

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 530 022**

51 Int. Cl.:

C12Q 1/68

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **13.02.2012 E 12709357 (3)**

97 Fecha y número de publicación de la concesión europea: **10.12.2014 EP 2675915**

54 Título: **Firma genética de células T CD4+ para la artritis reumatoide (AR)**

30 Prioridad:

**14.02.2011 GB 201102563
25.05.2011 GB 201108818**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
25.02.2015

73 Titular/es:

**UNIVERSITY OF NEWCASTLE UPON TYNE
(100.0%)
Kings Gate, Newcastle Upon Tyne
Tyne&Wear NE1 7RU, GB**

72 Inventor/es:

**ISAACS, JOHN D. y
PRATT, ARTHUR G.**

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 530 022 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

Firma genética de células T CD4+ para la artritis reumatoide (AR)

5 La presente invención se refiere a métodos para la identificación y el diagnóstico de la Artritis reumatoide (AR), en particular para el diagnóstico de la AR negativa para anticuerpos anti-péptido citrulinado (AAPC). Muy concretamente la invención se refiere a una firma de expresión genética que comprende 12 biomarcadores para su uso en el pronóstico o el diagnóstico de la AR.

10 La Artritis reumatoide (AR) es una enfermedad autoinmunitaria incapacitante, crónica con predilección por las articulaciones periféricas(1). La importancia de una terapia inmediata modificadora de la enfermedad en la mejora de los resultados clínicos es reforzada por las pautas de manejo internacionales(2). No obstante, aproximadamente
 15 40% de los pacientes con artritis inflamatoria de reciente inicio tienen una enfermedad que no es clasificable al principio, y se dice que tienen una artritis indiferenciada (AI)(3). Recientemente, se ha desarrollado una "regla de predicción" validada para su uso entre los pacientes con AI, por medio de la cual una puntuación compuesta derivada de datos clínicos y serológicos pronostica el riesgo de progreso a AR(4). El sistema de puntuación depende en gran medida de los autoanticuerpos y, en particular, del estado del anticuerpo anti-péptido citrulinado (AAPC), destacando la especificidad del AAPC circulante para la AR(5). No obstante, el diagnóstico de la AR negativa para AAPC sigue siendo un desafío en la clínica de la artritis temprana, retrasándose con frecuencia a pesar de la aplicación de la regla de predicción(6).

20 Los avances tecnológicos y computacionales han permitido rutas "impulsadas por el descubrimiento", de alto rendimiento para la identificación de biomarcadores en entornos clínicos a través del perfilado de la transcripción del genoma completo(7). El análisis del transcriptoma en la AR ha estado limitado normalmente a comparaciones transversales con controles normales(8, 9), con excepciones destinadas a predecir la sensibilidad a agentes biológicos en una enfermedad establecida(10). Los trabajos recientes han demostrado el potencial de que las células mononucleares de sangre periférica (PBMC) proporcionen "firmas genéticas" de pronóstico clínicamente
 25 relevante en las enfermedades autoinmunitarias(11). La aplicación de un futuro enfoque, similar para el descubrimiento de biomarcadores predictivos en la AI debería cumplimentar algoritmos de diagnóstico existentes, a la vez que proporcionaría nuevas perspectivas de la patogénesis de la enfermedad(12). Sin embargo, el uso de PBMC para el análisis transcripcional puede dar como resultado datos que están influidos por la abundancia relativa de subconjuntos(13). Para abordar esto, se han validado protocolos para la selección positiva ex vivo rápida de subconjuntos con el fin de perfilar la transcripción(14), permitiendo el escrutinio de células patofisiológicamente
 30 relevantes en el aislamiento.

Aunque no hay un solo tipo de células implicado exclusivamente en la AR, muchas de las asociaciones genéticas establecidas y emergentes de la afección implican a las células T CD4+ como pieza clave, y las anomalías en el fenotipo de las células T CD4+ de sangre periférica están bien documentadas(15,16,51). Por ejemplo, además de la asociación reconocida hace tiempo de la enfermedad con los alelos concretos del MHC de clase II que codifican una
 35 secuencia conservada en el surco de unión del péptido ("epítipo compartido")(12), análisis de asociación genómica amplia recientes han implicado la proteína tirosina fosfatasa 22 (involucrada en la señalización de receptores de células T), el receptor de IL2, las moléculas co-estimuladoras CD28, CTLA-4 y CD40, y las moléculas transductoras de señales que definen potencialmente el linaje y activadoras de la transcripción 4 (STAT4)(17). Los autores de la presente invención han deducido por lo tanto que el transcriptoma de las células T CD4+ de sangre periférica (PB)
 40 podría representar por lo tanto un sustrato plausible para el descubrimiento de biomarcadores predictivos en la artritis temprana.

A lo largo de todo este documento se utilizan los siguientes términos.

Una muestra es cualquier material biológico obtenido de un individuo.

45 Un polinucleótido es una forma polimérica de nucleótidos de cualquier longitud. Los nucleótidos pueden ser o bien ribonucleótidos o bien desoxirribonucleótidos. El término abarca, pero no se limita a, ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN) de hebra sencilla, doble o múltiple, ARNm, ADN genómico, ADNc, híbridos de ADN-ARN, o un polímero que comprende bases nucleotídicas naturales, modificadas químicamente o bioquímicamente, no naturales. Tales polinucleótidos pueden incluir modificaciones tales como las requeridas para permitir el anclaje a un soporte sólido.

50 Un gen es una secuencia de polinucleótidos que comprende las secuencias que son expresadas en una célula en forma de ARN y las secuencias de control necesarias para la producción de un transcrito o precursor.

Un producto de expresión génica puede ser codificado por una secuencia codificante completa o por cualquier porción de la secuencia codificante.

55 Una sonda puede ser una molécula de ADN tal como ADN genómico o un fragmento del mismo, una molécula de ARN, una molécula de ADNc o un fragmento de la misma, un producto de PCR, un oligonucleótido sintético, o

cualquier combinación de los mismos. Dicha sonda puede ser un derivado o una variante de una molécula de ácido nucleico, tal como, por ejemplo, una molécula de ácido peptidonucleico.

5 Una sonda puede ser específica para una diana cuando comprende un tramo continuo de nucleótidos que son totalmente complementarios a una secuencia de nucleótidos diana (generalmente un producto de ARN de dicho gen, o un producto de ADNc del mismo). No obstante también se puede considerar que una sonda es específica si comprende un tramo continuo de nucleótidos que son parcialmente complementarios a una secuencia de nucleótidos diana. En este caso se puede interpretar que parcialmente significa que un máximo de 10% de los nucleótidos de un tramo continuo de al menos 20 nucleótidos difiere de la correspondiente secuencia de nucleótidos de un producto de ARN de dicho gen. El término complementario es bien conocido y hace referencia a una secuencia que está relacionada por medio de las reglas de emparejamiento de bases con la secuencia diana. Las sondas se diseñarán generalmente para minimizar la hibridación no específica.

10 Cuando se hace referencia a "uno o más" o "12 o más" o "X o más" genes, se puede entender que esto es a efectos de ilustración y no de limitación (aunque puede ilustrar las opciones mejores o preferidas).

15 Las listas de genes 1 - 9 referidas en la siguiente descripción se proporcionan al final de la descripción. En las listas 2-5, la columna de "ID Illumina" contiene el número de dirección de la sonda sobre Illumina WG6 (v3) BeadChip (http://www.illumina.com/support/annotation_files.ilmn). Cuando las sondas Illumina "expresadas diferencialmente" >1 que aparecen en una lista dada correspondían a una única entidad génica, los duplicados fueron eliminados. Los valores de p no corregidos se proporcionan en las listas 2 - 5. Los símbolos de genes oficiales y los números de acceso RefSeq se proporcionan con fines de identificación. Se proporcionan los valores no linealizados de la tasa de cambio, los valores >1 indican genes regulados al alza en grupos con AR con respecto a grupos sin AR en cualquier comparación proporcionada. (Valores <1 = regulados a la baja). Para los valores regulados a la baja, se pueden obtener datos linealizados obteniendo el recíproco negativo del valor no linealizado; esto es FC de 0,75 => |FC| de -1,33. Las listas de genes se clasifican en función de FC. Cualquier información específica de la lista adicional se proporciona en las páginas relevantes.

25 Con el fin de proporcionar una mayor claridad al lector, ciertas secuencias se proporcionan en su totalidad en la presente memoria. En particular, los siguientes datos de secuencias son referidos en la presente memoria;

Gen	Núm. de Acceso	ID Secuencia
BCL3	NM_005178.2	SEQ ID No 1
SOCS3	NM_003955.3	SEQ ID No 2
30 PIM1	NM_002648.2	SEQ ID No 3
SBNO2	NM_014963.2	SEQ ID No 4
LDHA	NM_005566.1	SEQ ID No 5
CMAH	NR_002174.2	SEQ ID No 6
NOG	NM_005450.2	SEQ ID No 7
35 PDCD1	NM_005018.1	SEQ ID No 8
IGFL2	NM_01002915.1	SEQ ID No 9
LOC731186	XM_001128760.1	SEQ ID No 10
MUC1	NM_001044391.1	SEQ ID No 11
GPRIN3	CR743148	SEQ ID No 12
40 CD40LG	NM_000074.2	SEQ ID No 13

De acuerdo con la presente invención se proporciona un método de diagnóstico de Artritis reumatoide en un paciente, comprendiendo el método:

obtener una muestra que comprende células T CD4+ del paciente; y

determinar los niveles de expresión de todos los genes del grupo que consiste en

45 BCL3
SOCS3

PIM1
SBNO2
LDHA
CMAH
5 NOG
PDCD1
IGFL2
LOC731186
MUC1
10 GPRIN3; y

comparar dichos niveles de expresión con niveles de expresión de referencia, en donde una diferencia en la expresión de dichos genes indica una mayor probabilidad de que el paciente tenga Artritis reumatoide.

Opcionalmente el grupo consiste adicionalmente en CD40LG.

15 Generalmente los niveles de expresión de referencia son representativos de los niveles encontrados en muestras que comprenden células de un paciente que no tiene AR.

Se ha descubierto que un incremento en la expresión cuando se compara con los niveles de expresión de referencia indica un aumento de la probabilidad de que el paciente tenga artritis reumatoide.

20 El trabajo de los autores de la presente invención ha confirmado la utilidad de la firma donde se utilizan células T CD4+ de una pureza de >95%, y los datos preliminares sugieren que existe cierto solapamiento cuando se utiliza como sustrato ARN de sangre completa (procedente de células no purificadas).

El grupo puede ser referido como una "firma de 12 genes"

Opcionalmente el grupo comprende adicionalmente el gen CD40LG.

Este grupo puede ser referido como una "firma de 13 genes"

25 Se ha demostrado que la una diferencia en la expresión cuando se compara con los niveles de expresión de referencia de todos los genes mencionados indica un aumento de la probabilidad de que el paciente tenga Artritis reumatoide

De acuerdo con la presente invención, se proporciona un método in vitro para tipificar una muestra de un individuo, comprendiendo el método:

30 obtener una muestra de células T CD4+ del individuo; y
determinar los niveles de expresión de todos los genes del grupo que consiste en

BCL3
SOCS3
PIM1
SBNO2
35 LDHA
CMAH
NOG
PDCD1
IGFL2
40 LOC731186

MUC1

GPRIN3; y

tipificar dicha muestra basándose en los niveles de expresión determinados; en donde dicha tipificación proporciona información pronóstica relacionada con el riesgo de que el individuo tenga Artritis reumatoide (AR).

5 Opcionalmente, el grupo comprende adicionalmente el gen CD40LG.

Muy preferiblemente los niveles de expresión se determinan determinando los niveles de ARN.

Los métodos para determinar los niveles de ARNm están bien establecidos, estando descritos algunos en la presente memoria.

10 La muestra comprende células T CD4+.

Preferiblemente la muestra es sangre completa periférica.

Preferiblemente los métodos incluyen la etapa de separar las células T CD4+ de la sangre completa periférica.

Preferiblemente los métodos incluyen extraer ARN de las células T CD4+.

15 Muy preferiblemente el método es para diagnosticar artritis reumatoide negativa para el anticuerpo anti-péptido citrulinado (AAPC).

Preferiblemente los niveles de expresión de todos los genes del grupo se determinan y se comparan con un conjunto de niveles de expresión de referencia.

20 Opcionalmente el método comprende adicionalmente la etapa de combinar los resultados de la firma de 12 genes con los resultados de análisis de predicción conocidos. Se podría utilizar la firma de 13 en lugar de la firma de 12 genes.

Preferiblemente el análisis de predicción conocido es la regla de predicción de Leiden (Referencia: van der Helm-van Mil 2008 Arthritis and Rheumatism

Utilizando una combinación de la firma de 12 genes (o firma de 13 genes)/ensayo de predicción de Leiden se maximizan la especificidad, la precisión y la sensibilidad del ensayo.

25 De acuerdo con otro aspecto de la presente invención se proporciona un método de diagnóstico de artritis reumatoide en un paciente, comprendiendo el método:

obtener una muestra de sangre del paciente; y

determinar los niveles de expresión/ARNm de los 12 genes del grupo que consiste en BCL3, SOCS3, PIM1, SBNO2, LDHA, CMAH, NOG, PDCD1, IGFL2, LOC 731186, MUC1, GPRIN3,

30 comparar dichos niveles de expresión/ARNm con un conjunto de niveles de expresión/ARNm de referencia, en donde una diferencia en la expresión de dichos genes indica una mayor probabilidad de que el paciente tenga Artritis reumatoide.

De acuerdo con otro aspecto de la presente descripción se proporciona un método de diagnóstico de Artritis reumatoide en un paciente, comprendiendo el método:

35 obtener una muestra de sangre del paciente; y

determinar los niveles de Interleuquina-6 (IL-6); y

comparar dichos niveles con un conjunto de niveles de IL-6 de referencia, en donde una diferencia en la expresión de IL-6 indica una mayor probabilidad de que el paciente tenga Artritis reumatoide.

40 Se ha descubierto que un incremento en la expresión de IL-6 indica una mayor probabilidad de que el paciente tenga Artritis reumatoide.

Notablemente, la IL-6 en suero es notablemente sensible, por ejemplo, a la variación diurna, y los autores de la presente invención identificaron que es útil sistematizar el procedimiento de toma de muestras - todas las muestras se tomaron entre las 13,00 y las 16,30 horas, y se congelaron a -80 en las 4 horas siguientes a la extracción de sangre, experimentando no más de 1 ciclo de congelación-descongelación, por ejemplo.

Muy preferiblemente el método es para diagnosticar la artritis reumatoide negativa para el anticuerpo anti-péptido citrulinado (AAPC).

Preferiblemente los resultados de los análisis de expresión de IL-6 se combinan con los resultados del análisis de predicción conocido.

5 Adicionalmente se describen en la presente memoria:

Una matriz que comprende (a) un sustrato y (b) 12 o más elementos diferentes, comprendiendo cada elemento al menos un polinucleótido que se une a un transcrito de ARNm específico, siendo dicho transcrito de ARNm de un gen seleccionado del grupo definido en la LISTA DE GENES 2.

10 Una matriz que comprende (a) un sustrato y (b) uno o más elementos diferentes, comprendiendo cada elemento al menos un polinucleótido que se une a un transcrito de ARNm específico, siendo dicho transcrito de ARNm de un gen seleccionado del grupo que comprende.

BCL3

SOCS3

PIM1

15 SBNO2

LDHA

CMAH

NOG

PDCD1

20 IGFL2

LOC731186

MUC1

GPRIN3

Opcionalmente, el grupo comprende adicionalmente el gen CD40LG.

25 Una matriz que comprende (a) un sustrato y (b) 12 elementos, comprendiendo cada elemento al menos un polinucleótido que se une a un transcrito de ARNm, comprendiendo dicha matriz un elemento de unión para el ARNm de cada uno de los siguientes grupos de genes

BCL3

SOCS3

30 PIM1

SBNO2

LDHA

CMAH

NOG

35 PDCD1

IGFL2

LOC731186

MUC1

GPRIN3

Opcionalmente la matriz comprende adicionalmente un elemento adicional que comprende al menos un polinucleótido que se une a un transcrito de ARNm para CD40LG.

Preferiblemente el sustrato es un sustrato sólido,

Un kit que comprende una matriz como se ha descrito anteriormente e instrucciones para su uso.

- 5 El uso de un conjunto de sondas que comprende polinucleótidos específicos para 12 o más de los genes enumerados en la LISTA DE GENES 2.

La invención se refiere adicionalmente al uso de un conjunto de sondas que consiste solamente en polinucleótidos específicos para los genes;

BCL3

- 10 SOCS3

PIM1

SBNO2

LDHA

CMAH

- 15 NOG

PDCD1

IGFL2

LOC731186

MUC1

- 20 GPRIN3

para determinar el riesgo de un individuo de padecer artritis reumatoide.

Opcionalmente el conjunto de sondas comprende adicionalmente un polinucleótido específico para CD40LG.

Adicionalmente se describe en la presente memoria:

- 25 El uso de un conjunto de sondas que comprende cebadores específicos para uno o más de los genes seleccionados de la lista;

BCL3

SOCS3

PIM1

SBNO2

- 30 LDHA

CMAH

NOG

PDCD1

IGFL2

- 35 LOC731186

MUC1

GPRIN3

para determinar el riesgo de un individuo de padecer artritis reumatoide.

Opcionalmente el conjunto de sondas comprende adicionalmente un cebador específico para CD40LG.

El uso de un conjunto de sondas que comprende cebadores específicos para los genes seleccionados de la lista;

BCL3

SOCS3

5 PIM1

SBNO2

LDHA

CMAH

NOG

10 PDCD1

IGFL2

LOC731186

MUC1

GPRIN3

15 para determinar el riesgo de un individuo de padecer artritis reumatoide.

Opcionalmente el conjunto de sondas comprende adicionalmente un cebador específico para CD40LG.

Muy preferiblemente el uso de un conjunto de sondas es para determinar el riesgo de un individuo de padecer artritis reumatoide negativa para el anticuerpo anti-péptido citrulinado (AAPC).

20 De acuerdo con un aspecto adicional de la presente descripción, se proporciona un bloqueador del receptor de IL-6 para el tratamiento de la AR.

25 Se esperaría que esta clase de biomarcador tuviera utilidad en la estratificación de pacientes con AR temprana en subgrupos de trascendencia terapéutica. Por ejemplo, los pacientes con una IL-6 de partida elevada (y potencialmente también genes inducibles por STAT3 relativamente muy desregulados en las células T CD4+ circulantes, como consecuencia), podrían ser potencialmente manejados de manera más eficaz utilizando un bloqueador de la señalización de IL-6 (tal como tocilizumab) o un inhibidor de Jak1/3.

Opcionalmente el bloqueador del receptor de IL-6 es tocilizumab.

Opcionalmente el bloqueador del receptor de IL-6 es un inhibidor de Jak1/3.

