestra comparativa, por ejemplo, no superior a 39 %, 38 %, 37 %, 36 %, 35 %, 34 %, 33 %, 32 %, 31 %, 30 %, 29 %, 28 %, 27 %, 26 %, 25 %, 24 %, 23 %, 22 %, 21 %, 20 %, 19 %, 18 %, 17 %, 16 %, 15 %, 14 %, 13 %, 12 %, 11 %, 10 %, 9 %, 8 %, 7 %, 6 %, 5 %, 4 %, 3 %, 2 %, 1 % o 0 %.

- 45 En una realización adicional o alternativa, se miden los biomarcadores en las células proporcionadas en la etapa (a) antes y después de la exposición al agente de ensayo, y en la que la diferencia en la expresión entre los biomarcadores antes y después de la exposición al agente de ensayo es indicativa de la potencia de sensibilización de la piel del agente de ensayo de la etapa (b). Por lo tanto, las células proporcionadas en la etapa (a) pueden proporcionar tanto el control negativo como el resultado de ensayo.

- 50 En una realización adicional o alternativa, los biomarcadores se miden en las células proporcionadas en la etapa (a) antes y después de la exposición al agente de ensayo, y en la que la diferencia en la expresión entre los biomarcadores antes y después de la etapa (c) es indicativa de la potencia de sensibilización de la piel del agente de ensayo de la etapa (b). Por lo tanto, las células proporcionadas en la etapa (a) pueden proporcionar tanto el control negativo como el resultado de ensayo.

- 55 En una realización adicional o alternativa, el procedimiento comprende las etapas adicionales de:

- (d) proporcionar una población adicional de células dendríticas o una población de células de tipo dendrítico;
- (e) exponer las células proporcionadas en la etapa (d) a:

- i. uno o más agentes de control negativo que no sensibiliza la piel humana; y/o
- ii) uno o más agentes de control positivo que sensibilizan la piel humana; y/o
- iii) uno o más vehículo de control;

- (f) incubar durante el mismo período y aparte de la exposición a un agente el uno o más agentes de control, en las mismas condiciones que las células proporcionadas en la etapa (a);
- (g) medir en las células de la etapa (e) la expresión de los dos o más biomarcadores medidos en la etapa (c);

en el que la diferencia en la expresión entre los biomarcadores medidos en la etapa (c) y los biomarcadores medidos en la etapa (e) es indicativa de la potencia de sensibilización de la piel del agente de ensayo.

En una realización adicional o alternativa, el procedimiento comprende las etapas adicionales de:

- (d) proporcionar una población adicional de células dendríticas o una población de células de tipo dendrítico;
- (e) exponer las células proporcionadas en la etapa (d) a:

- i. un agente de control negativo que no sensibiliza la piel humana;
- ii) un vehículo de control; o
- iii) incubación durante el mismo periodo y aparte de la exposición a un agente de potencia predeterminada de sensibilización de la piel, en las mismas condiciones que las células proporcionadas en la etapa (a);

- (f) medir en las células de la etapa (e) la expresión de los biomarcadores medidos en la etapa (c);

en el que la diferencia en la expresión entre los biomarcadores medidos en la etapa (c) y los biomarcadores medidos en la etapa (e) es indicativa de la potencia de sensibilización de la piel del agente de ensayo.

En una realización adicional o alternativa, el vehículo de control comprende DMSO.

En "diferencia en la expresión" se incluye que la presencia y/o cantidad en una primera muestra (p. ej., una muestra de agente de ensayo) difiere de la de una segunda muestra (p. ej., una muestra de agente de control). Preferentemente, la presencia y/o cantidad no es superior a 40 % de la de la muestra de comparación, por ejemplo, no superior a 39 %, 38 %, 37 %, 36 %, 35 %, 34 %, 33 %, 32 %, 31 %, 30 %, 29 %, 28 %, 27 %, 26 %, 25 %, 24 %, 23 %, 22 %, 21 %, 20 %, 19 %, 18 %, 17 %, 16 %, 15 %, 14 %, 13 %, 12 %, 11 %, 10 %, 9 %, 8 %, 7 %, 6 %, 5 %, 4 %, 3 %, 2 %, 1 % o 0 %.

En una realización adicional o alternativa, el procedimiento comprende las etapas adicionales de:

- (h) proporcionar una población adicional de células dendríticas o una población de células de tipo dendrítico;

- (i) exponer las células proporcionadas en la etapa (h) a un agente sensibilizador de la piel de potencia predeterminada;

- (j) medir en las células de la etapa (i) la expresión de los biomarcadores medidos en la etapa (c);

en el que la correspondencia de la expresión entre los biomarcadores medidos en la etapa (c) y los biomarcadores medidos en la etapa (j) es indicativa de la potencia sensibilizadora de la piel del agente de ensayo.

En "correspondencia de la expresión" se incluye que la presencia y/o cantidad en una primera muestra (p. ej., una muestra de agente de ensayo) es similar o idéntica a la presencia y/o cantidad en una segunda muestra (p. ej., una muestra de control). Preferentemente, la presencia y/o cantidad es al menos 60 % de la de la muestra de control, por ejemplo, al menos 65 %, 66 %, 67 %, 68 %, 69 %, 70 %, 71 %, 72 %, 73 %, 74 %, 75 %, 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 %.

En una realización adicional o alternativa, el procedimiento comprende la etapa adicional de:

- (k) identificar la potencia de sensibilización de la piel del agente de ensayo.

La identificación se puede realizar utilizando cualquier procedimiento estadístico adecuado o algoritmo de aprendizaje automático conocido en la técnica, como bosque aleatorio (RF), máquina de vectores de soporte (SVM), Análisis de componente principal (PCA), mínimos cuadrados ordinarios (OLS), regresión de mínimos cuadrados parciales (PLS), regresión de mínimos cuadrados parciales ortogonales (O-PLS) y otros análisis estadísticos multivariados (p. ej., modelo de regresión logística escalonada hacia atrás). Para una revisión análisis estadísticos multivariado, véase, por ejemplo, Schervish, Mark J. (noviembre de 1987). "A Review of Multivariate Analysis". Statistical Science 2 (4): 396-41. Preferentemente se utiliza un modelo de bosque aleatorio (tal como se describe en detalle en el Ejemplo A).

- En una realización adicional o alternativa de los procedimientos de la invención, la potencia de sensibilización se determina con una AUC ROC de al menos 0,45, por ejemplo con una AUC ROC de al menos, 0,50, 0,55, 0,60, 0,65, 0,70, 0,75, 0,80, 0,85, 0,86, 0,87, 0,88, 0,89, 0,90, 0,91, 0,92, 0,93, 0,94, 0,95, 0,96, 0,97, 0,98, 0,99 o con una ROC AUC de 1,00. Preferentemente, la potencia de sensibilización de la piel se determina con una AUC ROC de al menos 0,80, más preferentemente al menos 0,85, aún más preferentemente al menos 0,88.
- Preferentemente, los procedimientos de la invención tienen una precisión de al menos 60 %, por ejemplo, 61 %, 62 %, 63 %, 64 %, 65 %, 66 %, 67 %, 68 %, 69 %, 70 %, 71 %, 72 %, 73 %, 74 %, 75 %, 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de precisión.
- Preferentemente, los procedimientos de la invención tienen una sensibilidad de al menos 60 %, por ejemplo, 61 %, 62 %, 63 %, 64 %, 65 %, 66 %, 67 %, 68 %, 69 %, 70 %, 71 %, 72 %, 73 %, 74 %, 75 %, 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de sensibilidad.
- Preferentemente, los procedimientos de la invención tienen una especificidad de al menos 60 %, por ejemplo, 61 %, 62 %, 63 %, 64 %, 65 %, 66 %, 67 %, 68 %, 69 %, 70 %, 71 %, 72 %, 73 %, 74 %, 75 %, 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de especificidad.
- Según los autores de la invención "precisión" significa la proporción de resultados correctos de un procedimiento, "sensibilidad" significa la proporción de todos los productos químicos positivos que se clasifican correctamente como positivos, y "especificidad" significa la proporción de todos los productos químicos negativos que se clasifican correctamente como negativos.
- En una realización adicional o alternativa, el procedimiento comprende la etapa adicional de:
- (l) una o más repeticiones de las etapas (h) a (j) (y, si está presente, la etapa (k)).
- En una realización adicional o alternativa, cada repetición de las etapas (a) a (c) (y cuando están presentes, las etapas (d) - (f/g)) se realizan simultáneamente (es decir, con las etapas originales (a) a (c) (y cuando están presentes las etapas (d) - (f/g))). En una realización adicional o alternativa, cada repetición de las etapas (a) a (c) (y cuando están presentes, las etapas (d) - (f/g)) se realiza consecutivamente.
- En una realización adicional o alternativa, la etapa (j) comprende o consiste en una o más repeticiones en las que:
- (I) el mismo agente de potencia de sensibilización de la piel predeterminada tal como se proporciona en la etapa (i) para una o más repeticiones; o
- (II) un agente diferente de potencia de sensibilización de la piel predeterminada tal como se proporciona en la etapa (i) para una o más repeticiones.
- En una realización adicional o alternativa, el agente diferente de potencia de sensibilización de la piel predeterminada proporcionado en la etapa (i) se selecciona del grupo que consiste en:
- (I) un agente de la misma potencia de sensibilización que el agente diferente de potencia de sensibilización de la piel predeterminada proporcionado en la etapa (i); o
- (II) un agente de una potencia de sensibilización diferente del agente diferente de la potencia de sensibilización de la piel predeterminada proporcionado en la etapa (i).
- En una realización adicional o alternativa, las etapas (a) a (c) (y cuando están presentes, las etapas (f/g)) se repiten al menos una vez, por ejemplo, ≥ 2 , ≥ 3 , ≥ 4 , ≥ 5 , ≥ 6 , ≥ 7 , ≥ 8 , ≥ 9 , ≥ 10 , ≥ 11 , ≥ 12 , ≥ 13 , ≥ 14 , ≥ 15 , ≥ 16 , ≥ 17 , ≥ 18 , ≥ 19 , ≥ 20 , ≥ 21 , ≥ 22 , ≥ 23 , ≥ 24 , ≥ 25 , ≥ 26 , ≥ 27 , ≥ 28 , ≥ 29 , ≥ 30 , ≥ 40 , ≥ 50 , ≥ 60 , ≥ 70 , ≥ 80 , ≥ 90 , ≥ 100 , ≥ 150 , ≥ 200 , ≥ 250 , ≥ 300 , ≥ 350 , ≥ 400 , ≥ 450 , ≥ 500 , ≥ 600 , ≥ 700 , ≥ 800 , ≥ 900 o ≥ 1000 repeticiones.
- En una realización adicional o alternativa, cada etapa o repetición de etapas puede repetirse idénticamente para controlar el error experimental, por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9 o 10 repeticiones.
- En una realización adicional o alternativa, la expresión de los biomarcadores medidos se mide para agentes de potencia de sensibilización de la piel predeterminada a lo largo de un intervalo completo de potencias de sensibilización de la piel.
- Los autores de la invención incluyen en "intervalo completo de potencias de sensibilización de la piel" la evaluación de los sensibilizadores del grado más alto y más bajo de la categoría de sensibilizadores de la piel del sistema de categorización utilizado. Preferentemente, los sensibilizadores se evalúan en cada categoría de sensibilizador del sistema de categorización que se utiliza.
- En una realización adicional o alternativa, el agente sensibilizador de la piel de potencia predeterminada comprende o consiste en uno o más agentes seleccionados del grupo que consiste en el grupo definido en la Tabla 2. Por

ejemplo, el procedimiento puede comprender la medición de ≥ 2 , ≥ 3 , ≥ 4 , ≥ 5 , ≥ 6 , ≥ 7 , ≥ 8 , ≥ 9 , ≥ 10 u 11, de los agentes definidos en la Tabla 2.

En una realización adicional o alternativa, el agente sensibilizador de la piel de potencia predeterminada comprende o consiste en agentes seleccionados del grupo que consiste en el grupo definido en la Tabla 8, Tabla 9 y/o Tabla 8, por ejemplo, solo la Tabla 8, solo la Tabla 9, solo la Tabla 10, las Tablas 8 y 9, las Tablas 8 y 10, las Tablas 9 y 10 o las Tablas 8, 9 y 10. Por ejemplo, el procedimiento puede comprender medir ≥ 2 , ≥ 3 , ≥ 4 , ≥ 5 , ≥ 6 , ≥ 7 , ≥ 8 , ≥ 9 , ≥ 10 , ≥ 11 , ≥ 12 , ≥ 13 , ≥ 14 , ≥ 15 , ≥ 16 , ≥ 17 , ≥ 18 , ≥ 19 , ≥ 20 , ≥ 21 , ≥ 22 , ≥ 23 , ≥ 24 , ≥ 25 , ≥ 26 , ≥ 27 , ≥ 28 , ≥ 29 , ≥ 30 , ≥ 31 , ≥ 32 , ≥ 33 , ≥ 34 , ≥ 35 , ≥ 36 , ≥ 37 , ≥ 38 , ≥ 39 , ≥ 40 , ≥ 41 , ≥ 42 , ≥ 43 , ≥ 44 , ≥ 45 , ≥ 46 , ≥ 47 , ≥ 48 , ≥ 49 , ≥ 50 , ≥ 51 , ≥ 52 , ≥ 53 , ≥ 54 , ≥ 55 , ≥ 56 , ≥ 57 , ≥ 58 , ≥ 59 , ≥ 60 , ≥ 61 , ≥ 62 , ≥ 63 , ≥ 64 , ≥ 65 , ≥ 66 , ≥ 67 , ≥ 68 , ≥ 69 , ≥ 70 , ≥ 71 , ≥ 72 , ≥ 73 , ≥ 74 , ≥ 75 , ≥ 76 , ≥ 77 , ≥ 78 , ≥ 79 , ≥ 80 , ≥ 81 , ≥ 82 , ≥ 83 , ≥ 84 , ≥ 85 , ≥ 86 , ≥ 87 , ≥ 88 , ≥ 89 , ≥ 90 , ≥ 91 , ≥ 92 , ≥ 93 , ≥ 94 , ≥ 95 , ≥ 96 , ≥ 97 , ≥ 98 , ≥ 99 , ≥ 100 , ≥ 101 , ≥ 102 , ≥ 103 , ≥ 104 , ≥ 105 , ≥ 106 , ≥ 107 , ≥ 108 , ≥ 109 , ≥ 110 , ≥ 111 , ≥ 112 , ≥ 113 , ≥ 114 , ≥ 115 , ≥ 116 , ≥ 117 , ≥ 118 , ≥ 119 , ≥ 120 , ≥ 121 o 122 de los agentes de control definidos en las Tablas 8, 9 y 10.

En una realización adicional o alternativa, el agente sensibilizador de la piel de potencia predeterminada comprende o consiste en agentes seleccionados del grupo que consiste en 1-butanol, ácido 4-aminobenzoico, benzaldehído, clorobenceno, ftalato de dietilo, dimetilformamida, etil vanilina, glicerol, isopropanol, ácido láctico, salicilato de metilo, ácido octanoico, propilen glicol., fenol, ácido p-hidroxibenzoico, permanganato de potasio, ácido salicílico, dodecil sulfato de sodio, Tween 80, sulfato de zinc, 2,4-dinitroclorobenceno, oxazolona, dicromato de potasio, Kathon CG (MC/MCI), formaldehído, 2-aminofenol, 2-nitro-1,4-fenilendiamina, p-fenilendiamina, aldehído hexilcinámico, acrilato de 2-hidroxietilo, 2-mercaptobenzotiazol, glioxal, cinamaldehído, isoeugenol, etilen diamina, resorcinol, alcohol cinámico, eugenol, penicilina G, geraniol y DMSO.

En una realización preferente, la etapa (c) comprende o consiste en la medición de la expresión de una molécula de ácido nucleico de uno o más de los biomarcadores. La molécula de ácido nucleico puede ser una molécula de ADNc o una molécula de ARNm. Preferentemente, la molécula de ácido nucleico es una molécula de ARNm. Sin embargo, la molécula de ácido nucleico puede ser una molécula de ADNc.

En una realización, la medición de la expresión de uno o más de los biomarcadores en la etapa (c) se realiza aplicando un procedimiento seleccionado del grupo que consiste en hibridación Southern, hibridación Northern, reacción en cadena de la polimerasa (PCR), PCR de transcriptasa inversa (RT-PCR), PCR cuantitativa en tiempo real (qRT-PCR), nanomatriz, micromatriz, macromatriz, autorradiografía e hibridación *in situ*. Preferentemente, la expresión de uno o más biomarcadores se mide utilizando una micromatriz de ADN.

En una realización adicional o alternativa, el uno o más biomarcadores medidos en la etapa (c) se miden utilizando una matriz (p.ej., una matriz de ADN). En una realización adicional o alternativa, el uno o más biomarcadores medidos en la etapa (c) se miden utilizando una matriz de genoma completa (p.ej., la matriz Affymetrix Human Gene 1.0 ST o la matriz Affymetrix Human Gene 2.0 ST). En una realización alternativa o adicional, se utiliza el sistema Nanostring nCounter (p.ej., conjuntos de códigos personalizados de Nanostring nCounter basados en la selección de una matriz de genoma completo (p.ej., matriz Affymetrix Human Gene 1.0 ST o matriz Affymetrix Human Gene 2.0 ST).

El procedimiento puede comprender medir la expresión de uno o más biomarcadores en la etapa (c) utilizando uno o más restos de unión, cada uno capaz de unirse selectivamente a una molécula de ácido nucleico que codifica uno de los biomarcadores identificados en la Tabla A.

En una realización, cada uno del uno o más restos de unión comprende o consiste en una molécula de ácido nucleico. En una realización adicional, cada uno del uno o más restos de unión comprende o consiste en ADN, ARN, ANP, ANL, GNA, TNA o PMO. Preferentemente, cada uno del uno o más restos de unión comprende o consiste en ADN. En una realización, cada uno del uno o más restos de unión tiene una longitud de 5 a 100 nucleótidos. Sin embargo, en una realización alternativa, tienen una longitud de 15 a 35 nucleótidos.

El uno o más restos de unión puede comprender o consistir en una o más sondas de la matriz Human Gene 1.0 ST (Affymetrix, Santa Clara, CA, Estados Unidos). Los números de identificación de la sonda se proporcionan en la Tabla A en el presente documento, en las que se denominan "ID de agrupación de transcritos".

Los agentes de unión adecuados (también denominados moléculas de unión o restos de unión) pueden seleccionarse o explorarse desde una biblioteca en función de su capacidad para unirse a un motivo de ácido nucleico, proteína o aminoácido dado, tal como se explica más adelante.

En una realización preferente, el resto de unión comprende un resto detectable.

Los autores de la invención incluyen en "resto detectable" un resto que permite determinar su presencia y/o cantidad relativa y/o emplazamiento (por ejemplo, el emplazamiento en una matriz), ya sea directa o indirectamente.

Los restos detectables adecuados son muy conocidos dentro de la técnica.

Por ejemplo, el resto detectable puede ser un resto fluorescente y/o luminiscente y/o quimioluminiscente que puede

detectarse cuando se expone a condiciones específicas. Es posible que se requiera exponer a radiación (es decir, luz) dicho resto fluorescente a una longitud de onda e intensidad específicas para provocar la excitación del resto fluorescente, lo cual permite que emita fluorescencia detectable a una longitud de onda específica que puede detectarse.

- 5 Alternativamente, el resto detectable puede ser una enzima que es capaz de convertir un sustrato (preferentemente no detectable) en un producto detectable que se puede visualizar y/o detectar. Los ejemplos de enzimas adecuadas se explican con más detalle más adelante en relación con los ensayos ELISA, por ejemplo.

Por lo tanto, el resto detectable puede seleccionarse del grupo que consiste en: un resto fluorescente; un resto luminiscente; un resto quimioluminiscente; un resto radiactivo (por ejemplo, un átomo radiactivo); o un resto enzimático. Preferentemente, el resto detectable comprende o consiste en un átomo radiactivo. El átomo radiactivo se puede seleccionar del grupo que consiste en tecnecio-99m, yodo-123, yodo-125, yodo-131, indio-111, flúor-19, carbono-13, nitrógeno-5, oxígeno-17, fósforo-32, azufre-35, deuterio, tritio, renio-186, renio-188 e itrio-90.

Claramente, el agente que se detecte (como, por ejemplo, el uno o más biomarcadores en la muestra de ensayo y/o la muestra de control descritas en el presente documento y/o una molécula de anticuerpo para su uso en la detección de una proteína seleccionada) debe tener suficientes isótopos atómicos apropiados para que pueda detectarse fácilmente el resto detectable.

En una realización preferente alternativa, el resto detectable del resto de unión es un resto fluorescente.

Pueden incorporarse radiomarcadores u otros marcadores a los biomarcadores presentes en las muestras de los procedimientos de la invención y/o los restos de unión de la invención de formas conocidas. Por ejemplo, si el agente de unión es un polipéptido, se puede biosintetizar o se puede sintetizar por síntesis química de aminoácidos utilizando precursores de aminoácidos adecuados que implican, por ejemplo, flúor-19 en lugar de hidrógeno. Pueden unirse por ejemplo marcadores como ^{99m}Tc , ^{123}I , ^{186}Rh , ^{188}Rh y ^{111}In a través de restos de cisteína en el resto de unión. Puede unirse itrio-90 a través de un resto de lisina. El procedimiento IODOGEN (Fraker y col. (1978) Biochem. Biophys. Res. Comm. 80, 49-57) puede utilizarse para incorporar ^{125}I . La referencia ("Monoclonal Antibodies in Immunoscintigraphy", JF Chatal, CRC Press, 1989) describe otros procedimientos en detalle. Los procedimientos para conjugar otros restos detectables (tales como restos enzimáticos, fluorescentes, luminiscentes, quimioluminiscentes o radioactivos) a proteínas son muy conocidos en la técnica.

Las personas expertas en la materia apreciarán que los biomarcadores en la(s) muestra(s) que se analicen pueden marcarse con un resto que ayuda indirectamente a determinar la presencia, cantidad y/o emplazamiento de dichas proteínas. Por lo tanto, el resto puede constituir un componente de un resto detectable multicomponente. Por ejemplo, los biomarcadores en la(s) muestra(s) que se analicen pueden marcarse con biotina, lo cual permite su posterior detección utilizando estreptavidina fusionada o unida de otro modo a un marcador detectable.

El procedimiento proporcionado en el primer aspecto de la presente invención puede comprender o consistir en la etapa (c), determinar la expresión de la proteína de uno o más biomarcadores definidos en la Tabla A. El procedimiento puede comprender medir la expresión de uno o más biomarcadores en la etapa (c) utilizando uno o más restos de unión, cada uno capaz de unirse selectivamente a uno de los biomarcadores identificados en la Tabla A. El uno o más restos de unión pueden comprender o consistir en un anticuerpo o un fragmento de unión a antígeno del mismo, como por ejemplo un anticuerpo monoclonal o fragmento del mismo.

El término "anticuerpo" incluye cualquier anticuerpo sintético, anticuerpos recombinantes o híbridos de anticuerpos, tales como, pero sin limitación, una molécula de anticuerpo de cadena simple producida por despliegue en fago de regiones constantes y/o variables y/o variables de cadena ligera y/o pesada de inmunoglobulina u otras moléculas inmunointeractivas capaces de unirse a un antígeno en un formato de inmunoensayo conocido entre las personas expertas en la materia.

Se incluye asimismo el uso de agentes de unión de tipo anticuerpo, como anticuerpos y aptámeros.

45 Se puede encontrar una revisión general de las técnicas relacionadas con la síntesis de fragmentos de anticuerpos que retienen sus sitios de unión específicos en Winter y Milstein (1991) Nature 349, 293-299.

Adicionalmente o alternativamente, una o más de las primeras moléculas de unión puede ser un aptámero (véase Collett y col., 2005, Procedimientos 37: 4-15).

Pueden utilizarse bibliotecas moleculares, como bibliotecas de anticuerpos (Clackson y col., 1991, Nature 352, 624-628; Marks y col., 1991, J Mol Biol 222 (3): 581-97), bibliotecas de péptidos (Smith, 1985, Science 228 (4705): 1315-7), bibliotecas de ADNc expresado (Santi y col., (2000) J Mol Biol 296 (2): 497-508), bibliotecas en otros andamios distintos del marco de anticuerpos, como affibodies (Gunneriusson y col., 1999, Appl Environ Microbiol 65 (9): 4134-40) o bibliotecas basadas en aptámeros (Kenan y col., 1999, Methods Mol Biol 118, 217-31) como una fuente de la cual se seleccionan moléculas de unión que son específicas para un motivo dado para su uso en los procedimientos de la invención.

Las bibliotecas moleculares pueden expresarse *in vivo* en células procariotas (Clackson y col., 1991, op. Cit; Marks y col., 1991, op. Cit.), o células eucariotas (Kieckhefer y col., 1999, Proc Natl Acad Sci EE.UU., 96 (10): 5651-6) o pueden expresarse *in vitro* sin la participación de células (Hanes & Pluckthun, 1997, Proc Natl Acad Sci EE.UU. 94 (10): 4937-42; He & Taussig, 1997, Nucleic Acids Res 25 (24): 5132-4; Nemoto y col., 1997, FEBS Lett, 414 (2): 405-8).

- 5 En los casos en los que se utilizan bibliotecas a base de proteínas, los genes que codifican las bibliotecas de posibles moléculas de unión suelen estar empaquetados en virus y la posible molécula de unión se despliega en la superficie del virus (Clackson y col., 1991, supra; Marks y col., 1991, supra; Smith, 1985, supra).

- 10 Quizás el sistema de despliegue más utilizado es fragmentos de anticuerpos que despliegan bacteriófago filamentoso en sus superficies, expresándose los fragmentos de anticuerpos como una fusión a la proteína de cubierta menor del bacteriófago (Clackson y col., 1991, supra; Marks y col., 1991, supra). Sin embargo, otros sistemas adecuados para el despliegue incluyen el uso de otros virus (patente europea EP 39578), bacterias (Gunneriusson y col., 1999, supra; Daugherty y col., 1998, Protein Eng 11 (9): 825-32; Daugherty y col., 1999, Protein Eng 12 (7): 613-21) y levadura (Shusta y col., 1999, J Mol Biol 292 (5): 949-56).

- 15 Por otra parte, se han desarrollado sistemas de despliegue que utilizan el enlace del producto polipeptídico con su ARNm codificador en los llamados sistemas de despliegue de ribosomas (Hanes y Pluckthun, 1997, supra; He y Taussig, 1997, supra; Nemoto y col., 1997, supra) o, alternativamente, la unión del producto polipeptídico con el ADN codificante (véase la patente de Estados Unidos Nº 5.856.090 y la patente internacional WO 98/37186).

- 20 Los dominios variable pesado (V_H) y variable ligero (V_L) del anticuerpo están relacionados con el reconocimiento de antígeno, un hecho reconocido en primer lugar por los primeros experimentos de digestión de proteasas. Se observó una confirmación adicional por "humanización" de anticuerpos de roedor. Los dominios variables de origen roedor pueden fusionarse con dominios constantes de origen humano de modo que el anticuerpo resultante retenga la especificidad antigénica del anticuerpo parental de roedor (Morrison y col (1984) Proc. Natl. Acad. Sci. EE.UU 81, 6851 -6855).

- 25 Se sabe que la especificidad antigénica es conferida por dominios variables y es independiente de los dominios constantes a partir de experimentos que implican la expresión bacteriana de fragmentos de anticuerpos, conteniendo todos ellos uno o más dominios variables. Estas moléculas incluyen moléculas de tipo Fab (Better y col. (1988) Science 240, 1041); Moléculas Fv (Skerra y col (1988) Science 240, 1038); moléculas Fv monocatenarias (ScFv) donde los dominios asociados V_H y V_L están unidos a través de un oligopéptido flexible (Bird y col. (1988) Science 242, 423; Huston y col. (1988) Proc. Natl. Acad. Sci. EE.UU. 85, 5879) y anticuerpos de dominio único (dAbs) que comprenden dominios V aislados (Ward y col., (1989) Nature 341, 544). Se puede encontrar una revisión general de las técnicas relacionadas con la síntesis de fragmentos de anticuerpos que retienen sus sitios de unión específicos en Winter & Milstein (1991) Nature 349, 293-299.

- 35 El anticuerpo o el fragmento de unión a antígeno se puede seleccionar del grupo que consiste en anticuerpos intactos, fragmentos Fv (p.ej., Fv monocatenario y Fv unido por disulfuro), fragmentos de tipo Fab (p.ej., fragmentos Fab, fragmentos Fab' y fragmentos F(ab')₂), dominios variables únicos (p.ej., dominios V_H y V_L) y anticuerpos de dominio (dAbs, incluyendo formatos simples y duales [es decir, dAb-engarce-dAb]). Preferentemente, el anticuerpo o fragmento de unión a antígeno es un Fv monocatenario (scFv).

El uno o más restos de unión pueden comprender o consistir alternativamente en un agente de unión de tipo anticuerpo, por ejemplo, un anticuerpo o un aptámero.

- 40 "Moléculas scFv" significa moléculas en las que los dominios asociados V_H y V_L están unidos a través de un oligopéptido flexible.

- 45 Las ventajas del uso de fragmentos de anticuerpos, en lugar de anticuerpos completos, son varias. El tamaño más reducido de los fragmentos puede conllevar mejores propiedades farmacológicas, como una mejor penetración en el tejido sólido. Se eliminan las funciones efectoras de los anticuerpos completos, como la unión del complemento. Los fragmentos de anticuerpos Fab, Fv, ScFv y dAb pueden expresarse y secretarse de *E. coli*, permitiendo así la fácil producción de grandes cantidades de dichos fragmentos.