Con el fin de proporcionar una mejor comprensión de la presente invención, se proporcionarán detalles y ejemplos adicionales más abajo con referencia a las siguientes figuras y tablas;

30 Figura 1. La expresión en células T CD4+ de sangre periférica de la firma de 12 genes es discriminatoria para la AR temprana. A. Clasificación jerárquica de muestras del conjunto de entrenamiento basándose en la similitud en la expresión génica. Se representan 111 muestras por medio de columnas y se indican los genes individuales por medio de filas; el color en cada coordenada indica el cambio de expresión por genes con respecto a la mediana, de acuerdo con la escala de color a la derecha de la Figura. Las muestras de marcas de barras de color subyacentes por diagnóstico inicial, se confirmaron en cada caso con un seguimiento de >1 año. B. Gráfico ROC de un intervalo de cortes para una métrica de riesgo de AR derivada de valores de expresión génica normalizados en la cohorte de entrenamiento (véase el texto). Área bajo la curva = 0,85; Error típico de la media = 0,04; $p < 0,001$. C. Clasificación jerárquica del conjunto de muestras de AI de validación basándose en correlaciones de patrones de expresión de los mismos genes (interpretación como para la Figura 1A). D. Curvas ROC que comparan el valor discriminatorio de la regla de predicción de Leiden original (línea de color gris) con una medida modificada que incorpora una firma de 12 genes (véase el texto). La medida modificada confiere valor añadido a la puntuación de predicción de Leiden original: AU curva ROC (regla de predicción de Leiden original) = 0,74; ETM = 0,08, frente a AU curva ROC (medida modificada que incorpora la firma génica) = 0,84; ETM = 0,06. $p < 0,001$ en ambos casos.

45 Figura 2. Análisis funcional de datos de la matriz. Las listas no redundantes de genes expresados diferencialmente (cambio de expresión >1,2; $p < 0,05$) entre OA y 3 grupos de comparadores inflamatorios separados se solaparon en un diagrama de Venn (véase el texto, y las listas de Genes 2-4 para composiciones detalladas de las listas). Los genes desregulados únicamente en la AR (negativa para AAPC, positiva para AAPC o ambas) pudieron ser

identificados de ese modo y sometidos a análisis de la ruta utilizando el soporte lógico IPA. Se muestran las 2 principales funciones biológicas representadas en exceso identificadas para los 3 conjuntos indicados, junto con la proporción del conjunto asociado con la función en cuestión, y un valor de p relacionado con la probabilidad de que se produzcan proporciones dadas por casualidad (prueba exacta de Fisher). Las listas de genes 5 - 7 resumen los genes relacionados funcionalmente identificados de ese modo. Los 3 conjuntos indicados se combinaron para identificar las rutas canónicas representadas en exceso entre los genes expresados diferencialmente entre la AR y OA en general. Se enumeran las rutas de particular interés en el contexto biológico (los genes en cuestión se enumeran en la Lista de genes 8), *valores de p hipergeométricos (exacto de Fisher) en cada caso <0,01.

Figura 3. A-B. Perfiles de expresión de células T CD4+ de PB de genes regulados por STAT3 indicados en 4 grupos de comparadores; véase la Figura 8 para más ejemplos, y la Tabla 6 para las características de los grupos de comparadores) C. Comparación de medidas de IL-6 en suero, cuando se encuentran disponibles, entre los grupos de comparadores (n=131). Cuando la lectura del ELISA fue <2,6 pg/ml de umbral de detección (línea discontinua), se registró un valor arbitrario de 1,5 pg/ml. D. Comparación de las mediciones de CRP entre los grupos de comparadores (n=173). Cuando la lectura fue <5 se registró un valor arbitrario de 2,5. A-F. Los valores de P mostrados derivan de análisis de varianza no paramétricos (Kruskall-Wallis); para los análisis post-hoc, los asteriscos 1, 2 y 3 indican p< 0,05, 0,01 y 0,001 respectivamente (análisis de comparación múltiple de Dunn).

Figura 4. (Véase la Figura 9 para más ejemplos). A-D. Las concentraciones de IL-6 en suero se corresponden con la expresión del gen inducible por STAT3 en células T CD4+ de PB. Los datos se muestran para 131 individuos en los que se encontraban disponibles muestras contemporáneas, pareadas; se muestran la R de Pearson y los valores de p asociados.

Figura 5. A. Titulación de cóctel patentado de suero no humano (Heteroblock; véase el texto) frente a recuperación de enriquecimiento de IFN- γ en muestra de suero humano RF+ ilustrativa. En ausencia de Heteroblock la diferencia en la lectura entre las muestras enriquecidas y no enriquecidas ("recuperación del enriquecimiento") es significativamente mayor que la cantidad de IFN- γ incorporada (>100%), indicando una lectura del análisis espuria debido a la presencia de RF heterófilo. La adición de ≥ 3 mg/ml de concentración final de Heteroblock neutraliza este efecto heterófilo. B. El gráfico Bland-Altman de las lecturas de IL-6 para las muestras de suero 24 RF+ y 56 RF- obtenidas utilizó la plataforma quimioluminiscente MSD, que comparaba los análisis realizados en presencia/ausencia de 3 μ g/ml finales [Heteroblock]. No se observa una discrepancia significativa entre las muestras RF+ y RF- en lo que se refiere a diferencia de lectura media de los 2 análisis. Esto indica que es poco probable que la presencia de anticuerpos potencialmente heterófilos afecte a la lectura del análisis en este sistema.

Figura 6. Análisis de citometría de flujo del producto aislado de selección positiva CD4+ antes (A) y después (B) de la etapa de agotamiento de monocitos descrita en Métodos. El grado de contaminación de monocitos CD4+ CD14+ varía, pero puede ser tan alto como 15%, como en este ejemplo.

Figura 7. Los rendimientos para los datos de expresión normalizada de 16.205 genes que pasaron el filtro se muestran entre 173 muestras antes y después de la corrección por lotes utilizando el método de Johnston et al. (paneles izquierdo y derecho, respectivamente) (referencia 23, texto principal). A. Clasificación jerárquica no supervisada de muestras basada en las correlaciones en los patrones de expresión génica (correlación convencional, conexión media, representada por medio de dendrograma). Se representan 173 muestras por medio de columnas y los genes individuales por medio de filas; el color en cada coordenada indica el cambio de expresión por genes con respecto a la mediana, de acuerdo con la escala de color a la derecha de la Figura. Muestras con marcas de barras de color azul, rojo y amarillo subyacentes de acuerdo con el número de miembros del lote de la fase (n=2), lote de amplificación de ARN (n=6) y categoría de resultados clínicos de interés (n=4; AR negativa para AAPC, AR positiva para AAPC, controles inflamatorios o no inflamatorios). La agrupación en relación a artefactos de acuerdo con los parámetros técnicos (fase de estudio o en el marco del lote de amplificación de ARN) se elimina por medio de la corrección del lote, que no desenmascara por sí misma la agrupación basada en el resultado clínico de interés. B. Listas de genes que variaban significativamente (ANOVA p<0,05) de acuerdo con el número de miembros de la muestra del lote de la fase (azul), lote de amplificación de ARN (rojo) o resultado clínico de interés (amarillo). Se generaron categorías entre los 16.205 genes pasados, y se solaparon en un diagrama de Venn. Sin la corrección del lote, virtualmente todos los genes parecen asociarse con el resultado clínico y están influenciados simultáneamente por los parámetros técnicos. Esta fuente potencial de sesgo técnico se elimina en 91% de los genes relacionados con el resultado mediante el procedimiento de corrección de lotes. Todos los genes nombrados y comentados en este manuscrito entran dentro de este 91%.

Figura 8. Perfiles de expresión de células T CD4+ de PB de genes indicados en 4 grupos de comparadores, continuación de la Figura 3; Véase la Tabla 6 para las características de los grupos de comparadores. Los valores de P mostrados derivan de análisis de varianza no paramétricos (Kruskall-Wallis); para los análisis post-hoc, los asteriscos 1, 2 y 3 indican p< 0,05, 0,01 y 0,001 respectivamente (análisis de comparación múltiple de Dunn).

Figura 9. Las concentraciones de IL-6 en suero se corresponden con la expresión del gen inducible por STAT3 en células T CD4+ de PB, continuación de la Figura 4. Los datos se muestran para 131 individuos en los que se encontraban disponibles muestras contemporáneas, pareadas; se muestran la R de Pearson y los valores de p asociados.

Figura 10. A-C. Sin relación entre los analitos en suero indicados y el resultado de diagnóstico entre 80 pacientes con artritis temprana. Test de Kruskal Wallis; $p > 0,1$ en todos los casos. D-F. Las concentraciones de analito en suero indicadas no se corresponden con la expresión del gen STAT3 (se muestra SOCS3 ilustrativo). Correlación del rango de Spearman; $p > 0,1$ en todos los casos; y

5 Figura 11. Una curva ROC para toda la cohorte, incluyendo individuos positivos para AAPC, con independencia de si se pudo asignar un diagnóstico al inicio o no. La cohorte incluye 131 pacientes (todos participantes clínicos con AT AAPC-, incluyendo aquellos con resultados definidos al inicio); tanto AAPC+ como AAPC-; y

10 Figura 12 - Curva ROC para individuos negativos para AAPC, pero también con independencia de si se pudo asignar un diagnóstico al inicio o no. 102 pacientes (todos participantes clínicos con AT AAPC-, incluyendo aquellos con resultados definidos al inicio). Entre todos los participantes con artritis temprana negativa para AAPC, una [IL-6] de ≥ 10 pg/ml tiene aproximadamente una especificidad de 0,89 y una sensibilidad de 0,65 para un resultado de AR; y

Figura 13 - Curva ROC para pacientes con AI, ya sea positiva para AAPC o negativa para AAPC. 61 pacientes (pacientes con AI solamente; tanto AAPC+ como AAPC-); y

15 Figura 14 - Curva ROC para pacientes con AI, negativos para AAPC solamente. 48 pacientes (pacientes con AI, solamente AAPC-). Entre todos los pacientes con AI negativa para AAPC, una [IL-6] de ≥ 10 pg/ml tiene aproximadamente una especificidad de 0,92 y una sensibilidad de 0,58 para un resultado de AR.

Estos ejemplos no se deben considerar limitantes.

PACIENTES Y MÉTODOS

20 Pacientes. Los pacientes con síntomas de artritis de comienzo reciente que no habían recibido fármacos anti-reumáticos modificadores de la enfermedad (FAME) ni corticoesteroides, se reclutaron de la clínica de artritis temprana (CAT) del Hospital Freeman, Newcastle upon Tyne, UK, entre Septiembre de 2006 y Diciembre de 2008. Se acometió una evaluación clínica detallada de cada paciente, incluyendo la determinación del estado del AAPC (prueba anti-CCP2, Axis-Shield), junto con la toma de muestras rutinaria de sangre periférica en el punto de partida.

25 Se asignó un diagnóstico de trabajo inicial a cada paciente de acuerdo con un "diagnóstico de trabajo proforma" (Tabla 3). La AR se diagnosticó solamente cuando se cumplieron de manera inequívoca los criterios de clasificación de la ACR de 1987(18), y la AI fue definida como una "sospecha de artritis inflamatoria donde quedaba posibilidad de AR, pero donde seguían sin cumplirse los criterios de clasificación establecidos para cualquier afección reumatológica". Este diagnóstico de trabajo fue actualizado por el especialista en reumatología en cada visita clínica

30 posterior durante la duración del estudio - una media de 28 meses y más de 12 meses en todos los casos. De este modo se averiguó el resultado de diagnóstico de los pacientes con AI al inicio, excluyendo los individuos cuya artritis seguía siendo indiferenciada al final del estudio. Los pacientes se beneficiaron del cuidado clínico rutinario a lo largo de la duración de la investigación, y todos dieron su consentimiento informado por escrito antes de su inclusión en el estudio, que fue aprobado por el Comité Ético Regional Local.

35 Tabla 3. Clasificación de los diagnósticos de trabajo utilizados entre los pacientes con artritis temprana en el inicio y el seguimiento durante el transcurso de este estudio. Se pidió a los especialistas en reumatología que marcaran una casilla en cada visita clínica, indicando la mejor descripción de su opinión experta del diagnóstico en un momento dado. Véase el texto.

• AR			<input type="checkbox"/>
• AI			<input type="checkbox"/>
• Sin AR:	"Inflamatoria"	Artritis psoriásica	<input type="checkbox"/>
		Artritis inflamatoria reactiva/auto-limitante	<input type="checkbox"/>
		Espondilitis anquilosante	<input type="checkbox"/>
		Artritis enteropática	<input type="checkbox"/>
		Espondiloartritis no diferenciada (no AR)	<input type="checkbox"/>
		CTD	<input type="checkbox"/>
		Cristal	<input type="checkbox"/>
		Otra	<input type="checkbox"/>
	"No inflamatoria"	Osteoartritis	<input type="checkbox"/>
		Artralgia no inflamatoria/otra.	<input type="checkbox"/>

Preparación de ARN de células T CD4+. Entre 13,00 hrs y 16,30 hrs durante la cita de CAT del paciente, se recogieron 15 ml de sangre completa periférica en tubos con EDTA (Greiner Bio-One, Austria) y se almacenaron a la temperatura ambiente durante un máximo de 4 horas antes de su procesamiento. En primer lugar se agotaron los monocitos por medio de formación de rosetas inmunorreactivas (cóctel de agotamiento de Monocitos Humanos Rosettesep®, Stemcell Technologies Inc., Vancouver, Canadá), y el resto de las células experimentaron selección positiva utilizando los reactivos del kit de selección positiva para CD4+ de sangre completa Easisep® junto con el separador celular automatizado Robosep® (Stemcell). La pureza de las células T CD4+ se determinó utilizando técnicas de citometría de flujo convencionales; se utilizaron anticuerpos anti-CD4 conjugados con FITC y anti-CD14 conjugados con PE (Beckton Dickinson, New Jersey, USA). El ARN se extrajo inmediatamente de productos aislados de células T CD4+ utilizando kits RNeasy MINI® (Qiagen GmbH, Alemania), que incorporan una etapa de digestión de ADN "en columna".

Micromatrices. Se llevaron a cabo experimentos con micromatrices en 2 fases (fase I, 95 muestras; fase II, 78 muestras). En cada caso, se evaluó la calidad del ARN total utilizando un Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA) de acuerdo con los protocolos convencionales(19). Se sometieron a transcripción inversa 250 ng de ARN a ARNc, y se marcaron con biotina-UTP, utilizando el Kit de Amplificación de ARN TotalPrep de Illumina (Ambion, Texas). El ARNc se hibridó con el BeadChip® Whole Genome 6 (versión 3) de Illumina (Illumina, San Diego, CA), siguiendo el protocolo del fabricante. Cada BeadChip midió la expresión de 48.804 genes (archivo de anotación en http://www.illumina.com/support/annotation_files.ilmn) y se visualizó utilizando un BeadArray Reader (Illumina).

Medición de citoquinas en suero. Durante la evaluación clínica en el momento inicial, se extrajo sangre en tubos de suero/gel (Greiner Bio-One, Austria), y el suero se separó y se congeló a -80°C hasta su uso. La concentraciones de IL-6, sIL6R, TNF-a, leptina y G-CSF en suero se midieron utilizando una plataforma de análisis de inmunoabsorbancia que incorpora un sistema de detección de electro-quimioluminiscencia altamente sensible (Meso Scale Discovery [MSD], Gaithersburg, Maryland) de acuerdo con las instrucciones del fabricante. El potencial de los factores reumatoides heterófilos (FR) en suero para entrecruzar anticuerpos de captura y detección y contribuir a lecturas falsas (20, 21) se excluyó durante el trabajo piloto (Métodos piloto; Figura 5).

qRT-PCR. Las muestras de ARN totales de células T CD4+ se sometieron a transcripción inversa utilizando transcriptasa inversa Superscript II® y hexámeros al azar de acuerdo con las instrucciones del fabricante (Invitrogen, Carlsbad, CA). Para la replicación de los hallazgos con micromatrices se realizaron reacciones de PCR en tiempo real para los transcritos referidos como parte de un TaqMan Low Density Array (7900HT sistema de PCR en tiempo real, Applied Biosystems, Foster City, CA) hecho a medida. Los datos de partida se normalizaron y se expresaron con respecto al gen doméstico de beta-actina (BACT) como valores 2^{-ΔCt} (22). BACT se seleccionó de un panel de 9 genes domésticos potenciales, que habían demostrado una estabilidad óptima para este propósito.

Análisis bioinformáticos y estadísticos generales. Los datos de micromatrices de partida se importaron a un soporte lógico GeneSpring GX 7.3.1 (Agilent Technologies), con el que se realizaron todos los análisis estadísticos excepto cuando se indique. Las fases I y II del estudio se normalizaron independientemente en 2 etapas: cada medición de la sonda se dividió primero por el percentil 50 de todas las mediciones en su matriz, antes de centrarla en su propia expresión de la mediana en todas las muestra en la fase. El efecto esperado de los lotes observado entre las fases durante su combinación, además de los efectos de los lotes entre fases minoritarios referentes a una de las etapas de Illumina TotalPrep RNA Amplification, se corrigió en el entorno informático estadístico de R (<http://www.r-project.org/>) utilizando el método de Bayes empírico de Johnson et al.(23). Los datos de partida y transformados están disponibles para su revisión en la dirección Gene Expression Omnibus (GEO): http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=bvftkociimgsnk&acc=G_SE20098. Los genes expresados detectablemente (valor p de detección <0,01 (24)) en ≥1 muestra de cada fase de estudio pasaron el filtro de los datos normalizados y corregidos por lotes, y se incluyeron en análisis posteriores (16.205 genes). Para definir la expresión diferencial en este estudio se combinó un corte arbitrario en la tasa de cambio de 1,2 entre los grupos de comparadores con un corte en el nivel de significación de p<0,05 (prueba t de Welch), corregida para pruebas múltiples utilizando el método de la tasa de falsos descubrimientos (FDR) de Benjamini et al.(25). Los genes identificados de esta manera se utilizaron para entrenar un modelo de clasificación mediante una máquina de soporte vectorial (SVM) (núcleo Gaussiano) basado en resultados conocidos entre un conjunto de muestras "de entrenamiento" (26). La exactitud, sensibilidad y especificidad del modelo como herramienta de predicción se evaluaron a continuación entre un conjunto de muestras de "validación" independiente. Con el fin de obtener listas más grandes de genes expresados diferencialmente para análisis de rutas biológicas, se relajaron a continuación los umbrales de significación a través de la omisión de la corrección para pruebas múltiples. Se utilizó soporte lógico Ingenuity Pathways Analysis (Ingenuity Systems, Redwood City, CA) para la mayoría de estos análisis. Una lista derivada objetivamente de conjuntos de genes inducibles por STAT3 se creó para pruebas estadísticas hipergeométricas adicionales combinando listas de dos bases de datos de acceso público (lista completa proporcionada en la Lista de Genes 1; fuentes; [http://www.broadinstitute.org/gsea/msigdb/geneset.page.jsp?geneSetName=V\\$STAT3_02&keywords=stat3](http://www.broadinstitute.org/gsea/msigdb/geneset.page.jsp?geneSetName=V$STAT3_02&keywords=stat3) [http://www.broadinstitute.org/gsea/msigdb/geneset.page.jsq?geneSetName=V\\$STAT3_01&keywords=stat3](http://www.broadinstitute.org/gsea/msigdb/geneset.page.jsq?geneSetName=V$STAT3_01&keywords=stat3) http://rulai.cshl.edu/cgi-bin/TRED/tred.cgi?process=searchTFGene&sel_type=factor_name&factor_organism=any&tx_search_terms=STAT3&target_organism=human&prom_quality=1&prom_quality=2&prom_quality=3&prom_qualit=4&prom_quality=5&bind

quality=0&submit=SEARCH). Las pruebas hipergeométricas en este caso se realizaron utilizando el recurso en Internet Stat Trek (<http://stattrek.com>). Los análisis de la varianza paramétricos y no paramétricos (ANOVA), las pruebas U de Mann-Whitney, los coeficientes de correlación de Pearson, las correlaciones intraclase, los análisis multivariados y la construcción de curvas características operativas del receptor (ROC) se realizaron utilizando un SPSS versión 15 (SPSS inc., Chicago IL).