Los anticuerpos completos y los fragmentos F(ab')₂ son "bivalentes". "Bivalente" significa que dichos anticuerpos y los fragmentos F(ab')₂ tienen dos sitios de combinación de antígeno. Por el contrario, los fragmentos Fab, Fv, ScFv y dAb son monovalentes y tienen solo un sitio de combinación de antígeno.

- 50 Los anticuerpos pueden ser monoclonales o policlonales. Los anticuerpos monoclonales adecuados pueden prepararse a través de técnicas conocidas, por ejemplo, las desveladas en "Monoclonal Antibodies: A manual of techniques", H Zola (CRC Press, 1988) y en "Monoclonal Hybridoma Antibodies: Techniques and applications", JGR Hurrell (CRC Press, 1982).

- 55 Cuando se seleccionan las posibles moléculas de unión de las bibliotecas, generalmente se emplean uno o más péptidos selectores que tienen motivos definidos. En el diseño de motivos para péptidos selectores pueden emplearse restos de aminoácidos que proporcionan estructura, disminuyendo la flexibilidad en el péptido o las

cadenas laterales cargadas, polares o hidrófobas que permiten la interacción con la molécula de unión. Por ejemplo:

- (i) Prolina puede estabilizar una estructura peptídica ya que su cadena lateral está unida tanto al carbono alfa como al nitrógeno;
- (ii) Fenilalanina, tirosina y triptófano tienen cadenas laterales aromáticas y son altamente hidrófobas, mientras que leucina e isoleucina tienen cadenas laterales alifáticas y también son hidrófobas;
- (iii) Lisina, arginina e histidina tienen cadenas laterales básicas y se cargarán positivamente a un pH neutro, mientras que aspartato y glutamato tienen cadenas laterales ácidas y se cargarán negativamente a pH neutro;
- (iv) Asparagina y glutamina son neutros a pH neutro pero contienen un grupo amida que puede participar en enlaces de hidrógeno;
- (v) Las cadenas laterales de serina, treonina y tirosina contienen grupos hidroxilo, que pueden participar en enlaces de hidrógeno.

Normalmente, la selección de moléculas de unión puede implicar el uso de tecnologías y sistemas de matriz para analizar la unión a puntos correspondientes a los tipos de moléculas de unión.

El uno o más restos de unión a proteínas pueden comprender un resto detectable. El resto detectable puede seleccionarse del grupo que consiste en un resto fluorescente, un resto luminiscente, un resto quimioluminiscente, un resto radioactivo y un resto enzimático.

En una realización más de los procedimientos de la invención, la etapa (c) puede realizarse utilizando un ensayo que comprende un segundo agente de unión capaz de unirse a una o más proteínas, comprendiendo también el segundo agente de unión un resto detectable. Los segundos agentes de unión adecuados se han descrito en detalle anteriormente en relación con los primeros agentes de unión.

Por lo tanto, las proteínas de interés en la muestra que se analice pueden aislarse y/o inmovilizarse primero utilizando el primer agente de unión, tras lo cual pueden determinarse la presencia y/o la cantidad relativa de dichos biomarcadores utilizando un segundo agente de unión.

En una realización, el segundo agente de unión es un anticuerpo o fragmento de unión a antígeno del mismo; normalmente, un anticuerpo recombinante o fragmento del mismo. Convenientemente, el anticuerpo o fragmento del mismo se selecciona del grupo que consiste en: scFv; Fab; un dominio de unión de una molécula de inmunoglobulina. Los anticuerpos y fragmentos adecuados, así como los procedimientos para prepararlos, se han descrito en detalle anteriormente.

Alternativamente, el segundo agente de unión puede ser un agente de unión de tipo anticuerpo, como por ejemplo un anticuerpo o un aptámero.

Alternativamente, cuando el resto detectable en la proteína en la muestra que se analice comprende o consiste en un miembro de un par de unión específico (p.ej., biotina), el segundo agente de unión puede comprender o consistir en el miembro complementario del par de unión específico (p.ej., estreptavidina).

Cuando se utiliza un ensayo de detección, preferentemente el resto detectable se selecciona del grupo que consiste en: un resto fluorescente; un resto luminiscente; un resto quimioluminiscente; un resto radiactivo; un resto enzimático. Los ejemplos de restos detectables adecuados para su uso en los procedimientos de la invención se han descrito anteriormente.

Los ensayos preferentes para detectar proteínas séricas o plasmáticas incluyen ensayos inmunosorbente ligado a enzimas (ELISA), radioinmunoensayo (RIA), ensayos inmunorradiométricos (IRMA) y ensayos inmunoenzimáticos (IEMA), incluyendo ensayos sándwich utilizando anticuerpos monoclonales y/o policlonales. David y col., describen ejemplos de ensayos tipo sándwich en las patentes estadounidenses Nos. 4.376.110 y 4.486.530. Puede emplearse tinción de anticuerpos de células sobre portaobjetos en procedimientos perfectamente conocidos en pruebas de diagnóstico de laboratorio de citología, muy conocidos también entre las personas expertas en la materia.

Por lo tanto, en una realización, el ensayo es un ELISA (ensayo inmunosorbente ligado a enzimas) que implica normalmente el uso de enzimas que dan un producto de reacción coloreado, generalmente, en ensayos de fase sólida. Está muy extendido el uso de enzimas como peroxidasa de rábano picante y fosfatasa. Una forma de amplificar la reacción de fosfatasa es utilizar NADP como sustrato para generar NAD que actúa ahora como una coenzima para un segundo sistema enzimático. La pirofosfatasa de *Escherichia coli* proporciona un buen conjugado porque la enzima no está presente en los tejidos, es estable y proporciona un buen color de reacción. También se pueden utilizar sistemas quimioluminiscentes a base de enzimas como luciferasa.

Con frecuencia, se utiliza la conjugación con la vitamina biotina, ya que puede detectarse fácilmente por su reacción con avidina o estreptavidina ligada a enzimas a las que se une con gran especificidad y afinidad.

En una realización alternativa, el ensayo utilizado para la detección de proteínas es convenientemente un ensayo fluorométrico. Por lo tanto, el resto detectable del segundo agente de unión puede ser un resto fluorescente, como un fluoróforo Alexa (por ejemplo, Alexa-647).

Preferentemente, las etapas (c), (g), (f) y/o (j) de los procedimientos descritos en el primer aspecto se llevan a cabo utilizando una matriz. La matriz puede ser una matriz a base de perlas o una matriz de superficie. La matriz se puede seleccionar del grupo que consiste en: macromatriz, micromatriz y noanomatriz.

Las matrices en sí se conocen perfectamente en la técnica. Normalmente, están formadas por una estructura lineal o bidimensional que tiene regiones separadas (es decir, discretas) ("puntos"), teniendo cada una un área finita, formada en la superficie de un soporte sólido. Una matriz también puede ser una estructura de perla en la que puede identificarse cada perla mediante un código molecular o un código de color o identificarse en un flujo continuo. El análisis también se puede realizar secuencialmente cuando se pasa la muestra sobre una serie de puntos, absorbiendo cada uno de ellos la clase de moléculas de la solución. El soporte sólido es normalmente vidrio o un polímero, siendo los polímeros más utilizados celulosa, poliácridamida, nilón, poliestireno, policloruro de vinilo o polipropileno. Los soportes sólidos pueden estar en forma de tubos, perlas, discos, chips de silicio, microplacas, membrana de difluoruro de polivinilideno (PVDF), membrana de nitrocelulosa, membrana de nilón, otra membrana porosa, membrana no porosa (p.ej., plástico, polímero, perspex, silicio, entre otros), una pluralidad de pasadores poliméricos o una pluralidad de pocillos de microtitulación o cualquier otra superficie adecuada para inmovilizar proteínas, polinucleótidos y otras moléculas adecuadas y/o realizar un inmunoensayo. Los procedimientos de unión son muy conocidos en la técnica y generalmente consisten en reticulación de una molécula de proteína, polinucleótido o similar que se une covalentemente o adsorbe físicamente al soporte sólido. Alternativamente, se puede emplear el acoplamiento por afinidad de las sondas mediante etiquetas de afinidad o construcciones similares. Mediante el uso de técnicas muy conocidas, como la impresión por contacto o sin contacto, el enmascaramiento o la fotolitografía, se puede definir la ubicación de cada punto. Para revisiones ver Jenkins, RE, Pennington, SR (2001, *Proteonomics*, 2, 13-29) y Lal y col., (2002, *Dru Dicov Today* 15; 7 (18 Suppl.): S 143-9).

Normalmente, la matriz es una micromatriz. En "micromatriz" se incluye el significado de una serie de regiones que tienen una densidad de regiones independientes de al menos aproximadamente 100/cm², y preferentemente al menos aproximadamente 1000/cm². Las regiones en una micromatriz tienen dimensiones típicas, p.ej., diámetro, en el intervalo de aproximadamente 0-250 µm, y están separadas de otras regiones en la matriz por aproximadamente la misma distancia. La matriz puede ser alternativamente una macromatriz o una nanomatriz.

Una vez identificadas y aisladas, las moléculas de unión adecuadas (explicadas anteriormente), la persona experta puede fabricar una matriz aplicando procedimientos perfectamente conocidos en la técnica de la biología molecular.

En una realización adicional o alternativa, uno o más biomarcadores medidos en la etapa (c) comprende o consiste en uno o más productos génicos homólogos expresados por células humanas. En una realización adicional o alternativa, uno o más biomarcadores medidos en la etapa (c) es una proteína o polipéptido. En una realización adicional o alternativa, uno o más biomarcadores medidos en la etapa (c) es un ácido nucleico (p.ej., ADN, ARNm o ADNc, etc.).

Se desvela también en el presente documento que el procedimiento se realiza *in vitro*, *in vivo*, *ex vivo* o *in silico*. Por ejemplo, el procedimiento puede realizarse en particular *in vitro*.

En una realización adicional o alternativa, la sensibilización de la piel es una respuesta de hipersensibilidad (p.ej., una respuesta de hipersensibilidad mediada por células). En una realización adicional o alternativa, la respuesta de hipersensibilidad es una respuesta de hipersensibilidad de tipo IV. En una realización adicional o alternativa, la respuesta de hipersensibilidad es dermatitis alérgica de contacto (DAC).

En "agente de ensayo" se incluye cualquier entidad química (o mezcla de entidades químicas) para la que se va a determinar la potencia de sensibilización y/o el estatus de sensibilización.

En "estatus de sensibilización" se incluye o significa si una entidad química (o una mezcla de entidades químicas) es un sensibilizador o no (p.ej., un sensibilizador de la piel y/o un sensibilizador respiratorio).

"Agente sensibilizador de la piel" significa cualquier agente capaz de inducir y desencadenar una reacción de hipersensibilidad de tipo retardado de tipo IV en un mamífero. Preferentemente, la reacción de hipersensibilidad de tipo retardado de tipo IV está mediada por DC.

En una realización, el uno o más "agentes sensibilizadores de la piel" es un agente capaz de inducir y desencadenar una reacción de hipersensibilidad de tipo retardado de tipo IV en un sitio de contacto epidérmico en un mamífero.

El mamífero puede ser cualquier animal doméstico o de granja. Preferentemente, el mamífero es una rata, un ratón, una cobaya, un gato, un perro, un caballo o un primate. Lo más preferentemente, el mamífero es un ser humano. Tal como se ha explicado, los procedimientos *in vivo* para determinar la sensibilización son conocidos en la técnica. Un procedimiento preferente es el ensayo de ganglios linfáticos locales (para más detalles, véase Basketter, DA, y col., *Local lymph node assay – validation, conduct and use in practice*. Food Chem Toxicol., 2002. 40 (5): p. 593- 8) Otro procedimiento adecuado, pero menos preferente, es la prueba de maximización con cobayas (para más detalles, véase Magnusson, B. y AM Kligman, *The identification of contact allergens by animal assay*. The guinea pig maximization assay. J Invest Dermatol, 1969. 52 (3): p. 268-76).

En una realización adicional o alternativa, la población de células dendríticas o la población de células de tipo dendrítico es una población de células de tipo dendrítico. En una realización adicional o alternativa, las células de tipo dendrítico son células de tipo dendrítico mieloides. En una realización adicional o alternativa, las células de tipo dendrítico mieloides se derivan de células dendríticas mieloides. En una realización adicional o alternativa, las células derivadas de células dendríticas mieloides son células derivadas de leucemia mieloides. En una realización adicional o alternativa, las células derivadas de leucemia mieloides se seleccionan del grupo que consiste en KG-1, THP-1, U-937, HL-60, Monomac-6, AML-193 y MUTZ-3. En una realización adicional o alternativa, las células de tipo dendrítico son células MUTZ-3. Las células MUTZ-3 son células de leucemia mielomonocíticas agudas humanas que están disponibles en el Deutsche Sammlung für Mikroorganismen und Zellkulturen GmbH (DSMZ), Braunschweig, Alemania (www.dsmz.de; DMSZ No. ACC 295).

"Células de tipo dendrítico" significa células no dendríticas que presentan características funcionales y fenotípicas específicas de las células dendríticas, tales como características morfológicas, expresión de moléculas co-estimuladoras y moléculas de MHC de clase II y la capacidad de pinocitosis para macromoléculas y para activar linfocitos T en reposo.

En una realización, las células de tipo dendrítico, después de la estimulación con citocina, presentan antígenos a través de CD1d, MHC de clase I y II y/o inducen una proliferación específica de células T.

En una realización, las células de tipo dendrítico son progenitores de células dendríticas CD34⁺. Opcionalmente, los progenitores de células dendríticas CD34⁺ pueden adquirir, con la estimulación con citocinas, los fenotipos de los antígenos de presentación a través de CD1d, MHC clase I y II, inducir la proliferación específica de linfocitos T y/o despliegan un perfil transcripcional y fenotípico maduro tras la estimulación con mediadores inflamatorios (es decir, fenotipos similares a células dendríticas inmaduras o células dendríticas de tipo Langerhans).

En una realización, la población de células dendríticas o la población de células de tipo dendrítico es una población de células dendríticas. Preferentemente, las células dendríticas son células dendríticas primarias. Preferentemente, las células dendríticas son células dendríticas mieloides.

Las células dendríticas pueden reconocerse según la función, el fenotipo y/o el patrón de expresión génica, particularmente según el fenotipo de superficie celular. Estas células se caracterizan por su morfología diferenciada, por altos niveles de expresión de MHC de clase II de superficie y por su capacidad para presentar antígeno para linfocitos T CD4⁺ y/o CD8⁺, particularmente a linfocitos T no tratados previamente (Steinman y col. (1991) Ann. Rev. Immunol.9: 271).

La superficie celular de las células dendríticas es inusual, con proyecciones características de tipo velo, y se caracteriza por la expresión de marcadores de superficie celular CD11c y MHC clase II. La mayoría de las CD son negativas para los marcadores de otros linajes de leucocitos, incluyendo linfocitos T, linfocitos B, monocitos/macrófagos y los granulocitos. Las subpoblaciones de células dendríticas también pueden expresar marcadores adicionales incluyendo 33D1, CCR1, CCR2, CCR4, CCR5, CCR6, CCR7, CD1 a-d, CD4, CD5, CD8alpha, CD9, CD11b, CD24, CD40, CD48, CD54, CD58, CD80, CD83, CD86, CD91, CD117, CD123 (IL3Ra), CD134, CD137, CD150, CD153, CD162, CXCR1, CXCR2, CXCR4, DCIR, DC-LAMP, DC-SIGN, DEC205, E-cadherina, Langerina, receptor de manosa, MARCO, TLR2, TLR3, TLR4, TLR5, TLR6, TLR9 y varias lectinas.

Los patrones de expresión de estos marcadores de la superficie celular pueden variar junto con la madurez de las células dendríticas, su tejido de origen y/o su especie de origen. Las células dendríticas inmaduras expresan bajos niveles de MHC de clase II, pero son capaces de endocitosis de proteínas antigénicas y procesarlas para su presentación en un complejo con moléculas de MHC de clase II. Las células dendríticas activadas expresan altos niveles de MHC clase 1.1, ICAM-1 y CD86, y son capaces de estimular la proliferación de células T alogénicas no tratadas previamente, por ejemplo, en una reacción mixta de leucocitos (MLR).

Funcionalmente, las células dendríticas o células de tipo dendrítico pueden identificarse a través de cualquier ensayo conveniente para determinar la presentación del antígeno. Dichos ensayos pueden incluir determinar la capacidad de estimular linfocitos T cebados con antígeno y/o no tratados previamente mediante la presentación de un antígeno de ensayo, seguido de la determinación de la proliferación de linfocitos T, la liberación de IL-2 y similares.

Los procedimientos para detectar y/o medir la concentración de proteína y/o ácido nucleico son muy conocidos entre las personas expertas en la materia, véase, por ejemplo, Sambrook y Russell, 2001, Cold Spring Harbor Laboratory Press.

Los procedimientos preferentes para la detección y/o medición de proteínas incluyen transferencia de Western, transferencia Northwestern, ensayos inmunosorbentes (ELISA), micromatriz de anticuerpo, micromatriz de tejido (TMA), inmunoprecipitación, hibridación *in situ* y otras técnicas de inmunohistoquímica, radioinmunoensayo (RIA), ensayos inmunoradiométricos (IRMA) y ensayos inmunoenzimáticos (IEMA), incluyendo los ensayos sándwich utilizando anticuerpos monoclonales y/o policlonales. David y col., en las patentes estadounidenses Nos. 4.376.110 y 4.486.530, describen ejemplos de ensayos de sandwich. Se puede recurrir a la tinción de anticuerpos de células sobre portaobjetos en procedimientos muy conocidos en pruebas de diagnóstico de laboratorio citológicas, tal como

conocen perfectamente las personas expertas en la materia.

Normalmente, ELISA implica el uso de enzimas que dan un producto de reacción coloreado, generalmente, en ensayos de fase sólida. El uso de enzimas como peroxidasa de rábano picante y fosfatasa está muy extendido. Una forma de amplificar la reacción de fosfatasa es utilizar NADP como sustrato para generar NAD que actúa ahora como coenzima para un segundo sistema enzimático. La pirofosfatasa de *Escherichia coli* proporciona un buen conjugado ya que la enzima no está presente en los tejidos, es estable y proporciona un buen color de reacción. Se pueden utilizar asimismo sistemas quimioluminiscentes a base de enzimas como la luciferasa.

Con frecuencia, se utiliza la conjugación con la vitamina biotina ya que puede detectarse fácilmente por su reacción con avidina o estreptavidina ligada a enzimas a las que se une con gran especificidad y afinidad.

Un segundo aspecto de la invención proporciona una matriz para su uso en el procedimiento de acuerdo con el primer aspecto de la invención, consistiendo los restos de unión de la matriz para cada uno de los biomarcadores definidos en la Tabla A, y en la que los restos de unión de la matriz comprenden restos de unión tal como se definen en el primer aspecto de la invención.

Se desvela también en el presente documento que la matriz comprende uno o más restos de unión para cada uno de los biomarcadores, tal como se define en el primer aspecto de la invención.

En una realización adicional o alternativa, el uno o más restos de unión están inmovilizados.

En una realización adicional o alternativa, la matriz es una matriz a base de perlas. En una realización adicional o alternativa, la matriz es una matriz de superficie.

En una realización adicional o alternativa, la matriz se selecciona del grupo que consiste en: macromatriz, micromatriz y nanomatriz.

La matriz del segundo aspecto de la invención puede comprender una o más, preferentemente dos o más, restos de unión, en las que cada uno de los restos de unión es capaz de unirse selectivamente a un biomarcador, tal como se define en el primer aspecto. Por lo tanto, la matriz puede comprender o consistir en una selección en particular de restos de unión específicos de biomarcador, lo cual se correlaciona con cualquier selección de biomarcadores en particular, tal como se define en el primer aspecto.

Un tercer aspecto de la invención proporciona el uso de los biomarcadores enumerados en la Tabla A para determinar la potencia de sensibilización de la piel de un agente.

Se desvela también en el presente documento el uso de un biomarcador seleccionado del grupo definido en la Tabla A para determinar la potencia de sensibilización de la piel de un agente, preferentemente, en el que el biomarcador se selecciona del grupo definido en la Tabla A (i) y/o la Tabla A (ii).

En una realización adicional o alternativa, se proporciona el uso de restos de unión con especificidad para cada uno de los biomarcadores enumerados en la Tabla A para determinar la potencia de sensibilización de la piel de un agente.

Un cuarto aspecto de la invención proporciona un kit de análisis para su uso en un procedimiento de acuerdo con el primer aspecto de la invención que comprende:

- (a) una matriz de acuerdo con el segundo aspecto de la invención; e
- (b) instrucciones para realizar el procedimiento como se define en el primer aspecto de la invención (opcional).

En una realización adicional o alternativa, el kit de análisis comprende además uno o más agentes de control, tal como se define en el primer aspecto de la invención.

Se desvela también en el presente documento un procedimiento para tratar o prevenir una reacción de hipersensibilidad de tipo I (como por ejemplo DAC) en un paciente que comprende las etapas de:

- (a) proporcionar uno o más agentes de ensayo a los que está o ha estado expuesto el paciente;
- (b) determinar si el uno o más agentes de ensayo proporcionados en la etapa (a) es un sensibilizador de la piel y/o su potencia como sensibilizador aplicando un procedimiento de la presente invención; y
- (c) cuando se identifica uno o más agentes de ensayo como sensibilizador de la piel y/o sensibilizador de la piel de una potencia en particular, reducir o evitar la exposición del paciente al uno o más agentes de ensayo y/o proporcionar un tratamiento adecuado para los síntomas de sensibilización.

Preferentemente, el uno o más agentes de ensayo a los que está o ha estado expuesto el paciente es un agente al que está expuesto actualmente el paciente al menos una vez al mes, por ejemplo, al menos una vez cada dos semanas, al menos una vez a la semana, o al menos una vez al día.

Los tratamientos de los síntomas de sensibilización pueden incluir, por ejemplo, esteroides tópicos.

Preferentemente, el procedimiento de tratamiento está en consonancia con el procedimiento descrito en el primer aspecto de la invención, y una o más de las realizaciones descritas en él.

Se desvela también en el presente documento un programa informático para operar los procedimientos de la invención, por ejemplo, para interpretar los datos de expresión de la etapa (c) (y las etapas de medición de expresión subsiguientes) y determinar de este modo si uno o más agentes de ensayo es un sensibilizador respiratorio y/o la potencia de sensibilización del uno o más agentes de ensayo. El programa informático puede ser un SVM programado. El programa informático puede grabarse en un soporte adecuado legible por ordenador conocido entre las personas expertas en la materia. Los soportes legibles por ordenador adecuados pueden incluir discos compactos (incluyendo CD-ROM, DVD, Blue Rays y similares), disquetes, unidades de memoria flash, ROM o unidades de disco duro. El programa informático puede instalarse en un ordenador adecuado para ejecutar el programa informático.

Las personas expertas podrán apreciar que es posible utilizar en combinación todas las realizaciones que no entren en conflicto. Por lo tanto, las realizaciones de un aspecto de la invención pueden aplicarse igualmente a un segundo aspecto de la invención.

A continuación, se describirán ejemplos preferentes, no exhaustivos, que representan ciertos aspectos de la invención, haciendo referencia a las siguientes figuras:

Figura 1. Predicciones binarias utilizando el ensayo GARD. Evaluación ROC para (A) predicciones binarias de productos químicos de referencia (línea continua) y de 37 productos químicos nuevos (línea discontinua). (B) Los valores de decisión GARD SVM (DV) se correlacionan con la potencia CLP (37 productos químicos, 11 1A, 19 1B, 7 sin cat). El aumento de la potencia se asocia con el aumento de DV.

Figura 2. Visualización del conjunto de datos de entrenamiento utilizado para desarrollar el modelo de bosque aleatorio, utilizando las 52 variables. (A) Gráfico PCA del conjunto de entrenamiento con réplicas biológicas separadas coloreadas de acuerdo con las clasificaciones CLP de los productos químicos. (B) Mapa de calor del conjunto de entrenamiento con repeticiones de las muestras agrupadas jerárquicamente, en las que la escala de grises representa la intensidad relativa de la expresión génica.

Figura 3. Visualización del conjunto de datos de ensayo utilizando las 52 variables. (A) Gráfico de PCA del conjunto de ensayo con repeticiones biológicas separadas coloreadas de acuerdo con las clasificaciones CLP. El PCA se creó sobre el conjunto de entrenamiento y el conjunto de ensayo trazado sin influir en el PCA. (B) Mapa de calor del conjunto de ensayo con muestras agrupadas jerárquicamente, en las que la escala de grises representa la intensidad relativa de la expresión génica.

Figura 4. El modelo de potencia CLP contiene información relacionada con la potencia humana. (A) Gráfico PCA de muestras de ensayo y de entrenamiento de repetición con clasificaciones de potencia humana disponibles coloreadas de acuerdo con la potencia humana. El PCA se basa en las 52 variables del bosque aleatorio como entrada y el PCA se construyó en el conjunto de entrenamiento. (B) Gráfico PCA de visualización de datos de ensayo solamente.

Figura 5. Análisis de ruta basado en una entrada de los 883 genes más significativos de una comparación multigrupo de las tres clases de CLP.

Figura 6. Rutas únicas para cada grupo de reactividad de proteínas, identificado según el análisis de ruta.

Figura 7. Diagramas de Venn [82] de genes comunes (A) y rutas biológicas (B) para las tres clases diferentes de reactividad de proteína. MA = aceptores de Michael; SB = formación de base de Schiff; SN = sustitución nucleófila bi-molecular/sustitución aromática nucleófila.

Figura 8: Detección rápida de alérgenos genómicos (GARD): flujo de trabajo. El modelo SVM calcula los valores de decisión para cada muestra de ARN derivada de células tratadas químicamente sobre la base de los datos de expresión génica de 200 genes que representan la firma de predicción GARD. Los valores de decisión se utilizan después para la clasificación binaria de los productos químicos [3]

Figura 9: Relación entre los valores de decisión de GARD y las clases de potencia humana. Los productos químicos se han clasificado con arreglo a la clase de potencia humana y el correspondiente valor de decisión SVM promedio derivado de triplicados se puede encontrar en el eje Y. El número en el eje x se refiere al número del producto químico en la Tabla 8.

Figura 10: Gráfico de dispersión del modelo O-PLS (Y = potencia humana y sensibilizador/no sensibilizador) basado en 197 muestras. Clases de potencia humana, tal como se describe en Basketter y col. (2014), clase 1 que representa la potencia más alta, 6 que representa no sensibilizadores.

Figura 11: Gráfico de permutación del modelo O-PLS. Comparación de los beneficios del ajuste y la

predicción del modelo original con varios modelos, en la que se ha permutado aleatoriamente el orden de las observaciones en Y. El gráfico indica rotundamente que el modelo no se obtuvo por casualidad.

Figura 12: Valores Y observados indicados en gráfico frente a valores Y pronosticados. El RMSEE (estimación de la raíz del error cuadrático medio) indica el error de ajuste de las observaciones al modelo. RMSEcv es una medida similar, pero se calcula mediante validación cruzada.

Figura 13: gráficos PCA y mapas de calor. PCA del conjunto de entrenamiento (A), conjunto de entrenamiento y ensayo (B) y conjunto de ensayo solamente (C) basado en una entrada de 18 variables identificadas según el modelo de bosque aleatorio. Cada esfera representa un producto químico, coloreado de acuerdo con las clasificaciones CLP (amarillo = 1A, rosa = 1B, azul = sin categoría). (D) Mapa de calor del conjunto de entrenamiento y (E) mapa de calor del conjunto de ensayo basado en la expresión de los 18 biomarcadores identificados.

EJEMPLO A

Introducción

La sensibilización de la piel y las enfermedades asociadas, como la alergia de contacto, afectan a una parte sustancial de la población general con una prevalencia estimada del 15-20 % en los países industrializados [1]. La dermatitis de contacto alérgica (DAC), una reacción de hipersensibilidad de tipo IV, es común entre ciertos grupos laborales, como los que están expuestos regularmente a productos químicos o relacionados con trabajos húmedos [2]. Sin embargo, también los cosméticos y los productos domésticos pueden contener numerosos productos químicos sensibilizadores de la piel. A través de diferentes marcos legales, la Unión Europea ha prohibido los ensayos con animales para cosméticos y sus ingredientes [3] y ha impuesto el requisito de ensayos para >60.000 productos químicos en el contexto de REACH [4]. Se debe proporcionar información sobre la capacidad de sensibilización de la piel y la potencia de un producto químico para satisfacer los requisitos reglamentarios para la clasificación y subcategorización. Se necesitan urgentemente ensayos alternativos sin animales que satisfagan estos requisitos.

Los eventos moleculares que conducen a la sensibilización de la piel y, en consecuencia, a la DAC pueden caracterizarse por una serie de eventos clave secuenciales (KE) desencadenados por un producto químico, y se han resumido en una ruta de resultados adversos (AOP) tal como describe la OCDE [5]. El evento iniciador (KE1) se define como la modificación de la proteína covalente producida por el producto químico sensibilizador de la piel una vez que accede a las capas más profundas de la piel. El siguiente KE2 representa las respuestas inflamatorias tras la activación de los queratinocitos. KE3 corresponde a la activación de las células dendríticas, lo que a su vez conlleva la activación y proliferación de linfocitos T (KE4). Tras la re-exposición al sensibilizador, se puede desencadenar el desarrollo de CAD, que como resultado adverso se caracteriza por lesiones cutáneas inducidas por linfocitos T CD8⁺ y Th1 específicos. Si bien los KE en el AOP están descritos perfectamente, queda pendiente aún una comprensión mecánica detallada de la biología que subyace tras los eventos clave individuales [5].