Obtención de métricas de riesgo para AI negativa para AAPC. Las puntuaciones de predicción de Leiden se calcularon para cada miembro de la cohorte de entrenamiento de acuerdo con los datos clínicos y de laboratorio del período inicial como se describe en la referencia (5). Las métricas de riesgo basadas en la "firma" de 12 genes de AR fueron la suma de los valores de expresión normalizados para los genes allí, asignando carga negativa al valor para NOG (que estaba regulado a la baja en la AR). Dentro del conjunto de datos de entrenamiento, ambas puntuaciones se introdujeron como variables continuas independientes en un análisis de regresión logística con resultados de AR frente a no AR como variable dependiente (Tabla 4). En el modelo resultante, la probabilidad de un resultado de AR se relaciona con ambas variables a través de la métrica modificada: $B_1x_1+B_2x_2$, donde B_1 y B_2 son los coeficientes de regresión para la puntuación de predicción de Leiden y la métrica de riesgo de 12 genes respectivamente (valores de B en la Tabla 4), y x_1 y x_2 son los valores para cada uno de los pacientes individuales. Por tanto, para un paciente dado la métrica modificada es igual a: $(0,98x[\text{puntuación de predicción de Leiden}]) + (0,36x[\text{métrica de riesgo/firma de 12 genes}])$.

Tabla 4. Resultados del análisis de regresión logístico para diagnósticos de AR frente a no AR entre 111 pacientes de AT en la cohorte de entrenamiento. B: coeficientes de regresión; ET(B): error típico para B; OR: razón de oportunidades "odds ratio"; IC: intervalo de confianza. La métrica de riesgo/firma de 12 genes para un paciente dado es la suma de los valores de expresión normalizados para 12 genes en la supuesta firma de AR (valor para NOG restado; Véase el texto). La regla de predicción de Leiden se calcula de acuerdo con la referencia (5). Ambas puntuaciones tienen un valor predictivo independiente en la discriminación de resultados clínicos de interés. Se utilizan coeficientes de regresión para cada uno para el cálculo de las métricas de riesgo modificadas entre la cohorte independiente de pacientes de AI negativa para AAPC (véase el texto).

Variable	B	ET (B)	Wald	valor p	OR (IC 95%)
Métrica de riesgo/firma de 12 genes	0,36	0,1	11,0	0,001	1,4 (1,2-1,8);
Regla de predicción de Leiden	0,98	0,2	21,4	<0,001	2,5 (1,7-3,7);
Constante	-10,2	1,8	30,8	<0,001	-

Métodos de Estudio Piloto. El potencial de los factores reumatoides heterófilos (FR) en suero para entrecruzar anticuerpos de captura y detección y contribuir a lecturas falsas, se investigó en un trabajo piloto para este estudio; Los autores de la presente invención confirmaron en primer lugar que una actividad de un cóctel patentado, asequible comercialmente de sueros no humanos (Heteroblock, Omega Biologicals Inc., Boseman, Montanna) podría neutralizar satisfactoriamente la actividad heterófila demostrable de FR nativo en suero humano. La muestra se "enriqueció" con una concentración final conocida de interferón-gamma (IFN- γ) recombinante y, mediante la comparación de la diferencia calculada en una lectura ELISA sándwich convencional (BD Pharmingen, New Jersey, USA) entre muestras enriquecidas y no enriquecidas con la concentración de IFN- γ añadida real, se podría determinar el grado de actividad heterófila, y determinar el efecto neutralizador de concentraciones variables de Heteroblock determinadas (Figura 5A)(referencia 21, texto principal). Los autores de la presente invención midieron a continuación la concentración de IL-6 en 24 muestras de suero FR+ (mediana de RF por nefelometría = 165 UI) y 56 muestras negativas para FR, utilizando la plataforma MSD, en cada caso ejecutando análisis paralelos con y sin concentración final optimizada de Heteroblock. Para las muestras FR+, se observó una correlación excelente entre análisis realizados con y sin Heteroblock (coeficiente de correlación intraclase = 0,98 [IC 95%=0,95-0,99]). Un gráfico de Bland-Altman confirmó que cualquiera de tales discrepancias que existiera no sería menos evidente en las muestras negativas para FR, sugiriendo que la interferencia por los FR heterófilos en los sueros analizados utilizando esta plataforma sería irrelevante (Figura 5B). Todas las mediciones en suero referidas en el estudio actual se llevaron a cabo por lo tanto utilizando la plataforma MSD en ausencia de Heteroblock.

RESULTADOS

Grupos de pacientes. Se seleccionaron retrospectivamente 173 muestras de pacientes para el análisis de micromatrices. 111 de estas procedentes de pacientes que podrían ser diagnósticos definitivos asignados al inicio, que fueron confirmados con una mediana de seguimiento de 28 meses (mínimo 1 año); se obtuvo una "firma" discriminatoria de AR frente a no AR a partir de esta "cohorte de entrenamiento" sola. Las 62 muestras restantes, que representaban todas pacientes de AI, formaron una "cohorte de validación" independiente para someter a ensayo la utilidad de la "firma" de acuerdo con los resultados de diagnóstico a medida que evolucionaban durante el mismo período de seguimiento. Como se esperaba, las características de la cohorte de AI con respecto a la edad, la

respuesta de la fase aguda, los recuentos articulares etc. se encontraron entre las mediciones equivalentes en los conjuntos de muestras de AR y de control dentro de la cohorte de entrenamiento (Tabla 1). Para posteriores análisis de las rutas, las 173 muestras se reunieron antes de dividir las en cuatro categorías basándose en el resultado de diagnóstico al final del estudio (Tabla 5).

- 5 Tabla 1. Las características clínicas de los grupos comparadores de AR y de Control se utilizaron para generar la lista de genes expresados diferencialmente, que constituyen juntos una cohorte de entrenamiento para el aprendizaje de la máquina (total n=111), y la cohorte de validación independiente de AI (n=62). Los valores son la media (1 intervalo DT), la mediana (IQR) o % para datos distribuidos normalmente, sesgados o dicótomos respectivamente. ^APruebas estadísticas para diferencia significativa entre grupos con AR y sin AR; prueba t, U de Mann-Whitney o prueba exacta de Fisher para datos distribuidos normalmente, sesgados o dicótomos respectivamente. Espond seroneg: espondiloartropatía seronegativa; CRP: proteína C reactiva; FR: factor reumatoide; DAS28: puntuación de actividad de enfermedad (que incorpora 28 recuentos de articulaciones hinchadas/sensibles).

	Cohorte de entrenamiento		Cohorte de ensayo	
	AR	Sin AR		AI
	(n=47)	(n=64)	p ^A	(n=62)
Edad (años; media, intervalo de DT)	60 (46-74)	48 (34-62)	<0,01	52 (37-67)
% Mujeres	65	61	NS	77
% Caucasianos blancos	96	92	NS	90
Duración de los síntomas (semanas; mediana, IQR)	12 (8-24)	21 (10,5-52)	0,026	14 (12-26)
Recuento de articulaciones sensibles (mediana, IQR)	10 (4-15)	7 (2-14)	0,246	8 (3-16,5)
Recuento de articulaciones inflamadas (mediana, IQR)	6 (2-10)	0 (0-2)	<0,001	1 (0-3)
Rigidez matutina (horas; mediana, IQR)	1 (0,75-2)	0,75 (0,1-2)	0,007	1 (0,5-2)
ESR (s; mediana, IQR)	56 (30-78)	24 (14-52)	<0,001	30 (18-60)
CRP (g/l; mediana, IQR)	17 (9-62)	5 (2,5-19)	<0,001	8,5 (0-17)
%AAPC+	69	0	-	21
%FR+	77	6	-	32
DAS28 (mediana, IQR)	5,37	n/a	-	
Puntuación de predicción de Leiden (mediana, IQR)	n/a	n/a		6,4 (5-7.6)
Diagnóstico de resultado (%)				
AR	100	0	-	40
Espond seroneg	-	34	-	13
Inflam. auto-limitante	-	19	-	15
Otra inflamación	-	5	-	3
OA/sin inflam.	-	42	-	29

- 15 Tabla 5. Características clínicas de sujetos utilizados en el análisis de rutas de conjuntos de muestras reunidas (n=173), divididas en 4 grupos de comparadores por el resultado en un seguimiento de >1 año: grupos con AR negativa para AAPC, AR positiva para AAPC, de control inflamatorio y no inflamatorio. Los valores son la media (1 intervalo DT), la mediana (IQR) o % para datos distribuidos normalmente, sesgados o dicótomos respectivamente. Pruebas estadísticas para una varianza significativa entre 3 grupos de comparadores inflamatorios (AR negativa para AAPC, AR positiva para AAPC y artritis inflamatoria distinta de AR); ANOVA, Krukskall-Wallis o prueba de Chi cuadrado de distribución normal, para datos sesgados o dicótomos respectivamente. B pruebas estadísticas para una varianza significativa entre los 4 grupos de comparadores inflamatorios; ANOVA, Krukskall-Wallis o prueba de Chi cuadrado de distribución normal, para datos sesgados o dicótomos respectivamente.

20

	AAPC- neg AR	AAPC- pos AR	Sin AR Inflam^y.	Sin AR (OA / sin inflam^y.)	p^A (3xInflam ^y .)	p^B (4xgrupos)
	(n=31)	(n=41)	(n=56)	(n=45)		
Edad	61	56	44	52		
(años; media, DT)	(46-77)	(44-70)	(30-60)	(40-64)	<0001	<0,0001
% Mujeres	66	61	62	80	NS	NS
Durac. síntomas	12	12	12	32		
(semana; mediana, IQR)	(10-20)	(9-22)	(8-25)	(20-89)	NS	<0,0001
Recuento articulaciones sensibles	10,5	10	5	9		
(mediana, IQR)	(5-15,5)	(3,5-16,5)	(2-13)	(2,5-19)	NS	NS
Recuento articulaciones inflamadas	4	3	1	0		
(mediana, IQR)	(1-4)	(0,5-7,5)	(0-4)	(0-0,5)	<0,001	<0,0001
Rigidez matutina	1	1	1	0.5		
(hrs; mediana, IQR)	(1-3,6)	(0,6-2,5)	(0,25-2)	(0,2-1,6)	NS	0.005
ESR	48	54	34	20		
(s; mediana, IQR)	(27-68)	(27-73)	(20-72)	(9,5-30)	NS	<0,0001
CRP	18	10	13	2.5		
(g/l; mediana, IQR)	(10-57)	(5-35)	(5-23)	(2,5-6)	NS	<0,0001
%AAPC+	0	100	0	0	-	-
%FR+	25	93	11	12	-	-
DAS28	4,9	5,2				
(mediana, IQR)	(4,5-5,9)	(4,1-6,0)	-	-	-	-

5 Pureza de células T CD4+ y control de calidad. El análisis citométrico de flujo se completó para 148/173 (86%) de las muestras, y se logró una mediana de pureza de CD4+ CD14- de 98,9% (intervalo 95 - 99,7%), con una contaminación por monocitos CD4+ CD14+ mínima (mediana 0,32%; intervalo 0,01 - 2,98%). El trabajo piloto había demostrado que la incorporación de la etapa de agotamiento de monocitos descrita era necesaria para lograr esto (Figura 6A y B). Se calcularon los números de integridad del ARN (RIN) para las 173 muestras basándose en el Agilent 2100 Bioanalyser(19), y todos tuvieron la calidad adecuada para su inclusión en los experimentos de micromatrices (mediana de número RIN 9,5). Tras la normalización de los datos brutos y la filtración de los genes expresados, se eliminó con éxito el sesgo técnico relacionado con los lotes de procesamiento utilizando el método de Johnson et al. (Figura 7)(23).

15 "Firma" de transcripción de AR muy exacta en AI negativa para AAPC. Utilizando un umbral de significación robusto para la corrección del ensayo múltiple (tasa de descubrimiento falsa p<0,05)(25), se demostró que 12 genes no redundantes eran expresados de manera diferencial (>1,2 veces) en células T CD4+ de PB entre 47 pacientes de una CAT en una "cohorte de entrenamiento" con un diagnóstico confirmado de AR, y 64 que pudieron ser asignados a diagnósticos distintos de AR (Tabla 2). Se proporciona una lista ampliada, obtenible omitiendo la corrección de ensayo múltiple en la Lista de Genes 2. El análisis de agrupación jerárquica supervisado del conjunto de datos multidimensional resultante (111 muestras, 12 genes), demostró una clara tendencia para los pacientes de la CAT con diagnóstico de AR a agruparse juntos basándose en este perfil de transcripción (Figura 1A). Se utilizó la PCR en tiempo real cuantitativa (qRT-PCR) para analizar la expresión de siete de los genes expresados de manera diferencial en un subconjunto de 73 muestras. A pesar de la reducida potencia para detectar cambio en este conjunto de datos más pequeño, se confirmó la expresión diferencial robusta para seis de los siete genes (Tabla 2).

25 Tabla 2. Tasa de cambio y el nivel de significación para los genes expresados diferencialmente al inicio entre células T CD4+ de PB entre pacientes con CAT con diagnóstico de inicio de AR y sin AR (confirmada en >1 año; mediana de seguimiento de 28 meses). El símbolo del gen oficial y el número de acceso RefSeq se proporcionan como identificadores, en negrita para los 12 genes incluidos en la "firma de AR" más robusta estadísticamente, y en texto normal para los genes regulados por STAT3 adicionales referidos en el texto. NR: no realizado. [TC]: expresión de la

- 5 tasa de cambio linealizada en la AR con respecto a No AR (es decir los valores negativos representan los genes regulados a la baja en la AR con respecto a no AR en n veces). ^ACálculos basados en valores de expresión normalizados de los datos de matrices; prueba t de Welch, se proporcionan valores p de partida y corregidos para pruebas múltiples (véanse los métodos). ^BCálculos basados en datos de expresión normalizados al gen doméstico de la beta-actina (2- $\Delta\Delta$ Ct); Prueba U de Mann-Whitney (véanse los métodos). Consérvese que el transcrito CR743148 (Illumina Probe ID 6370082) ha sido retirado de NCBI, pero la etiqueta de secuencia expresada corresponde a una o varias variantes de corte y empalme dentro del gen GPRIN3 (cromosoma 4.90).

Gen (Núm. Acc.)	Datos de micromatrices (47 AR vs. 64 sin AR)			Datos de qRT-PCR (32 AR vs. 41 sin AR)	
	TC	p ^A no corr.	p ^A corr.	TC	p ^B
Firma AR de 12 Genes:					
BCL3 (NM_005178)	1,59	2,6x10 ⁻⁵	0,03	2,15	0,005
SOCS (NM_003955)	1,55	3,4x10 ⁻⁶	0,03	1,83	0,002
P1M1 (NM_002648)	1,52	6,8x10 ⁻⁶	0,03	1,67	0,001
SBNO2 (NM_014963)	1,47	1,2x10 ⁻⁵	0,03	1,13	0,158
LDHA (NM_005566)	1,23	3,8x10 ⁻⁵	0,04	1,25	0,003
CMAH (NR_002174)	1,2	1,7x10 ⁻⁵	0,03	1,40	0,003
NOG (NM_005450)	-1,32	3,1x10 ⁻⁵	0,03	-1,59	0,004
PDCD1 (NM_005018)	1,42	1,0x10 ⁻⁵	0,03	NR	NR
IGFL2 (NM-001002915)	1,31	1,1x10 ⁻⁷	0,002	NR	NR
LOC731186 (XM_001128760)	1,28	2,3x10 ⁻⁵	0,03	NR	NR
MUC1 (NM_001044391)	1,26	2,0x10 ⁻⁵	0,03	NR	NR
GPRIN3 (CR743148) ^C	1,32	2,1x10 ⁻⁴	0,049	NR	NR
<i>Regulado por STAT3 Adicional:</i>					
ID3 (NM_002167)	-1,3	5,2x10 ⁻⁴	0,16	NR	NR
MYC (NM_002467)	1,2	0,04	0,75	1,29	0,01

- 10 Para obtener una métrica que indique riesgo de progreso de AR, se calculó la suma de los valores de expresión normalizados para la "firma" de AR de 12 genes para cada individuo en la cohorte de entrenamiento (véanse los métodos). A continuación se construyó una curva característica operativa del receptor (ROC), trazando la sensibilidad frente [1-especificidad] para un rango de cortes de esta métrica de riesgo, el área bajo la cual (0,85; error típico de la media [ETM]=0,04) sugirió una utilidad discriminatoria prometedora (Figura 1B). Cuando el valor de corte discriminatorio óptimo para esta métrica basada en la cohorte de entrenamiento se aplicó para clasificar los miembros de la cohorte de validación, la AR se podría pronosticar entre los pacientes con AI con una exactitud de sensibilidad, especificidad, razones de probabilidad positiva y negativa (IC de 95%) de 0,64 (0,45-0,80), 0,70 (0,54-0,82), 2,2 (1,2-3,8) y 0,5 (0,3-0,9) respectivamente. También se sometió a ensayo una metodología de aprendizaje de la máquina alternativa, una máquina de soporte vectorial (SVM), como herramienta de clasificación en las cohortes de los autores de la presente invención. El uso del modelo de predicción con SVM condujo a una modesta mejora en la exactitud de la clasificación de la AI sobre la del modelo ROC, con exactitud de sensibilidad, especificidad, razones de probabilidad positiva y negativa (IC de 95%) de 0,68 (0,48-0,83), 0,70 (0,60-0,87), 2,2 (1,2-3,8) y 0,4(0,2-0,8) respectivamente. Sin embargo, los autores de la presente invención observaron que de 13 pacientes de AI positivos para AAPC, 12 progresaron a AR, indicando que el estado de autoanticuerpos solo era un predictor mucho más sensible de AR en este subconjunto. En contraste, cuando se aplicó exclusivamente al subconjunto negativo para AAPC de la cohorte de validación de AI (n=49), el modelo de clasificación SVM proporcionó una sensibilidad de 0,85 (0,58-0,96) y una especificidad de 0,75 para la progresión a AR, funcionando de ese modo mejor en este grupo de pacientes diagnósticamente más difícil. El agrupamiento jerárquico de las muestras de AI negativas para AAPC basado en sus perfiles de expresión de la "firma" de AR de 12 genes ilustra adicionalmente las similitudes moleculares dentro del grupo con resultado de AR negativa para AAPC (Figura 1C).
- 20
- 25
- 30 Los autores de la presente invención han encontrado también que se podría incluir un tercer gen para producir una firma de 13 genes. Efectivamente, todos los genes están incluidos como para la firma de 13 genes original, pero

también se incluye un gen CD40LG regulado a la baja adicional; esto proporciona especificidad adicional al ensayo proporcionando un área bajo la curva ROC de 0,835.