El ensayo de ganglios linfáticos locales murinos (LLNA) [6] ha sido durante muchos años la alternativa preferente para las pruebas de sensibilización de la piel, ya que tiene capacidad de proporcionar datos para la identificación y caracterización del peligro, incluyendo la información sobre la potencia del sensibilizador de la piel. Sin embargo, se caracteriza por ciertas limitaciones, como la susceptibilidad a los efectos del vehículo y cuestiones sobre resultados falsos positivos [7]. Se han desarrollado varios procedimientos de predicción sin animales para reducir la experimentación con animales que se utiliza para analizar productos químicos, entre los que se incluyen enfoques informáticos para integrar datos de diferentes plataformas de ensayo para la identificación de peligros, tal como se ha revisado recientemente [8]. Se aceptan tres procedimientos de ensayo en cuanto a la sensibilización de la piel como directrices de ensayo en la OCDE; el procedimiento de ensayo de luciferasa ARE-NRF2 (KeratoSens[™]) [9, 10], el ensayo directo de reactividad peptídica (DPRA) [11] y el ensayo de activación de línea celular humana (h-CIAT) [12]. Además de la identificación de peligros, la información sobre la potencia del sensibilizador de la piel es imprescindible para permitir la evaluación cuantitativa del riesgo y definir los límites de exposición. Se han publicado algunos enfoques para la predicción de la potencia del sensibilizador de la piel y se han revisado recientemente en [8], como los ensayos dirigidos a KE2 (ensayo de potencia del sensibilizador epidérmico equivalente [13], SENS-IS [14]) y el ensayo U-SENS de modelación de KE3 [15]. Asimismo, se han descrito modelos *in silico*, que suelen combinar la información de varios procedimientos *in vitro*; por ejemplo, QSAR [16], red neuronal artificial [7], modelos probabilísticos y enfoques de estrategia de prueba integrada (ITS) que incluyen un modelo de Bayesian [18-21].

El ensayo alternativo de detección rápida de alérgenos genómicos (GARD) para la evaluación de la capacidad de sensibilización de la piel de los productos químicos se basa en el análisis transcriptómico global de la expresión diferencial en una línea de células mieloides humanas, inducida por productos químicos sensibilizadores en comparación con controles no sensibilizadores. La firma de biomarcador resultante, la firma de predicción GARD (GPS), consiste en 200 transcritos, que se utilizan como entrada en un modelo de máquina de vectores de soporte (SVM) entrenado en un conjunto de productos químicos de referencia [22]. Los cambios en la transcripción pueden estar relacionados con la maduración y activación de las células dendríticas (KE3) durante la sensibilización. En un estudio interno basado en 26 productos químicos ocultos, la precisión del ensayo se estimó en 89 %

[23]. Observaciones anteriores indicaron que el ensayo GARD es capaz de proporcionar información relevante también para la evaluación de la potencia. En primer lugar, se regularon las rutas de señalización de forma diferenciada dependiendo de la potencia de un subconjunto de grupos de reactividad química [24]. En segundo lugar, se observó que a los sensibilizadores más potentes se les asignó generalmente valores de decisión GARD SVM más altos en comparación con los sensibilizadores más débiles, lo cual indica que había genes dentro de la firma que contribuyeron a la información de la potencia (observaciones no publicadas). Sin embargo, la información en la firma de predicción GARD no fue suficiente para estratificar completamente los productos químicos en grupos de potencia bien definidos, tal como se define en el Reglamento de Clasificación, Etiquetado y Envasado (CLP) [25].

El Reglamento CLP se basa en el Sistema Globalmente Armonizado [25] y utiliza tres categorías para la clasificación de productos químicos; sin categoría (sin cat.) para no sensibilizadores, categoría 1B para sensibilizadores débiles y 1A para sensibilizadores fuertes. A la luz de las observaciones que se han descrito, se planteó la hipótesis de que GARD puede desarrollarse aún más como herramienta para la predicción de la potencia como sensibilizador de la piel de un producto químico, teniendo como objetivo las categorías CLP. Dado que el modelo de máquina de vectores de soporte de GARD establecido no se puede aplicar a problemas de varias clases, se utilizó otro enfoque basado en el modelado de bosques aleatorio [26]. El bosque aleatorio es un procedimiento basado en árboles de decisión y muy adecuado para datos de micromatrices [27]. Divide el conjunto de datos interna y repetidamente en un conjunto de entrenamiento y ensayo mediante submuestreo aleatorio (bootstrapping). Las muestras del conjunto de ensayo, denominadas muestras fuera de bolsa, comprenden aproximadamente una tercera parte del conjunto de datos entero y se utilizan para calcular el error fuera de la bolsa (OOB), es decir el error de clasificación.

En el presente documento, se presenta un nuevo enfoque para predecir la potencia como sensibilizador de la piel de acuerdo con las categorías CLP basadas en el aprendizaje supervisado utilizando un modelo de bosque aleatorio. En primer lugar, se utilizaron los datos de expresión génica global de un conjunto entrenado que comprende 68 productos químicos únicos y 2 muestras de vehículo de control como entrada en un modelo de bosque aleatorio. El modelo de bosque aleatorio se combinó posteriormente con un algoritmo para la eliminación de variables hacia atrás. El algoritmo clasificó inicialmente la importancia variable de cada gen de las micromatrices y después ajustó iterativamente nuevos bosques aleatorios, al mismo tiempo que eliminaba las variables menos importantes de la iteración anterior. Al utilizar esta estrategia, se pudo identificar un conjunto de 52 genes con la tasa de error de OOB mínima al predecir las muestras fuera de bolsa del conjunto de entrenamiento. Se desafió el rendimiento predictivo de los 52 genes con un conjunto de ensayos independiente que contenía 18 productos químicos inéditos para el modelo. Se pudo predecir los productos químicos de este conjunto de ensayo con una precisión general del 78 %. Además del modelo predictivo, se demostró también la versatilidad de analizar transcriptomas completos de células realizando análisis de rutas para mejorar aún más la comprensión de la mecánica de la potencia de sensibilización de la piel a nivel celular, confirmando la hipótesis de que diferentes clases de reactividad química inducen rutas de señalización diferenciadas.

Materiales y procedimientos

Células y citometría de flujo

La línea de células mieloides utilizada en este estudio se derivó de MUTZ-3 (DSMZ, Braunschweig, Alemania) y se mantuvo como se describe en [22, 28]. Se realizó un análisis de control fenotípico de las células antes de cada experimento por citometría de flujo para confirmar el estado inmaduro de las células. Se utilizaron los siguientes anticuerpos monoclonales: CD1a (DakoCytomation, Glostrup, Dinamarca), CD34, CD86, HLA-DR (BD Biosciences, San José, EE.UU.), todos conjugados con FITC; CD14 (DakoCytomation), CD54, CD80 (BD Biosciences), todos conjugados con PE. FITC e IgG1 de ratón conjugado con PE (BD Biosciences) sirvió como control de isotipo y el yoduro de propidio como marcador de células no viables (BD Biosciences). Se expusieron tres lotes de células durante 24 horas en experimentos independientes y se evaluó la viabilidad y la expresión de CD86 por citometría de flujo. Se analizaron todas las muestras de FACS en un instrumento FACSCanto II con el software FACS Diva para la adquisición de datos. Se adquirieron 10 000 eventos y se realizó un análisis adicional en FCS Express V4 (De Novo Software, Los Ángeles, CA). Se lisaron las células para la extracción de ARN en TRIzol® (Life Technologies Thermo Fisher Scientific, Waltham, EE.UU.) y se almacenaron hasta su uso posterior a -20 °C.

Productos químicos y estimulaciones

Todos los productos químicos se compraron a Sigma Aldrich (Saint Louis, EE.UU.) en calidad de alta pureza o fueron proporcionados por Cosmetics Europe. Se almacenaron todos los productos químicos de acuerdo con las recomendaciones del proveedor. La estimulación química de las células se realizó como se ha descrito anteriormente [28]. Brevemente, las concentraciones de entrada de GARD se definieron por las características de solubilidad y citotoxicidad de los productos químicos. Se fijó como objetivo una concentración final de 500 µM para productos químicos no citotóxicos y solubles y la concentración máxima posible para productos químicos con solubilidad limitada (inferior a 500 µM en medio). Se utilizaron productos químicos citotóxicos en una concentración establecida como objetivo para una viabilidad relativa de las células del 90 %. La mayoría de los productos químicos se utilizaron a partir de una pre-dilución de 1000x en dimetilsulfóxido (DMSO) o agua MilliQ en autoclave. La concentración de DMSO como vehículo nunca excedió el 0,1 %. Se incluyeron muestras de DMSO y MilliQ como controles de vehículo en este estudio y por tanto pertenecen al grupo o de muestras no sensibilizadoras.

Extracción de ARN, ADNc e hibridación de matriz

Se llevó a cabo el aislamiento de ARN de las células lisadas en TRIzol® de acuerdo con las instrucciones del fabricante. Se sintetizó ADN codificante marcado de acuerdo con los protocolos de Affymetrix (Affymetrix, Cleveland, EE.UU.) utilizando los kits y controles recomendados. Se hibridó el ADNc con matrices Human Gene 1.0 ST (Affymetrix) y se procesó y escaneó adicionalmente según lo recomendado por el proveedor.

Clasificaciones binarias

Se realizaron las clasificaciones binarias de los 37 productos químicos resumidos en la Tabla 1 en sensibilizadores o no sensibilizadores con el modelo previamente establecido basado en SVM, utilizando datos de expresión normalizados por SCAN [29, 30] del GPS como entrada variable en el algoritmo de aprendizaje [22]. Antes de la construcción del modelo, se eliminaron los posibles efectos de lote entre el conjunto de entrenamiento y los productos químicos de ensayo al ajustar la escala de los valores de expresión de la matriz para los productos químicos de ensayo con respecto al conjunto de entrenamiento. Se generó un factor de escala calculando la relación del valor de expresión promedio para cada gen en muestras de control de vehículo DMSO del conjunto de entrenamiento y el valor de expresión promedio para el mismo gen en muestras de DMSO en el lote originario del producto químico de ensayo. El factor de escala para cada gen se multiplicó después por los valores de expresión génica para el gen correspondiente en el producto químico de ensayo. Se realizaron predicciones SVM tal como se ha descrito anteriormente [22, 31]. Brevemente, se entrenó un modelo SVM basado en un núcleo lineal en estimaciones químicas de referencia del conjunto de entrenamiento original utilizado durante la identificación del GPS [22]. Posteriormente, se aplicó el modelo entrenado para asignar a cada producto químico de ensayo un valor de decisión SVM (SVM DV). Se utilizaron los DV SVM resultantes para todos los productos químicos de ensayo para construir una curva de características operativas del receptor (ROC) y se utilizó el área resultante bajo la curva (AUC) como medida de clasificación [32]. Se realizaron el modelado SVM y las visualizaciones de la curva ROC en el entorno estadístico R, utilizando los paquetes adicionales e1071 [33] y ROCR [34]. Antes de evaluar el rendimiento predictivo final del modelo, se calibraron los DV SVM para cada réplica individual de los productos químicos de ensayo con respecto al límite para un rendimiento predictivo máximo obtenido durante la clasificación de las muestras de referencia en la Tabla 2, tal como se describe en [31]. Los SV SVM calibrados se utilizaron posteriormente para clasificaciones finales, y los productos químicos de ensayo se clasificaron como sensibilizadores si el valor medio de salida de las réplicas >0. La precisión, la sensibilidad y la especificidad se estimaron utilizando estadísticas de cobre [35]. Se realizó la prueba no paramétrica de dos muestras de Wilcoxon para determinar si las distribuciones SVM DV entre las categorías CLP 1A, 1 B y sin diferían significativamente.

Gestión de datos y análisis estadístico

Con el fin de construir un modelo de bosque aleatorio, se definieron 68 productos químicos y dos muestras de vehículos de control como un conjunto de entrenamiento y 18 productos químicos, se incluyeron seis de cada categoría CLP, en el conjunto de ensayo independiente. Los productos químicos en el conjunto de ensayo no se incluyeron en la construcción del modelo. El objetivo era obtener un conjunto de entrenamiento equilibrado que representara las tres categorías CLP (Tabla 3) y diferentes grupos de reactividad química (Tabla 4). La mayoría de los productos químicos en el conjunto de ensayo (14 de 8) se originó de la última campaña experimental (Tabla 1), que comprendió 37 productos químicos no investigados previamente utilizando el ensayo GARD. En el conjunto de entrenamiento, aproximadamente un tercio de las muestras (23 de 68) provenían de este último conjunto de datos. Las muestras de vehículos fueron parte de todos los proyectos y, por lo tanto, están presentes con números de réplica más altos.

Se fusionaron los nuevos datos de micromatrices con datos históricos [22, 23] y se sometieron a control de calidad. Se eliminaron cuatro matrices debido a la mala calidad; sin embargo, no hubo productos químicos presentes en menos de duplicados biológicos. Los datos de la matriz se importaron al entorno estadístico R y se normalizaron utilizando el algoritmo SCANfast [29, 30]. Como tuvieron que combinarse varias campañas experimentales, se normalizó este conjunto de datos utilizando el procedimiento ComBat [36, 37] para eliminar los efectos por lotes entre las muestras. En este momento, las muestras en el conjunto de entrenamiento se separaron de las muestras en el conjunto de ensayo. Para evitar el sobreajuste, solo se utilizaron muestras en el conjunto de entrenamiento designado durante la identificación de la firma predictiva de biomarcadores y para el ajuste de parámetros al clasificador y se reservaron las muestras del conjunto de ensayo para validar el rendimiento de la firma identificada y el clasificador especificado. La firma del biomarcador predictivo se identificó mediante la alimentación de intensidades de transcrito normalizadas y corregidas por lotes de muestras individuales en el conjunto de entrenamiento en un modelo de bosque aleatorio [26] combinado con un procedimiento de eliminación hacia atrás en el paquete varSelRF [38] en R/Bioconductor versión 3.1.2. El bosque inicial utilizado para la clasificación de importancia variable creció hasta 2000 árboles y todos los demás parámetros se mantuvieron en la configuración predeterminada por defecto. El paquete ajusta y evalúa de forma iterativa modelos de bosque aleatorio, en cada iteración omitiendo el 20 % de las variables menos importantes. Se seleccionó el conjunto de variables con mejor rendimiento en función de las tasas de error OOB de todos los bosques aleatorios ajustados como el número mínimo de genes dentro de un error típico de la solución de error mínimo (es decir, la regla 1.s.e). Se validó el procedimiento de selección de variables estimando la tasa de error de predicción mediante el procedimiento de submuestreo aleatorio .632+ del paquete varSelRF utilizando 100 muestras de submuestreo aleatorio, y se validó la importancia

de las transcripciones individuales en la firma del biomarcador según la frecuencia de aparición en las muestras de submuestreo aleatorio (denominado frecuencia de cuestionario de validación (VCF)). Se validó el rendimiento predictivo de la firma de biomarcadores identificados mediante la construcción de un nuevo bosque en el paquete Random Forest [39], basado en parámetros anteriores, utilizando solo las muestras en el conjunto de entrenamiento y los transcritos seleccionados en la firma del biomarcador como entrada variable. El modelo se aplicó para asignar cada muestra replicada individual en el conjunto de ensayo a una categoría CLP, y el voto mayoritario a través de los estímulos de réplica biológica para cada producto químico se aceptó como la categoría predicha. Se construyeron los diagramas de mapas de calor y los análisis de componentes principales (PCA) en Qlucore Omics Explorer (Qlucore AB, Lund, Suecia).

10 *Análisis de ruta*

Se realizó el análisis de ruta con la herramienta Key Pathway Advisor (KPA) [40] versión 16.6, que proporciona un flujo de trabajo de análisis de ruta para investigar, por ejemplo, datos de expresión génica. Éste asocia genes expresados distintivamente con procedimientos tanto aguas arriba como aguas abajo que dan cabida a la interpretación biológica. Se filtró primero por varianza el conjunto de datos investigado, que consiste en los valores de expresión normalizados por SCANfast y ComBat del conjunto de ensayo y el conjunto de entrenamiento, en total 308 muestras, para eliminar variables con varianza consistentemente baja hasta que quedara aproximadamente un tercio (10009 variables, $\sigma/\sigma_{\max} = 0,1478$). Luego se aplicó una comparación multigrupo (ANOVA) que compara muestras pertenecientes a CLP sin cat., 1B y 1A para identificar los transcritos que estaban regulados de manera diferenciada. Los 883 genes más significativos (tasa de descubrimiento falso FDR = 10^{-9} ; $p = 8,53 \times 10^{-11}$) se utilizaron como entrada en KPA (ID de Exon de Affymetrix y su correspondiente valor p, análisis de sobreconectividad). Para identificar las rutas asociadas con la reactividad de las proteínas, se filtró el mismo conjunto de datos con filtro de varianza basado en comparaciones de dos grupos (pruebas t, no sensibilizadores de la clase de unión Toxtree "sin unión" (81 muestras) frente a sensibilizadores "aceptores Michael" (MA, 63 muestras), "sin unión" frente a "formación de base de Schiff" (SB, 29 muestras), "sin unión" frente a "sustitución nucleófila bi-molecular/sustitución aromática nucleófila" combinada (SN, 25 muestras). A continuación, se introdujeron las listas de las 500 variables más significativamente reguladas de cada comparación, junto con el valor p y el cambio de pliegue (análisis de razonamiento causal), en la herramienta KPA para cada grupo de reactividad de proteínas. Se alcanzó el valor p mínimo al comparar las muestras MA con la "no unión" (FDR = $3,65 \times 10^{-7}$), seguido de SB (FDR = $2,3 \times 10^{-5}$) y (FDR = $2,44 \times 10^{-5}$).

30 *Secuencias de comandos*

A continuación se enumeran las secuencias de comandos utilizadas en cuanto a la normalización y las predicciones de potencia:

Secuencia de comandos 1: Utilizada para combatir los efectos de lotes entre diferentes proyectos antes de la construcción del modelo:

```
# ¡Esta secuencia de comandos se utiliza para combatir los efectos de lote entre
diferentes proyectos antes de la construcción de modelo!
library(sva)
library(caret)
data set <- read.delim("/home/andy/allfiles-scannorm.txt",stringsAsFactors =
F,header = F)
# Crear conjunto de datos numéricos con nombres de columna/nombres de fila y
matrices de diseño y factores apropiados
annotation headers <- data set[1:17,1] #Header comprises row 1-17
annotations <- data set[1:17,2:ncol(data set)]
colnames(annotations) <- annotations[1,]
rownames(annotations) <- annotation headers
rownames data set <- data set[,1]
data set <- data set[,-1]
# Convertir los datos a valores numéricos
data set <- data.frame(sapply(data set[18:nrow(data set),],as.numeric))
colnames(data set) <- annotations[1,] # The colnames corresponds to the
IDs
rownames(data set) <- rownames data set[18:length(rownames data set)] projects <-
factor(as.character(annotations[3,]))
# ¿Se utiliza CLP USENS en el modelo combat?
groups <- as.character(annotations[7,])
#Hacer NA desconocido
groups[is.na(groups)] <- "unknown" groups <- factor(groups)
group model <- model.matrix(~groups)
# Separar sondas de variancia cercana a cero (podría interferer potencialmente con
ComBat)
nzv <- nearZeroVar(t(data set))
```

```

data set <- data set[-nzv,]
#Los datos consisten en 5 proyectos diferentes con efectos de lote pronunciados
# ComBat se utiliza para eliminar los efectos de lote
normalized data <- ComBat(data set,projects,mod = group model) normalized data <-
rbind(annotations,normalized data)
write.table(normalized data,"/home/andy/processed data sets/ComBat/allfiles SCAN
ComBat.txt",col.names=F,sep= "\t",quote = F)

```

Secuencia de comando 2: Utilizada para la construcción del modelo de potencia inicial

```

#Construcción del modelo inicial de potencia
require(pROC)
require(ROCR)
require(varSelRF)
require(ggplot2)
require(reshape)
require(randomForest)
require(caret)
library(sva)
library(parallel)
library(doMC)
library(foreach)
##Cargar conjunto de datos
data1<-read.table(file="allfiles-scannorm-
cleaned ComBat.txt",stringsAsFactors = FALSE,row.names=1,header=F,
sep="\t")
##Definir conjunto de entrenamiento. La fila 10 corresponde a anotaciones con info
entrenamiento/ensayo
trainingset<-as.matrix(data1[,which(data1[10,]==,"training set")])
trainingsetannot<-as.vector(trainingset[7,])
trainingsetannot[which(trainingsetannot=="no cat")]<-0
trainingsetclass<-as.factor(trainingsetannot)
trainingset<-as.matrix(trainingset[-(1:17),])
class(trainingset) <-"numeric"
training<-as.data.frame(t(trainingset))
##Cargar conjunto de datos de ensayo 1
testset1<-as.matrix(data1[,which(data1[10,]=="test set 1")])
testset1annot<-as.vector(testset1[7,])

```

(continuación)

```

testset1annot[which(testset1annot=="no cat")]<-0
sampleannot<-as.vector(testset1[2,])
testset1<-as.matrix(testset1[-(1:17),])
class(testset1)<-"numeric"
testset1<-as.data.frame(t(testset1))
testsetclass<-as.factor(testset1annot)
##cargar conjunto de datos de ensayo 2
testset2<-as.matrix(data1[,which(data1[10,]=="test set 2")])
sampleannottest2<-as.vector(testset2[2,])
testset2annot<-as.vector(testset2[7,])
testset2annot[which(testset2annot=="no cat")]<-0
testset2annot<-as.factor(testset2annot)
testset2<-as.matrix(testset2[-(1:17),])
class(testset2) <-"numeric"
testset2<-as.data.frame(t(testset2))
##Construir el modelo
set.seed(2)
vsrf<-
varSelRF(training,Class=trainingsetclass,ntree=2000,ntreeliterat=2000,vars.d
rop.frac=0.2,keep.forest = T,whole.range=TRUE,verbose = TRUE)
RFmodel<-randomForest(y=trainingsetclass, x=subset(training, select =
vsrf$selected.vars), ntree=vsrf$ntreeIterat,importance=TRUE)
rfpr.pred<-predict(RFmodel,newdata=subset(testset1, select =
vsrf$selected.vars))
confusionMatrix(rfpr.pred,testsetclass)
a<-data.frame (stimulant=sampleannot, group=testsetclass,predicted=rfpr.pred)
write.table(a, file="VarSelRF combat Test1.txt", sep=" ", row.names = TRUE)
rfpr.pred2<-predict(RFmodel,newdata=subset(testset2, select =
vsrf$selected.vars))
b<-
data.frame(stimulant=sampleannottest2,group=testset2annot,predicted=rfpr.pr
ed2)
write.table(b, file="VarSelRF combat Test2.txt", sep=" ", row.names = TRUE)
#Extraer importancia de variable
write.table(vsrf$selected.vars, file="VarSelRF ComBat.txt", sep=" ",
row.names = TRUE)
##Arranque de procedimiento de selección variable
forkCL <- makeForkCluster(12)
clusterSetRNGStream(forkCL, iseed = (100))
clusterEvalQ(forkCL, library(varSelRF))
vsrf.vsb <-
varSelRFBoot(training,trainingsetclass,usingCluster=T,srf=vsrf,TheCluster=f
orkCL,ntree=2000,ntreeIterat=2000,vars.drop.frac=0.2,bootnumber = 100)
stopCluster(forkCL)
#####Mezclar entrenamiento y ensayo#####
#Mezcla de conjunto de entrenamiento y ensayo 18 veces para asegurarse de que la
selección primaria no está sesgada
#Combinar conjunto de datos de entrenamiento y ensayo
notest2<-as.matrix(data1[,~which(data1[10,]=="test set 2")])
notest2annot<-as.vector(notest2[7,])
notest2annot[which(notest2annot=="no cat")]<-0
notest2classes<-as.factor(notest2annot)

```

(continuación)

```

## notest2compounds<-levels(factor(unlist(as.vector(notest [2,]))))
## notest2<-as.matrix(notest2[-(1:17),])
## class(notest2)<-"numeric"
#;Generar una nueva matriz de muestreo sin DMSO ni Agua! Estos son controles y no
deberian utilizarse para muestreo para el conjunto de ensayo.
notestDMSO <-as.matrix(notest2[,~which(notest2[2,]=="dimethyl sulfoxide")])
notestDMSO Unstim <-as.matrix(notestDMSO[,~
which(notestDMSO[2,]=="unstimulated")])
notest2DU <-as.vector(notestDMSO Unstim[7,])
notest2DU[which(notest2DU=="no cat")]<-0
notest2DUclasses<-as.factor(notest2DU)
#####
##Construir bucles de mezcla
##;Mezclar entrenamiento y ensayo 20 veces! Los resultados se almacenan en testcomp,
que se accede tecleando testcomp[[i]]
testcomp <- c()
looptest <- NULL
looptrain <- NULL
looptests <- list()
trainingsetannot<-list()
trainingsetclasses<-list()
testsetannot<-list()
testsetclasses<-list()
testsetnames<-list()
vsrf<-list ()
rfvsrf<-list()
rfpr.pred<-list()
cf<-list ()
results<-list()
testcomploc<-list()
rfpr.predtest2<-list()
resultstest2<-list()
vsrf.vsb <-list()
#cl = makeCluster(4)
#registerDoSNOW(cl)
registerDoMC(18)
startTime = Sys.time()
writeLines(paste('Starting shuffle: ',startTime, sep = ''))
loop results = foreach(i = 1:18, .packages =
c('varSelRF', 'caret', 'randomForest')) %dopar% {
set.seed((10+i))
testcomp <-
(sample(sort(unique(levels(factor(notestDMSO Unstim[2,which(notest2DUclasse
s=="1A")]))),method="radix"),6))
testcomp <-
c(testcomp,sample(sort(unique(levels(factor(notestDMSO Unstim[2,which(notes
t2DUclasses=="1B")]))),method="radix"),6))
testcomp <-
c(testcomp,sample(sort(unique(levels(factor(notestDMSO Unstim[2,which(notes
t2DUclasses=="0")]))),method="radix"),6))

```

(continuación)

```

testcomploc <- c()
for(n in 1:length(testcomp))
testcomploc <- c(testcomploc ,which(notest2[2,]==testcomp[n]))
## Obtener ids de matriz original que contienen agua y muestras de DMSO.
# testcomploc <- unlist(testcomploc) looptest <-notest2[,testcomploc] looptrain <-
notest2[,~testcomploc]
trainingsetannot <- as.vector(looptrain[7,])
trainingsetannot[which(trainingsetannot=="no cat")]<-0 trainingsetclasses <-
as.factor(trainingsetannot) looptrain <- as.matrix(looptrain[-(1:17),])
class(looptrain) <- "numeric" looptrain <- as.data.frame(t(looptrain))
testsetannot <- as.vector(looptest[7,])
testsetannot[which(testsetannot=="no cat")]<-0 testsetclasses <-
as.factor(testsetannot) testsetnames <-as.vector(looptest[2,]) looptest <-
as.matrix(looptest[-(1:17),]) class(looptest) <- "numeric" looptest <-
as.data.frame(t(looptest)) vsrf <-
varSelRF(looptrain,Class=trainingsetclasses,ntree=2000,
ntreeIterat=2000 ,vars.drop.frac=0.2,keep.forest =
T,whole.range=TRUE,verbose = FALSE)
rfvsrf <- randomForest(y=trainingsetclasses, x=subset(looptrain,
select=vsrf$selected.vars),ntree=vsrf$ntreeIterat,importance=TRUE) rfpr.pred <-
predict(rfvsrf,newdata=subset(looptest, select = vsrf$selected.vars))
cf <- confusionMatrix(rfpr.pred,testsetclasses) results <-
t(rbind(stimulant=testsetnames,group=testsetclasses,predicted=rfpr.pred))
# La lista contiene todos los resultados # Los resultados se pueden encontrar a
través de loop_results[[i]]$
list(Results = results,VarSelRF = vsrf, RFValSelRF = rfvsrf,Predictions =
rfpr.pred,Confusion.Matrix = cf, TestSet = testsetnames,TestClasses =
testsetclasses, Train=looptrain,Trainclass=trainingsetclasses)
print(loop_results[[1]]$TestSet)
print(loop_results[[2]]$TestSet)
#stopCluster(cl)
##Ejecutar arranque
vsrf.vsb<-list()
for(i in 1:18){
forkCL <- makeForkCluster(12)
clusterSetRNGStream(forkCL, iseed = (100+i))
clusterEvalQ(forkCL, library(varSelRF))
vsrf.vsb[[i]]<-
varSelRFBoot(loop_results[[i]]$Train,Class=loop_results[[i]]$Trainclass,usi
ngCluster=T,srf=loop_results[[i]]$VarSelRF,TheCluster=forkCL,ntree=2000,ntr
eeIterat=2000,vars.drop.frac=0.2,bootnumber = 100)
stopCluster(forkCL)
print(Sys.time())
}

```

Resultados

Clasificaciones binarias de 37 productos químicos.

- 5 Se seleccionó un nuevo conjunto de datos que comprendía 37 productos químicos perfectamente caracterizados (Tabla 1) para complementar los datos históricos de GARD para 51 productos químicos y representar una selección relevante de productos químicos, equilibrada por lo que respecta a la clase de reactividad química y su uso en productos de consumo. Los nuevos productos químicos fueron seleccionados basándose en las bases de datos de CLP de la Agencia Química Europea y la literatura [15, 41] y en cooperación con el Grupo de Trabajo sobre
- 10 Tolerancia de la Piel de Cosméticos de la Comisión Europea (CE), que amablemente proporcionó 27 productos químicos. Los 37 productos químicos nuevos se predijeron como sensibilizadores o no sensibilizadores utilizando el GPS y protocolos establecidos previamente basados en clasificaciones SVM. El modelo SVM se aplicó para asignar a cada muestra replicada individual un DV SVM. Antes de las clasificaciones finales, se calibraron en primer lugar los DV SVM de las 37 muestras frente a 11 muestras de referencia (Tabla 2) incluidas en el mismo lote de muestra
- 15 que los productos químicos de ensayo. Con el fin de evaluar las predicciones binarias, en este punto se decidió priorizar los datos humanos [41] cuando estuvieron disponibles, en los que las clases uno a cuatro corresponden a sensibilizadores y cinco y seis a no sensibilizadores, en lugar de las clasificaciones CLP. El rendimiento del modelo que predice los 37 productos químicos se resume según una AUC ROC de 0,88, lo cual indica una buena capacidad discriminatoria y, tal como se ilustra en la Fig. 1A (productos químicos de referencia-línea continua; 37 productos químicos-línea discontinua). La sensibilidad, especificidad y precisión basadas en las estadísticas de Cooper se
- 20 estimaron en 73 %, 80 % y 76 %, respectivamente. En combinación con datos publicados previamente [23], la

precisión predictiva actualizada de GARD para la clasificación binaria de sensibilizadores de la piel se estima en 84 % sobre la base a un conjunto de datos que comprende un total de 74 productos químicos. Cuando los productos químicos en la Tabla 1 se agruparon de acuerdo con la clasificación CLP, y se resumieron los correspondientes valores SVM DV obtenidos durante la clasificación de los 37 productos químicos en un diagrama de caja presentado en la Fig. 1 B, surgió un gradiente de potencia, a medida que se asignaron a los sensibilizadores más fuertes DV SVM más altos en comparación con los sensibilizadores más débiles en la categoría 1B y los no sensibilizadores en la sin cat. De acuerdo con las pruebas de Wilcoxon de dos muestras no paramétricas con las que se comparan las distribuciones de muestras SVM DV, estos grupos difirieron significativamente (sin cat. frente a 1B: $p = 2,8^{-5}$; 1B frente a 1A $p = 2,8^{-6}$, sin cat. frente a 1A $p = 3,5^{-12}$). Aunque las diferencias entre los grupos fueron significativas, hubo cierta superposición entre los productos químicos individuales, lo cual indica que la información no fue suficiente para estratificar completamente muestras en grupos de potencia bien definidos.

Un modelo de bosque aleatorio para la predicción de categorías CLP

Para establecer una firma de biomarcador para la predicción de las categorías CLP, se fusionó el conjunto de datos que comprendía los 37 productos químicos nuevos con el conjunto de datos históricos que comprendía 51 productos químicos. En total, el conjunto de datos consistió en 86 productos químicos únicos (Tabla 4) y dos vehículos de control, equilibrados con respecto a las categorías 1A y 1 B y no sensibilizadores (sin cat.), tal como se describe en CLP (Tabla 3). Para cuatro productos químicos, la clasificación CLP 1B se cambió a ninguna categoría/no sensibilizador de acuerdo con las fuentes indicadas en la Tabla 4 por una entre tres razones: i) para retener la coherencia con proyectos GARD anteriores (benzaldehído, xileno), ii) para su utilización como vehículo en concentración no sensibilizadora (DMSO), y por ser un falso positivo perfectamente descrito en el LLNA (dodecil sulfato de sodio).

Se desarrolló un modelo de bosque aleatorio para la predicción de tres categorías CLP sobre la base de un conjunto de entrenamiento que consiste en 70 muestras únicas, que incluyen dos controles de vehículos. Se identificaron 52 variables predictivas (Tabla 5) como óptimas para la clasificación CLP. La tasa de error de predicción del modelo que se deriva del submuestreo aleatorio se estimó en 0.225, lo cual proporciona una indicación del rendimiento del modelo. Para visualizar el conjunto de datos utilizado para desarrollar el modelo, se realizó un análisis de componentes principales (PCA). Se utilizaron como entrada las 52 variables identificadas por bosque aleatorio sobre la base de un análisis de matriz de genoma completo y se construyó el PCA sobre el conjunto de entrenamiento (Fig. 2A). La Fig. 2A se basa en productos químicos con replicados biológicos coloreados de acuerdo con las categorías CLP y se puede observar un claro gradiente desde la ausencia de cat. hasta los sensibilizadores fuertes (1A) a lo largo del primer componente principal. El mapa de calor del conjunto de entrenamiento con agrupamiento jerárquico de las variables en la Fig. 2B ilustra la regulación de los genes en relación con la correspondiente categoría química y CLP.

Predicción de un conjunto de ensayos independiente

Se evaluó el rendimiento del modelo además por la predicción de las categorías CLP de un conjunto de ensayos independiente, que comprendió 18 productos químicos previamente inéditos para el modelo. Se visualizó el conjunto de ensayo coloreado de acuerdo con las categorías CLP en el gráfico de PCA (Fig. 3A), sin influir en los componentes de PCA sobre la base de las 52 variables identificadas. La Fig. 3B visualiza la regulación de los 52 genes en el conjunto de ensayo normalizado en forma de mapa de calor. Cuando se predijeron las réplicas por separado y se utilizó el voto mayoritario para clasificar los correspondientes productos químicos, 14 de los 18 productos químicos se asignaron a la categoría CLP correcta (Tabla 6, Supl. Tabla 1), lo cual resultó en una precisión general del 78 % (Tabla 7). Los cuatro productos químicos mal clasificados fueron maleato de dietilo, éter de butil glicidilo, Lyril y cloruro de cianúrico. La única predicción falsa negativa fue cloruro cianúrico, que se clasifica como 1A en CLP y sin cat., en el modelo de la invención, mientras que los tres productos químicos restantes se clasifican como CLP 1B pero se pronosticaron como 1A. En una etapa posterior para confirmar que las selecciones de entrenamiento y conjunto de ensayos no eran sesgados y que el modelo de predicción no dependía por completo de la composición del conjunto de entrenamiento, se construyeron 18 modelos de bosques aleatorios alternativos, en los que se barajó al azar la composición de productos químicos del conjunto de entrenamiento y ensayo. Para cada nuevo modelo, se repitió el proceso completo de selección de variables, tal como se ha descrito anteriormente. Dependiendo del número de muestras replicadas disponibles para cada estimulación química, el número total de muestras en cada entrenamiento y conjunto de ensayo varió, pero el número de productos químicos en cada conjunto y su distribución de CLP se mantuvieron constantes. Los modelos alternativos fueron todos significativos y la tasa de error de predicción promedio obtenida del procedimiento de submuestreo aleatorio fue idéntica al modelo inicial en 0.22, lo cual corrobora que no se obtuvo el modelo presentado debido a la selección sesgada de los conjuntos de entrenamiento y de ensayo.

El modelo de potencia CLP contiene información relevante para la predicción de la potencia humana

A continuación, se utilizó PCA para investigar el comportamiento de las 52 variables (Tabla 5) utilizando información relacionada con las categorías de potencia humana [41]. Se colorearon las 70 muestras de entrenamiento y 18 muestras de ensayo de acuerdo con la potencia humana (clase 1-6, 6 = verdadero no sensibilizador, 1 = el sensibilizador más fuerte) y se eliminaron las muestras para las que no había disponible una categoría de potencia

humana (ver Tabla 4). Aunque se ha desarrollado el modelo para predecir las categorías de potencia CLP, también contiene información relacionada con la potencia humana, tal como se ilustra en las Fig. 4A, B.

Identidad de las variables del modelo de bosque aleatorio

Las 52 variables identificadas por bosque aleatorio (Tabla 5) representan transcritos que pertenecen a diferentes compartimentos celulares y diferentes funciones funcionales. Cinco de ellos se superponen con la firma de predicción GARD. Cinco de los primeros 10 marcadores, que fueron seleccionados con mayor frecuencia en el proceso de submuestreo aleatorio y que tienen las frecuencias de llamada de validación máximas (Tabla 5), son miembros de la agrupación de histonas 1, como HIST H2AB [42]. Las histonas están altamente conservadas y desempeñan un papel importante no solo para mantener la estructura de la cromatina sino también en la regulación génica [43]. PFAS y PAICS tienen relación con la biosíntesis de purina [44, 45] y TMEM97 es un regulador de los niveles de colesterol [46], que según se ha descrito está relacionado con la regulación del ciclo celular, la migración celular y la invasión en un modelo celular de glioma de acuerdo con experimentos de interferencia de ARN [47]. DHCR24 es una enzima multifuncional localizada en el retículo endoplásmico (ER) y cataliza la etapa final en la síntesis de colesterol [48], pero posee también actividad antiapoptótica como se muestra, por ejemplo, para las células neuronales bajo estrés ER [49]. Se ha demostrado que PLK1, una quinasa, se fosforila como respuesta a la activación de TLR y los resultados de la interferencia de ARN indicaron que la señalización de PLK1 estaba relacionada con la respuesta inflamatoria inducida por TLR [50]. Además, se ha notificado que PLK1 está relacionada con la regulación del ciclo celular pues inhibe la expresión de ciclina D1 inducida por TNF y podría reducir la activación de NF- κ B inducida por TNF [51]. Muchos de los transcritos restantes son proteínas nucleares y, por tanto, probablemente relacionados con procedimientos dependientes del ADN, como replicación, transcripción, empalme y regulación del ciclo celular. Hay más transcritos en la firma que codifican proteínas conocidas por su participación en las respuestas inmunitarias y/o sensibilización, como NQO1, que está perfectamente descrita por su papel en la respuesta celular a los sensibilizadores de la piel [52]. CD53 pertenece a la familia de la tetraspanina, proteínas transmembrana que tienen múltiples funciones, p.ej., adhesión celular, migración y señalización y, según se ha demostrado, es elevado en las CD de piel con Fc ϵ RI positivo de pacientes con dermatitis atópica [53], y de forma similar en monocitos derivados de sangre periférica de pacientes con eccema atópico [54] en comparación con los correspondientes controles sanos. CD44 es una glucoproteína de la superficie celular, receptor de adhesión y hialuronano [55] expresado por numerosos tipos de células y que está relacionado por ejemplo con las respuestas inflamatorias [56], p.ej., al mediar la migración de leucocitos en tejidos inflamados, lo cual se ha demostrado en un modelo de ratón de dermatitis alérgica [57].

Las rutas reguladas comunes y únicas son inducidas por sensibilizadores que difieren en su reactividad de proteínas

Las 33 rutas clave identificadas con una entrada de los 883 genes regulados más significativamente después de una comparación multigrupo (FDR = 10^{-9}) entre las categorías CLP en el análisis KPA (Fig. 5), reflejan varios de los grupos funcionales de las 52 variables definidas por bosque aleatorio, como por ejemplo regulación de genes, control del ciclo celular y metabolismo. Las rutas asociadas a la respuesta inmunitaria como "reguladores inducidos por IL-4 de crecimiento, supervivencia, diferenciación y metabolismo de células" y "señalización IL-3 de respuesta inmunitaria a través de JAK/STAT, p38, JNK y NF- κ B" estuvieron entre los 50 % regulados de manera más significativa.

Los análisis descritos a continuación se centraron en los tres grupos de reactividad química más grandes del presente conjunto de datos; sustitución nucleófila (SN), adición de Michael (MA) y formación de base de Schiff (SB). Entre los productos químicos incluidos, la mayoría de los productos químicos etiquetados como "sin cat." no poseían propiedades de unión a proteínas; sin embargo, hubo algunas formaciones de SB y productos químicos SN presentes. En la categoría 1B, estaban representados casi todos los tipos de reactividad de proteínas, mientras que había un claro dominio de productos químicos MA en la categoría 1A.

Para cada reactividad de proteína asociada, se pudieron identificar rutas únicas para sensibilizar productos químicos que pertenecen al correspondiente grupo de reactividad de proteínas, tal como se presenta en la Fig. 6. Estos resultados combinan genes regulados de forma diferenciada de los datos de entrada con los llamados centros clave, moléculas que pueden regular el nivel de expresión de los genes de entrada. No pueden identificarse necesariamente ellos mismos mediante experimentos de expresión génica, ya que su regulación puede ser visible en otros niveles biológicos, como los cambios de actividad (p.ej., para las cinasas) o los cambios pueden ser a corto plazo o de baja magnitud. En total, 173 genes fueron comunes para los tres grupos de reactividad (Fig. 7A) y estuvieron presentes seis rutas en los tres grupos de reactividad (Fig. 7B); "Ciclo celular: papel de APC en la regulación del ciclo celular", "Ciclo celular: Papel de complejo SCF en la regulación del ciclo celular", "Desarrollo: Regulación de transcripción de desarrollo de granulocito", "Ciclocelular: Ciclo celular (esquema genérico)", "Daño de ADN: regulación ATM/ATR de punto de control G1/S" y "Acción mitogénica de Estradiol/ESR1 (nuclear) en cáncer de mama". También en este caso, se representaron altamente las rutas de ciclo. Se identificaron las respuestas al estrés oxidativo como parte de los resultados de la ruta clave solamente para productos químicos MA y SB (Fig. 6). En el grupo de productos químicos sensibilizadores MA, se observó como centros claves KEAP1 y NRF2, así como sus genes diana NQO1 y HMOX1 [52, 58] y AHR [59, 60]. Para productos químicos MA, los genes diana NQO1, HMOX1 y CES1 [61] estuvieron presentes en el nivel de gen de entrada. En el nivel de entrada, estuvo presente CEST1 para productos químicos SN también, pero se identificaron solamente NQO1, NRF2 y AHR como centros

clave. No se encontró KEAP1 como centro clave en el análisis SB y SN KPA, y AHR fue el único centro clave identificado para los tres grupos de reactividad de proteínas. Las subunidades NF- κ B (RelB o p52) fueron centros clave pronosticados en todos los grupos de reactividad excepto en los productos químicos MA.

- 5 En suma, parece haber respuestas de la mecánicas comunes a las exposiciones químicas en sí, como el ciclo celular y relacionadas con el daño de ADN, pero los resultados del análisis de la ruta también corroboran la hipótesis de que las diferentes clases de reactividad química inducen rutas de señalización distintas, tal como hemos observado anteriormente [24]] y en otros sistemas experimentales [14, 21, 62, 63]. Varias rutas están ligadas a procedimientos conocidos por ser relevantes en la sensibilización de la piel.

Análisis

- 10 La cantidad de producto químico por área de piel expuesta que induce la sensibilización varía significativamente [41]; por lo tanto, la información de la potencia del sensibilizador de la piel es imprescindible para una evaluación precisa del riesgo. Las personas que han desarrollado procedimientos de ensayo alternativos dependen de los datos clínicos humanos para conseguir una alta capacidad de predicción de la sensibilización humana, sin embargo, este tipo de datos es bastante escaso y la mayoría de los datos disponibles se derivan del LLNA [64]. A pesar del hecho
- 15 de que los modelos animales reflejan la complejidad de enfermedades sistémicas, como sensibilización de la piel, los datos *in vitro* han demostrado hasta el momento una buena correlación y un rendimiento incluso mejor que los modelos animales, especialmente cuando se combinan en un ITS [65]. Asimismo, los sistemas de ensayo alternativos pueden proporcionar información de la mecánica que los ensayos con animales completos no pueden proporcionar [66].
- 20 En el presente documento, se muestra un enfoque para predecir la potencia del sensibilizador de la piel utilizando el sistema CLP basado en un modelo de células dendríticas (DC) y caracterización de transcritos. Las categorías de CLP son categorías determinadas empíricamente y definidas arbitrariamente, que no representan la diversidad de diferentes productos químicos, sus características moleculares y mecanismos responsables de sus características de sensibilización o la falta de las mismas. Sin embargo, son lo que la legislación exige actualmente para clasificar y
- 25 etiquetar los productos químicos. Por lo tanto, se investigó 37 productos químicos adicionales que no se habían analizado previamente en el ensayo GARD convencional, para combinar estos nuevos datos con conjuntos de datos históricos. En las clasificaciones binarias de estos nuevos productos químicos de acuerdo con el modelo GARD establecido, cuatro productos químicos sensibilizadores mal clasificadas estuvieron cerca del límite, tal como se define en las muestras de referencia, concretamente, anilina, benzocaína, limoneno y butil glicidil éter. Tres de estos
- 30 pertenecen a la clase 4 de potencia humana, lo cual demuestra que el límite del modelo es crítico para traducir los valores de SVM, que a menudo se correlacionan bien con la potencia, en clasificaciones precisas. En combinación con las predicciones históricas, GARD sigue mostrando una alta precisión general del 84 % para las clasificaciones binarias.

- A continuación, los autores de la invención utilizaron tanto datos nuevos como históricos para desarrollar un modelo
- 35 de bosque aleatorio para cada categoría CLP, que muestra precisiones equilibradas [67] de 96 % para ausencia de categoría, 79 % para la categoría 1A y 75 % para la categoría 1B (sobre la base de los votos mayoritarios, para el rendimiento en forma replicada, véase Tabla 7). Se clasificaron erróneamente butil glicidil éter, maleato de dietilo, cloruro cianúrico y lyral; la única predicción falsa negativa fue sin categoría en lugar de 1A para el cloruro cianúrico. Sin embargo, cloruro cianúrico reacciona exotérmicamente con agua formando cloruro de hidrógeno y
- 40 posiblemente otros productos de reacción. Debido a esta reacción de hidrolización, que probablemente ya ocurre en DMSO (que contiene agua), se desconoce la cantidad de cloruro cianúrico y los productos de reacción presentes en el ensayo. Este producto químico puede quedar fuera del dominio de aplicabilidad de los ensayos basados en la plataforma GARD. Sin embargo, nada en el control de calidad u otros análisis de modelado previo motivaron la eliminación de estas muestras. Maleato de dietilo y lyral se clasifican como 1B por CLP, pero como 1A según el
- 45 modelo de los autores de la invención, que también en este caso parece ajustarse más a su categoría de potencia humana, categoría 2, tal como describe Basketter y col. [41] El cuarto producto químico éter de butil glicidilo mal clasificado es una categoría de potencia humana 3. Evidentemente, predecir 1B, es decir, sensibilizadores débiles, se manifestó como la parte más desafiante. También en el LLNA, las predicciones de potencia de los sensibilizadores débiles varían más que las de los sensibilizadores fuertes [8, 68, 69]. Asimismo, 1B es un grupo
- 50 muy heterogéneo, teniendo en cuenta tanto el intervalo de concentraciones de LLNA EC3 como las categorías de potencia humana asociadas a los productos químicos resumidos en la categoría 1B.

- El ensayo U-SENS™, anteriormente MUSST, utiliza otra línea celular mieloide, U937 y mediciones de CD86 para distinguir sensibilizadores y no sensibilizadores. Cuando los autores combinaron CD86 con datos de citotoxicidad y
- 55 ciertos niveles límite para predecir las categorías de CLP, se registraron predicciones correctas del 82 % de Cat. 1 A (41/50) y 73 % de Cat. 1 B/No Cat (85/116) [15]. Sin embargo, queda pendiente de aclarar cómo resultaría la discriminación más desafiante entre sin cat. y 1B. Cottrez y col. [70] han publicado recientemente un estudio, en el que notifican que su ensayo alternativo SENS-IS, un modelo basado en epidermis reconstituido en 3D, funciona muy bien para la predicción de la potencia del sensibilizador de la piel; sin embargo, no se dirigen a las categorías CLP. A juzgar por la Fig. 4, puramente basada en la entrada variable 52, que se definió para predecir categorías CLP, el
- 60 modelo de los autores de la invención parece contener información relevante para la clasificación de la potencia humana. Una vez que hayan recibido clasificaciones de potencia humana más productos químicos, debería ser

posible desarrollar sin problemas un modelo de potencia humana basado en la plataforma GARD.

El análisis de la ruta de KPA identificó eventos biológicamente relevantes en el conjunto de datos presentado ya que varias rutas reguladas tienen un papel conocido en la sensibilización de la piel, p.ej., la señalización de citocinas y las respuestas al estrés oxidativo (Figuras 5-6). Aunque las DC no son el objetivo principal para la modificación de proteínas *in vivo*, se planteó la hipótesis de que diferentes clases de reactividad de proteínas influyen en el transcriptoma de DC de manera diferenciada. La reactividad de proteínas es una de las características más importantes de los productos químicos que definen su capacidad y potencia de sensibilización de la piel con ciertas limitaciones [63]. Los patrones específicos de reactividad de proteínas fueron detectables tal como lo revela la comparación de los genes regulados más significativamente inducidos por los grupos de reactividad MA, SB y SN. Cabe destacar que las subunidades NF-KB fueron centros clave pronosticados para todos los grupos de reactividad, excepto para los productos químicos MA, lo cual puede reflejar el efecto inhibitorio descrito sobre la señalización de NF-kB de este tipo de productos químicos [71]. Aunque algunas rutas no parecen encajar en el contexto, como la "Acción mitogénica del estradiol/ESR1 (nuclear) en el cáncer de mama", una mirada más de cerca de las moléculas reguladas revela que son ciertamente relevantes también para otras rutas. En este caso, por ejemplo, participaron p21, c-myc, E2F1, SGOL2 (shugoshin 2) y CAD (carbamoil fosfato sintetasa), entre los cuales los primeros son reguladores del ciclo celular/factores de transcripción conocidos [72] y desempeñan un papel en segregación cromosómica (SGOL2) [73]. CAD, una enzima, que limita la velocidad en la biosíntesis de nucleótidos de pirimidina, por otro lado, ha sido relacionada más recientemente también con cooperación con las rutas de señalización celular [74] y parece inhibir la función antibacteriana del sensor bacteriano NOD2 (dominio de oligomerización de unión a nucleótidos 2) en células epiteliales intestinales humanas [75]. Estos ejemplos pueden ilustrar que los datos transcriptómicos de los autores de la invención merecen una mayor atención y análisis más detallados, y este tipo de análisis, utilizando diferentes herramientas bioinformáticas y, finalmente, análisis funcionales, puede contribuir a dilucidar los mecanismos que subyacen tras los procedimientos biológicos y las enfermedades.

Tal como se ha explicado anteriormente, la asignación de clases de potencia correctas a productos químicos con potencia débil o intermedia parece ser un problema más general. Benigni y col. [76] presentaron datos que demuestran que incluso los sistemas experimentales *in vivo*, aunque en general se correlacionan bien con los datos humanos, funcionan peor para los sensibilizadores de potencia intermedia. Sostienen además que la etapa de modificación de proteínas es la etapa de limitación de la velocidad de todo el proceso de sensibilización y que las pruebas *in vitro* dirigidas a otros eventos de AOP no añaden mucha información. Por otro lado, solo existe una relación débil entre la constante de velocidad de los sensibilizadores de MA según lo determina el perfil cinético con un péptido modelo y su potencia en el LLNA [71]. Se ha descrito que esto está relacionado con el efecto antiinflamatorio de productos químicos MA al inhibir la señalización de NF-kB, que aumenta con la reactividad. Sin embargo, teniendo en cuenta los eventos clave en el AOP de sensibilización de la piel, el proceso de sensibilización como tal puede entenderse como un continuo y, por lo tanto, no solo se caracterizaría por eventos aislados. Por supuesto, en realidad el proceso debe comenzar con la penetración del producto químico en la piel. En este contexto, debe tenerse en cuenta que se ha demostrado recientemente como erróneo el concepto de que la capacidad de un producto químico para penetrar eficientemente en el estrato córneo es crucial para su capacidad y su potencia de sensibilización de la piel [77]. Además, el acceso a las capas inferiores de la piel puede verse facilitado en realidad en gran medida por la función de barrera cutánea deteriorada, debido, p.ej., al trabajo húmedo y pequeñas heridas. Cabe destacar que muchos sensibilizadores fuertes poseen propiedades irritantes, que se correlacionan con la citotoxicidad, y la citotoxicidad a su vez parece contribuir a la potencia del sensibilizador [18], que también se refleja en el conjunto de datos de la invención (datos no mostrados). La citotoxicidad también está relacionada con la reactividad de proteína del producto químico: los productos químicos, que son fuertemente reactivos a la cisteína, en general son citotóxicos y, por lo tanto, pueden interferir con la función enzimática vital [18, 78]. Al mismo tiempo, los efectos irritantes pueden generar señales de peligro como el ATP extracelular o los productos de degradación del ácido hialurónico [79, 80], que pueden servir para activar las DC y, en consecuencia, para cebar linfocitos T no tratados previamente. Por otra parte, pueden ser difíciles de implementar en cualquier sistema de ensayo otros factores como la inflamación preexistente y la exposición conjunta a otras sustancias, que sin duda desempeñan un papel en la sensibilización alérgica. Dado que los ensayos *in vitro* deben afrontar la demanda de ser rentables de fácil funcionamiento, existen limitaciones evidentes en lo que se puede lograr *in vitro*, pero es alentador hasta el momento el rendimiento de los ensayos [76]. Aunque la reactividad de las proteínas puede ser muy importante, los sistemas a base de células deberían ser capaces de recapitular ciertos eventos adicionales por encima de la reactividad del péptido. Lo más probable es que el rendimiento de los ensayos alternativos se mejore aún más a medida que se revelen detalles más de la mecánica de la sensibilización de la piel, lo cual permitirá identificar tanto los dominios de aplicación como las dificultades más fácilmente.

En conclusión, los autores de la invención han identificado una firma de biomarcador predictivo que comprende 52 transcritos para la clasificación de compuestos sensibilizadores de la piel en grupos CLP, tal como lo exige la legislación vigente. Cuando se desafió con 18 compuestos de ensayo independientes, el ensayo proporcionó resultados precisos para el 78 % de las predicciones de potencia. Además identificó 11/12 sensibilizadores correctamente, lo cual indica que es bastante conservador, es decir, evita predicciones falsas negativas. Dado que la firma de biomarcadores presentada está optimizada para las predicciones de potencia, y no solo para las clasificaciones de peligro binario, los autores de la invención indican una posible aplicación de su modelo dentro de

una Estrategia de Pruebas Integradas para predicciones precisas de potencia, similar a lo indicado [18]. Por otra parte, los resultados ilustran efectivamente la flexibilidad y versatilidad de la configuración GARD. La medición de los transcriptomas completos de las células proporciona la oportunidad de realizar estudios basados en el modo de acción para identificar las rutas de sensibilización, pero también para identificar las firmas predictivas de biomarcador, previamente demostradas para las predicciones binarias de sensibilización de la piel y la sensibilización respiratoria. Por lo tanto, se presenta en el presente documento una prueba inicial del concepto de un modelo de potencia dirigido a los grupos CLP, que puede modificarse y mejorarse a medida que se analicen más muestras y surjan datos de referencia humanos más precisos.