5 La firma genética añade valor a las herramientas existentes en el diagnóstico de la AI negativa para AAPC. A continuación, los autores de la presente invención sometieron a ensayo el valor diagnóstico aditivo potencial de la firma de 12 genes de los autores de la presente invención en comparación con la "regla de predicción de Leiden" existente como un predictor de AR entre los pacientes de AI (4). Si bien la utilidad discriminatoria alcanzada por la regla de predicción en la cohorte de AI de los autores de la presente invención fue comparable a la referida anteriormente (n=62; AU curva ROC=0,86; ETM=0,05, datos no mostrados), su rendimiento disminuyó entre la sub-cohorte negativa para AAPC (n=49; AU curva ROC=0,74; ETM=0,08; Figura 1D. Empleando una métrica de riesgo de 12 genes como se ha descrito anteriormente, se encontró una utilidad discriminatoria equivalente en esta sub-cohorte (AU curva ROC= 0,78; ETM=0,08, datos no mostrados). Sin embargo, al obtener una métrica de riesgo modificada, que combinaba todas las características de la regla de predicción de Leiden con la firma de riesgo de 12 genes de los autores de la presente invención (o firma de 13 genes) (véanse los métodos), y al aplicarla a la cohorte de AI negativa para AAPC independiente, los autores de la presente invención podrían mejorar la utilidad de la regla de predicción en este grupo de pacientes más difícil diagnósticamente (AU ROC=0,84; ETM=0,06; Figura 1D).

20 Es más destacado un perfil de transcripción STAT3 en la AR negativa para AAPC. Los 173 pacientes estudiados se agruparon ahora en 4 categorías basadas en la diagnosis por resultados sola: AR positiva para AAPC, AR negativa para AAPC, controles sin AR inflamatorios y osteoartritis (OA); sus características demográficas y clínicas se presentan para la comparación en la Tabla 5. A continuación se podrían generar tres listas de genes expresados diferencialmente comparando cada uno de los grupos "inflamatorios" (que exhibieron por sí mismos respuestas de fase aguda comparables) con el grupo de OA (tasa de cambio >1,2; p<0,05 no corregido; Listas de genes 3-5). Las 3 listas se solaparon en un diagrama de Venn (Figura 2).

25 Se identificó una representación en exceso altamente significativa de genes implicados en el ciclo celular en asociación con la AR positiva para AAPC (24/46; p<1,0x10⁻⁵); Figura 2. Lista de genes 6). Además, los genes implicados en la regulación de la apoptosis estuvieron particularmente representados en exceso en los pacientes con AR negativa para AAPC, y la AR estuvo caracterizada en general por genes con papeles funcionales en la maduración de células T (Figura 2; Listas de genes 6-9). Resulta interesante el hecho de que, dentro de la "firma" de AR de 12 genes altamente significativa se observaron varios genes (PIM1, SOCS3, SBNO2, BCL3 y MUC1) que eran inducibles por STAT3 basándose en las fuentes de la bibliografía (27-32). La mayoría de estos fueron expresados diferencialmente muy marcadamente en la AR negativa para AAPC en comparación con la AR positiva para AAPC (Figuras 3A-B y Figuras 8A-C). Los genes inducibles por STAT3 adicionales (MYC, IL2RA; (27, 33, 34)) exhibieron patrones de expresión similares, y hubo una tendencia a que el propio STAT3 estuviera regulado al alza en la AR negativa para AAPC en comparación con la AR positiva para AAPC (Figuras 8D-F). Por otra parte, se observó un patrón de expresión recíproco a través de los grupos de resultados para el inhibidor del gen que codifica la proteína hélice-bucle-hélice negativa dominante de Unión 3 al ADN (ID3) (Figura 7G), coincidente con su supuesto papel regulador en la señalización de STAT3 (35). También se observó sin embargo que MYC e ID3, aunque no incluidos en la firma de AR discriminatoria bajo los umbrales de significación rigurosos utilizados, mostraban una expresión diferencial robusta entre pacientes con AR y sin AR dentro de la cohorte de entrenamiento sola (Tabla 2). Finalmente, con respecto tanto a la firma de 12 genes, como a la lista ampliada de genes exclusivamente desregulados en la AR negativa para AAPC (Lista de Genes 7), el solapamiento con conjuntos de genes inducibles por STAT3 pronosticados independientemente (métodos y Lista de Genes 1) confirmó una preponderancia de genes inducibles por STAT3 (valores p hipergeométricos <0,005 en ambos casos) - que no se observó para genes desregulados solo en la AR positiva para AAPC (p=0,19).

45 La IL-6 en suero es la más alta en la AR negativa para AAPC, y pronostica independientemente la expresión génica inducible por STAT3 de CD4+. Puesto que un mecanismo clásico de fosforilación de STAT3 es a través de la ligación del co-receptor gp130 (36), los autores de la presente invención plantearon la hipótesis de que el aumento de los niveles sistémicos de un ligando de gp130 clave y citoquina proinflamatoria, IL-6, puede ser responsable del programa transcripcional mediado por STAT3 en pacientes con AR temprana. La IL-6 en suero en el punto de partida se midió en 131/173 pacientes de CAT, agrupados con posterioridad de acuerdo con su última diagnosis (AR negativa para AAPC, AR positiva para AAPC, artropatía inflamatoria distinta de AR u OA). Los niveles de IL-6 fueron bajos en conjunto (generalmente <100 pg/ml), pero fueron los más altos en el grupo de AR negativa para AAPC (Figura 3C). Por supuesto, a diferencia del marcador genérico de proteína C reactiva de inflamación sistémica (CRP), IL-6 tenía un valor discriminatorio para la AR negativa para AAPC en comparación con las artritis inflamatorias distintas de AR (Figuras 3C y 3D). Además, entre los individuos para los que se obtuvieron las muestras de IL-6 y de ARN de células T CD4+ de PB pareadas y contemporáneas, estuvo presente una clara correlación entre IL-6 y la expresión normalizada de un rango de genes inducibles por STAT3 (Figuras 4A-D; Figuras 9A-D); por ejemplo, las mediciones de IL-6 en suero se correlacionaron con la expresión normalizada de SOCS3: R de Pearson=0,57, p<0,001 (Figura 4A). Para excluir la posibilidad de que esta observación reflejara meramente inflamación sistémica, se llevó a cabo un análisis multivariado para medir la contribución relativa de tres variables en suero relacionadas sobre la expresión génica inducible por STAT3. Por tanto, de CRP, IL-6 y la citoquina proinflamatoria alternativa factor de necrosis tumoral alfa (TNF- α , que no señala a través de STAT3), únicamente IL-6 pronosticó independientemente la expresión de SOCS3 de células T CD4+ de PB entre 131 pacientes de artritis

temprana ($\beta=0,53$; $p<0,001$. Tabla 6). Finalmente, los autores de la presente invención no encontraron relaciones similares entre los ligandos de gp130 alternativos medibles en sueros (G-CSF y leptina) y la expresión de genes inducibles por STAT3 de células T CD4+ de PB (Figura 10).

5 Tabla 6. Resultados del análisis de regresión lineal estándar para identificar las variables en suero relacionadas asociadas independientemente con la expresión del gen inducible STAT-3 entre 131 pacientes clínicos con AT. La variable dependiente fue Log10 (expresión del gen SOCS3 normalizada). ET(B): error típico para B; CI: intervalo de confianza. Todas las variables experimentaron una transformación previa con el fin de satisfacer las condiciones de normalidad de la regresión lineal estándar. Solamente la [IL-6] en suero está asociada independientemente con la expresión de SOCS3 en células T CD4+ ($p<0,001$; Véase el texto).

Variable en suero	Coeficientes no Estandarizados:		Coeficientes Estandarizados:	valor p	IC 95% (B) (inferior, superior)
	B	ET (B)	β		
Log ₁₀ [IL-6]	0,21	0,05	0,53	<0,001	0,12, 0,30
Log ₁₀ [CRP]	0,06	0,04	0,13	0,18	-0,03, 0,15
Log ₁₀ [TNF α]	-0,09	0,09	-0,08	0,32	-0,27, 0,09
Constante	-0,12	0,05	-	0,026	-0,23, -0,02

10

DISCUSIÓN.

Los autores de la presente invención presentan un análisis único del transcriptoma de las células T CD4+ de PB ex vivo en una cohorte de inicio bien caracterizada de pacientes con artritis temprana. Los autores de la presente invención han minimizado la confusión incluyendo solamente pacientes que no habían recibido previamente terapia modificadora de la enfermedad, centrándose en un único subconjunto de células de PB, recogiendo y procesando las muestras de manera expeditiva en condiciones normalizadas, y empleando un control de calidad cuidadoso. En términos de una herramienta de diagnóstico potencial, resulta satisfactorio que la "firma de expresión de AR" de 12 genes de los autores de la presente invención (Tabla 2) funcionara mejor entre el grupo de pacientes con AI negativa para AAPC que suponía un reto desde el punto de vista del diagnóstico. Estos descubrimientos apoyan la implicación de las células T CD4+ en la enfermedad tanto positiva como negativa para AAPC. La observación de que ambos serotipos de AR diferían de un grupo de control no inflamatorio en un grado mayor que un grupo de control inflamatorio sin AR (Figura 2), apoya adicionalmente este concepto.

La sensibilidad y especificidad de la firma (0,85 y 0,75) para pronosticar una posterior AR en pacientes con AI seronegativos equivale a una razón de probabilidad positiva (LR+) de 3,4, indicando que una probabilidad previa de 25% para el progreso de la AR entre esta cohorte (13/49 pacientes progresaron a AR) se duplica a 53% para un individuo asignado a una clasificación SVM positiva (probabilidad posterior; $[3,4 \times \{0,25/0,75\}]/[1+\{3,4 \times (0,25/0,75)\}](37)$). Por otra parte, de los 13 pacientes con AI negativa para AAPC que progresaron a AR en la cohorte de los autores de la presente invención, 8 entraron en la categoría de riesgo "intermedio" para el progreso a AR de acuerdo con la puntuación de predicción de Leiden validada(4), permaneciendo de ese modo sujetos a un diagnóstico retardado. Resulta alentador que todos menos uno de estos pacientes fueran correctamente clasificados basándose en sus perfiles de expresión de 12 genes. La prueba de concepto de los autores de la presente invención de que este enfoque podría añadir valor a los algoritmos existentes en el diagnóstico de la AI negativa para AAPC está apoyada adicionalmente por la construcción de curvas ROC que comparan la regla de predicción de Leiden con una métrica de riesgo modificada que amalgama las características de la firma de genes de los autores de la presente invención con las de la regla de predicción (Figura 1D). La validación adicional de la firma de AR en poblaciones de AI negativa para AAPC bien definidas es ahora una prioridad.

Los datos de los autores de la presente invención indican que las células T CD4+ de PB en la AR temprana se caracterizan por una regulación al alza predominante de rutas biológicas implicadas en el progreso del ciclo celular (positiva para AAPC) y la supervivencia (negativa para AAPC) (Figura 2 y Listas de Genes 6 - 7). El análisis de la ruta también sugirió que el desarrollo y la diferenciación de las células T eran des-regulados en ambos serotipos de AR (Lista de genes 8). Estos descubrimientos concuerdan con observaciones previas de homeostasis deteriorada de las células T en AR, caracterizada por un incremento del recambio, un acortamiento del telómero e inmunosenescencia (38). Curiosamente, tales observaciones se pueden asociar con el transporte de alelos de epítipo compartido HLA-DRB1(39), que se han definido posteriormente como factores de riesgo para la positividad

40

de AAPC (40), en coherencia con el programa de ciclo de células T CD4+ más marcado en individuos seropositivos sugerido por el estudio de los autores de la presente invención (*Figura 2*).

Dada la importancia bien caracterizada de la ruta de señalización de STAT3 tanto en la ruta de oncogénesis como en la de supervivencia de células T, fue notable que 5 genes de la firma de AR de 12 genes estadísticamente robusta de los autores de la presente invención sean aparentemente inducidos después de la fosforilación de STAT3 (27-32). Esta regulación al alza fue generalmente muy pronunciada en AR negativa para AAPC (*Figuras 3A-B y Figura 8A-C*), explicando potencialmente por qué la utilidad predictiva de la firma de 12 genes fue óptima en este subconjunto de la enfermedad. Otros genes inducibles por STAT3 (incluyendo IL2RA y MYC(27, 33, 34)), junto con el propio STAT3, mostraron patrones de expresión similares, aunque estadísticamente menos robustos en los grupos de comparadores clínicos, y se observó un patrón recíproco para la expresión de ID3, coherente con una función reguladora propuesta de su producto con respecto a la señalización de STAT3 (35) (*Figura 8D-G*; véase también la *Lista de Genes 7*). La observación de los autores de la presente invención de que el incremento de los niveles de IL-6 en suero entre los asistentes a clínicas de artritis temprana puede predecir un diagnóstico de AR frente a artritis alternativas es coherente con los descubrimientos de estudios con biomarcadores anteriores (41, 42), pero la de los autores de la presente invención es, para su conocimiento, la primera demostración de una asociación concreta con la enfermedad negativa para AAPC (*Figura 3C*).

Se han observado correlaciones sorprendentes entre la expresión de células T CD4+ de PB de varios genes inducibles por STAT3 y concentraciones de IL-6 en suero contemporáneas, emparejadas (*Figuras 4 A-D; Figuras 9A-D*). Aunque las mediciones de IL-6 también se correspondían con la inflamación generalizada en general (medida en forma de CRP), así como con niveles en suero de una citoquina pro-inflamatoria no señalizada por STAT3 alternativa, TNF- α (datos no mostrados), los análisis multivariados confirmaron que IL-6 era el único predictor independiente de la expresión del gen STAT3 (*Tabla 6*). La fosforilación de STAT3 y la transcripción aguas abajo se inician por ligación del co-receptor gp130 de la superficie celular por una gama de ligandos incluyendo IL-6 (43). Los autores de la presente invención midieron la IL-6 en particular debido a su papel reconocido como citoquina pro-inflamatoria en la AR(44). No obstante, dado que se piensa que solamente 30-50% de las células T CD4+ de PB expresan IL6R unido a membrana(45), era posible que el determinante crítico *in vivo* de al menos cierta expresión de genes inducibles por STAT-3 en este entorno pudieran ser los niveles de sIL6-R circulante en lugar de IL-6. La señalización en trans de IL-6 a través de la forma soluble del receptor es crucial para las respuestas mediadas por IL-6 en células T negativas para IL-6R. Por lo tanto, los autores de la presente invención midieron las concentraciones de sIL-6R en suero en el momento inicial en un subconjunto de 80 pacientes con artritis temprana del estudio actual, que comprendía 20 de cada resultado diagnóstico definido en la *Tabla 2*. Al contrario que IL-6, no se detectó relación con el resultado diagnóstico, y tampoco hubo correlación entre la concentración de sIL-6R en suero y la expresión de genes inducibles por STAT3 (*Figuras 10A y D*).

Los autores de la presente invención también estudiaron otros dos ligandos de gp130, buscando un papel potencial para ellos en la inducción de la ruta de STAT3; El factor estimulador de las colonias de granulocitos (G-CSF) y la leptina han sido implicados ambos en la patogénesis de la AR(46, 47), pero sus niveles en suero del mismo subconjunto de pacientes del estudio no se correlacionaban con el resultado diagnóstico ni con la expresión del gen STAT3 (*Figuras 10 B-C, E-F*). Finalmente, la IL-10, que también es sabido que señala a través de STAT3(48), fue indetectable en la mayor parte de los sueros (*datos no mostrados*). Por lo tanto parece probable que los descubrimientos en relación con una firma de la expresión génica inducible por STAT3 como parte de un biomarcador de artritis temprana para AR seronegativa sean en gran medida específicos para la señalización de IL-6.

Los autores de la presente invención también revisaron las curvas ROC para examinar la utilidad discriminatoria de diversos sistemas de puntuación en su cohorte, y los subconjuntos de la misma. Se excluyeron 42 pacientes del análisis para los cuales no se encontraban disponibles las mediciones de IL-6, sin embargo, solamente uno de estos se presentaba con AI.

Figuras 11-14, examinar:

Figura 11 - Toda la cohorte, incluyendo individuos positivos para AAPC, con independencia de si se pudo asignar un diagnóstico al inicio o no.

Figura 12 - Individuos negativos para AAPC, pero también con independencia de si se pudo asignar un diagnóstico al inicio o no.

Figura 13 - Pacientes con AI, ya sean positivos para AAPC o negativos para AAPC.

Figura 14 - Pacientes con AI, solamente negativos para AAPC.

En cada cohorte/sub-cohorte, se comparan 4 curvas ROC: la regla de predicción de Leiden, la métrica de riesgo de 12 genes que comentan los autores de la presente invención, la combinación de Leiden/medida de 12 genes mencionada en el manuscrito, e IL-6 sola.

En las diapositivas 12 y 14 (excluyendo los pacientes positivos para AAPC), se proporcionan sensibilidades/especificidades para un corte ilustrativo de 10 pg/ml de [IL-6] en suero.

5 Los resultados sugieren que la IL-6 es un parámetro útil para predecir el resultado en una enfermedad negativa para AAPC temprana en particular. La predicción más eficaz parece ser proporcionada por una combinación de la puntuación de predicción de Leiden y la métrica de 12 genes.

10 En conclusión, los datos proporcionan una fuerte evidencia para la inducción del programa de transcripción de STAT3 mediado por IL-6 en células T CD4+ de PB de pacientes con AR temprana, que es muy prominente en individuos negativos para AAPC, y que contribuye a una "firma" de la expresión génica que puede tener utilidad en el diagnóstico. Semejante patrón de expresión génica entre las células T CD4+ en esta fase temprana crítica en la historia natural de la artritis inflamatoria podría tener un papel definitorio en el cambio de inflamación potencialmente auto-limitante a autoinmunidad crónica perpetuada en células T - un modelo que puede no estar limitado al ejemplo de la AR. En cualquier evento, los descubrimientos podrían preparar el camino para un tratamiento novedoso paradigma en la artritis temprana, por medio del cual los fármacos que se dirigen al "eje" *IL-6-gp130-STAT3* encuentren un nicho racional como agentes biológicos de primera elección en el manejo de la AR negativa para AAPC. Uno de tales agentes, ya disponible en la medicina clínica, es el bloqueador del receptor de IL-6 tocilizumab, cuya eficacia ya ha sido establecida en la AR(49); otros incluyen los inhibidores de quinasas Janus actualmente experimentando pruebas clínicas de fase III para la enfermedad (50). Los estudios tales como los de los autores de la presente invención deben contribuir por último a la realización de la auténtica "medicina personalizada" en la artritis inflamatoria temprana, en la que la heterogeneidad compleja está estratificada en subconjuntos patofisiológicamente y terapéuticamente relevantes, con claros beneficios en términos de resultados clínicos y coste.

Lista de Genes 1 - GENES INDUCIBLES POR STAT3 sobre una MATRIZ ILLUMINA

Símbolo (Illumina)	RefSeq	Símbolo (base de datos de origen, si fuera diferente)
A2M	NM_000014.4	
ACCN1	NM_001094.4	
ACCN4	NM_018674.3	
ADM2	NM_024866.4	
ALKBH6	NM_198867.1	MGC14376
AP1S2	NM_003916.3	
AP2B1	NM_001282.2	
APBA1	NM_001163.2	
ARF3	NM_001659.1	
ARHGAP8	NM_181335.2	
ARL6IP6	NM_152522.3	MGC33864
ARX	NM_139058.1	
ASXL1	NM_015338.4	
ATG3	NM_022488.3	
AZIN1	NM_015878.4	OAZIN
B3GAT3	NM_012200.2	

ES 2 530 022 T3

Símbolo (Illumina)	RefSeq	Símbolo (base de datos de origen, si fuera diferente)
BCAM	NM_005581.3	LU
BCL2	NM_000633.2	
BCL2L1	NM_138578.1	
BCL7A	NM_001024808.1	
BIRC5	NM_001168.2	
BMI1	NM_005180.5	
BMP4	NM_130851.1	
BNC1	NM_001717.2	
BTBD1	NM_001011885.1	
C14orf179	NM_052873.1	MGC16028
C16orf85	NM_001001682.1	FLJ45530
C17orf91	NM_001001870.1	MGC 14376
C5orf41	NM_153607.1	LOC153222
CA10	NM_020178.3	
CALU	NM_001219.2	
CAPZA1	NM_006135.1	
CCL2	NM_002982.3	
CCND1	NM_053056.2	
CCND3	NM_001760.2	
CD40	NM_152854.2	TNFRSF5
CDKN1A	NM_000389.2	
CEBPB	NM_005194.2	
CENTD1	NM_139182.1	
CHRM1	NM_000738.2	
CISH	NM_145071.1	
CLDN5	NM_003277.2	

ES 2 530 022 T3

Símbolo (Illumina)	RefSeq	Símbolo (base de datos de origen, si fuera diferente)
COL4A3BP	NM_005713.1	
CPA4	NM_016352.2	
CPLX2	NM_006650.3	
CRTAC1	NM_018058.4	
CSRP1	NM_004078.1	
CTGF	NM_001901.2	
CXorf36	NM_024689.1	FLJ14103
CYP19A1	NM_000103.2	
DDIT3	NM_004083.4	
DERL2	NM_016041.3	F-LANA
EGR1	NM_001964.2	
EGR3	NM_004430.2	
EHHADH	NM_001966.2	
EIF4E	NM_001968.2	
EIF4G1	NM_198244.1	
E1F5A	NM_001970.3	
ELMO1	NM_130442.2	
EPHA7	NM_004440.2	
EXOC3	NM_007277.4	SEC6L1
FAS	NM_152872.1	TNFRSF6
FASN	NM_004104.4	
FBN2	NM_001999.3	
FBXL3	NM_012158.1	FBXL3A
FCGR1A	NM_000566.2	
FLJ33387	NM_182526.1	
FLRT1	NM_013280.4	

ES 2 530 022 T3

Símbolo (Illumina)	RefSeq	Símbolo (base de datos de origen, si fuera diferente)
FOS	NM_005252.2	
FOSB	NM_006732.1	
FOXO4	NM_005938.2	MLLT7
FUT8	NM_178156.1	
GABRB1	NM_000812.2	
GEN1	NM_182625.2	FLJ40869
GPC3	NM_004484.2	
GPHN	NM_001024218.1	
GRIN2D	NM_000836.2	
HEYL	NM_014571.3	
HMOX1	NM_002133.1	
HNRPR	NM_005826.2	
HOXB13	NM_006361.5	
HOXB4	NM_024015.3	
HOXB9	NM_024017.3	
HOXC4	NM_014620.4	
HOXC6	NM_153693.3	
HS6ST3	NM_153456.2	
HSP90AA1	NM_001017963.2	HSPCA
HSP90AB1	NM_007355.2	HSPCB
ICAM1	NM_000201.1	
IGF1	NM_000618.2	
IL10	NM_000572.2	
IL18BP	NM_005699.2	
IL2RA	NM_000417.1	
IL6	NM_000600.1	

ES 2 530 022 T3

Símbolo (Illumina)	RefSeq	Símbolo (base de datos de origen, si fuera diferente)
IL6ST	NM_002184.2	
IRF1	NM_002198.1	
IRX5	NM_005853.4	
JAK3	NM_000215.2	
JUN	NM_002228.3	
KAZALD1	NM_030929.3	
KCNH3	NM_012284.1	
KCNN2	NM_021614.2	
KCNN3	NM_002249.4	
KCNT2	NM_198503.2	SLICK
KIAA0146	NM_001080394.1	
KIAA0913	NM_015037.2	
KIRREL3	NM_032531.2	
KPNB1	NM_002265.4	
LBP	NM_004139.2	
LRP2	NM_004525.2	
LTA	NM_000595.2	
LTBP1	NM_000627.2	
MAFF	NM_012323.2	
MAML1	XM_937023.1	
MATN4	NM_030592.1	
MBD6	NM_052897.3	
MCL1	NM_021960.3	
MEIS2	NM_172315.1	
MIA2	NM_054024.3	
MID1IP1	NM_021242.4	MIG12

ES 2 530 022 T3

Símbolo (Illumina)	RefSeq	Símbolo (base de datos de origen, si fuera diferente)
MIS12	NM_024039.1	
MLL	NM_005933.2	
MNT	NM_020310.2	
MOBKL2C	NM_201403.2	
MTMR14	NM_001077525.1	FLJ22405
MUC1	NM_002456.4	
MUC4	NM_018406.3	
MYC	NM_002467.3	
MYT1	NM_004535.2	
NAPB	NM_022080.1	
NAV2	NM_145117.3	
NCAM1	NM_001076682.2	
NCOA5	NM_020967.2	
NDST2	NM_003635.2	
NELL2	NM_006159.1	
NFAM1	NM_145912.5	
NOL3	NM_003946.3	
NOS2A	NM_000625.3	
NPAS4	NM_178864.2	NXF
NR1D1	NM_021724.2	
NR4A1	NM_002135.3	
OSM	NM_020530.3	
OXTR	NM_000916.3	
PAPD1	NM_018109.2	
PCBP4	NM_033009.1	
PGF	NM_002632.4	

ES 2 530 022 T3

Símbolo (Illumina)	RefSeq	Símbolo (base de datos de origen, si fuera diferente)
PIM1	NM_002648.2	
PPFIA2	NM_003625.2	
PRF1	NM_005041.4	
PROS1	NM_000313.1	
PTMS	NM_002824.4	
RBPJ	NM_203283.1	PRBPSUH
RBPJL	NM_014276.2	RBPSUHL
REG1A	NM_002909.3	
REM2	NM_173527.2	FLJ38964
RGS3	NM_134427.1	
RIMS1	NM_014989.3	
RND1	NM_014470.2	
RNF213	NM_020914.3	C17ORF27
RORA	NM_002943.2	
RPUSD4	NM_032795.1	FLJ14494
RSPO4	NM_001040007.1	R-SPONDIN
S100A14	NM_020672.1	
SCUBE3	NM_152753.2	
SDC1	NM_002997.4	
SENP3	NM_015670.4	
SERPING1	NM_001032295	
SET	NM_003011.2	
SGMS1	NM_147156.3	TMEM23
SHOX2	NM_006884.2	
SLC35A5	NM_017945.2	
SLC38A5	NM_033518.1	

ES 2 530 022 T3

Símbolo (Illumina)	RefSeq	Símbolo (base de datos de origen, si fuera diferente)
SLCO5A1	NM_030958.1	
SMG7	NM_201569.1	C10RF16
SOCS1	NM_003745.1	
SOCS3	NM_003955.3	
SOS1	NM_005633.2	
SP6	NM_199262.2	
SPON1	NM_006108.2	
SPTBN2	NM_006946.1	APG3
ST7L	NM_138729.2	
STRA13	NM_144998.2	
SV2B	NM_014848.3	
TAOK1	NM_020791.1	TAO1
TCF7L2	NM_030756.2	
TIMP1	NM_003254.2	
TIMP3	NM_000362.4	
TJAP1	NM_080604.1	TJP4
TLR2	NM_003264.3	
TM9SF1	NM_006405.5	
TMEM158	NM_015444.2	
TMEM180	NM_024789.3	C10ORF77
TMEM37	NM_183240.2	PR1
TNF	NM_000594.2	
TNFRSF8	NM_001243.3	
TNFSF18	NM_005092.2	
TNRC6A	NM_014494.2	TNRC6
TRAF4	NM_004295.3	

ES 2 530 022 T3

Símbolo (Illumina)	RefSeq	Símbolo (base de datos de origen, si fuera diferente)
TRH	NM_007117.1	
TRIB2	NM_021643.1	
TRIP10	NM_004240.2	
TSC22D4	NM_030935.3	THG-1
UBE4B	NM_006048.2	
UBR1	NM_174916.1	
UBR5	NM_015902.4	DD5
UBTF	NM_14233.2	
UPK2	NM_006760.2	
VCL	NM_014000.2	
VEGFA	NM_003376.4	VEGF
VEZF1	NM_007146.2	ZNF161
VIP	NM_194435.1	
VSNL1	NM_003385.4	
WDR81	NM_152348.1	FLJ33817
WEE1	NM_003390.2	
WNT4	NM_030761.3	
YY1	NM_003403.3	
ZBTB11	NM_014415.2	
ZBTB17	NM_003443.1	ZNF151
ZBTB25	NM_006977.2	ZNF46
ZBTB9	NM_152735.3	
ZC3H18	NM_144604.2	LOC124245
ZFP112	NM_001083335.1	ZNF228
ZFYVE9	NM_007324.2	
ZHX2	NM_014943.3	

ES 2 530 022 T3

Símbolo (Illumina)	RefSeq	Símbolo (base de datos de origen, si fuera diferente)
ZNF296	NM_145288.1	ZNF342
ZNF395	NM_018660.2	PBF

Nota, para probabilidades hipergeométricas, el número total de genes "no redundantes" utilizado como "tamaño de población" = 37.847

Lista de genes 2: AR vs No-AR, Muestras de entrenamiento (n=111)

***Entradas en Rojo/Negrita/Cursiva => p<0,05 cuando se corrigen para el ensayo múltiple (FDR)**

****el transcrito CR743148 (ID de Sonda Illumina 6370082) ha sido retirado de NCBI, pero la EST corresponde a una o varias variantes de corte y empalme dentro del gen GPRIN3 (cromosoma 4.90).**

ID de Illumina	Símbolo	RefSeq	TC	p* no corregido
2070168	CX3CR1	NM_001337.3	1,657	0,0308
6330725	<i>BCL3</i>	<i>NM_005178.2</i>	1,59	2,63E-05
4230102	<i>SOCS3</i>	<i>NM_003955.3</i>	1,55	3,36E-06
430438	MIAT	NR_003491.1	1,531	0,000239
3130301	<i>PIM1</i>	<i>NM_002648.2</i>	1,515	6,80E-06
3190609	<i>SBNO2</i>	<i>NM_014963.2</i>	1,465	1,15E-05
6280170	<i>PDCD1</i>	<i>NM_005018.1</i>	1,423	1,04E-05
6220288	PRDM1	NM_001198.2	1,415	0,000242
60470	STX11	NM_003764.2	1,386	0,00117
6620689	MTHFD2	NM_001040409.1	1,375	7,88E-05
1510553	DACT1	NM_016651.5	1,37	0,00763
4670193	PRF1	NM_005041.4	1,351	0,0465
1710070	ITGAM	NM_000632.3	1,339	0,0249
5700753	CEACAM1	NM_001024912.1	1,334	0,000235
160292	APOBEC3H	NM_181773.2	1,332	0,0017
6220195	BATF	NM_006399.2	1,326	0,000953
6370082	<i>GPRIN3**</i>	<i>CR743148</i>	1,321	0,00021
7200301	ARID5A	NM_212481.1	1,321	0,00388

ES 2 530 022 T3

*Entradas en Rojo/Negrita/Cursiva => p<0,05 cuando se corrigen para el ensayo múltiple (FDR)				
**el transcrito CR743148 (ID de Sonda Illumina 6370082) ha sido retirado de NCBI, pero la EST corresponde a una o varias variantes de corte y empalme dentro del gen GPRIN3 (cromosoma 4.90).				
ID de Illumina	Símbolo	RefSeq	TC	p* no corregido
4730523	IGFL2	NM_001002915.1	1,307	1,12E-07
4060358	ABCA1	NM_005502.2	1,299	0,00162
6550600	MYC	NM_002467.3	1,295	0,0425
6770673	SOCS2	NM_003877.3	1,287	0,011
670576	LOC731186	XM_001128760.1	1,283	2,28E-05
650452	PLCH2	NM_014638.2	1,275	0,00242
7560731	SNORA64	NR_002326.1	1,265	0,000326
1430598	FBXO32	NM_058229.2	1,264	0,00127
2070037	ICOS	NM_012092.2	1,263	0,00318
5490068	MCOLN2	NM_153259.2	1,26	0,0131
3800647	UGCG	NM_003358.1	1,258	0,00242
4230228	CDK5RAP3	NM_176095.1	1,257	0,0126
7650026	MUC1	NM_001044391.1.1	1,255	1,95E-05
2070288	MT1E	NM_175617.3	1,253	0,0171
1410408	ARID5B	NM_032199.1	1,248	0,000716
1510424	S100P	NM_005980.2	1,246	0,000653
3420128	AP3M2	NM_006803.2	1,246	0,00305
3190148	DDIT4	NM_019058.2	1,245	0,0281
160494	AQP9	NM_020980.2	1,243	0,0275
870202	TNFSF10	NM_003810.2	1,24	0,00949
2320129	CSDA	NM_003651.3	1,235	0,0405
2100215	MAF	NM_001031804.1	1,233	0,00222
6420731	SLC20A1	NM_05415.3	1,231	9,77E-05
10333	LOC731682	XM_001129369.1	1,229	0,0356

ES 2 530 022 T3

*Entradas en Rojo/Negrita/Cursiva => p<0,05 cuando se corrigen para el ensayo múltiple (FDR)				
**el transcrito CR743148 (ID de Sonda Illumina 6370082) ha sido retirado de NCBI, pero la EST corresponde a una o varias variantes de corte y empalme dentro del gen GPRIN3 (cromosoma 4.90).				
ID de Illumina	Símbolo	RefSeq	TC	p* no corregido
4540376	FAM13A1	NM_001015045.1	1,228	0,043
1850554	NPDC1	NM_015392.2	1,227	0,000194
3190092	LDHA	NM_005566.1	1,226	3,82E-05
4260372	GTSCR1	XM_496277.2	1,225	0,049
6250010	GPRIN3	NM_198281.2	1,223	0,0136
3840470	ST6GALNAC1	NM_018414.2	1,222	0,0141
4590446	MSL3L1	NM_078628.1	1,22	0,00221
5890524	LINS1	NM_181740.1	1,22	0,000796
7570600	FLJ33590	NM_173821.1	1,22	0,00361
3940390	TBXAS1	NM_001061.2	1,218	0,0186
450615	MT2A	NM_005953.2	1,216	0,000379
520278	FAM100B	NM_182565.2	1,214	0,00367
7050326	CDKN2D	NM_079421.2	1,214	0,000125
1400601	C20orf100	NM_032883.1	1,212	0,0142
5130382	CLDN5	NM_003277.2	1,212	0,0321
5700735	PARP9	NM_031458.1	1,211	0,00126
5910364	TYMS	NM_001071.1	1,211	0,00678
5900471	PTGER2	NM_000956.2	1,21	0,0115
2850291	GARNL4	NM_015085.3	1,209	0,00831
4180301	ZNF365	NM_014951.2	1,209	0,0192
4200541	FAM113B	NM_138371.1	1,209	0,000487
5090754	KIAA0101	NM_014736.4	1,208	0,00994
7100372	PRPF4B	NM_003913.3	1,206	0,0112
5420538	TP53INP1	NM_033285.2	1,205	0,016

ES 2 530 022 T3

*Entradas en Rojo/Negrita/Cursiva => p<0,05 cuando se corrigen para el ensayo múltiple (FDR)				
**el transcrito CR743148 (ID de Sonda Illumina 6370082) ha sido retirado de NCBI, pero la EST corresponde a una o varias variantes de corte y empalme dentro del gen GPRIN3 (cromosoma 4.90).				
ID de Illumina	Símbolo	RefSeq	TC	p* no corregido
1510364	GBP5	NM_052942.2	1,204	0,0182
3800168	SLC2A3	NM_006931.1	1,204	0,0331
3840053	UGP2	NM_006759.3	1,204	0,000214
4610201	SNORA10	NR_002327.1	1,203	8,50E-05
6200168	PIM2	NM_006875.2	1,203	0,000135
2190689	OSBPL5	NM_020896.2	1,202	0,0208
2760112	P2RY5	NM_005767.4	1,201	0,00774
4670603	ELMO2	NM_133171.2	1,201	0,0101
3460008	TMEM173	NM_198282.1	1,2	0.000813
3780161	TMEM70	NM_017866.4	1,2	0.00548
5550066	CMAH	NR_002174.2	1,2	1,69E-05
3170703	LY9	NM_001033667.1	0,837	0,0043
5810746	MATN2	NM_002380.3	0,837	0,00284
5080615	IL16	NM_172217.2	0,835	0,0262
160242	C13orf15	NM_014059.2	0,834	0,00866
6200019	KLRB1	NM_002258.2	0,831	0,0393
6280504	LOC100008589	NR_003287.1	0,83	0,0364
6280243	DNTT	NM_001017520.1	0,829	0,000589
5130692	DDX17	NM_030881.3	0,821	0,0105
2970730	MYADM	NM_001020820.1	0,819	0,0171
870056	FAM119B	NM_015433.2	0,816	0,00801
1580477	C11orf74	NM_138787.2	0,815	7,68E-05
2710309	ELA1	NM_001971.4	0,783	0,00177
50706	CD40LG	NM_000074.2	0,78	0,000441