Referencias

1. Peiser, M., y col., Allergic contact dermatitis: epidemiology, molecular mechanisms, in vitro methods and regulatory aspects. Current knowledge assembled at an international workshop at BfR, Alemania. *Cell Mol Life Sci*, 2012. 69(5): p. 763-81.
2. Behroozy, A. and T.G. Keegel, Wet-work Exposure: A Main Risk Factor for Occupational Hand Dermatitis. *Saf Health Work*, 2014. 5(4): p. 175-80.
3. European Parliament, C.o.t.E.U., REGULATION (EC) No 1223/2009 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. 2009.
4. Hartung, T. and C. Rovida, Chemical regulators have overreached. *Nature*, 2009. 460(7259): p. 1080-1.
5. OECD, The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 1: Scientific Evidence. 2012: p. 1-59.
6. Gerberick, G.F., y col., Local lymph node assay (LLNA) for detection of sensitization capacity of chemicals. *Methods*, 2007. 41(1): p. 54-60.
7. Anderson, S.E., P.D. Siegel, and B.J. Meade, The LLNA: A Brief Review of Recent Advances and Limitations. *J Allergy (Cairo)*, 2011. 2011: p. 424203.
8. Ezendam, J., H.M. Braakhuis, and R.J. Vandebruiel, State of the art in non-animal approaches for skin sensitization testing: from individual assay methods towards ensaying strategies. *Arch Toxicol.*, 2016.
9. Andreas, N., y col., The intra- and inter-laboratory reproducibility and predictivity of the KeratinoSens assay to predict skin sensitizers in vitro: Results of a ring-study in five laboratories. *Toxicology in Vitro*, 2011. 25(3): p. 733-744.
10. Natsch, A. and R. Emter, Skin sensitizers induce antioxidant response element dependent genes: application to the in vitro ensaying of the sensitization potential of chemicals. *Toxicol Sci*, 2008. 102(1): p. 110-9.
11. Gerberick, G.F., y col., Development of a peptide reactivity assay for screening contact allergens. *Toxicol Sci*, 2004. 81(2): p. 332-43.
12. Ashikaga, T., y col., Development of an in vitro skin sensitization test using human cell lines: the human Cell Line Activation Ensayo (h-CLAT). I. Optimization of the h-CLAT protocol. *Toxicol In Vitro*, 2006. 20(5): p. 767-73.
13. Teunis, M.A., y col., International ring trial of the epidermal equivalent sensitizer potency assay: reproducibility and predictive-capacity. *ALTEX*, 2014. 31(3): p. 251-68.
14. Cottrez, F., y col., Genes specifically modulated in sensitized skins allow the detection of sensitizers in a reconstructed human skin model. Development of the SENS-IS assay. *Toxicol In Vitro*, 2015. 29(4): p. 787-802.
15. Piroid, C., y col., The Myeloid U937 Skin Sensitization Test (U-SENS) addresses the activation of dendritic cell event in the adverse outcome pathway for skin sensitization. *Toxicol In Vitro*, 2015. 29(5): p. 901-16.
16. Dearden, J.C., y col., Mechanism-Based QSAR Modeling of Skin Sensitization. *Chem Res Toxicol.*, 2015. 28(10): p. 1975-86.
17. Tsujita-Inoue, K., y col., Skin sensitization risk assessment model using artificial neural network analysis of data from multiple in vitro assays. *Toxicol In Vitro*, 2014. 28(4): p. 626-39.
18. Jaworska, J.S., y col., Bayesian integrated ensaying strategy (ITS) for skin sensitization potency assessment: a decision support system for quantitative weight of evidence and adaptive ensaying strategy. *Arch Toxicol.*, 2015. 89(12): p. 2355-83.
19. Jaworska, J., y col., Bayesian integrated testing strategy to assess skin sensitization potency: from theory to practice. *J Appl Toxicol.*, 2013. 33(11): p. 1353-64.
20. Luechtefeld, T., y col., Probabilistic hazard assessment for skin sensitization potency by dose-response modeling using feature elimination instead of quantitative structure-activity relationships. *J Appl Toxicol.*, 2015. 35(11): p. 1361-71.
21. Natsch, A., y col., Predicting skin sensitizer potency based on in vitro data from KeratinoSens and kinetic peptide binding: global versus domain-based assessment. *Toxicol Sci*, 2015. 143(2): p. 319-32.
22. Johansson, H., y col., A genomic biomarker signature can predict skin sensitizers using a cell-based in vitro alternative to animal tests. *BMC Genomics*, 2011. 12: p. 399.
23. Johansson, H., y col., Genomic allergen rapid detection in-house validation--a proof of concept. *Toxicol Sci*, 2014. 139(2): p. 362-70.
24. Albrekt, A.S., y col., Skin sensitizers differentially regulate signaling pathways in MUTZ-3 cells in relation to their individual potency. *BMC Pharmacol Toxicol.*, 2014. 15: p. 5.
25. European Parliament, C.o.t.E.U., <http://echa.europa.eu/sv/regulations/clp/>. accessed July 13th, 2016.
26. Breiman, L., Random forests. *Machine Learning*, 2001. 45.
27. Diaz-Uriarte, R. and S. Alvarez de Andrés, Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 2006. 7(1): p. 1-13.

28. Johansson, H., y col., The GARD assay for assessment of chemical skin sensitizers. *Toxicol In Vitro*, 2013. 27(3): p. 1163-9.
29. Piccolo, S.R., y col., A single-sample microarray normalization method to facilitate personalized-medicine workflows. *Genomics*, 2012. 100(6): p. 337-44.
- 5 30. Stephen R. Piccolo, A.H.B., W. Evan Johnson. <https://www.bioconductor.org/packages/release/bioc/html/SCAN.UPC.html>. [citado el 14 de octubre de; Bioconductor version: Release (3.3):]
31. Forreryd, A., y col., From genome-wide arrays to tailor-made biomarker readout - Progress towards routine analysis of skin sensitizing chemicals with GARD. *Toxicol In Vitro*, 2016.
32. Lasko, T.A., y col., The use of receiver operating characteristic curves in biomedical informatics. *J Biomed Inform*, 2005. 38(5): p. 404-15.
- 10 33. Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., Weingessel, A., e1071: Misc functions of the Department of statistics (q1071). TU Wien. R package version 1.6, 2011. <http://CRAN.R-project.org/package=e1071>.
34. Sing, T., y col., ROCR: visualizing classifier performance in R. *Bioinformatics*, 2005. 21(20): p. 3940-1.
35. Cooper, J.A., 2nd, R. Saracci, and P. Cole, Describing the validity of carcinogen screening tests. *Br J Cancer*, 1979. 39(1): p. 87-9.
- 15 36. Jeffrey T. Leek, W.E.J., Hilary S. Parker, Andrew E. Jaffe, John D. Storey, sva: Surrogate Variable Analysis. R package version 3.10.0., 2014.
37. Johnson, W.E., C. Li, and A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, 2007. 8(1): p. 118-27.
- 20 38. Diaz-Uriarte, R., GeneSrf and varSelRF: a web-based tool and R package for gene selection and classification using random forest. *BMC Bioinformatics*, 2007. 8: p. 328.
39. Wiener, A.L.a.M., Classification and Regression by randomForest. *R News*, 2002. 2(3), 18--22.
40. Key Pathway Advisor by Clarivate Analytics (Formerly the IP & Science business of Thomson Reuters), <http://ipscience.thomsonreuters.com/product/metacore/>. 2016.
- 25 41. Basketter, D.A., y col., Categorization of chemicals according to their relative human skin sensitizing potency. *Dermatitis*, 2014. 25(1): p. 11-21.
42. Singh, R., y col., Increasing the complexity of chromatin: functionally distinct roles for replication-dependent histone H2A isoforms in cell proliferation and carcinogenesis. *Nucleic Acids Res*, 2013. 41(20): p. 9284-95.
43. Harshman, S.W., y col., H1 histones: current perspectives and challenges. *Nucleic Acids Res*, 2013. 41(21): p. 9593-609.
- 30 44. Lane, A.N. and T.W.-M. Fan, Regulation of mammalian nucleotide metabolism and biosynthesis. *Nucleic Acids Research*, 2015.
45. Li, S.X., y col., Octameric structure of the human bifunctional enzyme PAICS in purine biosynthesis. *J Mol Biol*, 2007. 366(5): p. 1603-14.
- 35 46. Bartz, F., y col., Identification of cholesterol-regulating genes by targeted RNAi screening. *Cell Metab*, 2009. 10(1): p. 63-75.
47. Qiu, G., y col., RNA interference against TMEM97 inhibits cell proliferation, migration, and invasion in glioma cells. *Tumour Biol*, 2015. 36(10): p. 8231-8.
48. Waterham, H.R., y col., Mutations in the 3beta-hidroxiesterol Delta24-reductase gene cause desmosterolosis, an autosomal recessive disorder of cholesterol biosynthesis. *Am J Hum Genet*, 2001. 69(4): p. 685-94.
- 40 49. Lu, X., y col., 3 beta-hidroxiesteroid-Delta 24 reductase (DHCR24) protects neuronal cells from apoptotic cell death induced by endoplasmic reticulum (ER) stress. *PLoS One*, 2014. 9(1): p. e86753.
50. Hu, J., y col., Polo-like kinase 1 (PLK1) is involved in toll-like receptor (TLR)-mediated TNF-alpha production in monocytic THP-1 cells. *PLoS One*, 2013. 8(10): p. e78832.
- 45 51. Higashimoto, T., y col., Regulation of I(kappa)B kinase complex by fosforylation of (gamma)-binding domain of I(kappa)B kinase (beta) by Polo-like kinase 1. *J Biol Chem*, 2008. 283(51): p. 35354-67.
52. Ade, N., y col., HMOX1 and NQO1 genes are upregulated in response to contact sensitizers in dendritic cells and THP-1 cell line: role of the Keap1/Nrf2 pathway. *Toxicol Sci*, 2009. 107(2): p. 451-60.
53. Peng, W.M., y col., Tetraspanins CD9 and CD81 are molecular partners of trimeric FcγepsilonRI on human antigen-presenting cells. *Allergy*, 2011. 66(5): p. 605-11.
- 50 54. Jockers, J.J. and N. Novak, Different expression of adhesion molecules and tetraspanins of monocytes of patients with atopic eczema. *Allergy*, 2006. 61(12): p. 1419-22.
55. Lee-Sayer, S.S., y col., The where, when, how, and why of hyaluronan binding by immune cells. *Front Immunol*, 2015. 6: p. 150.
56. Johnson, P. and B. Ruffell, CD44 and its role in inflammation and inflammatory diseases. *Inflamm Allergy Drug Targets*, 2009. 8(3): p. 208-20.
57. Gonda, A., y col., CD44, but not I-selectin, is critically involved in leucocyte migration into the skin in a murine model of allergic dermatitis. *Exp Dermatol*, 2005. 14(9): p. 700-8.
- 60 58. Natsch, A., The Nrf2-Keap1-ARE toxicity pathway as a cellular sensor for skin sensitizers--functional relevance and a hypothesis on innate reactions to skin sensitizers. *Toxicol Sci*, 2010. 113(2): p. 284-92.
59. Schulz, V.J., y col., Aryl hidrocarbon receptor activation affects the dendritic cell phenotype and function during allergic sensitization. *Immunobiology*, 2013. 218(8): p. 1055-62.
60. Kohle, C. and K.W. Bock, Coordinate regulation of Phase I and II xenobiotic metabolisms by the Ah receptor and Nrf2. *Biochem Pharmacol.*, 2007. 73(12): p. 1853-62.
- 65 61. Roberts, D.W., A.O. Aptula, and G. Patlewicz, Electrophilic chemistry related to skin sensitization. Reaction mechanistic applicability domain classification for a published dataset of 106 chemicals tested in the mouse local

lymph node assay. *Chem Res Toxicol.*, 2007. 20(1): p. 44-60.

62. Migdal, C., y col., Reactivity of chemical sensitizers toward amino acids in cellulo plays a role in the activation of the Nrf2-ARE pathway in human monocyte dendritic cells and the THP-1 cell line. *Toxicol Sci*, 2013. 133(2): p. 259-74.

63. Chipinda, I., J.M. Hettick, and P.D. Siegel, Haptenation: chemical reactivity and protein binding. *J Allergy (Cairo)*, 2011. 2011: p. 839682.

64. Basketter, D.A., y col., Nothing is perfect, not even the local lymph node assay: a commentary and the implications for REACH. *Contact Dermatitis*, 2009. 60(2): p. 65-9.

65. Urbisch, D., y col., Assessing skin sensitization hazard in mice and men using non-animal test methods. *Regul Toxicol Pharmacol.*, 2015. 71(2): p. 337-51.

66. Natsch, A., y col., Chemical basis for the extreme skin sensitization potency of (E)-4-(ethoxymetile)-2-phenyloxazol-5(4H)-one. *Chem Res Toxicol.*, 2010. 23(12): p. 1913-20.

67. Brodersen, K.H., y col., The Balanced Accuracy and Its Posterior Distribution, in *Proceedings of the 2010 20th International Conference on Pattern Recognition*. 2010, IEEE Computer Society. p. 3121-3124.

68. Dumont, C., y col., Analysis of the Local Lymph Node Assay (LLNA) variability for assessing the prediction of skin sensitisation potential and potency of chemicals with non-animal approaches. *Toxicol In Vitro*, 2016. 34: p. 220-8.

69. Hoffmann, S., LLNA variability: An essential ingredient for a comprehensive assessment of non-animal skin sensitization ensayo methods and strategies. *ALTEX*, 2015. 32(4): p. 379-83.

70. Cottrez, F., y col., SENS-IS, a 3D reconstituted epidermis based model for quantifying chemical sensitization potency: Reproducibility and predictivity results from an inter-laboratory study. *Toxicol in Vitro*, 2016. 32: p. 248-60.

71. Natsch, A., T. Haupt, and H. Laue, Relating skin sensitizing potency to chemical reactivity: reactive Michael acceptors inhibit NF-kappaB signaling and are less sensitizing than S(N)Ar- and S(N)2- reactive chemicals. *Chem Res Toxicol.*, 2011. 24(11): p. 2018-27.

72. Buchmann, A.M., S. Swaminathan, and B. Thimmapaya, Regulation of cellular genes in a chromosomal context by the retinoblastoma tumor suppressor protein. *Mol Cell Biol*, 1998. 18(8): p. 4565-76.

73. Xu, Z., y col., Structure and function of the PP2A-shugoshin interaction. *Mol Cell*, 2009. 35(4): p. 426-41.

74. Huang, M. and L.M. Graves, De novo synthesis of pyrimidine nucleotides; emerging interfaces with signal transduction pathways. *Cell Mol Life Sci*, 2003. 60(2): p. 321-36.

75. Richmond, A.L., y col., The nucleotide synthesis enzyme CAD inhibits NOD2 antibacterial function in human intestinal epithelial cells. *Gastroenterology*, 2012. 142(7): p. 1483-92 e6.

76. Benigni, R., C. Bossa, and O. Tcheremenskaia, A data-based exploration of the adverse outcome pathway for skin sensitization points to the necessary requirements for its prediction with alternative methods. *Regul Toxicol Pharmacol.*, 2016. 78: p. 45-52.

77. Fitzpatrick, J.M., D.W. Roberts, and G. Patlewicz, What determines skin sensitization potency: Myths, maybes and realities. The 500 molecular weight cut-off: An updated analysis. *J Appl Toxicol.*, 2016.

78. Bohme, A., y col., Kinetic glutathione chemoassay to quantify thiol reactivity of organic electrophiles--application to alpha,beta-unsaturated ketones, acrylates, and propiolates. *Chem Res Toxicol.*, 2009. 22(4): p. 742-50.

79. Esser, P.R., y col., Contact sensitizers induce skin inflammation via ROS production and hyaluronic acid degradation. *PLoS One*, 2012. 7(7): p. e41340.

80. Martin, S.F., y col., Mechanisms of chemical-induced innate immunity in allergic contact dermatitis. *Allergy*, 2011. 66(9): p. 1152-63.

81. Roggen, E.L. and B.J. Blaauboer, Sens-it-iv: A European Union project to develop novel tools for the identification of skin and respiratory sensitizers. *Toxicology in Vitro*, 2013. 27(3): p. 1121.

82. Heberle, H., y col., InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics*, 2015. 16: p. 169.

EJEMPLO 1

Introducción

Basándose en estudios anteriores, los autores de la invención plantearon la hipótesis de que el ensayo GARD es capaz de predecir la potencia de sensibilización de la piel. Se siguieron dos enfoques en paralelo para desarrollar modelos de predicción de potencia. En uno de ellos, se utiliza una máquina de vectores de soporte establecida (SVM) entrenada para proporcionar clasificaciones binarias [1-3]. El otro enfoque se basa en proyecciones de cuadrados mínimos parciales ortogonales para modelo de regresión lineal de estructuras latentes (O-PLS) (Simca, Umetrics, Suecia). O-PLS [5] es un procedimiento de proyección relacionado con análisis de componentes principales y por lo tanto se adapta bien para matrices con más variables que las observaciones (o muestras) como es el caso de los datos de micromatriz de ARN de genoma completo (>29.000 transcritos). Es una modificación del procedimiento PLS (Wold, 1975), diseñado para dividir la variabilidad estructurada en la matriz X en información predictiva (correlacionada con Y) y ortogonal (no correlacionada con Y), además de la variación residual. Esto puede mejorar la capacidad de interpretación de variables latentes (combinaciones lineales de las variables originales).

Resultados

La relación entre los valores de decisión SVM GARD y las clases de potencia humana para un conjunto de 34 productos químicos y los controles queda ilustrada en la **Figura 9**. El modelo SVM se ha desarrollado para clasificaciones binarias (sensibilizador frente a no sensibilizador).

Los autores de la invención investigaron varios modelos en un enfoque multivariado utilizando un software Simca. La potencia humana [6] es una clasificación de productos químicos basada en los datos humanos disponibles combinando sustancias químicas muy diferentes parcialmente en una posible categoría, según lo cual la clase 1 representa la potencia máxima y la clase 6 representa verdaderos no sensibilizadores.

En el presente documento, se presenta un modelo O-PLS, que está diseñado para predecir dos Y, concretamente potencia humana y sensibilizador/no sensibilizador. El diagrama de dispersión de la **Figura 10** ilustra la separación de sensibilizadores y no sensibilizadores y, la agrupación de sensibilizador a lo largo del segundo eje en una agrupación de alta y baja potencia.

La fracción de la variación total que se puede explicar según el modelo tras la validación cruzada (Q2cum) es 0,478, que es un valor considerado como aceptable/bueno. Una comparación de los beneficios de un ajuste y predicción del correspondiente modelo original con el ajuste de varios modelos con observaciones Y permutadas (**Figura 11**) indica sólidamente que el modelo original es válido. En la **Figura 12**, se traza en gráfico $Y_{\text{observado}}$ frente a $Y_{\text{pronosticado}}$.

Análisis

En los análisis, se observa una clara relación entre los datos de micromatriz de la invención y la potencia humana. Está en marcha otro modelo de desarrollo y se espera que mejore su rendimiento con el número y los tipos de productos químicos analizados. Asimismo los autores de la invención investigarán procedimientos de análisis multivalente como herramienta de selección de características, lo cual conllevará finalmente nuevas aclaraciones de los mecanismos asociados con la potencia de sensibilización y proporcionará medios para mejorar la predicción de potencia humana de productos químicos con mayor precisión.

Referencias del ejemplo 1

1. Johansson H y col. A genomic biomarker signature can predict skin sensitizers using a cell-based in vitro alternative to animal assays. BMC Genomics. 2011.
2. Johansson H y col. The GARD assay for assessment of chemical skin sensitizers. Toxicology in vitro 2013.
3. Johansson H y col. GARD in-house validation - A proof of concept. Toxicological Sciences 2014.
4. Albrekt y col. Skin sensitizers differentially regulate signaling pathways in MUTZ-3 cells in relation to their individual potency. BMC Pharmacology and Toxicology 2014.
5. Trygg and Wold. Orthogonal projections to latent structures (O-PLS). Journal of Chemometrics 2002.
6. Basketter y col. Categorization of chemicals according to their relative human skin sensitizing potency. Dermatitis 2014.

EJEMPLO 2

Introducción

Sobre la base de los estudios anteriores, los autores de la invención plantearon la hipótesis de que el ensayo GARD tiene capacidad para predecir la potencia de sensibilización de la piel. Se siguieron dos enfoques en paralelo para desarrollar modelos de predicción de potencia. Uno de ellos utiliza la máquina de vectores de soporte (SVM) establecida de la invención entrenada para proporcionar clasificaciones binarias [1-3] y presenta correlación entre los valores de decisión SVM GARD y las clases de potencia humana [4] para un conjunto de 34 productos químicos y controles, tal como se ilustra en la **Figura 9**. El modelo SVM se ha desarrollado para clasificaciones binarias (sensibilizador frente a no sensibilizador).

El otro enfoque se basa en modelos de bosque aleatorio (RF), un procedimiento basado en árbol de decisión que se adapta perfectamente a los conjuntos de datos con más variables que observaciones. Esto produce un rendimiento predictivo bueno incluso cuando las variables tienen ruido (no se requiere pre-selección) y reviste importancia variable [5].

Resultados

En lo que respecta a la potencia de sensibilización de productos químicos, la Agencia Europea de Productos Químicos (ECHA) propone una categorización de acuerdo con el reglamento (CE) No 1272/2008 sobre clasificación, etiquetado y envasado de sustancias y mezclas químicas (CLP) que consiste en: 1A (sensibilizadores fuertes), 1B (sensibilizadores débiles) y sin categoría (no sensibilizadores) [7]. En el presente documento se presenta un modelo RF, construido sobre la media aritmética de intensidades de transcrito desde replicados en los datos de entrenamiento utilizando el paquete varSelRF de bosque aleatorio [6] en la versión R/Bioconductor 3.1.2 con tasas

de error estimadas según el procedimiento de submuestreo aleatorio .632+ suministrado. Asimismo, se puede construir desde las intensidades de transcrito de replicados directamente. Se entrenó el modelo utilizando el conjunto de entrenamiento descrito en la **Tabla 9** (70 sustancias) y después se desafió con el conjunto de ensayo de la **Tabla 10** (18 sustancias). En la **Tabla 11** se presenta un resumen del rendimiento del modelo. La **Figura 13** comprende los resultados del análisis de componente principal y de mapa de calor del modelo de bosque aleatorio descrito. En la **Tabla 12** se presenta una lista de los biomarcadores de potencia identificados.

Análisis

En los análisis, los autores de la invención observan una clara relación entre los datos de micromatriz y la información de potencia, que comprende tanto potencia humana como CLP. Utilizando los procedimientos de análisis multivariables como herramienta de selección de características, se pueden aclarar mejor los mecanismos asociados con la potencia de sensibilizador. Asimismo, permite la predicción de la potencia de sensibilizador de productos químicos con una alta precisión.

Referencias del Ejemplo 2

7. Johansson H y col. A genomic biomarker signature can predict skin sensitizers using a cell-based in vitro alternative to animal assays. BMC Genomics. 2011.
8. Johansson H y col. The GARD assay for assessment of chemical skin sensitizers. Toxicology in vitro 2013.
9. Johansson H y col. GARD in-house validation - A proof of concept. Toxicological Sciences 2014.
10. Basketter y col. Categorization of chemicals according to their relative human skin sensitizing potency. Dermatitis 2014
11. Diaz-Uriarte R, Alvarez de Andrés S. Gene selection and classification of microarray data using random forest. BMC Bioinformatics. 2006.
12. Diaz-Uriarte R. GeneSrF and varSelRF: a web-based tool and R package for gene selection and classification using random forest. BMC Bioinformatics 2007
13. https://echa.europa.eu/documents/10162/13562/clp_en.pdf.

Tablas

Tabla A

ID agrupación transcrito	VCF (%)	Título de gen	Símbolo de gen	Asignación de gen
Tabla A(i)				
8117594	93	agrupación de histona 1, H2bm	HIST1H2BM	NM_003521
8124385	86	agrupación de histona 1, H4b	HIST1H4B	NM_003544
8124430	81	agrupación de histona 1, H1d	HIST1H1D	NM_005320
8095221	80	fosforribosilaminoimidazol carboxilasa, fosforribosilaminoimidazol succinocarboxamida sintetasa	PAICS	NM_001079524
8124413	69	agrupación de histona 1, H4d	HIST1H4D	NM_003539
8117608	56	agrupación de histona 1, H2al /// agrupación de histona 1, H2bn	HIST1H2AL/// HIST1H2BN	NM_003511
7994109	51	cinasa 1 tipo polo	PLK1	NM_005030
7904433	44	fosfoglicerato dehidrogenasa	PHGDH	ENST00000369407
8082350	44	minicomplejo de mantenimiento de cromosoma componente 2	MCM2	NM_004526
8141395	43	minicomplejo de mantenimiento de cromosoma componente 7	MCM7	NM_001278595
7903893	41	Molécula CD53	CD53	NM_000560
8118669	41	miembro de la familia cinesina C1	KIFC1	NM_002263
7938348	40	cinasa de punto de control WEE1 G2	WEE1	NM_001143976
7957737	34	Timopoyetina	TMPO	NM_001032283
8146357	34	Minicomplejo de mantenimiento de cromosoma componente 4	MCM4	NM_005914
7918300	33	arrollamiento rico en prolina/serina 1	PSRC1	NM_001005290
8054329	31	Proteína de dedo de anillo 149	RNF149	NM_173647
8055426	31	Minicomplejo de mantenimiento de cromosoma componente 6	MCM6	NM_005915
8072687	29	Minicomplejo de mantenimiento de cromosoma componente 5	MCM5	NM_006739

(continuación)

ID agrupación transcrito	VCF (%)	Título de gen	Símbolo de gen	Asignación de gen
8003503	20	Grupo de complementación de anemia Fanconi A	FANCA	NM_000135
Tabla A(ii)				
8040843	44	carbamoil fosfato sintetasa 2, aspartato transcarbamilasa y dihidroorotasa	CAD	NM_004341
7898549	42	Homólogo MRT4, factor de maduración de ribosoma	MRT04	NM_016183
7901091	41	diana de EGR1, miembro 1 (nuclear)	TOE1	NM_025077
7900699	40	Ciclo de división celular 20	CDC20	NM_001255
8121087	36	Dominio peptidasa M20 que contiene 2	PM20D2	NM_00101085
8084630	35	Dominio de la familia de tipo NmrA que contiene 1 pseudogeno	LOC344887	NR_033752
7958455	30	Uracil ADN glucosilasa	UNG	NM_003362
8119088	27	Inhibidor de cinasa dependiente de ciclina 1A (p21, Cip1)	CDKN1A	NM_000389
8117395	26	Agrupación de histona 1, H2bf	HIST1H2BF	NM_003522
8124527	25	Agrupación de histona 1, H1b	HIST1H1B	NM_005322
8097417	20	Dedo PHD de la familia jade 1	JADE1	NM_001287441
7977445	18	KIAA0125	KIAA0125	NR_026800
7985213	17	Receptor colinérgico, alfa nicotínico 5	CHRNA5	NM_000745
Tabla A(ii)				
8068478	17	factor de conjunto de cromatina 1, subunidad B (p60) /// anillo de zinc tipo CW3 familia MORC	CHAF1B /// MORC3	NM_005441
8099721	16	sel-1 supresor de 3 tipo lin-12 (C. elegans)	SEL1L3	NM_015187
7948192	14	Proteína de reconocimiento específico de estructura 1	SSRP1	NM_003146
7960340	14	Cabeza de horquilla box M1	FOXM1	NM_001243088
8107706	14	Lamina B1	LMNB1	NM_001198557
8124524	14	Agrupación de histona 1, H2ak	HIST1H2AK	NM_003510
8040712	11	Proteína centrómero A	CENPA	NM_001042426
8043602	10	Subunidad H de complejo condesina no-SMC I	NCAPH	NM_001281710
8124394	7	Agrupación de histona 1, H2bb	HIST1H2BB	NM_021062
8144931	7	ATPasa, H+ de transporte, lisosómica 56/58kDa, V1 subunidad B2	ATP6V1B2	NM_001693
7999025	5	Proteína 1 asociada a receptor TNF	TRAP1	NM_001272049
Tabla A(iii)				
8004804	83	Fosforribosilformilglicinamidina sintasa	PFAS	NM_012393
8005839	63	Proteína de membrana 97	TMEM97	NM_014573
7916432	61	24-dehidrocolesterol reductasa	DHCR24	NM_014762
7948656	30	Ferritina, polipéptido pesado 1	FTH1	NM_002032
8117408	30	Agrupación de histona 1, H2ae	HIST1H2AE	NM_021052
8002303	17	NAD(P)H dehidrogenasa, quinona 1	NQO1	NM_000903
7939341	8	Molécula CD44 (grupo de sangre india)	CD44	NM_000610

Tabla 1. 37 productos químicos nuevos con anotaciones CLP utilizados para complementar datos GARD existentes. S = sensibilizador, NS = no sensibilizador

Nombre	No. CAS	CLP	citotox	Entrada GARD [M]	Clase binaria HP*)	Predicción
2,4-Dinitrofluorobenceno	70-34-8	1A	sí	0,00001	S	S
3-Metilcatecol	488-17-5	1A	sí	0,00004	S	S
bisfenol A-éter de diglicidilo	1675-54-3	1A	sí	0,00005	S	S
clorpromacina	50-53-3	1A	sí	0,0000125	S	S
cloruro cianúrico	108-77-0	1A	sí	0,00005	S	NS
glutaraldehído	111-30-8	1A	sí	0,00002	S	S
salicilato de hexilo	6259-76-3	1A	sí	0,00007	S	S
butil carbamato de yodopropinilo	55406-53-6	1A	sí	0,00001	S	S
carbonato de metil heptina	111-12-6	1A	sí	0,0001	S	S
p-benzoquinona	106-51-4	1A	sí	0,00005	S	S
galato de propilo	121-79-9	1A	sí	0,000125	S	S
ácido abiético	514-10-3	1B	no	0,000125	S	S
alcohol amilcinámico	101-85-9	1B	sí	0,0003	S	S
anetol	104-46-1	1B	no	0,0005	NS	S
anilina	62-53-3	1B	no	0,0005	S	NS
alcohol anisílico	105-13-5	1B	no	0,0005	NS	NS
benzocaína	94-09-7	1B	no	0,0005	S	NS
benzoato de bencilo	120-51-4	1B	sí	0,0003	NS	S
éter butil glicidílico	2426-08-6	1B	sí	0,0005	S	NS
citral	5392-40-5	1B	sí	0,0000625	S	S
citronelol	106-22-9	1B	no	0,0005	NS	S
dietanolamina	111-42-2	1B	no	0,0005	NS	NS
imidazolidinil urea	39236-46-9	1B	sí	0,00005	S	S
miristato de isopropilo	110-27-0	1B	no	0,0005	NS	NS
lilial	80-54-6	1B	sí	0,0001875	S	S
limoneno	5989-27-5	1B	no	0,0005	S	NS
linalool	78-70-6	1B	no	0,0005	S	NS
lyral	31906-04-4	1B	sí	0,0001	S	S
pentaclorofenol	87-86-5	1B	no	0,0000625	NS	NS
piridina	110-86-1	1B	no	0,0005	NS	NS
1-bromobutano	109-65-9	sin cat.	no	0,0005	NS	NS
ácido benzoico	65-85-0	sin cat.	no	0,0005	NS	NS
alcohol bencílico	100-51-6	sin cat.	no	0,0005	NS	NS
ácido cítrico	77-92-9	sin cat.	no	0,0005	NS	NS
dextrano	9004-54-0	sin cat.	no	0,00003	NS	NS
canamicina A	25389-94-0	sin cat.	no	0,000125	NS	NS
ácido tartárico	87-69-4	sin cat.	no	0,0005	NS	NS
*basado en [41] cuando está disponible; HP 1-4=S; 5-6=NS. De lo contrario de acuerdo con CLP. Véase también Tabla 4.						