***Entradas en Rojo/Negrita/Cursiva => p<0,05 cuando se corrigen para el ensayo múltiple (FDR)**

****el transcrito CR743148 (ID de Sonda Illumina 6370082) ha sido retirado de NCBI, pero la EST corresponde a una o varias variantes de corte y empalme dentro del gen GPRIN3 (cromosoma 4.90).**

ID de Illumina	Símbolo	RefSeq	TC	p* no corregido
1470762	AUTS2	NM_015570.1	0,779	0,0224
7570324	ID3	NM_002167.2	0,775	0,000522
2320253	USMG5	NM_032747.2	0,774	0,0236
6770603	<i>NOG</i>	NM_005450.2	0,755	3,14E-05
130609	FCGBP	NM_003890.1	0,707	0,00124
5080192	SERPINE2	NM_006216.2	0,657	0,00192

Lista de genes 3: AR neg para AAPC vs OA, Conjunto de datos reunidos (n=173)

***Entradas en Rojo/Negrita/Cursiva => p<0,05 cuando se corrigen para el ensayo múltiple (FDR)**

****el transcrito CR743148 (ID de Sonda Illumina 6370082) ha sido retirado de NCBI, pero la EST corresponde a una o varias variantes de corte y empalme dentro del gen GPRIN3 (cromosoma 4.90).**

ID de Illumina	Símbolo	RefSeq	FC	p* no corregido
4230102	SOCS3	NM_003955.3	1,916	9,17E-09
6330725	BCL3	NM_005178.2	1,797	1,64E-05
3130301	PIM1	NM_002648.2	1,715	4,06E-06
510079	HLA-DRB4	NM_021983.4	1,701	0,0231
3190609	SBNO2	NM_014963.2	1,618	1,06E-05
1820594	HBEGF	NM_001945.1	1,607	0,00284
6220195	BATF	NM_006399.2	1,594	1,40E-05
6620689	MTHFD2	NM_001040409.1	1,561	1,42E-05
6290270	MNDA	NM_002432.1	1,558	0,0499
670010	LOC650298	XM_939387.1	1,555	0,033
60470	STX11	NM_003764.2	1,553	0,000765
6220288	PRDM1	NM_001198.2	1,531	0,000238
6550600	MYC	NM_002467.3	1,527	0,00645

ES 2 530 022 T3

*Entradas en Rojo/Negrita/Cursiva => p<0,05 cuando se corrigen para el ensayo múltiple (FDR)				
**el transcrito CR743148 (ID de Sonda Illumina 6370082) ha sido retirado de NCBI, pero la EST corresponde a una o varias variantes de corte y empalme dentro del gen GPRIN3 (cromosoma 4.90).				
ID de Illumina	Símbolo	RefSeq	FC	p* no corregido
6590377	RPS26	NM_001029.3	1,527	0,0216
3800647	UGCG	NM_003358.1	1,518	1,99E-05
4230201	CDKN1A	NM_000389.2	1,505	0,0136
1240152	CFD	NM_001928.2	1,499	0,0314
4670048	RPS26L	NR_002225.2	1,499	0,0372
1990300	SOCS1	NM_003745.1	1,49	0,0155
6270307	LOC644934	XM_930344.2	1,476	0,0265
6370082	<i>GPRIN3</i>	CR743148	1,457	8,53E-05
4060358	ABCA1	NM_005502.2	1,454	0,000771
7200301	ARID5A	NM_212481.1	1,448	0,00212
3780161	TMEM70	NM_017866.4	1,442	3,84E-06
6770673	SOCS2	NM_003877.3	1,441	0,00609
2070037	ICOS	NM_012092.2	1,438	0,000297
670576	LOC731186	XM_001128760.1	1,435	6,02E-06
430438	MIAT	NR_003491.1	1,428	0,00396
6560376	RPS26L1	NR_002309.1	1,423	0,0367
6250010	GPRIN3	NM_198281.2	1,419	0,00082
1440736	LDLR	NM_000527.2	1,415	0,00256
870202	TNFSF10	NM_003810.2	1,404	0,000859
3710397	EFNA1	NM_004428.2	1,402	0,000713
2650192	C6orf105	NM_032744.1	1,399	0,000885
5870692	GPR132	NM_013345.2	1,399	0,0278
50672	GSTM1	NM_000561.2	1,394	0,0462
1510553	DACT1	NM_016651.5	1,374	0,0373

ES 2 530 022 T3

*Entradas en Rojo/Negrita/Cursiva => $p < 0,05$ cuando se corrigen para el ensayo múltiple (FDR)

**el transcrito CR743148 (ID de Sonda Illumina 6370082) ha sido retirado de NCBI, pero la EST corresponde a una o varias variantes de corte y empalme dentro del gen GPRIN3 (cromosoma 4.90).

ID de Illumina	Símbolo	RefSeq	FC	p* no corregido
670255	GADD45A	NM_001924.2	1,373	0,0411
2320129	CSDA	NM_003651.3	1,369	0,0161
3940438	NCF1	NM_000265.4	1,369	0,0486
6960195	LOC650646	XM_942527.2	1,369	0,0456
6280458	BCL6	NM_001706.2	1,363	0,00517
520278	FAM100B	NM_182565.2	1,352	0,000191
160494	AQP9	NM_020980.2	1,349	0,0216
7400747	FAM89A	NM_198552.1	1,336	0,000375
6860347	FAM46C	NM_017709.3	1,334	0,0429
1430598	FBXO32	NM_058229.2	1,331	0,0017
2570291	IFNGR2	NM_005534.2	1,329	0,000131
3800168	SLC2A3	NM_006931.1	1,329	0,00763
3840470	ST6GALNAC1	NM_018414.2	1,329	0,00551
4670603	ELMO2	NM_133171.2	1,322	0,00404
270152	SLC7A5	NM_003486.5	1,321	0,0382
5700753	CEACAM1	NM_001024912.1	1,321	0,00359
1050068	F2RL1	NM_005242.3	1,316	0,00015
4810520	TRIB1	NM_025195.2	1,316	0,0366
5670465	ADM	NM_001124.1	1,313	0,0187
7050326	CDKN2D	NM_079421.2	1,313	2,36E-05
4730411	SFXN1	NM_022754.4	1,312	0,00143
5270097	LOC653853	XM_936029.1	1,309	0,00641
1510424	S100P	NM_005980.2	1,305	0,00177
3190092	LDHA	NM_005566.1	1,302	4,95E-05

ES 2 530 022 T3

***Entradas en Rojo/Negrita/Cursiva => $p < 0,05$ cuando se corrigen para el ensayo múltiple (FDR)**

****el transcrito CR743148 (ID de Sonda Illumina 6370082) ha sido retirado de NCBI, pero la EST corresponde a una o varias variantes de corte y empalme dentro del gen GPRIN3 (cromosoma 4.90).**

ID de Illumina	Símbolo	RefSeq	FC	p* no corregido
5890524	LINS1	NM_181740.1	1,298	0,000974
3420128	AP3M2	NM_006803.2	1,296	0,00528
1030102	RGS16	NM_002928.2	1,295	0,000337
3190148	DDIT4	NM_019058.2	1,291	0,0213
3460008	TMEM173	NM_198282.1	1,287	0,000248
6280672	TMEM49	NM_030938.2	1,284	0,000283
5820020	PRDX3	NM_006793.2	1,282	0,0019
2470348	NFKBIZ	NM_001005474.1	1,281	0,00621
2470358	IFNGR1	NM_000416.1	1,279	0,00205
7560731	SNORA64	NR_002326.1	1,278	0,00259
1230201	CTLA4	NM_005214.3	1,277	0,00417
450348	GNG10	NM_001017998.2	1,276	9,60E-05
380056	B3GNT2	NM_006577.5	1,274	0,000643
1990753	SLA	NM_006748.1	1,271	0,0131
2600735	TLR6	NM_006068.2	1,271	0,00765
3840053	UGP2	NM_006759.3	1,271	0,000201
2070288	MT1E	NM_175617.3	1,27	0,0397
6270554	LGALS8	NM_201545.1	1,27	0,000789
2640341	FKBP5	NM_004117.2	1,269	0,00778
6660630	TP53INP1	NM_033285.2	1,266	0,00444
4590446	MSL3L1	NM_078628.1	1,265	0,00248
4610201	SNORA10	NR_002327.1	1,265	0,00016
4010097	FBXO5	NM_012177.2	1,264	0,000131
1260086	ID2	NM_002166.4	1,263	0,014

ES 2 530 022 T3

*Entradas en Rojo/Negrita/Cursiva => p<0,05 cuando se corrigen para el ensayo múltiple (FDR)				
**el transcrito CR743148 (ID de Sonda Illumina 6370082) ha sido retirado de NCBI, pero la EST corresponde a una o varias variantes de corte y empalme dentro del gen GPRIN3 (cromosoma 4.90).				
ID de Illumina	Símbolo	RefSeq	FC	p* no corregido
7320041	GALNAC4S-6ST	NM_015892.2	1,262	0,0289
4670414	TMEM140	NM_018295.2	1,261	0,00276
3870706	FURIN	NM_002569.2	1,259	0,00518
1410408	ARID5B	NM_032199.1	1,258	0,00417
4200541	FAM113B	NM_138371.1	1,258	0,000767
4590228	GLRX	NM_002064.1	1,258	0,00719
20446	CEBPB	NM_005194.2	1,257	0,0108
1850554	NPDC1	NM_015392.2	1,257	0,00132
5550343	PDCL	NM_005388.3	1,253	0,000603
840554	RYBP	NM_012234.4	1,251	0,00531
1340075	BAG3	NM_004281.3	1,251	0,000391
4230554	REXO2	NM_015523.2	1,251	0,00173
5420564	NFIL3	NM_005384.2	1,251	0,0156
6220543	HIF1A	NM_001530.2	1,251	0,0068
70167	LY96	NM_015364.2	1,249	0,00105
4230619	GCA	NM_012198.2	1,249	0,0104
2760112	P2RY5	NM_005767.4	1,247	0,0114
6420731	SLC20A1	NM_005415.3	1,247	0,000324
7150280	LOC145853	XM_096885.9	1,247	1,21E-05
3420593	LMNB1	NM_005573.2	1,245	0,00222
4590349	ACVR2A	NM_001616.3	1,245	0,0009
630167	SDCBP	NM_001007067.1	1,244	0,0149
2750719	DDX21	NM_004728.2	1,243	0,00255
5130382	CLDN5	NM_003277.2	1,241	0,0235

ES 2 530 022 T3

***Entradas en Rojo/Negrita/Cursiva => p<0,05 cuando se corrigen para el ensayo múltiple (FDR)**

****el transcrito CR743148 (ID de Sonda Illumina 6370082) ha sido retirado de NCBI, pero la EST corresponde a una o varias variantes de corte y empalme dentro del gen GPRIN3 (cromosoma 4.90).**

ID de Illumina	Símbolo	RefSeq	FC	p* no corregido
1010653	POLR1C	NM_203290.1	1,239	0,00639
10630	IL21R	NM_181078.1	1,237	0,00733
5270167	GNL3	NM_206826.1	1,237	0,00236
6200168	PIM2	NM_006875.2	1,237	0,000751
3190112	SERPINB1	NM_0306662	1,236	0,000149
6280170	PDCD1	NM_005018.1	1,236	0,0285
670086	MXD1	NM_002357.2	1,235	0,0026
2230379	NAMPT	NM_005746.2	1,232	0,0132
3890326	SOD2	NM_001024465.1	1,232	0,0147
4050681	NDUFV2	NM_21074.1	1,232	0,00154
6370414	CLECL1	NM_172004.2	1,232	0,0459
2060615	ACVR1B	NM_020328.2	1,23	0,000616
2710709	FCGR1B	NM_001017986.1	1,229	0,0434
5270110	EIF4A3	NM_014740.2	1,228	0,000397
4920110	GADD45B	NM_015675.2	1,227	0,00136
6380112	GRAMD4	NM_015124.2	1,227	0,000299
2190452	PIM3	XM_938171.2	1,225	0,00047
5900274	EDA	NM_001005611.1	1,225	0,000448
2630400	CSTF2T	NM_015235.2	1,224	0,00143
7150176	MAT2A	NM_005911.4	1,224	0,00517
5080021	BIRC3	NM_001165.3	1,22	0,00923
4260019	NGRN	NM_016645.2	1,218	0,0008
4810615	SLC25A44	NM_014655.1	1,218	0,00342

ES 2 530 022 T3

***Entradas en Rojo/Negrita/Cursiva => $p < 0,05$ cuando se corrigen para el ensayo múltiple (FDR)**

****el transcrito CR743148 (ID de Sonda Illumina 6370082) ha sido retirado de NCBI, pero la EST corresponde a una o varias variantes de corte y empalme dentro del gen GPRIN3 (cromosoma 4.90).**

ID de Illumina	Símbolo	RefSeq	FC	p* no corregido
5870307	LOC440359	XM_496143.2	1,218	0,0223
1990630	TRIB3	NM_021158.3	1,217	0,0221
2600059	GNPDA1	NM_005471.3	1,215	0,0146
5900471	PTGER2	NM_000956.2	1,213	0,0365
3930390	SMAP2	NM_022733.1	1,212	0,00839
3420241	SLC2A14	NM_153449.2	1,211	0,0374
5360079	GIMAP5	NM_018384.3	1,211	0,00847
130452	GP5	NM_004488.1	1,21	0,000553
5810504	METRNL	NM_001004431.1	1,21	0,0219
4120131	KISS1R	NM_032551.3	1,209	0,00165
1030646	FLJ43692	NM_001003702.1	1,208	0,00793
4480504	ZNF828	NM_032436.1	1,207	0,00585
270601	HIAT1	NM_033055.2	1,206	0,0116
990735	RNF149	NM_173647.2	1,205	0,0033
3170091	GIMAP7	NM_153236.3	1,203	0,00101
3390612	TLR8	NM_016610.2	1,203	0,0404
1940524	STS-1	NM_032873.3	1,202	0,00732
4900575	PTRH2	NM_016077.3	1,202	0,0103
130021	IL2RA	NM_000417.1	1,2	0,00378
2100484	STAT3	NM_139276.2	1,2	0,00317
60079	DNAJB1	NM_006145.1	0,831	0,00833
3170703	LY9	NM_001033667.1	0,83	0,0178
7610440	XAF1	NM_199139.1	0,827	0,0389
4200475	MAST3	NM_015016.1	0,824	0,0359

ES 2 530 022 T3

*Entradas en Rojo/Negrita/Cursiva => $p < 0,05$ cuando se corrigen para el ensayo múltiple (FDR)

**el transcrito CR743148 (ID de Sonda Illumina 6370082) ha sido retirado de NCBI, pero la EST corresponde a una o varias variantes de corte y empalme dentro del gen GPRIN3 (cromosoma 4.90).

ID de Illumina	Símbolo	RefSeq	FC	p* no corregido
770161	C10orf73	XM_096317.11	0,813	0,00616
870056	FAM119B	NM_015433.2	0,811	0,0225
3610300	CCDC58	NM_001017928.2	0,811	0,0149
5080615	IL16	NM_172217.2	0,806	0,0306
3990170	IFI27	NM_005532.3	0,799	0,0267
6770603	NOG	NM_005450.2	0,779	0,000954
7570324	ID3	NM_002167.2	0,75	0,00109
50706	CD40LG	NM_000074.2	0,739	0,000205
2230538	LRRN3	NM_001099660.1	0,578	0,0308

Lista de genes 4: AR pos para AAPC vs OA, Conjunto de datos reunidos (n=173)

*Entradas en Rojo/Negrita/Cursiva => $p < 0,05$ cuando se corrigen para el ensayo múltiple (FDR)

ID de Illumina	Símbolo	RefSeq	TC	p* no corregido
5080692	HLA-A29.1	NM_001080840.1	1,961	0,0149
430438	MIAT	NR_003491.1	1,546	0,000406
3130301	PIM1	NM_002648.2	1,536	0,000107
6220288	PRDM1	NM_001198.2	1,485	0,000136
4230102	SOCS3	NM_003955.3	1,452	0,000601
6330725	BCL3	NM_005178.2	1,434	0,00186
6280170	PDCD1	NM_005018.1	1,427	1,73E-05
1400601	C20orf100	NM_032883.1	1,381	0,000272
4730523	IGFL2	NM_001002915.1	1,381	1,69E-08
6220195	BATF	NM_006399.2	1,372	0,00029
2070037	ICOS	NM_012092.2	1,37	7,70E-05

ES 2 530 022 T3

*Entradas en Rojo/Negrita/Cursiva => p<0,05 cuando se corrigen para el ensayo múltiple (FDR)				
ID de Illumina	Símbolo	RefSeq	TC	p* no corregido
3190609	SBNO2	NM_014963.2	1,369	0,000182
5090754	KIAA0101	NM_014736.4	1,337	0,000717
5910364	TYMS	NM_001071.1	1,334	0,00037
6620689	MTHFD2	NM_001040409.1	1,332	0,000143
6370082		CR743148	1,331	6,90E-05
1990300	SOCS1	NM_003745.1	1,328	0,0283
1230201	CTLA4	NM_005214.3	1,326	0,000122
60470	STX11	NM_003764.2	1,311	0,00495
2070520	CDCA7	NM_031942.4	1,311	0,00086
3800647	UGCG	NM_003358.1	1,308	0,000353
130022	CDCA5	NM_080668.2	1,305	0,000177
7560731	SNORA64	NR_002326.1	1,3	6,93E-05
5420538	TP53INP1	NM_033285.2	1,29	0,00336
6250010	GPRIN3	NM_198281.2	1,288	0,00228
2570253	BTN3A2	NM_007047.3	1,285	0,017
4060558	ASCA1	NM_005502.2	1,262	0,00522
4260368	UBE2C	NM_181800.1	1,278	0,00038
10333	LOC731682	XM_001129369.1	1,274	0,00834
160292	APOBEC3H	NM_181773.2	1,274	0,0101
4230228	CDK5RAP3	NM_176095.1	1,273	0,0141
5420095	MYC	NM_002467.3	1,273	0,00207
670576	LOC731186	XM_001128760.1	1,272	5,19E-06
1850554	NPDC1	NM_015392.2	1,272	0,000954
3990619	TOP2A	NM_001067.2	1,269	0,000693
5360070	CCNB2	NM_004701.2	1,258	0,000841

ES 2 530 022 T3

*Entradas en Rojo/Negrita/Cursiva => p<0,05 cuando se corrigen para el ensayo múltiple (FDR)				
ID de Illumina	Símbolo	RefSeq	TC	p* no corregido
3840470	ST6GALNAC1	NM_018414.2	1,256	0,00534
3780161	TMEM70	NM_017866.4	1,255	0,000225
520278	FAM100B	NM_182565.2	1,252	0,00317
1430598	FBXO32	NM_058229.2	1,246	0,00115
5890524	LINS1	NM_181740.1	1,246	0,000377
3610440	MAF	NM_005360.3	1,245	0,00787
6350189	MGC4677	NM_052871.3	1,239	0,0214
1500010	CDC20	NM_001255.2	1,236	0,000944
2320170	CDC45L	NM_003504.3	1,236	4,66E-05
5700753	CEACAM1	NM_001024912.1	1,236	0,0226
5340246	CRIP2	NM_001312.2	1,235	0,0179
1410408	ARID5B	NM_032199.1	1,234	0,00269
1690692	SOCS2	NM_003877.3	1,233	0,0309
2470348	NFKBIZ	NM_001005474.1	1,232	0,00421
4880646	FKSG30	NM_001017421.1	1,232	0,0191
1450056	CPA5	NM_080385.3	1,231	0,00564
1500553	NUSAP1	NM_018454.5	1,23	0,00215
1710019	ICA1	NM_004968.2	1,23	0,0016
4730411	SFXN1	NM_022754.4	1,225	0,000836
4810520	TRIB1	NM_025195.2	1,224	0,0324
4890750	DDX11	NM_030653.3	1,224	0,0376
2490161	CLEC2B	NM_005127.2	1,222	0,0378
10414	PTTG1	NM_004219.2	1,219	0,00173
1510364	GBP5	NM_052942.2	1,218	0,0161
380056	B3GNT2	NM_006577.5	1,216	0,000916

ES 2 530 022 T3

*Entradas en Rojo/Negrita/Cursiva => p<0,05 cuando se corrigen para el ensayo múltiple (FDR)				
ID de Illumina	Símbolo	RefSeq	TC	p* no corregido
2940110	UHRF1	NM_001048201.1	1,216	0,00165
3190092	LDHA	NM_005566.1	1,215	9,71E-05
3420241	SLC2A14	NM_153449.2	1,21	0,00873
6100408	NLRC5	NM_032206.3	1,21	0,00475
7570600	FLJ33590	NM_173821.1	1,21	0,012
7650026	MUC1	NM_001044391.1	1,21	0,0017
450615	MT2A	NM_005953.2	1,209	0,00164
1450280	NCAPG	NM_022346.3	1,208	0,000123
160097	MELK	NM_014791.2	1,207	2,95E-05
4730196	TK1	NM_003258.2	1,207	0,000283
5090528	CYorf15B	NM_032576.2	1,207	0,0415
6480053	ATF4	NM_001675.2	1,206	0,0166
4610189	HERPUD1	NM_001010990.1	1,205	0,0217
4830056	ARPC5L	NM_030978.1	1,204	6,82E-05
3800168	SLC2A3	NM_006931.1	1,203	0,0313
5260600	ZNF655	NM_001009957.1	1,203	0,00963
7200301	ARID5A	NM_212481.1	1,202	0,0349
2600735	TLR6	NM_006068.2	1,201	0,0157
2760112	P2RY5	NM_005767.4	1,201	0,0133
2970730	MYADM	NM_001020820.1	0,818	0,0113
3610300	CCDC58	NM_001017928.2	0,815	0,0182
50706	CD40LG	NM_000074.2	0,81	0,00637
6770603	NOG	NM_005450.2	0,802	0,00116
7570324	ID3	NM_002167.2	0,757	8,06E-05
130609	FCGBP	NM_003890.1	0,688	0,00193

ES 2 530 022 T3

*Entradas en Rojo/Negrita/Cursiva => p<0,05 cuando se corrigen para el ensayo múltiple (FDR)				
ID de Illumina	Símbolo	RefSeq	TC	p* no corregido
2230538	LRRN3	NM_001099660.1	0,679	0,0483
5080192	SERPINE2	NM_006216.2	0,628	0,0069
7050021	PRKAR1A	NM_002734.3	0,522	0,00561

Lista de genes 5: Inflam. sin AR vs OA, Conjunto de datos reunidos (n=173)

ID de Illumina	Símbolo	RefSeq	TC	p* no corregido
3610743	SF1	NM_201997.1	1,658	0,0427
6960661	FAM118A	NM_017911.1	1,595	0,0349
1430113	LOC728505	XM_001127580.1	1,366	0,000994
1690440	XIST	NR_001564.1	1,358	0,013
6560376	RPS26L1	NR_002309.1	1,319	0,045
4060358	ABCA1	NM_005502.2	1,307	0,00185
5420095	MYC	NM_002467.3	1,302	0,00377
6960195	LOC650646	XM_942527.2	1,301	0,0432
3710397	EFNA1	NM_004428.2	1,251	0,0055
2570253	BTN3A2	NM_007047.3	1,246	0,0272
1940632	NCAPG2	NM_017760.5	1,238	0,0261
3800647	UGCG	NM_003358.1	1,233	0,00373
6590377	RPS26	NM_001029.3	1,226	0,0457
5490408	CEBPD	NM_005195.3	1,222	0,0436
3130296	AMY2A	NM_000699.2	1,221	0,0485
160370	TPM2	NM_213674.1	1,209	0,0474
3130301	PIM1	NM_002648.2	1,203	0,0173
1440736	LDLR	NM_000527.2	1,2	0,0263
6250010	GPRIN3	NM_198281.2	0,833	0,0147

ES 2 530 022 T3

ID de Illumina	Símbolo	RefSeq	TC	p* no corregido
1260086	ID2	NM_002166.4	0,814	0,0158
5890730	RPS26L	XR_017804.1	0,805	0,0405

Lista de Genes 6 - Únicamente desregulados en AR positiva para AAPC vs OA

Símbolo	RefSeq	TC
HLA-A29.1	NM_001080840.1	1,961
C20orf100	NM_032883.1	1,381
IGFL2	NM_001002915.1	1,381
KIAA0101	NM_014736.4	1,337
TYMS	NM_001071.1	1,334
CDCA7	NM_031942.4	1,311
CDCA5	NM_080668.2	1,305
UBE2C	NM_181800.1	1,278
APOBEC3H	NM_181773.2	1,274
LOC731682	XM_001129369.1	1,274
CDK5RAP3	NM_176095.1	1,273
TOP2A	NM_001067.2	1,269
CCNB2	NM_004701.2	1,258
MGC4677	NM_052871.3	1,239
CDC20	NM_001255.2	1,236
CDC45L	NM_003504.3	1,236
CRIP2	NM_001312.2	1,235
FKSG30	NM_001017421.1	1,232
CPA5	NM_080385.3	1,231
ICA1	NM_004968.2	1,23
NUSAP1	NM_018454.5	1,23

ES 2 530 022 T3

Símbolo	RefSeq	TC
DDX11	NM_030653.3	1,224
CLEC2B	NM_005127.2	1,222
MAF	NM_001031804.1	1,219
PTTG1	NM_004219.2	1,219
GBP5	NM_052942.2	1,218
UHRF1	NM_001048201.1	1,216
FLJ33590	NM_173821.1	1,21
MUC1	NM_001044391.1	1,21
NLRC5	NM_032206.3	1,21
MT2A	NM_005953.2	1,209
NCAPG	NM_022346.3	1,208
CYorf15B	NM_032576.2	1,207
MELK	NM_014791.2	1,207
TK1	NM_003258.2	1,207
ATF4	NM_001675.2	1,206
HERPUD1	NM_001010990.1	1,205
ARPC5L	NM_030978.1	1,204
ZNF655	NM_001009957.1	1,203
MYADM	NM_001020820.1	0,818
FCGBP	NM_003890.1	0,688
SERPINE2	NM_006216.2	0,628
PRKAR1A	NM_002734.3	0,522

Funciones Biológicas: Se facilita la proporción de genes de una lista de genes dada asignada a una función biológica concreta, junto con el valor de p...

ES 2 530 022 T3

Subconjunto "Ciclo celular".	Subconjunto "Cáncer".
24/43 (p<10e-6)	21/43 (p<10e-6)
CCNB2	CCNB2
CDC20	CDC20
CDC45L	CDCA5
COCA5	CDCA7
CDCA7	FCGBP
CDK5RAP3 (incluye EG:80279)	KIAA0101
DDX11	MAF
FCGBP	MELK
KIAA0101	MT2A
MAF	MUC1
MELK	NCAPG (incluye EG:64151)
MT2A	PRKAR1A
MUC1	PTTG1
NCAPG (incluye EG:64151)	SERPINE2
NUSAP1	TK1
PRKAR1A	TOP2A
PTTG1	TP53INP1
SERPINE2	TYMS
TK1	UBE2C
TOP2A	UHRF1
TYMS	UHRF1
UBE2C	
UHRF1	

ES 2 530 022 T3

Subconjunto "Ciclo celular".	Subconjunto "Cáncer".
ZNF655	

Lista de Genes 7 - Únicamente desregulados en AR negativa para AAPC vs OA

Símbolo	RefSeq	TC
HLA-DRB4	NM_021983.4	1,701
HBEGF	NM_001945.1	1,607
MNDA	NM_002432.1	1,558
LOC650298	XM_939387.1	1,555
CDKN1A	NM_000389.2	1,505
CFD	NM_001928.2	1,499
LOC644934	XM_930344.2	1,476
TNFSF10	NM_003810.2	1,404
C6orf105	NM_032744.1	1,399
GPR132	NM_013345.2	1,399
GSTM1	NM_000561.2	1,394
DACT1	NM_016651.5	1,374
GADD45A	NM_001924.2	1,373
CSDA	NM_003651.3	1,369
NCF1	NM_000265.4	1,369
BCL6	NM_001706.2	1,363
RPS26L	XR_017804.1	1,362
AQP9	NM_020980.2	1,349
FAM46C	NM_017709.3	1,334
IFNGR2	NM_005534.2	1,329
SLC7A5	NM_003486.5	1,321
F2RL1	NM_005242.3	1,316

ES 2 530 022 T3

Símbolo	RefSeq	TC
ADM	NM_001124.1	1,313
LOC653853	XM_936029.1	1,309
S100P	NM_005980.2	1,305
AP3M2	NM_006803.2	1,296
CDKN2D	NM_001800.3	1,296
RGS16	NM_002928.2	1,295
DDIT4	NM_019058.2	1,291
TMEM173	NM_198282.1	1,287
TMEM49	NM_030938.2	1,284
PRDX3	NM_006793.2	1,282
IFNGR1	NM_000416.1	1,279
GNG10	NM_001017998.2	1,276
UGP2	NM_006759.3	1,271
MT1E	NM_175617.3	1,27
FKBP5	NM_004117.2	1,269
MSL3L1	NM_078628.1	1,265
SNORA10	NR_002327.1	1,265
FBXO5	NM_012177.2	1,264
GALNAC4S-6ST	NM_015892.2	1,262
SLA	NM_001045556.1	1,262
TMEM140	NM_018295.2	1,261
FURIN	NM_002569.2	1,259
FAM113B	NM_138371.1	1,258
GLRX	NM_002064.1	1,258
CEBPB	NM_005194.2	1,257
PDCL	NM_005388.3	1,253

ES 2 530 022 T3

Símbolo	RefSeq	TC
ELM02	NM_182764.1	1,252
BAG3	NM_004281.3	1,251
HIF1A	NM_001530.2	1,251
NFIL3	NM_005384.2	1,251
REXO2	NM_015523.2	1,251
RYBP	NM_012234.4	1,251
GCA	NM_012198.2	1,249
LY96	NM_015364.2	1,249
LOC145853	XM_096885.9	1,247
SLC20A1	NM_005415.3	1,247
ACVR2A	NM_001616.3	1,245
LMNB1	NM_005573.2	1,245
SDCBP	NM_001007067.1	1,244
DDX21	NM_004728.2	1,243
CLDN5	NM_003277.2	1,241
POLR1C	NM_203290.1	1,239
GNL3	NM_206826.1	1,237
IL21R	NM_181078.1	1,237
PIM2	NM_006875.2	1,237
SERPINB1	NM_030666.2	1,236
MXD1	NM_002357.2	1,235
CLECL1	NM_172004.2	1,232
NAMPT	NM_005746.2	1,232
NDUFV2	NM_021074.1	1,232
ACVR1B	NM_020328.2	1,23
FCGR1B	NM_001017986.1	1,229

ES 2 530 022 T3

Símbolo	RefSeq	TC
EIF4A3	NM_014740.2	1,228
GADD45B	NM_015675.2	1,227
GRAMD4	NM_015124.2	1,227
EDA	NM_001005611.1	1,225
PIM3	XM_938171.2	1,225
CSTF2T	NM_015235.2	1,224
MAT2A	NM_005911.4	1,224
BIRC3	NM_001165.3	1,22
LOC44035	NM_496143.2	1,218
NGRN	NM_016645.2	1,218
SLC25A44	NM_014655.1	1,218
TRIB3	NM_021158.3	1,217
GNPDA1	NM_005471.3	1,215
SOD2	NM_001024466.1	1,215
PTGER2	NM_000956.2	1,213
LGALS8	NM_006499.3	1,212
SMAP2	NM_022733.1	1,212
GIMAP5	NM_018384.3	1,211
FAM89A	NM_198552.1	1,21
GP5	NM_004488.1	1,21
METRNL	NM_001004431	1,21
KISS1R	NM_032551.3	1,209
FLJ43692	NM_001003702.1	1,208
ZNF828	NM_032436.1	1,207
HIAT1	NM_033055.2	1,206
RNF149	NM_173647.2	1,205

ES 2 530 022 T3

Símbolo	RefSeq	TC
GIMAP7	NM_153236.3	1,203
TLR8	NM_016610.2	1,203
PTRH2	NM_016077.3	1,202
STS-1	NM_032873.3	1,202
IL2RA	NM_000417.1	1,2
STAT3	NM_139276.2	1,2
DNAJB1	NM_006145.1	0,831
LY9	NM_001033667.1	0,83
XAF1	NM_199139.1	0,827
MAST3	NM_015016.1	0,824
C10orf73	XM_096317.11	0,813
FAM119B	NM_015433.2	0,811
IL16	NM_172217.2	0,806
IFI27	NM_005532.3	0,799

Lista de Genes 8 - Desregulados en (AR negativa para AAPC y positiva para AAPC) vs OA

Símbolo	RefSeq	AR negativa para AAPC vs OA	AR positiva para AAPC vs OA
SOCS3	NM_003955.3	1,916	1,452
BCL3	NM_005178.2	1,797	1,434
SBNO2	NM_014963.2	1,618	1,369
BATF	NM_006399.2	1,594	1,372
MTHFD2	NM_001040409.1	1,561	1,332
STX11	NM_003764.2	1,553	1,311
SOCS1	NM_003745.1	1,49	1,328
<i>GRIN3*</i>	CR743148	1,457	1,331
ARID5A	NM_212481.1	1,448	1,202

ES 2 530 022 T3

Símbolo	RefSeq	AR negativa para AAPC vs OA	AR positiva para AAPC vs OA
TMEM70	NM_017866.4	1,442	1,255
ICOS	NM_012092.2	1,438	1,37
LOC731186	XM_001128760.1	1,435	1,272
MIAT	NR_003491.1	1,428	1,546
SOCS2	NM_003877.3	1,403	1,233
FAM100B	NM_182565.2	1,352	1,252
FBXO32	NM_058229.2	1,331	1,246
SLC2A3	NM_006931.1	1,329	1,203
ST6GALNAC1	NM_018414.2	1,329	1,256
PRDM1	NM_182907.1	1,328	1,335
CEACAM1	NM_001024912.1	1,321	1,236
TRIB1	NM_025195.2	1,316	1,224
SFXN1	NM_022754.4	1,312	1,225
LDHA	NM_005566.1	1,302	1,215
LINS1	NM_181740.1	1,298	1,246
NFKBIZ	NM_001005474.1	1,281	1,232
SNORA64	NR_002326.1	1,278	1,3
CTLA4	NM_005214.3	1,277	1,326
B3GNT2	NM_006577.5	1,274	1,216
TLR6	NM_006068.2	1,271	1,201
TP53INP1	NM_033285.2	1,266	1,278
ARID5B	NM_032199.1	1,258	1,234
NPDC1	NM_015392.2	1,257	1,272
P2RY5	NM_005767.4	1,247	1,201
PDCD1	NM_005018.1	1,236	1,427
SLC2A14	NM_153449.2	1,211	1,21

ES 2 530 022 T3

Símbolo	RefSeq	AR negativa para AAPC vs OA	AR positiva para AAPC vs OA
CCDC58	NM_001017928.2	0,811	0,815
NOG	NM_005450.2	0,779	0,802
ID3	NM_002167.2	0,75	0,757
CD40LG	NM_000074.2	0,739	0,81
LRRN3	NM_001099660.1	0,578	0,679
*el transcrito CR743148 (ID de Sonda Illumina 6370082) ha sido retirado de NCBI, pero EST corresponde a una o varias variantes de corte y empalme dentro del gen GPRIN3 (cromosoma 4.90).			

Funciones Biológicas: Se facilita la proporción de genes de una lista de genes dada asignada a una función biológica concreta, junto con el valor de p...

Desarrollo celular	Diferenciación de linfocitos T	Desarrollo de linfocitos T
14/40 (p<10e-10)	7/40 (2.6e-7)	9/40 (3.14e-7)
BATF	BCL3	BCL3
BCL3	CD40LG	CD40LG
CD40LG	CTLA4	CTLA4
CEACAM1	ICOS	ICOS
CTLA4	ID3	ID3
ICOS	NOG	NOG
ID3	SOCS3	PDCD1
NOG		SOCS1
PDCD1		SOCS3
PRDM1		
SFXN1		
SOCS1		
SOCS2		
SOCS3		

Lista de Genes 9 Listas de Genes relacionados funcionalmente basándose en el Análisis de la ruta de las Listas 5, 6 y 7 combinadas (n=197) (Únicamente desregulados en AR vs OA, pero no en controles inflamatorios)

5 **Rutas Canónicas. La proporción de genes enumerados en la ruta concreta que aparece se proporciona en cada caso, junto con el valor de p para la significación ...**

Célula T Coadyuvante	Ciclo celular	Señalización de Interferón
Diferenciación		
6/41 (p=2,63e-4)	(reg. punto de control del daño de ADN G2/M) 4/43(p=3.25e-4)	3/30 (p=1.8e-3)
ICA1	CCB2	IFNGR1
IFNGR1	CDKN1A	IFNGR2
IFNGR2	GADD45A	SOCS1
SOCS1	TOP2A	
SOCS2		
SOCS3		
Señalización IL-9	Señalización JAK/STAT	
3/37 (p=2,44e-03)	4/64 (p=1,97e-03)	
BCL3	CDKN1A	
SOCS2	SOCS1	
SOCS3	SOCS2	
	SOCS3	

Funciones Biológicas: Se facilita la proporción de genes de una lista de genes dada asignada a una función biológica concreta, junto con el valor de p...