Tabla 2. 11 productos químicos de referencia

Producto químico	CAS	CLP	Clase binaria	HP	Entrada GARD [M]
2,4-dinitroclorobenceno	97-00-7	1A	sens	1	0,000004
p-fenilendiamina	106-50-3	1A	sens	1	0,000075
acrilato de 2-hidroxietilo	818-61-1	1A	sens	na	0,0001
2-nitro-1,4-fenilendiamina	5307-14-2	1A	sens	2	0,0003
2-aminofenol	95-55-6	1A	sens	2	0,0001
resorcinol	108-46-3	1B	sens	4	0,0005
geraniol	106-24-1	1B	sens	4	0,0005
aldehído hexilcinámico	101-86-0	1B	sens	5	0,000032
benzaldehído*	100-52-7	sin cat.	no sens	5	0,00025
clorobenceno	108-90-7	sin cat.	no sens	6	0,000098
1-butanol	71-36-3	sin cat.	no sens	6	0,0005
*no sens de acuerdo con [41]					

Tabla 3. Composición de conjunto de entrenamiento y ensayo

	Número total	CLP 1A	CLP 1B	CLP sin cat.
Conjunto de entrenamiento	70	23	25	22

(continuación)

	Número total	CLP 1A	CLP 1B	CLP sin cat.
Conjunto de ensayo	18	6	6	6

Tabla 4. Controles y 86 productos químicos únicos utilizados para entrenar y analizar el modelo de bosque aleatorio para la predicción de categorías CLP

Estimulación	HP	CLP	clase binaria	conjunto	clase de unión de proteína Toxtree
1-brombutano	na	sin cat.	no sens	ensayo	SN2
anetol	5	1B	sens	ensayo	MA
ácido benzoico	na	sin cat.	no sens	ensayo	Sin unión
benzoato de bencilo	5	1B	sens	ensayo	AT
bisfenol A-éter diglicídico	3	1A	sens	ensayo	SN2
éter butil glicídico	3	1B	sens	ensayo	SN2
ácido cítrico	na	sin cat.	no sens	ensayo	Sin unión
cloruro cianúrico	na	1A	sens	ensayo	SNAr
maleato de dietilo	2	1B	sens	ensayo	MA
ftalato de dietilo	6	sin cat.	no sens	ensayo	Sin unión
etil vanilina	nf	sin cat.	no sens	ensayo	SB
glutaraldehído	2	1A	sens	ensayo	SB
yodopropinilo carbamato de butilo	4	1A	sens	ensayo	AT
linalool	4	1B	sens	ensayo	Sin unión
lyral	2	1B	sens	ensayo	SB
p-benzoquinona	na	1A	sens	ensayo	MA
galato de propilo	2	1A	sens	ensayo	MA
xileno ¹	6	sin cat.	no sens	ensayo	Sin unión
1-butanol	6	sin cat.	no sens	entrenamiento	Sin unión
2,4-dinitroclorobenceno	1	1A	sens	entrenamiento	SNAr
2,4-dinitrofluorobenceno	na	1A	sens	entrenamiento	SNAr
2-aminofenol	2	1A	sens	entrenamiento	MA
acrilato de 2-hidroxietilo	3	1A	sens	entrenamiento	MA
2-mercaptobenzotiazol	3	1A	sens	entrenamiento	AT
2-nitro-1,4-fenilendiamina	2	1A	sens	entrenamiento	MA
3-metilcatecol	na	1A	sens	entrenamiento	MA
sulfato de 4-metilaminofenol	3	1A	sens	entrenamiento	MA
bromuro de 4-nitrobencilo	na	1A	sens	entrenamiento	SN2
ácido abiético	3	1B	sens	entrenamiento	Sin unión
alcohol amilcinámico	4	1B	sens	entrenamiento	MA
anilina	4	1B	sens	entrenamiento	Sin unión
alcohol anisílico	5	1B	sens	entrenamiento	MA/SN2
benzaldehído ²	5	sin cat.	no sens	entrenamiento	SB
benzocaina	4	1B	sens	entrenamiento	Sin unión
alcohol bencílico	na	sin cat.	no sens	entrenamiento	Sin unión
cloroanilina	na	1B	sens	entrenamiento	Sin unión
clorobenceno	na	sin cat.	no sens	entrenamiento	Sin unión
clorpromacina	3	1A	sens	entrenamiento	SB
cinamalaldehído	2	1A	sens	entrenamiento	MA
alcohol cinámico	3	1B	sens	entrenamiento	MA
citral	3	1B	sens	entrenamiento	SB
citronelol	5	1B	sens	entrenamiento	Sin unión
dextrano	6	sin cat.	no sens	entrenamiento	SB
dietanolamina	5	1B	sens	entrenamiento	Sin unión
dimetil formamida	nf	sin cat.	no sens	entrenamiento	nf
sulfóxido de dimetilo ³	6	sin cat.	no sens	entrenamiento	Sin unión
difenilciclopropenona	1	1A	sens	entrenamiento	MA
etilendiamina	3	1B	sens	entrenamiento	SB
eugenol	3	1B	sens	entrenamiento	MA
formaldehído	2	1A	sens	entrenamiento	SB
geraniol	4	1B	sens	entrenamiento	SB
glicerol	6	sin cat.	no sens	entrenamiento	Sin unión

(continuación)

Estimulación	HP	CLP	clase binaria	conjunto	clase de unión de proteína Toxtree
glioxal	2	1A	sens	entrenamiento	Sin unión
hexano	6	sin cat.	no sens	entrenamiento	Sin unión
salicilato de hexilo	4	1A	sens	entrenamiento	Sin unión
aldehído hexilcinámico	5	1B	sens	entrenamiento	MA
hidroquinona	3	1A	sens	entrenamiento	MA
hidroxicitronelal	4	1B	sens	entrenamiento	SB
imidazolidinil urea	3	1B	sens	entrenamiento	AT
isoeugenol	2	1A	sens	entrenamiento	MA
isopropanol	5	sin cat.	no sens	entrenamiento	Sin unión
miristato de isopropilo	5	1B	sens	entrenamiento	Sin unión
canamicina A	6	sin cat.	no sens	entrenamiento	Sin unión
Kathon CG	1	1A	sens	entrenamiento	nf
ácido láctico	6	sin cat.	no sens	entrenamiento	Sin unión
galato de laurilo	2	1A	sens	entrenamiento	MA
lilia I	4	1B	sens	entrenamiento	SB
limoneno	4	1B	sens	entrenamiento	Sin unión
carbonato de metil heptina	2	1A	sens	entrenamiento	MA
sicilato de metilo	5	sin cat.	no sens	entrenamiento	Sin unión
metildibromo glutaronitrilo	2	1A	sens	entrenamiento	MA/SN2
ácido octanoico	6	sin cat.	no sens	entrenamiento	Sin unión
pentaclorofenol	5	1B	sens	entrenamiento	SNAr
fenol	6	sin cat.	no sens	entrenamiento	Sin unión
benzoato de fenilo	3	1B	sens	entrenamiento	AT
fenil acetaldehído	na	1B	sens	entrenamiento	SB
ácido p -hidroxibenzoico	nf	sin cat.	no sens	entrenamiento	Sin unión
dicromato dipotásico	1	1A	sens	entrenamiento	Sin unión
Permanganato potásico	nf	sin cat.	no sens	entrenamiento	nf
p-fenilendiamina	1	1A	sens	entrenamiento	MA
piridina	5	1B	sens	entrenamiento	Sin unión
resorcinol	4	1B	sens	entrenamiento	MA
ácido salicílico	6	sin cat.	no sens	entrenamiento	Sin unión
docecil sulfato sódico ⁴	6	sin cat.	no sens	entrenamiento	SN2
ácido tartárico	na	sin cat.	no sens	entrenamiento	Sin unión
disulfuro de tetrametiluram	3	1B	sens	entrenamiento	Sin unión
Tween 80	6	sin cat.	no sens	entrenamiento	na
no estimulado	6	sin cat.	no sens	entrenamiento	nf

Potencia no humana; ^{1,3,4} no sens Basketter y col.; ² no sens de acuerdo con proyecto sens-it-iv [81] MA-aceptor de Michael; SB-formación de base Schiff; AT-agente de transferencia de acilo; sustitución nucleófila bimolecular SN2; sustitución aromático nucleófilo SNAr; na-no disponible; nf-no encontrado.

Tabla 5. Las 52 variables identificadas según el modelo de bosque aleatorio según lo óptimo para clasificación CLP. VCF_ frecuencia de cuestionario de validación

ID agrupación de transcrito	VCF (%)	Título de gen	Símbolo de gen
8117594	93	agrupación de histona 1, H2bm	HIST1H2BM
8124385	86	agrupación de histona 1, H4b	HIST1H4B
8004804	83	fosforribosilformilglicinamidina sintasa	PFAS
8124430	81	agrupación de histona 1, H1d	HIST1H1D
8095221	80	fosforribosilaminoimidazol carboxilasa, fosforribosilaminoimidazol usccinocarboxamida sintetasa	PAICS
8124413	69	agrupación de histona 1, H4d	HIST1H4D
8005839	63	proteína de membrana 97	TMEM97
7916432	61	24-dehidrocolesterol reductasa	DHCR24
8117608	56	agrupación de histona 1, H2al /// agrupación de histona 1, H2bn	HIST1H2AL/// HIST1H2BN
7994109	51	cinasa de tipo polo 1	PLK1
7904433	44	fosfoglicerato dehidrogenasa	PHGDH

(continuación)

ID agrupación de transcrito	VCF (%)	Título de gen	Símbolo de gen
8040843	44	carbamoil-fosfato sintetasa 2, aspartato transcarbamilasa y dihidrooroasa	CAD
8082350	44	minicomplejo de mantenimiento de cromosoma componente 2	MCM2
8141395	43	minicomplejo de mantenimiento de cromosoma componente 7	MCM7
7898549	42	homólogo de MRT4, factor de maduración de ribosoma	MRT04
7901091	41	diana de EGR1, miembro 1 (nuclear)	TOE1
7903893	41	molécula CD53	CD53
8118669	41	miembro de familia cinesina	C1 KIFC1
7900699	40	ciclo de división celular 20	CDC20
7938348	40	cinasa de punto de control WEE1 G2	WEE1
8121087	36	dominio peptidasa M20 que contiene 2	PM20D2
8084630	35	dominio de la familia tipo NmrA que contiene pseudogeno 1	LOC344887
7957737	34	timopoyetina	TMPO
8146357	34	minicomplejo de mantenimiento de cromosoma componente 4	MCM4
7918300	33	arrollamiento rico en prolina/serina 1	PSRC1
8054329	31	proteína de dedo de anillo 149	RNF149
8055426	31	minicomplejo de mantenimiento de cromosoma componente 6	MCM6
7948656	30	ferritina, polipéptido pesado 1	FTH1
7958455	30	uracil ADN glucosilasa	UNG
8117408	30	agrupación de histona 1, H2ae	HIST1H2AE
8072687	29	minicomplejo de mantenimiento de cromosoma componente 5	MCM5
8119088	27	inhibidor cinasa dependiente de ciclina 1A (p21, Cip1)	CDKN1A
8117395	26	agrupación de histona 1, H2bf	HIST1H2BF
8124527	25	agrupación de histona 1, H1b	HIST1H1B
7896697	21	---	---
8003503	20	grupo de complementación de anemia Fanconi A	FANCA
8097417	20	dedo PHD familia jade 1	JADE1
7977445	18	KIAA0125	KIAA0125
7985213	17	receptor colinérgico, nicotínico alfa 5	CHRNA5
8002303	17	NAD(P)H dehidrogenasa, quinona 1	NQO1
8068478	17	factor de conjunto cromatina 1, subunidad B (p60) /// familia MORC tipo CW dedo de zinc 3	CHAF1B /// MORC3
8099721	16	sel-1 supresor de lin-12-tipo 3 (C. elegans)	SEL1L3
7948192	14	proteína de reconocimiento específica de estructura 1	SSRP1
7960340	14	cabeza de horquilla Box M1	FOXM1
8107706	14	lamina B1	LMNB1
8124524	14	agrupación de histona 1, H2ak	HIST1H2AK
8040712	11	proteína centrómero A	CENPA
8043602	10	subunidad H de complejo I condensina no SMC	NCAPH
7939341	8	Molécula CD44 (grupo de sangre india)	CD44
8124394	7	agrupación de histona 1, H2bb	HIST1H2BB
8144931	7	ATPasa, H ⁺ de transporte, lisosómica 56/58kDa, subunidad V1 B2	ATP6V1B2
7999025	5	proteína 1 asociada a receptor TNF 1	TRAP1

Tabla 6. Predicciones de conjunto de ensayo utilizando votación de mayoría

Producto químico	CLP verdadero	CLP pronosticado GARD	Potencia humana [41]	Reactividad de proteína
1-brombutano	sin cat.	sin cat.	na	SN2
ácido benzoico	sin cat.	sin cat.	na	Sin unión
ácido cítrico	sin cat.	sin cat.	na	Sin unión
ftalato de dietilo	sin cat.	sin cat.	6	Sin unión
etil vanilina	sin cat.	sin cat.	nf	formación de base de Schiff
xileno	sin cat.	sin cat.	6	Sin unión
anetol	1B	1B	5	Aceptor de Michael

(continuación)

Producto químico	CLP verdadero	CLP pronosticado GARD	Potencia humana [41]	Reactividad de proteína
benzoato de bencilo	1B	1B	5	Agente de transferencia de acilo
linalool	1B	1B	4	Sin unión
lyral	1B	1A	2	Formación de base de Schiff
éter butil glicidílico	1B	1A	3	SN2
maleato de dietilo	1B	1A	2	Aceptor de Michael
cloruro cianúrico	1A	sin cat.	na	SNAr
galato de propilo	1A	1A	2	Aceptor de Michael
bisfenol A-éter de diglicidilo	1A	1A	3	SN2
glutaraldehído	1A	1A	2	Formación de base Schiff
butil carbamato de yodopropinilo	1A	1A	4	Agente de transferencia de acilo
p-benzoquinona	1A	1A	na	Aceptor de Michael

Tabla 7. Estadística según la clase para replicados independientes en predicciones de conjunto de ensayo

	sin cat.	1A	1B
Sensibilidad	0,889	0,833	0,556
Especificidad	0,917	0,806	0,917
Valor predictivo positivo	0,842	0,682	0,769
Valor predictivo negativo	0,943	0,902	0,805
Prevalencia	0,333	0,333	0,333
Tasa de detección	0,296	0,278	0,185
Prevalencia de detección	0,352	0,407	0,241
Precisión equilibrada	0,903	0,819	0,736

Suppl. Tabla 1. Predicciones de replicados en el conjunto de ensayo

Producto químico	CLP verdadero	CLP pronosticado
bisfenol A-éter de diglicidilo	1A	1A
bisfenol A-éter de diglicidilo	1A	1A
bisfenol A-éter de diglicidilo	1A	1A
cloruro cianúrico	1A	1B
cloruro cianúrico	1A	sin cat.
cloruro cianúrico	1A	sin cat.
glutaraldehído	1A	1A
glutaraldehído	1A	1A
glutaraldehído	1A	1A
butil carbamato de yodopropinilo	1A	1A
butil carbamato de yodopropinilo	1A	1A
butilcarbamato de yodopropinilo	1A	1A
p-benzoquinona	1A	1A
p- benzoquinona	1A	1A
p- benzoquinona	1A	1A
galato de propilo	1A	1A
galato de propilo	1A	1A
galato de propilo	1A	1A
anetol	1B	1B
anetol	1B	1B
anetol	1B	1B
benzoato de bencilo	1B	1B
benzoato de bencilo	1B	1B
benzoato de bencilo	1B	1B
éter butil glicidílico	1B	1A
éter butil glicidílico	1B	1B
éter butil glicidílico	1B	1A
maleato de dietilo	1B	1A
maleato de dietilo	1B	1A
maleato de dietilo	1B	1A
linalool	1B	sin cat.

(continuación)

Producto químico	CLP verdadero	CLP pronosticado
linalool	1B	1B
linalool	1B	1B
lyral	1B	1A
lyral	1B	1B
lyral	1B	1A
1-brombutano	sin cat.	sin cat.
1-brombutano	sin cat.	sin cat.
1-brombutano	sin cat.	sin cat.
ácido benzoico	sin cat.	sin cat.
ácido benzoico	sin cat.	sin cat.
ácido benzoico	sin cat.	sin cat.
ácido cítrico	sin cat.	sin cat.
ácido cítrico	sin cat.	1B
ácido cítrico	sin cat.	sin cat.
ftalato de dietilo	sin cat.	sin cat.
ftalato de dietilo	sin cat.	sin cat.
ftalato de dietilo	sin cat.	sin cat.
etil vanilina	sin cat.	sin cat.
etil vanilina	sin cat.	1B
etil vanilina	sin cat.	sin cat.
xileno	sin cat.	sin cat.
xileno	sin cat.	sin cat.
xileno	sin cat.	sin cat.

Tabla B

Título de gen	Símbolo de gen	ID agrupación de Transcrito	Clasificación
Tabla B(i)			
Ciclina A2	CCNA2	8102643	2
<i>Desconocido</i>	<i>Desconocido</i>	8151252	4
Biosíntesis de ancla de fosfatidilinositol glucano, clase W	PIGW	8006634	5
Polipéptido D1 de ribonucleoproteína nuclear pequeño 16kDa	SNRPD1	8020411	6
Proteína de activación de GTPasa 19 //ARHGAP19-SLIT1 Lectura (NMD candidato)	ARHGAP19 /// ARHGAP19-SLIT1	7935403	7
Agrupación de histona 1, H2ab	HIST1 H2AB	8124391	8
repeticiones ricas en leucina y dominio de homología de calponina (CH) que contiene 2	LRCH2	8174610	9
placofilina 4 /// placofilina 4	PKP4 /// PKP4	8045860	11
ribonucleoproteína, unión a PTB 2 /// ribonucleoproteína, unión PTB 2	RAVER2 /// RAVER2	7902023	12
desoxiuridina trifosfatasa /// desoxiuridina tifosfatasa	DUT /// DUT	7983594	13
aurora cinasa A	AURKA	8067167	14
<i>Desconocido</i>	<i>Desconocido</i>	8055309	15
NDC1 nucleoporina de transmembrana	NDC1	7916316	16
miembro de la familia cinesina 2 /// miembro de la familia cinesina 2	KIF22 /// KIF22	7994620	17
OK/SW-CL.58	OK/SW-CL.58	7970828	18
<i>Desconocido</i>	<i>Desconocido</i>	7994343	19
Tabla B(ii)			
Agrupación de histona 1, H2ab /// agrupación de histona 1, H2ae	HIST1H2AB /// HIST1H2AE	8117408	1
Grupo box 3 de alta movilidad	HMGB3	8170468	10

Tabla 8. Lista de productos químicos

<i>Clase 1 que representa la potencia máxima, 6 que representa no sensibilizadores</i>			
Número	Producto químico	SVM DV	Potencia humana
1	Dicromato potásico	8,45	1
2	Dinitroclorobenceno	6,31	1
3	PPD	5,32	1
4	Katon CG	3,76	1
5	2-Aminofenol	5,98	2
6	Formaldehído	2,32	2
7	Glioxal	1,08	2
8	Iso-eugenol	1,91	2
9	Acrilato de 2-hidroxietilo	8,87	3
10	Alcohol cinámico	2,94	3
11	2-Mercaptobenzotiazol	2,52	3
12	Etilendiamina	1,12	3
13	Penicilina G	1,06	3
14	Eugenol	1,02	3
15	Geraniol	1,76	4
16	Resorcinol	1,03	4
17	Aldehído hexilcinámico	1,24	5
18	Isopropanol	-1,37	5
19	Salicilato de metilo	-1,53	5
20	PABA	-1,04	5
21	Propilen glicol	-1,18	5
22	Benzaldehído	-2,13	5
23	Fenol	-1,08	6
24	Ácido octanoico	-1,19	6
25	Tween 80	-1,39	6
26	Ácido salicílico	-1,41	6
27	Lauril sulfato sódico	-1,52	6
28	Clorobenceno	-1,63	6
29	Glicerol	-2,05	6
30	1-Butanol	-2,07	6
31	Ftalato de dietilo	-2,15	6
32	Sin estimular	-2,20	6
33	DMSO	-2,31	6
34	Acido láctico	-2,38	6

Tabla 9: Productos químicos conjunto de entrenamiento

Producto químico	Cas	CLP	conjunto	vehículo	Entrada GARD [M]
2,4-dinitroclorobenceno	97-00-7	1A	conjunto entrenamiento	DMSO	0,000004
p-fenilendiamina	106-50-3	1A	conjunto entrenamiento	DMSO	0,0003
dodecil sulfato sódico	151-21-3	sin cat.	conjunto entrenamiento	H2O	0,0002
ácido salicílico	69-72-7	sin cat.	conjunto entrenamiento	DMSO	0,0005
fenol	108-95-2	sin cat.	conjunto entrenamiento	H2O	0,0005
glicerol	56-81-5	sin cat.	conjunto entrenamiento	H2O	0,0005
ácido láctico	50-21-5	sin cat.	conjunto entrenamiento	H2O	0,0005
clorobenceno	108-90-7	sin cat.	conjunto entrenamiento	DMSO	0,000098
ácido p-hidroxibenzoico	99-96-7	sin cat.	conjunto entrenamiento	DMSO	0,00025
benzaldehído	100-52-7	sin cat.	conjunto entrenamiento	DMSO	0,00025
ácido octanoico	124-07-2	sin cat.	conjunto entrenamiento	DMSO	0,0005
no estimulado		sin cat.	conjunto entrenamiento	H2O	vehículo
dimetil sulfóxido	67-68-5	sin cat.	conjunto entrenamiento	DMSO	vehículo
glicoxal	107-22-2	1A	conjunto entrenamiento	H2O	0,0003
2-mercaptobenzotiazol	149-30-4	1A	conjunto entrenamiento	DMSO	0,00025
resorcinol	108-46-3	1B	conjunto entrenamiento	H2O	0,0005
isoeugenol	97-54-1	1A	conjunto entrenamiento	DMSO	0,0003
eugenol	97-53-0	1B	conjunto entrenamiento	DMSO	0,0003
alcohol cinámico	104-54-1	1B	conjunto entrenamiento	DMSO	0,0005
geraniol	106-24-1	1B	conjunto entrenamiento	DMSO	0,0005
2-nitro-1,4-fenilendiamina	5307-14-2	1A	conjunto entrenamiento	DMSO	0,0003
isopropanol	67-63-0	sin cat.	conjunto entrenamiento	H2O	0,0005
Tween 80	9005-65-6	sin cat.	conjunto entrenamiento	DMSO	0,0005
acrilato de 2-hidroxietilo	818-61-1	1A	conjunto entrenamiento	H2O	0,0001
formaldehído	50-00-0	1A	conjunto entrenamiento	H2O	0,00008

Kathon CG	96118-96-6	1A	conjunto entrenamiento	H2O	0,0035%
aldehído hexilcinámico	101-86-0	1B	conjunto entrenamiento	DMSO	0,00003224
2-aminofenol	95-55-6	1A	conjunto entrenamiento	DMSO	0,0001
salicilato de metilo.	119-36-8	sin cat	conjunto entrenamiento	DMSO	0,0005
etilen diamina	107-15-3	1B	conjunto entrenamiento	H2O	0,0005
dicromato potásico	7778-50-9	1A	conjunto entrenamiento	H2O	0,0000015
dimetil formamide	68-12-2	sin cat.	conjunto entrenamiento	H2O	0,0005
1-butanol	71-36-3	sin cat.	conjunto entrenamiento	DMSO	0,0005
permanganato potásico	7722-64-7	sin cat.	conjunto entrenamiento	H2O	0,000038
fenilacetaldehído	122-78-1	1B	conjunto entrenamiento	DMSO	0,00005
hidroquinona	123-31-9	1A	conjunto entrenamiento	DMSO	0,00002
bromuro de 4-nitrobencilo	100-11-8	1A	conjunto entrenamiento	DMSO	0,0000015
difenilciclopropenona	886-38-4	1A	conjunto entrenamiento	DMSO	0,000005
hexano	110-54-3	sin cat.	conjunto entrenamiento	DMSO	0,0005
benzoato de fenilo	93-99-2	1B	conjunto entrenamiento	DMSO	0,0002- 0,0005
sulfato de 4-metilaminofenol	55-55-0	1A	conjunto entrenamiento	DMSO	0,000007
hidroxicitronelal	107-75-5	1B	conjunto entrenamiento	DMSO	0,0005
cloroanilina	106-47-8	1B	conjunto entrenamiento	DMSO	0,0005
disulfuro de tetrametiluram	137-26-8	1B	conjunto entrenamiento	DMSO	0,0000001
galato de laurilo	1166-52-5	1A	conjunto entrenamiento	DMSO	0,000003
cinnamaldehído	104-55-2	1A	conjunto entrenamiento	DMSO	0,00005
metildibromo glutaronitrilo	35691-65-7	1A	conjunto entrenamiento	DMSO	0,00002
2,4-dinitrofluorobenceno	70-34-8	1A		DMSO	0,00001
3-metilcatecol	488-17-5	1A	conjunto entrenamiento	DMSO	0,00004
ácido abiético	514-10-3	1B	conjunto entrenamiento	DMSO	0,000125
alcohol amilcinámico	101-85-9	1B	conjunto entrenamiento	DMSO	0,0003
anilina	62-53-3	1B	conjunto entrenamiento	DMSO	0,0005
alcohol anisílico	105-13-5	1B	conjunto entrenamiento	DMSO	0,0005
benzocaína	94-09-7	1B	conjunto entrenamiento	DMSO	0,0005
alcohol bencílico	100-51-6	sin cat.	conjunto entrenamiento	DMSO	0,0005
clorpromacina	50-53-3	1A	conjunto entrenamiento	DMSO	0,0000125
citral	5392-40-5	1B	conjunto entrenamiento	DMSO	0,0000625
citronelol	106-22-9	1B	conjunto entrenamiento	DMSO	0,0005
dextrano	9004-54-0	sin cat.	conjunto entrenamiento	DMSO	0,00003
dietanolamina	111-42-2	1B	conjunto entrenamiento	H2O	0,0005
salicilato de hexilo	6259-76-3	1A	conjunto entrenamiento	DMSO	0,00007
imidazolidinil urea	39236-46-9	1B	conjunto entrenamiento	DMSO	0,00005
miristato de isopropilo	110-27-0	1B	conjunto entrenamiento	DMSO	0,0005
canamicina A	25389-94-0	sin cat.	conjunto entrenamiento	H2O	0,000125
lilial	80-54-6	1B	conjunto entrenamiento	DMSO	0,0001875
limoneno	5989-27-5	1B	conjunto entrenamiento	DMSO	0,0005
carbonato de metil heptina	111-12-6	1A	conjunto entrenamiento	DMSO	0,0001
pentaclorofenol	87-86-5	1B	conjunto entrenamiento	DMSO	0,0000625
piridina	110-86-1	1B	conjunto entrenamiento	H2O	0,0005
ácido tartárico	87-69-4	sin cat	conjunto entrenamiento	DMSO	0,0005

Tabla 10: Productos conjunto de ensayo
Producto químico **Cas**

Producto químico	Cas	CLP	conjunto	vehículo	Entrada GARD [M]
ftalato de dietilo	84-66-2	sin cat.	conjunto de ensayo 1	DMSO	0,00005
etil vanilina	121-32-4	sin cat.	conjunto de ensayo 1	DMSO	0,0005
maleato de dietilo	141-05-9	1B	conjunto de ensayo 1	DMSO	0,00012
xileno	1330-20-7	sin cat.	conjunto de ensayo1	DMSO	0,0005
1-brombutano	109-65-9	sin cat.	conjunto de ensayo 1	DMSO	0,0005
anetol	104-46-1	1B	conjunto de ensayo 1	DMSO	0,0005
ácido benzoico	65-85-0	sin cat.	conjunto de ensayo 1	DMSO	0,0005
benzoato de bencilo	120-51-4	1B	conjunto de ensayo 1	DMSO	0,0003
bisfenol A-éter de diglicidilo	1675-54-3	1A	conjunto de ensayo 1	DMSO	0,00005
éter butil glicídico	2426-08-6	1B	conjunto de ensayo 1	DMSO	0,0005
ácido cítrico	77-92-9	sin cat.	conjunto de ensayo 1	DMSO	0,0005
cloruro cianúrico	108-77-0	1A	conjunto de ensayo 1	DMSO	0,00005
glutaraldehído	111-30-8	1A	conjunto de ensayo 1	H2O	0,00002
yodopropinilo carbamato de butilo	55406-53-6	1A	conjunto de ensayo 1	DMSO	0,00001

linalool	78-70-6	1B	conjunto de ensayo 1	DMSO	0,0005
lyral	31906-04-4	1B	conjunto de ensayo 1	DMSO	0,0001
p-benzoquinona	106-51-4	1A	conjunto de ensayo 1	DMSO	0,00005
galato de propilo	121-79-9	1A	conjunto de ensayo 1	DMSO	0,000125

Tabla 11: Rendimiento de modelo –conjunto de ensayo externo

	CLP 1A	CLP 1B	CLP sin cat,
Precisión	0,93	0,76	0,90
Sensibilidad	0,86	0,63	1,0
Especificidad	1,0	0,90	0,80

Tabla 12

Clasificación	Frecuencias variables	ID agrupación de Transcrito	Título de gen	Símbolo de gen
1	0,335	8117408	agrupación de histona 1, H2ab /// agrupación de histona 1, H2ae	HIST1H2AB /// HIST1H2AI
2	0,235	8102643	ciclina A2	CCNA2
3	0,22	8004804	fosforibosilformilglicinamida sintasa---	PFAS
4	0,21	8151252	---	---
5	0,175	8006634	biosíntesis de ancla de fosfatidilinositol glicano, clase W	PIGW
6	0,165	8020411	polipéptido D1ribonucleoproteína nuclear pequeña 16kDa	SNRPD1
7	0,16	7935403	Proteína 19 activadora de Rho GTPasa 19 ///lectura ARHGAP19-SLIT1 (NM	ARHGAP19 /// ARHGAP1S
8	0,145	8124391	agrupación de histona 1, H2ab	HIST1H2AB
9	0,1	8174610	repeticiones ricas en leucina – dominio 2 que contiene homología de calponina (CH)	LRCH2
10	0,085	8170468	grupo box 3 de alta movilidad	HMGB3
11	0,08	8045860	placofilina 4 /// placofilina 4	PKP4 /// PKP4
12	0,07	7902023	ribonucleo proteína, unión-PTB- 2 /// ribonucleo proteína, unión-PTB- 2	RAVER2 /// RAVER2
13	0,06	7983594	deoxiuridina trifosfatasa /// deoxiuridina trifosfatasa	DUT /// OUT
14	0,06	8067167	aurora cinasa A	AURKA
15	0,045	8055309	---	---
16	0,03	7916316	NDC1 nucleoporina de membrana	NDC1
17	0,03	7994620	miembro de la familia cinesina 22 /// miembro de la familia cinesina 22	KIF22 /// KIF22
18	0,025	7970828	OK/SW-CL.58	OK/SW-CL.58
19	0,195	7994343		
20	0,15	8141395	complejo de mantenimiento cromosoma componente 7	MCM7

Transcrito**ID Agrupación****ID Transcrito**

8117408	ENST00000303910 /// GENSCAN00000029843 /// BC093862 /// NM_021052 /// BC093836
8102643	ENST00000618014 /// ENST00000274026 /// GENSCAN00000033473 /// CR407692 /// BC104787
	/// BC104783 /// AK291931 /// NM_001237
8004804	ENST00000314666 /// ENST00000546020 /// ENST00000590356 /// BC031807 /// GENSCAN00000001671 /// BC167158 /// BC146768 /// AK295895 /// AK292804 /// XM_006721546
	/// AK292402 /// NM_012393
8151252	ENST00000521867 /// GENSCAN00000046362
8006634	ENST00000620233 /// ENST00000614443 /// ENST00000619326 /// ENST00000616581 /// ABO97818 /// NM_178517
8020411	ENST00000582475 /// ENST00000579618 /// ENST00000300413 /// BC001721 /// BC072427
	/// J03798 /// NM_001291916 /// NM_006938
7935403	ENST00000358308 /// ENST00000466484 /// ENST00000371027 /// ENST00000487035 /// ENST00000358531 /// ENST00000492211 /// ENST00000479633 /// DQ338460 /// GENSCAN00000038278 /// BC114490 /// BC113888 /// AY336750 /// AK303055 /// AK093316
	/// NR_

ES 2 745 580 T3

037909 /// AK090447 /// NM_001256423 /// NM_032900 /// NM_001204300
8124391 ENST00000615868 /// GENSCAN00000029813 /// AK311785 /// BC125140 /// NM_003513
8174610 ENST00000538422 /// ENST00000317135 /// GENSCAN00000029788 /// BC125224 ///
NM_020871
/// NM_001243963
8170468 ENST00000325307 /// BX537505 /// NM_005342
8045860 ENST00000480171 /// ENST00000421462 /// ENST00000452162 /// ENST00000389759 ///
ENST00000389757 /// ENST00000426248 /// GENSCAN00000017165 /// BC034473 ///
BC050308
/// AK124237 /// AK055823 /// AK054911 /// NM_001005476 /// NM_003628
7902023 ENST00000294428 /// ENST00000371072 /// GENSCAN0000026769 /// NM_018211 ///
BC065303

(continuación)

ID Agrupación	ID Transcrito
7983594	ENST00000558978 /// ENST00000455976 /// ENST00000558472 /// ENST00000331200 /// ENST00000559416 /// U90223 /// GENSCAN00000024989 /// U62891 /// U31930 /// CR541781 /// M89913 /// BC110377 /// CR541720 /// BC070339 /// BC033645 /// AK298464 /// AB049113 /// AK291515 /// NM_001948 /// NM_001025248
8067167	ENST00000395914 /// ENST00000456249 /// ENST00000422322 /// ENST00000441357 /// ENST00000395907 /// ENST00000395913 /// ENST00000395915 /// ENST00000312783 /// ENST00000395911 /// ENST00000347343 /// ENST00000371356 /// D84212 /// BC027464 /// BC006423 /// BC002499 /// BC001280 /// AK301769 /// AF011468 /// AF008551 /// NM_198437 /// NM_198434 /// NM_198435 /// NM_198433 /// NM_003600 /// NM_198436
8055309	---
7916316	DQ141696 /// ENST00000371429 /// BC003082 /// AK302910 /// AK298909 /// AK295000 /// NR_033142 /// NM_018087 /// XM_006710762 /// NM_001168551
7994620	ENST00000563263 /// ENST00000569382 /// ENST00000400751 /// ENST00000570173 /// ENST00000569636 /// ENST00000160827 /// ENST00000561482 /// L29096 /// GENSCAN00000029536 /// BT007259 /// BC004352 /// BC028155 /// AK316389 /// AK316050 /// NM_007317 /// AB017430 /// AK294380 /// NM_001256270 /// NM_001256269
7970828	AB064667
7994343	NONHSAT141477 /// ENST00000363268
8141395	ENST00000354230 /// ENST00000621318 /// ENST00000425308 /// ENST00000463722 /// ENST00000477372 /// ENST000003D3887 /// ENST00000485286 /// ENST00000343023 /// ENST00000489841 /// GENSCAN00000001629 /// D86748 /// D55716 /// D28480 /// BC013375 ///BC009398 /// NM_182776 /// AK293172 /// NM_005916 /// NM_001278595

REIVINDICACIONES

1. Un procedimiento *in vitro* para determinar la potencia de sensibilización de la piel de un agente que comprende o consiste en las etapas de:

- 5 (a) proporcionar una población de células dendríticas o una población de células de tipo dendrítico;
- (b) exponer las células proporcionadas en la etapa (a) a un agente de ensayo; y
- (c) medir en las células de la etapa (b) la expresión de todos los biomarcadores enumerados en la Tabla A;

en el que la expresión de los biomarcadores medidos en la etapa (c) es indicativa de la potencia de sensibilización de la piel del agente de ensayo de la etapa (b).

- 10 2. El procedimiento de acuerdo con la reivindicación 1 en el que ya se sabe o se sospecha que el agente de ensayo es capaz de inducir sensibilización de la piel.

3. El procedimiento de acuerdo con una cualquiera de las reivindicaciones anteriores en el que la sensibilización de la piel es una respuesta de hipersensibilidad, opcionalmente

- 15 en el que la respuesta de hipersensibilidad es una respuesta de hipersensibilidad mediada por célula, opcionalmente
- en el que la respuesta de hipersensibilidad es una respuesta de hipersensibilidad de tipo IV, opcionalmente
- en el que la respuesta de hipersensibilidad es dermatitis alérgica de contacto (DAC).

4. El procedimiento de acuerdo con una cualquiera de las reivindicaciones anteriores en el que la etapa (c) comprende medir la expresión de una molécula de ácido nucleico de los biomarcadores, opcionalmente

- 20 en el que la molécula de ácido nucleico es una molécula de ADNc o una molécula de ARNm, opcionalmente
- en el que la medición de la expresión de los biomarcadores en la etapa (c) se lleva a cabo aplicando un procedimiento seleccionado del grupo que consiste en hibridación de Southern, hibridación de Northern, reacción en cadena de la polimerasa (PCR), PCR de transcriptasa inversa (RT-PCR), PCR en tiempo real cuantitativa (qRT-PCR), nanomatriz, micromatriz, macromatriz, autorradiografía e hibridación *in situ*, opcionalmente
- 25 en el que la medición de la expresión de los biomarcadores en la etapa (c) se determina utilizando una micromatriz de ADN.

5. El procedimiento de acuerdo con una cualquiera de las reivindicaciones anteriores en el que la medición de la expresión de los biomarcadores en la etapa (c) se realiza utilizando uno o más restos de unión, cada uno de ellos capaz de unirse selectivamente a una molécula de ácido nucleico que codifica uno de los biomarcadores identificados en la Tabla A, opcionalmente,

- 30 en el que cada uno del uno o más restos de unión comprende o consiste en una molécula de ácido nucleico, opcionalmente,
- en el que cada uno del uno o más restos de unión comprende o consiste en ADN, ARN, APN, LNA, GNA, TNA o PMO, opcionalmente
- en el que el resto de unión comprende un resto detectable, opcionalmente
- 35 en el que el resto detectable se selecciona del grupo que consiste en: un resto fluorescente; un resto luminiscente; un resto quimioluminiscente; un resto radioactivo (por ejemplo un átomo radioactivo); o un resto enzimático.

6. El procedimiento de acuerdo con una cualquiera de las reivindicaciones 1 a 3 en el que la etapa (b) comprende o consiste en la medición de la expresión de la proteína de los biomarcadores definidos en la etapa (c), opcionalmente

- 40 en el que la medición de la expresión de los biomarcadores en la etapa (c) se lleva a cabo utilizando uno o más restos de unión, cada uno de ellos capaz de unirse selectivamente a uno de los biomarcadores identificados en la Tabla A, opcionalmente
- en el que el uno o más restos de unión comprende o consiste en un anticuerpo o un fragmento de unión a antígeno del mismo.

7. El procedimiento de acuerdo con una cualquiera de las reivindicaciones anteriores en el que la etapa (c) se realiza utilizando una matriz, opcionalmente,

- 45 en el que la matriz es una matriz basada en perlas, opcionalmente
- en el que la matriz es una matriz de superficie, opcionalmente
- en el que la matriz se selecciona del grupo que consiste en macromatriz; micromatriz; nanomatriz.

8. El procedimiento de acuerdo con una cualquiera de las reivindicaciones anteriores en el que se mide la expresión de los biomarcadores en las células proporcionadas en la etapa (a) antes y después de la exposición al agente de ensayo y en el que la diferencia en la expresión entre los biomarcadores antes y después de la exposición al agente de ensayo es indicativa de la potencia de sensibilización de la piel del agente de ensayo de la etapa (b).

9. El procedimiento de acuerdo con una cualquiera de las reivindicaciones anteriores que comprende además las etapas de:

- 55 (d) proporcionar una población adicional de células dendríticas o una población de células de tipo dendrítico;
- (e) exponer las células proporcionadas en la etapa (d) a:

- i. uno o más agentes de control negativo que no sensibiliza la piel humana; y/o
 - ii) uno o más agentes de control positivo que sensibilizan la piel humana; y/o
 - iii) uno o más vehículos de control;
- 5 (f) incubar durante el mismo período y, aparte de la exposición a un agente al uno o más agentes de control, en las mismas condiciones que las células proporcionadas en la etapa (a);
- (g) medir en las células de la etapa (e) la expresión de los dos o más biomarcadores medidos en la etapa (c);
- en el que la diferencia en la expresión entre los biomarcadores medidos en la etapa (c) y los biomarcadores medidos en la etapa (e) es indicativa de la potencia de sensibilización de la piel del agente de ensayo.
- 10 10. El procedimiento de acuerdo con una cualquiera de las reivindicaciones anteriores que comprende además las etapas de:
 - (h) proporcionar una población adicional de células dendríticas o una población de células de tipo dendrítico;
 - (i) exponer las células proporcionadas en la etapa (h) a un agente sensibilizador de la piel de potencia predeterminada;
 - (j) medir en las células de la etapa (i) la expresión de los biomarcadores medidos en la etapa (c);
- 15 en el que la correspondencia de la expresión entre los biomarcadores medidos en la etapa (c) y los biomarcadores medidos en la etapa (j) es indicativa de la potencia de sensibilización de la piel del agente de ensayo.
- 11. El procedimiento de acuerdo con una cualquiera de las reivindicaciones anteriores que comprende las etapas adicionales de:
 - (k) identificación de la potencia de sensibilización de la piel del agente de ensayo.
- 20 12. El procedimiento de acuerdo con una cualquiera de las reivindicaciones anteriores en el que la población de células dendríticas o la población de células de tipo dendrítico es una población de células de tipo dendrítico, opcionalmente,
- en el que las células de tipo dendrítico son células derivadas de leucemia mieloides como las seleccionadas del grupo que consiste en KG-1, THP-1, U-937, HL-60, Monomac-6, AML-193 y MUTZ-3.
- 25 13. Una matriz para su uso en el procedimiento de acuerdo con una cualquiera de las reivindicaciones 1-12 en el que los restos de unión de la matriz consisten en restos de unión para cada uno de los biomarcadores definidos en la Tabla A y en el que los restos de unión de la matriz comprenden restos de unión tal como se define en una cualquiera de las reivindicaciones 5-7.
- 30 14. Uso de los biomarcadores enumerados en la Tabla A para determinar la potencia de sensibilización de la piel de un agente.
- 15. Uso de restos de unión con especificidad para cada uno de los biomarcadores enumerados en la Tabla A para determinar la potencia de sensibilización de la piel de un agente.
- 16. Un kit de análisis para su uso en un procedimiento de acuerdo con una cualquiera de las reivindicaciones 1-12 que comprende:
 - 35 (a) una matriz de acuerdo con la reivindicación 13; y
 - (b) (opcionalmente) uno o más agentes de control.
 - (c) (opcionalmente) instrucciones para llevar a cabo el procedimiento tal como se define en una cualquiera de las reivindicaciones 1-12.

Figura 1

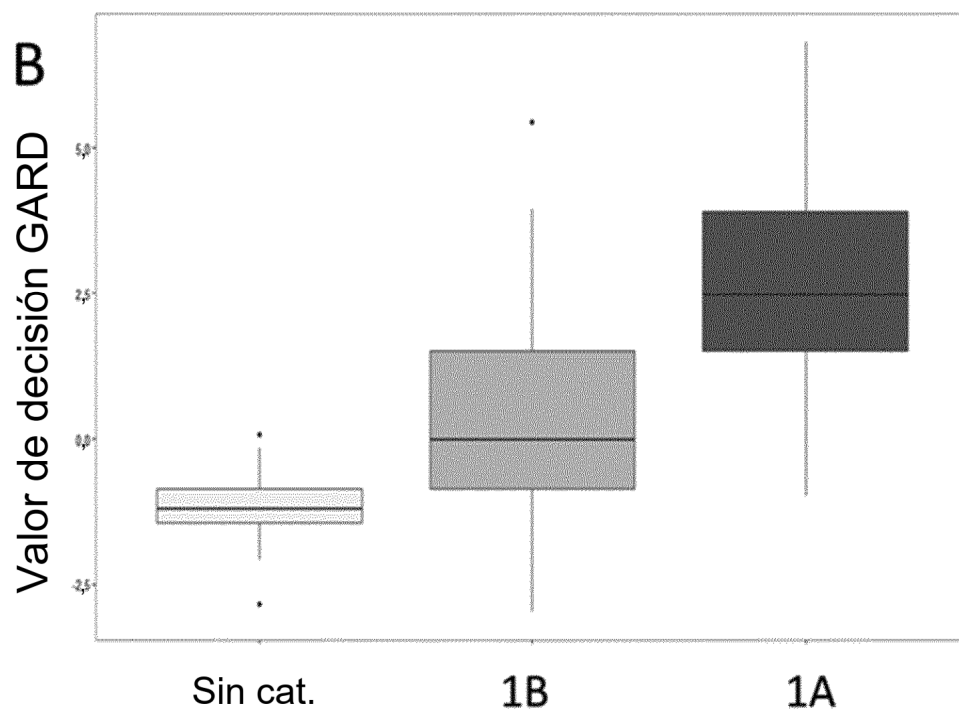
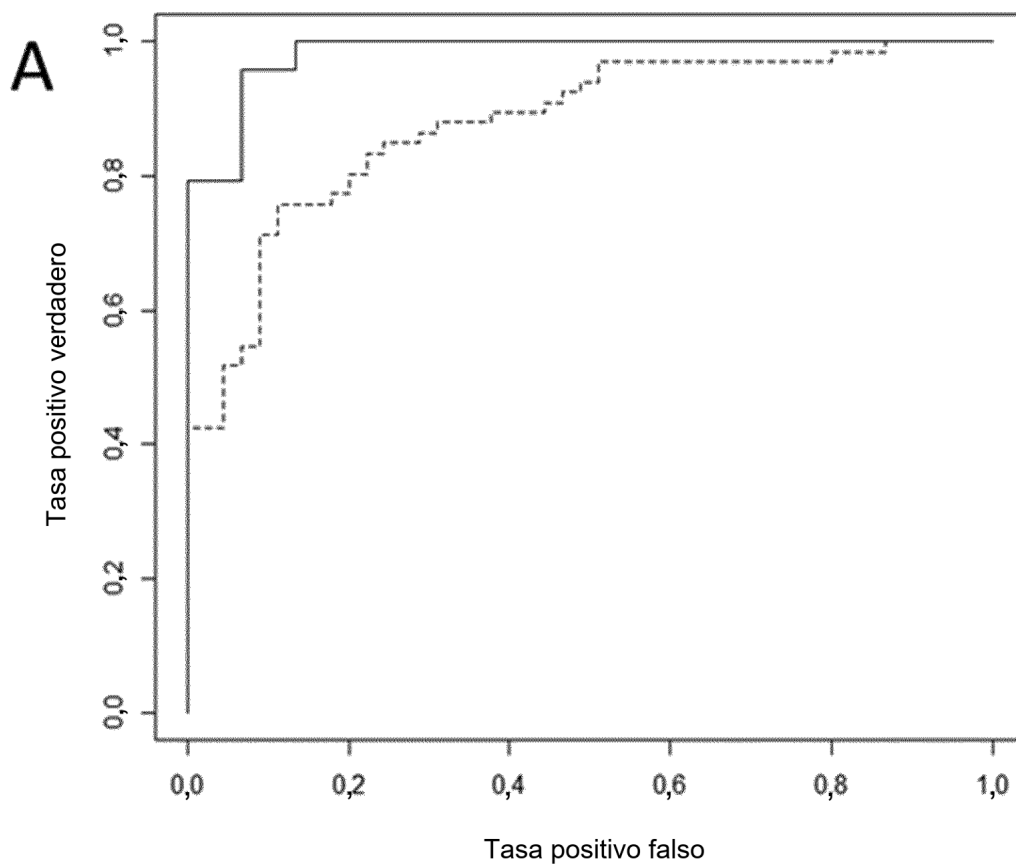


Figura 2

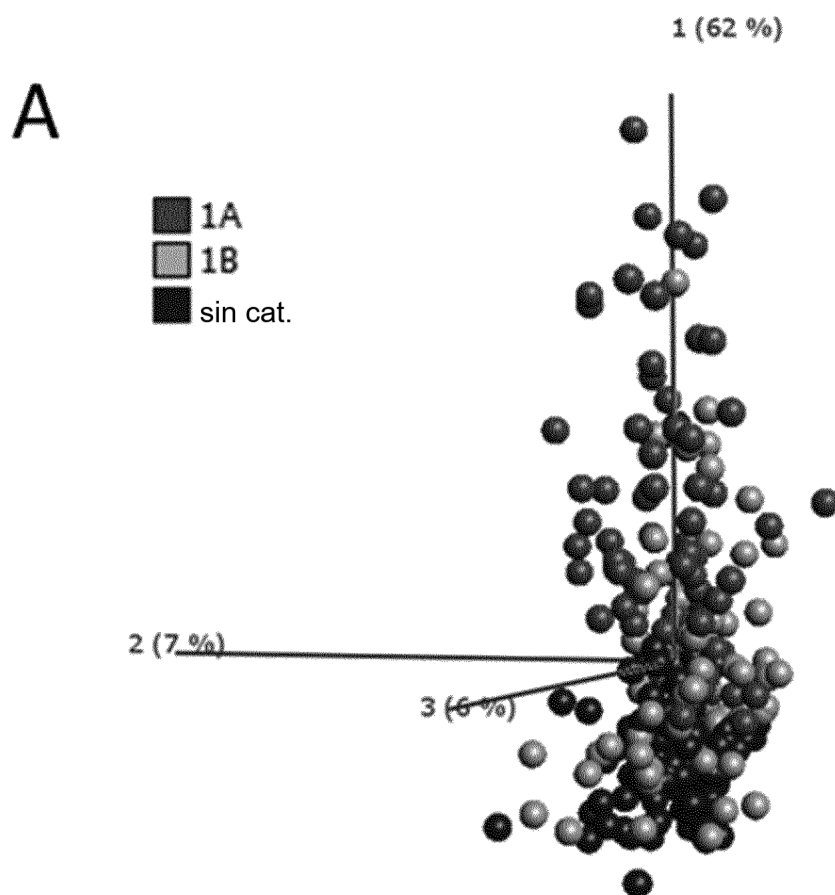


Figura 2 continuación

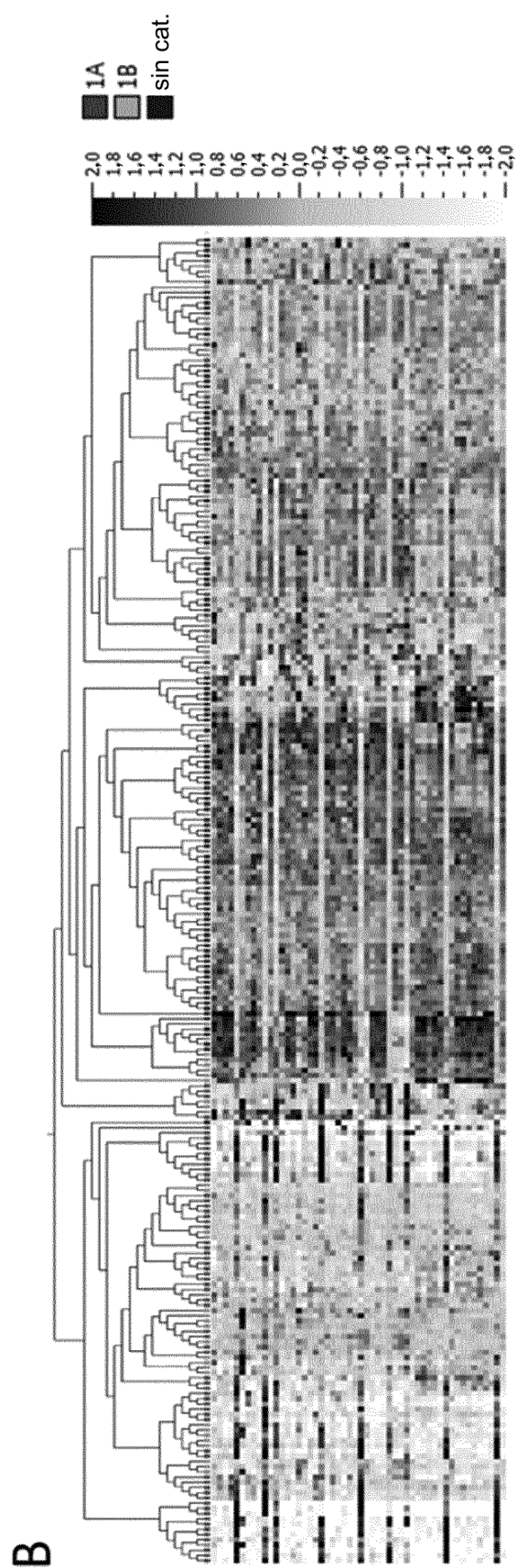


Figura 3

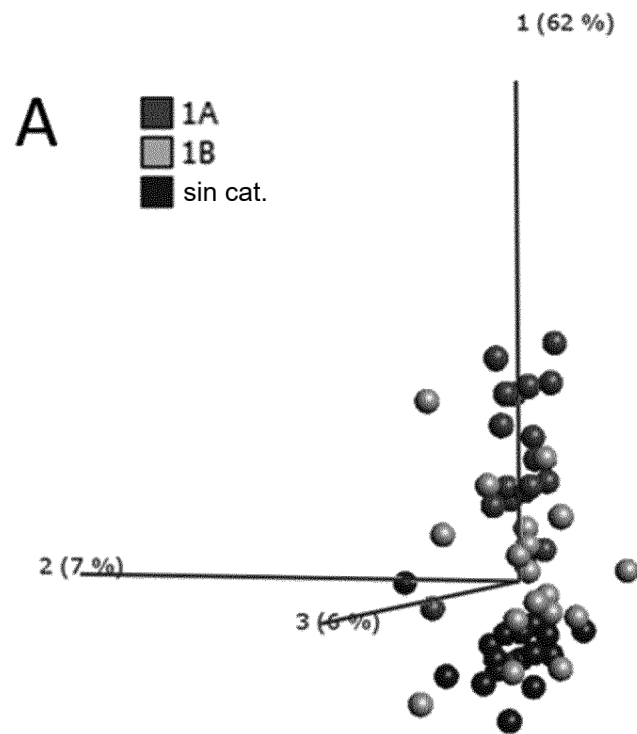


Figura 3 continuación

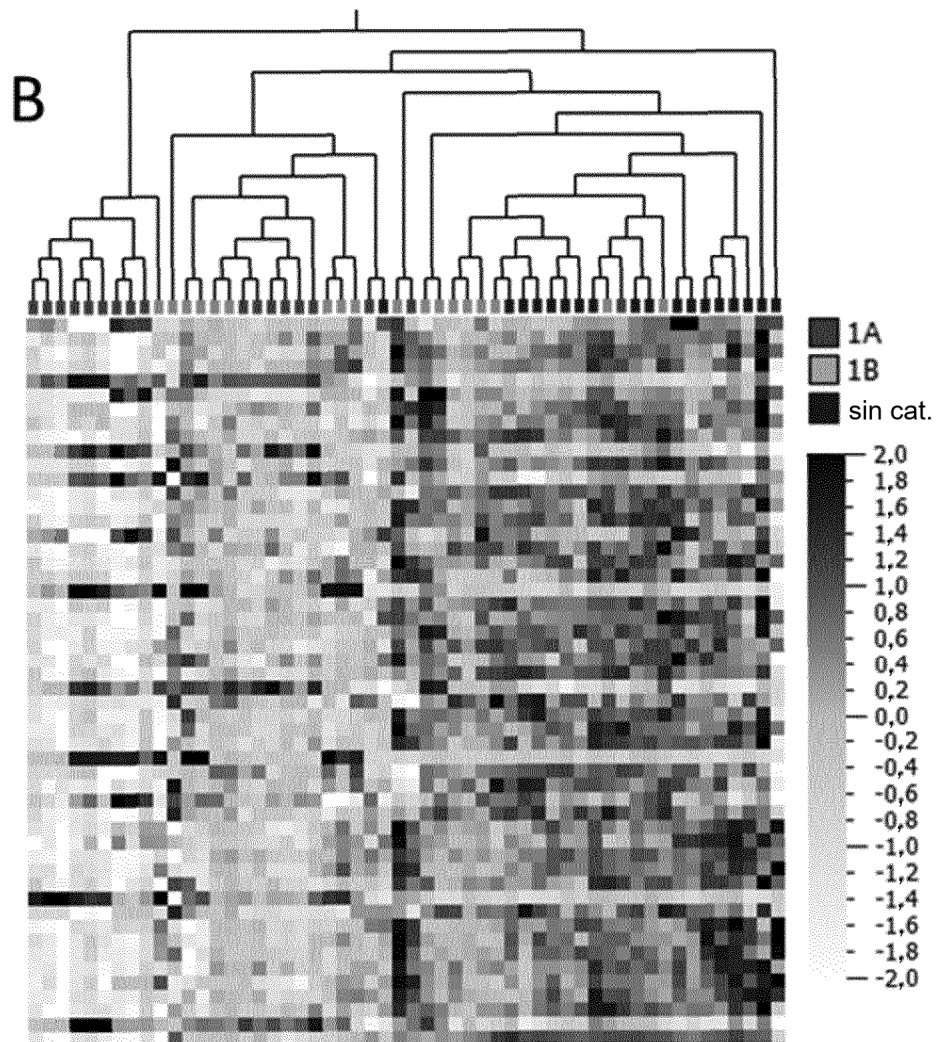


Figura 4

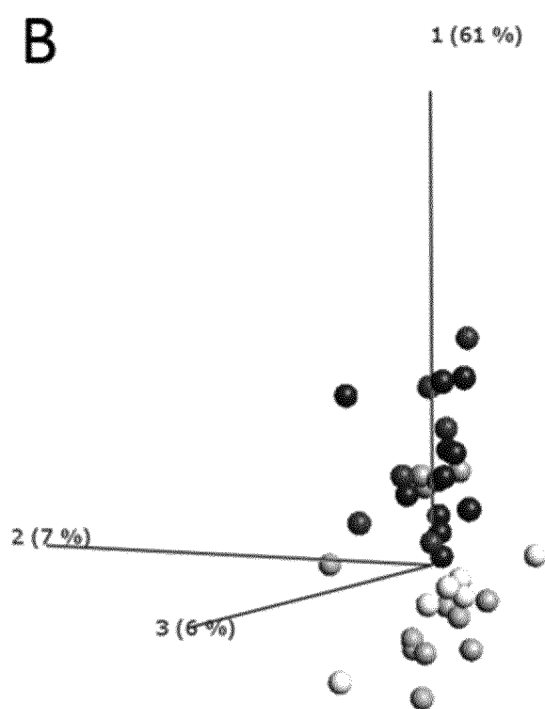
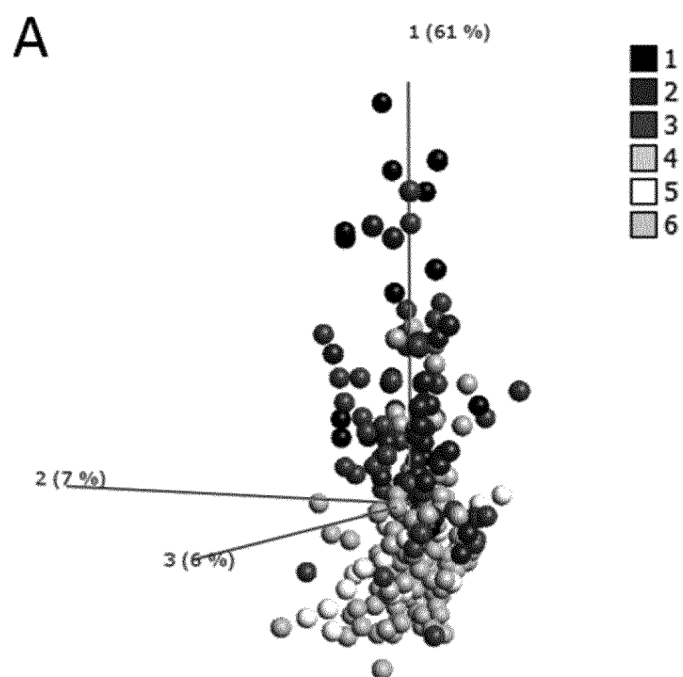


Figura 5



Figura 6

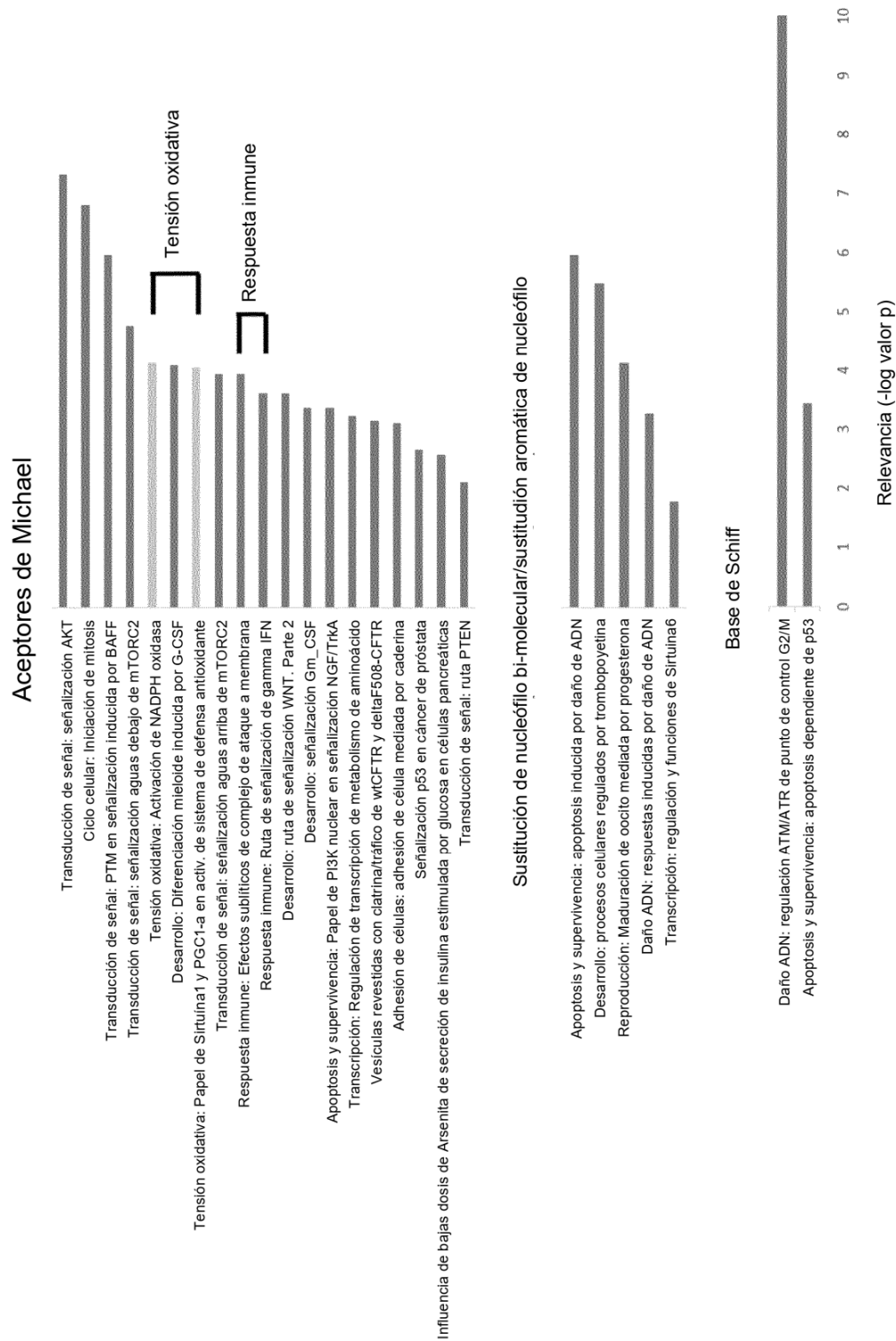
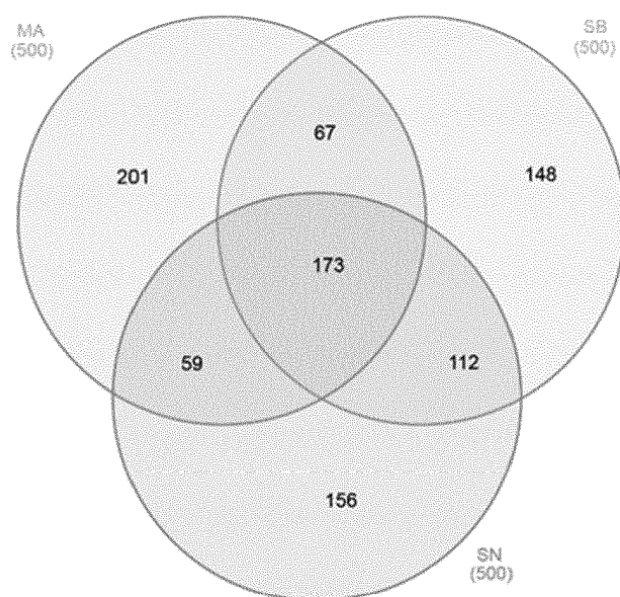


Figura 7

A



B

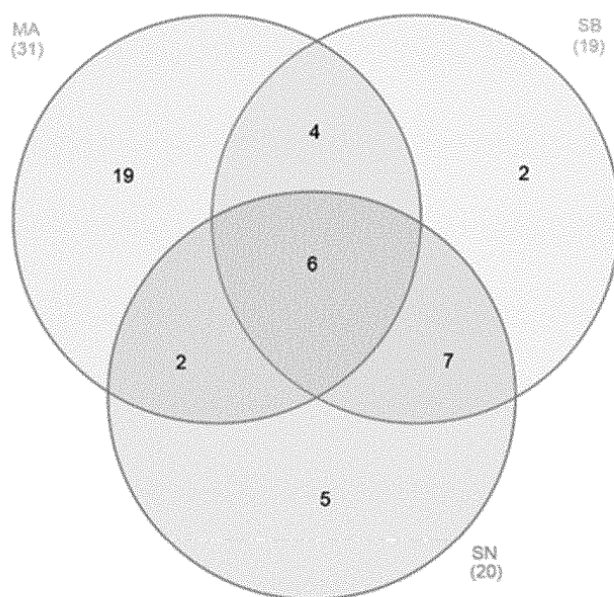


Figura 8

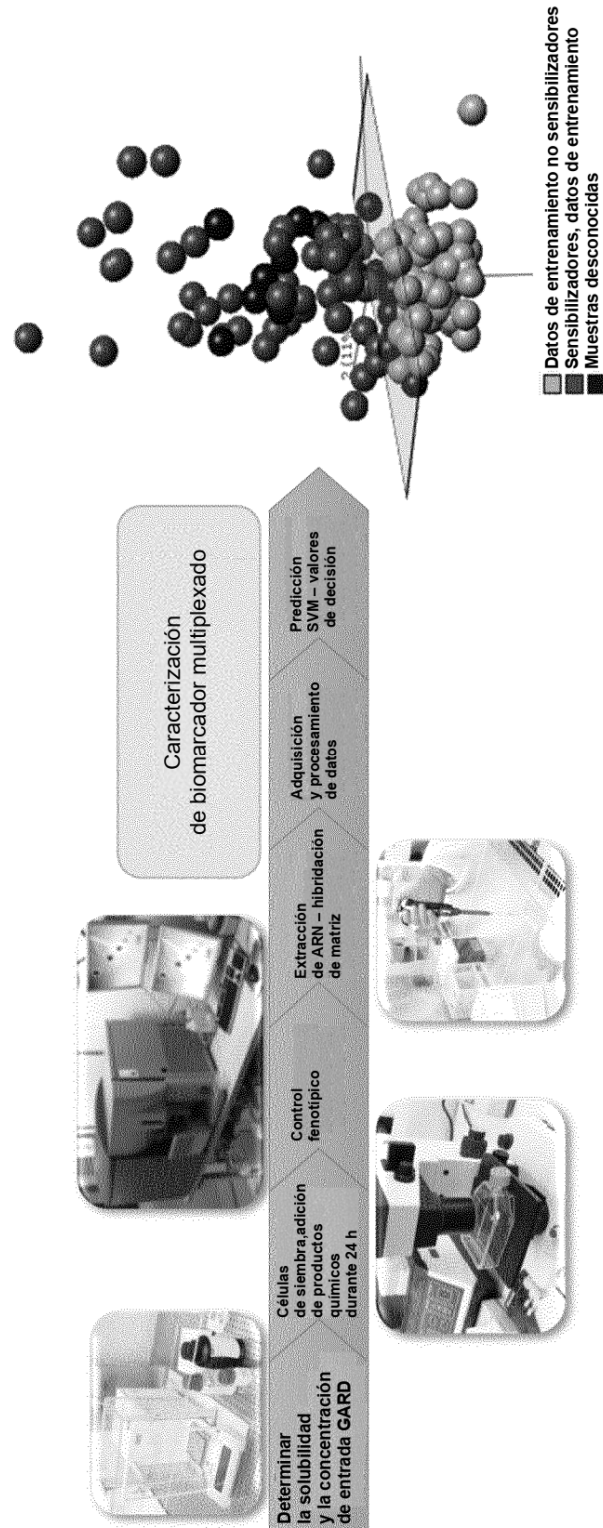


Figura 9

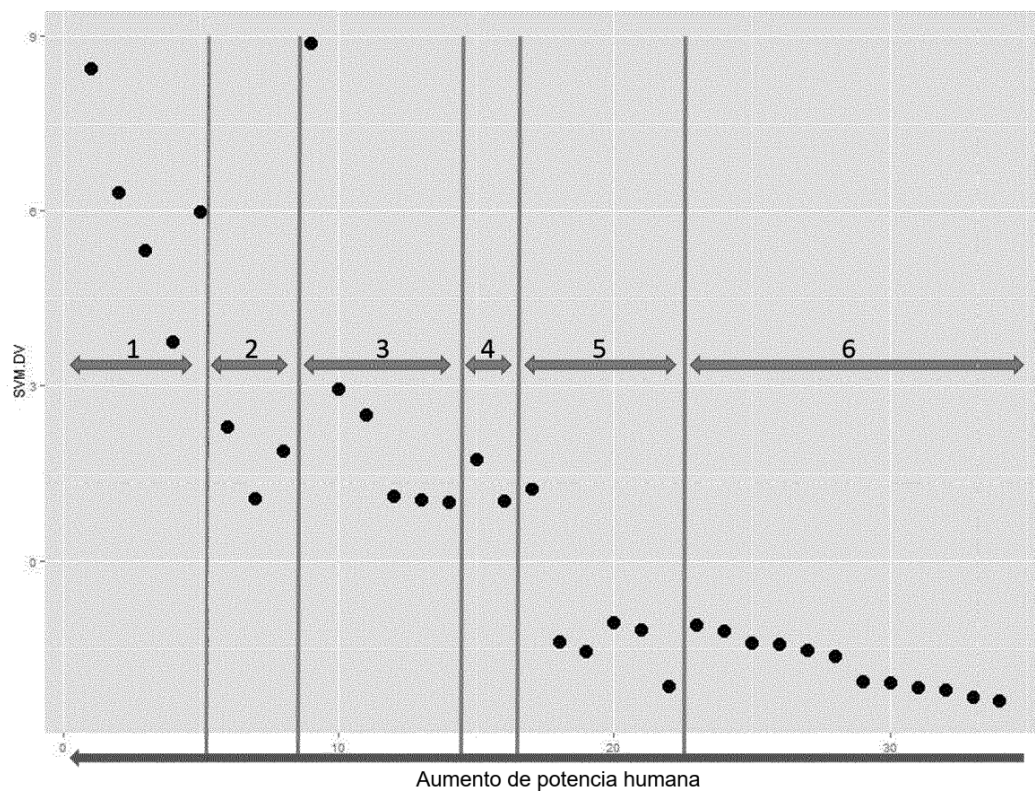


Figura 10

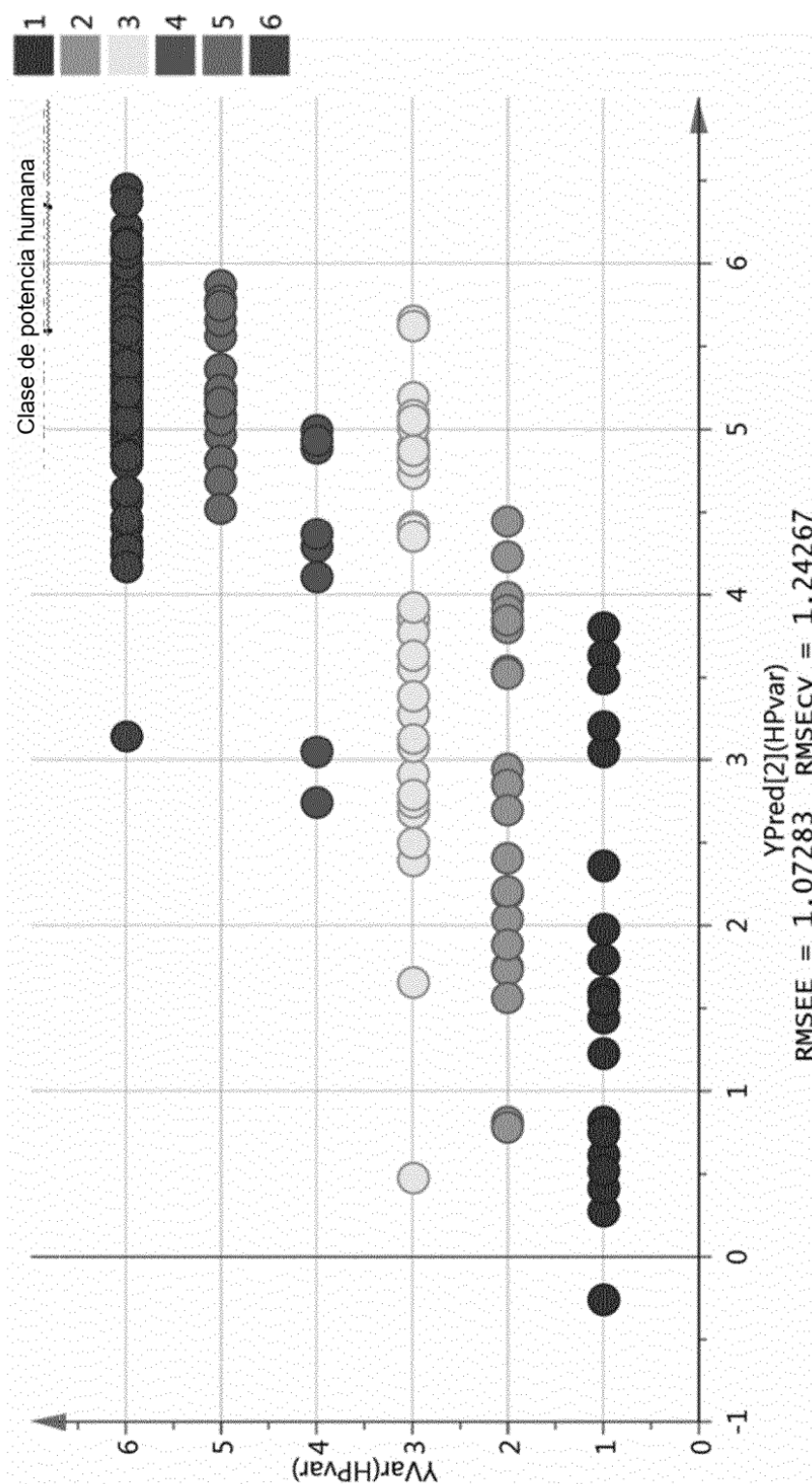


Figura 11

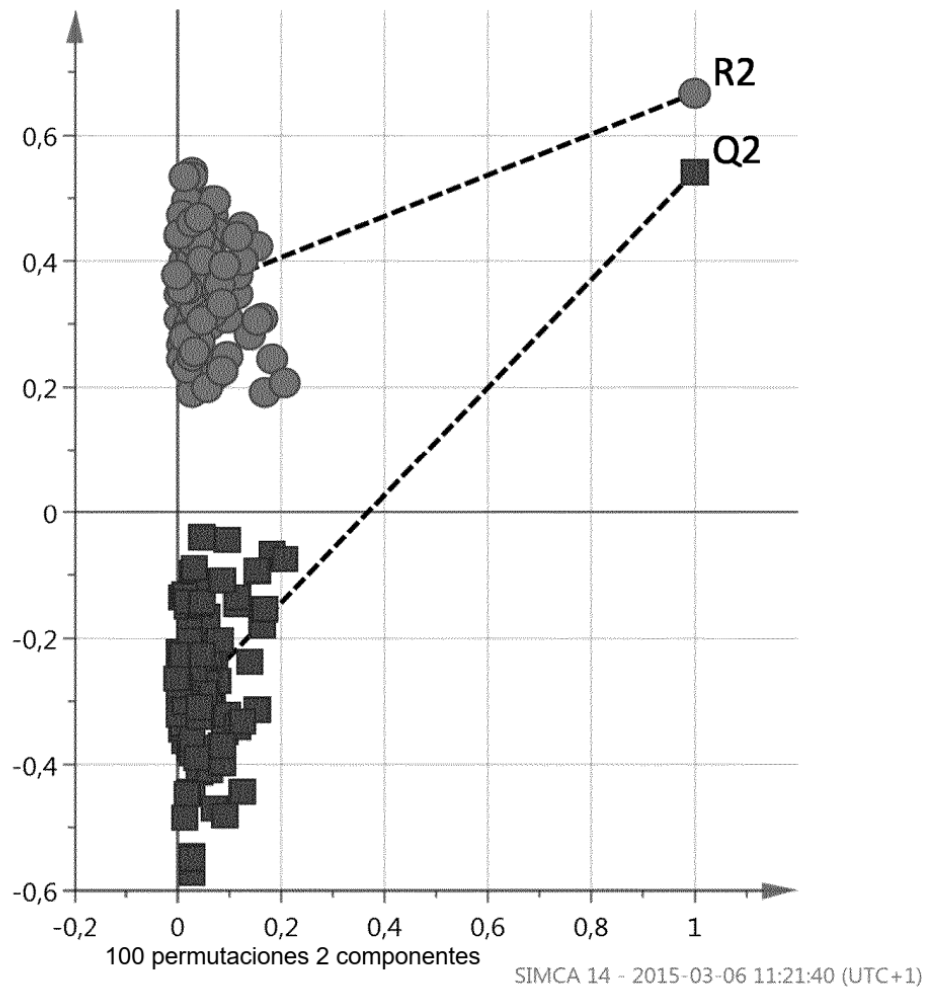


Figura 12

Res_norm_bySampleK.M27 (OPLS), Como M21 HP y NS/S coloreado
de acuerdo con Obs ID (potencia humana 2)

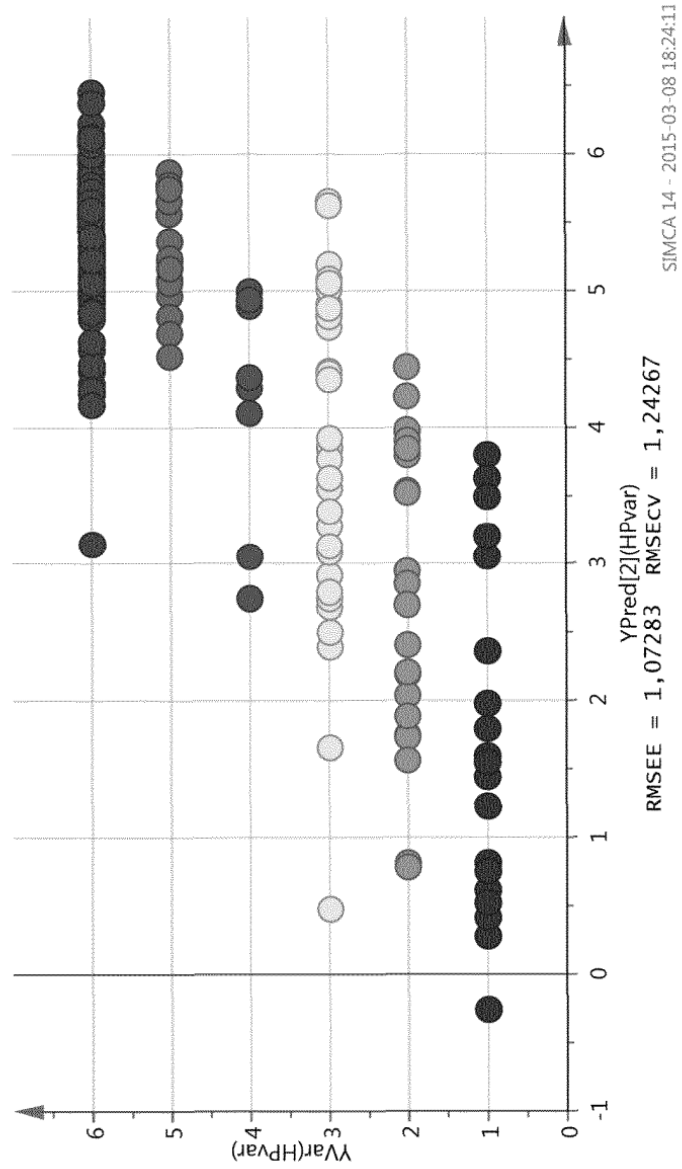
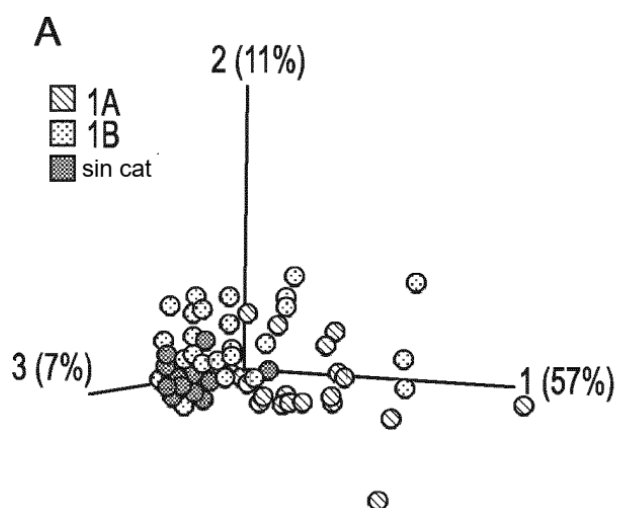


Figura 13



D

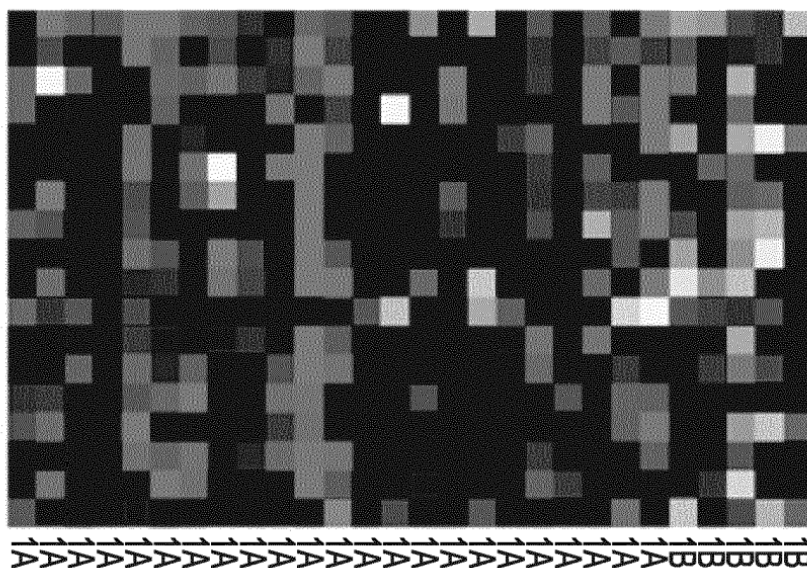
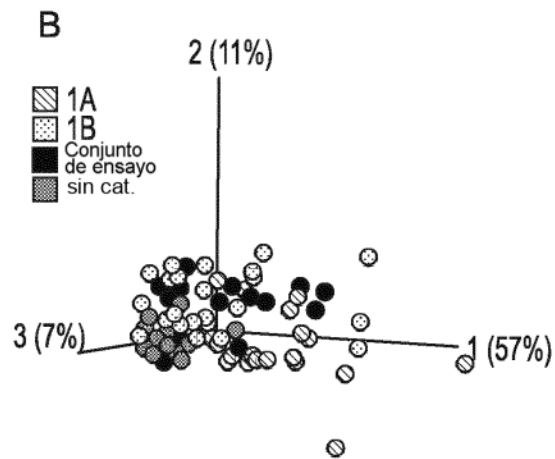


Figura 13 (continuación)



D continuación

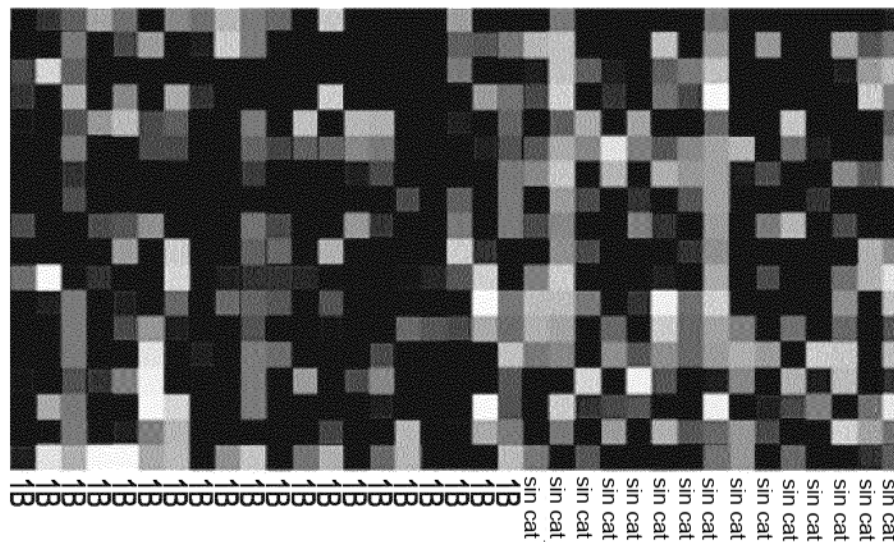
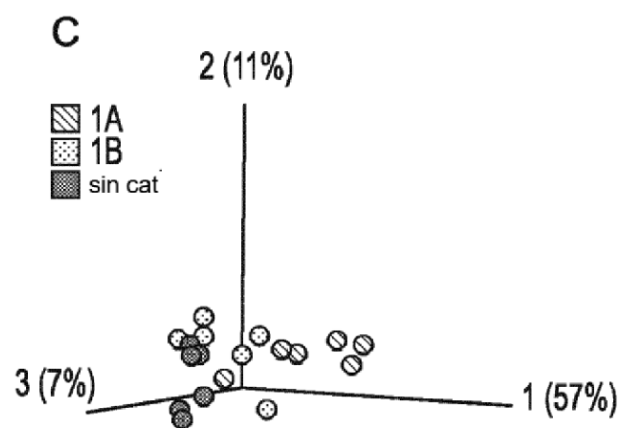
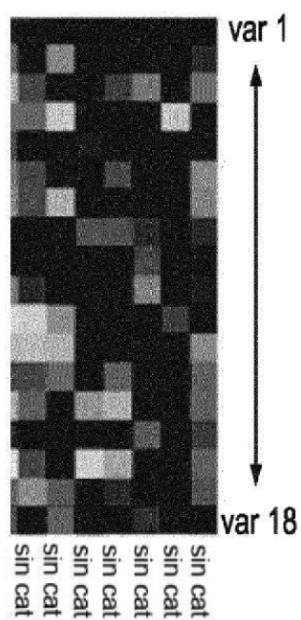


Figura 13 (continuación)



D continuación



E