Muerte celular 97/197 (p=2,31e-23)	Supervivencia celular 79/197 (p=2,97e-7)	Proliferación celular 67/197 (p=2,28e-20)	Proliferación celular 17/197 (p=2,27e-7)
ACVR1B	ACVR2A	ACVR2A	B3GNT2
ACVR2A	ADM	ADM	BATF

ES 2 530 022 T3

Muerte celular 97/197 (p=2,31e-23)	Supervivencia celular 79/197 (p=2,97e-7)	Proliferación celular 67/197 (p=2,28e-20)	Proliferación celular 17/197 (p=2,27e-7)
ADM	AP3M2	ARID5B	CD40LG
AP3M2	ATF4	ATF4	CDKN1A
BAG3	BCL3	B3GNT2	CEACAM1
BCL3	BCL6	BATF	CLECL1
BCL6	CCNB2	BCL3	CTLA4
BIRC3	CD40LG	BCL6	F2RL1
CCNB2	CDC20	CD40LG	GADD45A
CD40LG	CDCA5	CDC45L	ICOS
CDC20	CDCA7	CDCA7	IFNGR1
CDC45L	CDKN1A	CDKN1A	IL2RA
CDCA5	CDKN2D	CDKN2D	PDCD1
CDCA7	CEACAM1	CEACAM1	PRDM1
CDK5RAP3 EG:B0279)	(inct CEBPB	CEBPB	SOCS1
CDKN1A	CFD	CLECL1	SOCS3
CDKN2D	CTLA4	CRIP2	TNFSF10
CEACAM1	DDIT4	CSDA	
CEBPB	FCGBP	CTLA4	
CFD	FKBP5	DDX11	
CSDA	FURIN	DDX21	
CTLA4	GADD45A	F2RL1	
DDIT4	GLRX	FKBP5	
DNAJB1	GPR132	FURIN	
EDA	GSTM1	GADD45A	
FBXO32	HBEGF	GNL3	
FCGBP	HERPUD1	GPR132	

ES 2 530 022 T3

Muerte celular 97/197 (p=2,31e-23)	Supervivencia celular 79/197 (p=2,97e-7)	Proliferación celular 67/197 (p=2,28e-20)	Proliferación celular 17/197 (p=2,27e-7)
FKBP5	HIF1A	GSTM1	
FURIN	HLA-DRB4	HBEGF	
GADD45A	ICOS	HIF1A	
GIMAP5	IFI27	ICOS	
GLRX	IFNGR1	ID3	
GNL3	IFNGR2	IFNGR1	
GPR132	IL2RA	IL16	
GSTM1	KIAA0101	IL21R	
HBEGF	LDHA	IL2RA	
HERPUD1	LGALS8	KIAA0101	
HIF1A	MAF	KISS1R	
HLA-DRB4	MAT2A	LDHA	
ICOS	MELK	LY96	
ID3	MT1E	MT2A	
IFI27	MT2A	MUC1	
IFNGR1	MTHFD2	MXD1	
IFNGR2	MUC1	NAMPT	
IL2RA	MXD1	NCF1	
KIAA0101	NAMPT	NOG	
LDHA	NCAPG (incluye EG:64151)	NPDC1	
LGALS8	NDUFV2	PDCD1	
LMNB1	NFIL3	PIM2 (incluye EG:11040)	
MAF	NOG	PRDM1	
MAT2A	NPDC1	PRDX3	
MELK	PIM2 (incluye EG:11040)	PRKAR1A	
MT1E	PRDX3	PTGER2	

ES 2 530 022 T3

Muerte celular 97/197 (p=2,31e-23)	Supervivencia celular 79/197 (p=2,97e-7)	Proliferación celular 67/197 (p=2,28e-20)	Proliferación celular 17/197 (p=2,27e-7)
MT2A	PRKAR1A	PTTG1	
MTHFD2	PTGER2	S100P	
MUC1	PTTG1	SERPINE2	
MXD1	S100P	SLC7A5	
NAMPT	SDCBP	SOCS1	
NCAPG EG:64151)	(incluye SERPINB1	SOCS2	
NCF1	SERPINE2	SOCS3	
NDUFV2	SLC2A3	SOD2	
NFIL3	SLC2A14	TNFSF10	
NFKBIZ	SLC7A5	TP53INP1	
NOG	SOCS1	TRIB1	
NPDC1	SOCS2	TYMS	
PDCD1	SOCS3	UBE2C	
PIM2 (incluye EG:11040)	SOD2	UHRF1	
PRDM1	TK1		
PRDX3	TNFSF10		
PRKAR1A	TOP2A		
PTGER2	TP53INP1		
PTRH2	TRIB1		
PTTG1	TYMS		
RYBP	UBE2C		
S100P	UHRF1		
SDCBP	XAF1		
SERPINB1			
SERPINE2			

ES 2 530 022 T3

Muerte celular 97/197 ($p=2,31e-23$)	Supervivencia celular 79/197 ($p=2,97e-7$)	Proliferación celular 67/197 ($p=2,28e-20$)	Proliferación celular 17/197 ($p=2,27e-7$)
SLC2A3			
SLC2A14			
SLC7A5			
SOCS1			
SOCS2			
SOCS3			
SOD2			
TK1			
TLR6			
TMEM173			
TNFSF10			
TOP2A			
TP53INP1			
TRIB1			
TRIB3			
TYMS			
UBE2C			
UHRF1			
XAF1			
Diferenciación de células de la sangre 25/197 ($p=4,5e-12$)		Diferenciación células T 15/197 ($p=3,3e-09$)	
ACVR1B		BCL3	
ACVR2A		BCL6	
BCL3		CD40LG	
BCL26		CEBPB	
CD40LG		CTLA4	

ES 2 530 022 T3

Muerte celular 97/197 (p=2,31e-23)	Supervivencia celular 79/197 (p=2,97e-7)	Proliferación celular 67/197 (p=2,28e-20)	Proliferación celular 17/197 (p=2,27e-7)
CDKN2D		GIMAP5	
CEBPB		ICOS	
CTLA4		ID3	
GIMAP5		IFNGR2	
HIF1A		IL21R	
ICOS		IL2RA	
ID3		MAF	
IFNGR2		MUC1	
IL21R		NOG	
IL2RA		SOCS3	
MAF			
MUC1			
NOG			
PDCD1			
PRDM1			
PRDX3			
SFXN1			
SOCS1			
SOCS3			
TNFSF10			

Referencias

1. Klareskog L, Catrina AI, Paget S, Klareskog L, Catrina AI, Paget S. Rheumatoid arthritis. *Lancet* 2009; 373(9664): 659-72.
- 5 2. Combe B, Landewe R, Lukas C, Bolosiu HD, Breedveld F, Dougados M, et al. EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT).[véase el comentario]. *Annals of the Rheumatic Diseases* 2007;66(1):34-45.
- 10 3. van Gaalen FA, Linn-Rasker SP, van Venrooij WJ, de Jong BA, Breedveld FC, Verweij CL, et al. Autoantibodies to cyclic citrullinated peptides predict progression to rheumatoid arthritis in patients with undifferentiated arthritis: a prospective cohort study. *Arthritis & Rheumatism* 2004;50(3):709-15.
- 15 4. van Der Helm-van Mil AHM, Detert J, Cessie SL, Filer A, Bastian H, Burmester GR, et al. Validation of a prediction rule for disease outcome in patients with recent-onset undifferentiated arthritis: Moving toward individualized treatment decision-making. *Arthritis & Rheumatism* 2008;58(8):2241-7.
- 5 15 5. Nishimura K, Sugiyama D, Kogata Y, Tsuji G, Nakazawa T, Kawano S, et al. Meta-analysis: diagnostic accuracy of anti-cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis.[véase el comentario]. *Annals of Internal Medicine* 2007;146(11):797-808.
- 20 6. Pratt AG, Isaacs JD, Wilson G. The clinical utility of a rule for predicting rheumatoid arthritis in patients with early undifferentiated arthritis: comment on the article by van der Helm-van Mil et al. [comentario]. *Arthritis & Rheumatism* 2009;60(3):905. respuesta del autor 906.
- 25 7. van't Veer LJ, Bernards R, van't Veer LJ, Bernards R. Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature* 2008; 452(7187): 564-70.
8. Pascual V, Chaussabel D, Banchereau J. A genomic approach to human autoimmune diseases. *Annual Review of Immunology* 2010;28:535-71.
9. Toonen EJ, Barrera P, Radstake TR, van Riel PL, Scheffer H, Franke B, et al. Gene expression profiling in rheumatoid arthritis: current concepts and future directions. *Annals of the Rheumatic Diseases* 2008;67(12):1663-9.
10. Lequerre T, Gauthier-Jauneau A-C, Bansard C, Derambure C, Hiron M, Vittecoq O, et al. Gene profiling in white blood cells predicts infliximab responsiveness in rheumatoid arthritis. *Arthritis Research & Therapy* 2006;8(4):R105.
11. McKinney EF, Lyons PA, Carr EJ, Hollis JL, Jayne DR, Willcocks LC, et al. A CD8+ T cell transcription signature predicts prognosis in autoimmune disease. *Nature Medicine* 2010; 16(5): 586-91.
12. van Baarsen LGM, Bos WH, Rustenburg F, van der Pouw Kraan TCTM, Wolbink GJJ, Dijkmans BAC, et al. Gene expression profiling in autoantibody-positive patients with arthralgia predicts development of arthritis. *Arthritis & Rheumatism* 2010;62(3):694-704.
13. Batliwalla FM, Baechler EC, Xiao X, Li W, Balasubramanian S, Khalili H, et al. Peripheral blood gene expression profiling in rheumatoid arthritis. *Genes & Immunity* 2005;6(5):388-97.
14. Lyons PA, Koukoulaki M, Hatton A, Doggett K, Woffendin HB, Chaudhry AN, et al. Microarray analysis of human leucocyte subsets: the advantages of positive selection and rapid purification. *BMC Genomics* 2007;8:64.
15. Koetz K, Bryl E, Spickschen K, O'Fallon WM, Goronzy JJ, Weyand CM. T cell homeostasis in patients with rheumatoid arthritis. *Proceedings of the National Academy of Sciences of the United States of America* 2000;97(16):9203-8.
16. Ponchel F, Morgan AW, Bingham SJ, Quinn M, Such M, Verburg RJ, et al. Dysregulated lymphocyte proliferation and differentiation in patients with rheumatoid arthritis. *Blood* 2002; 100(13): 4550-6.
17. McInnes IB, O'Dell JR. State-of-the-art: rheumatoid arthritis. *Annals of the Rheumatic Diseases* 2010;69(11):1898-906.
18. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis & Rheumatism* 1988;31(3):315-24.
19. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology* 2006;7:3.

20. de Jager W, Prakken BJ, Bijlsma JWW, Kuis W, Rijkers GT. Improved multiplex immunoassay performance in human plasma and synovial fluid following removal of interfering heterophilic antibodies. *Journal of Immunological Methods* 2005;300(1-2):124-35.
- 5 21. Hueber W, Tomooka BH, Zhao X, Kidd BA, Drijfhout JW, Fries JF, et al. Proteomic analysis of secreted proteins in early rheumatoid arthritis: anti-citrulline autoreactivity is associated with up regulation of proinflammatory cytokines. *Annals of the Rheumatic Diseases* 2007;66(6):712-9.
22. Livak KJ, Schmittgen TD, Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (Duluth)* 2001; 25(4): 402-8.
- 10 23. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007; 8(1): 118-27.
24. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* 2008; 24(13): 1547-8.
25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 1995;57(1):289-300.
- 15 26. Cortes C, Vapnik V. Support-vector networks. *Machine Learning* 1995; 20(3): 273-97.
27. Owaki T, Asakawa M, Morishima N, Mizoguchi I, Fukai F, Takeda K, et al. STAT3 is indispensable to IL-27-mediated cell proliferation but not to IL-27-induced Th1 differentiation and suppression of proinflammatory cytokine production. *Journal of Immunology* 2008;180(5):2903-11.
- 20 28. Starr R, Willson TA, Viney EM, Murray LJ, Rayner JR, Jenkins BJ, et al. A family of cytokine-inducible inhibitors of signalling. *Nature* 1997; 387(6636): 917-21.
29. El Kasmī KC, Smith AM, Williams L, Neale G, Panopoulos AD, Watowich SS, et al. Cutting edge: A transcriptional repressor and corepressor induced by the STAT3-regulated anti-inflammatory signaling pathway.[Aparace una fe de erratas en *J Immunol.* 2008 Mar 1;180(5):3612 Nota: Panopolous, Athanasia [corregido para Panopoulos, Athanasia D]]. *Journal of Immunology* 2007;179(11):7215-9.
- 25 30. Brocke-Heidrich K, Ge B, Cvijic H, Pfeifer G, Loffler D, Henze C, et al. BCL3 is induced by IL-6 via Stat3 binding to intronic enhancer HS4 and represses its own transcription. *Oncogene* 2006; 25(55): 7297-304.
31. Richard M, Louahed J, Demoulin JB, Renaud JC, Richard M, Louahed J, et al. Interleukin-9 regulates NF-kappaB activity through BCL3 gene induction. *Blood* 1999; 93(12): 4318-27.
- 30 32. Gao J, McConnell MJ, Yu B, Li J, Balko JM, Black EP, et al. MUC1 is a downstream target of STAT3 and regulates lung cancer cell survival and invasion. *International Journal of Oncology* 2009;35(2):337-45.
33. Akaishi H, Takeda K, Kaisho T, Shineha R, Satomi S, Takeda J, et al. Defective IL-2-mediated IL-2 receptor alpha chain expression in Stat3-deficient T lymphocytes. *Nature Medicine* 1998; 10(11): 1747-51.
- 35 34. Matikainen S, Sareneva T, Ronni T, Lehtonen A, Koskinen PJ, Julkunen I, et al. Interferon-alpha activates multiple STAT proteins and upregulates proliferation-associated IL-2Ralpha, c-myc, and pim-1 genes in human T cells. *Blood* 1999; 93(6): 1980-91.
35. Nichane M, Ren X, Bellefroid EJ. Self-regulation of Stat3 activity coordinates cell-cycle progression and neural crest specification. *EMBO Journal*;29(1):55-67.
- 40 36. Hirano T, Ishihara K, Hibi M, Hirano T, Ishihara K, Hibi M. Roles of STAT3 in mediating the cell growth, differentiation and survival signals relayed through the IL-6 family of cytokine receptors. *Oncogene* 2000; 19(21): 2548-56.
37. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. *BMJ* 1994/309 (6947):102.
38. Goronzy JJ, Weyand CM. rheumatoid arthritis. *Immunological Reviews* 2005;204:55-73.
- 45 39. Schonland SO, Lopez C, Widmann T, Zimmer J, Bryl E, Goronzy JJ, et al. Premature telomeric loss in rheumatoid arthritis is genetically determined and involves both myeloid and lymphoid cell lineages. *Proceedings of the National Academy of Sciences of the United States of America* 2003; 100(23): 13471-6.
40. van der Helm-van Mil AHM, Verpoort KN, Breedveld FC, Huizinga TWJ, Toes REM, de Vries RRP. The HLA-DRB1 shared epitope alleles are primarily a risk factor for anti-cyclic citrullinated peptide antibodies and are not an independent risk factor for development of rheumatoid arthritis. *Arthritis & Rheumatism* 2006;54(4):1117-21.

41. Kokkonen H, Soderstrom I, Rocklov J, Hallmans G, Lejon K, Rantapaa Dahlqvist S. Up-regulation of cytokines and chemokines predates the onset of rheumatoid arthritis. *Arthritis & Rheumatism* ;62(2):383-91.
- 5 42. Karlson EW, Chibnik LB, Tworoger SS, Lee IM, Buring JE, Shadick NA, et al. Biomarkers of inflammation and development of rheumatoid arthritis in women from two prospective cohort studies. *Arthritis & Rheumatism* 2009;60(3):641-52.
43. Schindler CW. Series introduction. JAK-STAT signaling in human disease. *Journal of Clinical Investigation* 2002;109(9):1133-7.
44. Fonseca JE, Santos MJ, Canhao H, Choy E. Interleukin-6 as a key player in systemic inflammation and joint destruction. *Autoimmunity Reviews* 2009;8(7):538-42.
- 10 45. Nowell MA, Williams AS, Carty SA, Scheller J, Hayes AJ, Jones GW, et al. Therapeutic targeting of IL-6 trans signaling counteracts STAT3 control of experimental inflammatory arthritis. *Journal of Immunology* 2009; 182(1):613, -22.
46. Eyles JL, Hickey MJ, Norman MU, Croker BA, Roberts AW, Drake SF, et al. A key role for G-CSF-induced neutrophil production and trafficking during inflammatory arthritis. *Blood* 2008; 112(13): 5193-201.
- 15 47. Rho YH, Solus J, Sokka T, Oeser A, Chung CP, Gebretsadik T, et al. Adipocytokines are associated with radiographic joint damage in rheumatoid arthritis. *Arthritis & Rheumatism* 2009;60(7):1906-14.
48. El Kasmī KC, Smith AM, Williams L, Neale G, Panopoulos AD, Watowich SS, et al. Cutting edge: A transcriptional repressor and corepressor induced by the STAT3-regulated anti-inflammatory signaling pathway.[Aparece un fe de erratas en *J Immunol.* 2008 Mar 1;180(5):3612 Nota: Panopolous, Athanasia [corregido para Panopoulos, Athanasia D]]. *Journal of Immunology* 2007;179(11):7215-9.
- 20 49. Nishimoto N, Miyasaka N, Yamamoto K, Kawai S, Takeuchi T, Azuma J. Long-term safety and efficacy of tocilizumab, an anti-IL-6 receptor monoclonal antibody, in monotherapy, in patients with rheumatoid arthritis (the STREAM study): evidence of safety and efficacy in a 5-year extension study. *Annals of the Rheumatic Diseases* 2009;68(10):1580-4.
- 25 50. Cohen S, Fleischmann R. Kinase inhibitors: a new approach to rheumatoid arthritis treatment. *Current Opinion in Rheumatology* 2010;22(3):330-5.
51. Li J. et al. Altered microRNA expression profile with miR-146a upregulated in CD4+ Tcells from patients with rheumatoid arthritis. *Arthritis Research & Therapy* 2010;12::R81.

LISTADO DE SECUENCIAS

<110> Universidad de Newcastle Upon Tyne

5 120Firma genética de células T CD4+ para la Artritis Reumatoide AR.

<130> 119460.WO.01

<151> 2011-02-14

10 <151> 2011-02-14

<150> GB1108818.14

<151> 2011-05-25

<160> 13

15 <170> PatentIn versión 3.5

<210> 1

20 <211> 1813

<212> ADN

<213> Homo sapiens

<300>

25 <308> NM_005178.2

<309> 2007-01-28

<313> (1)..(1813)

<400> 1

ccgtcccggg	cggcccccattg	ccccgatgcc	ccgcggggggc	catggacgag	gggcccgtgg	60
acctgcgcac	ccggcccaag	gccgcccggac	tcccggggcgc	cgcgctgccg	ctccgcaagc	120
gcccgtctgg	cgcgcctctc	ccggagcccg	ccgctccccg	cggcgtctgg	ggccttgtctg	180
tccccctgga	ccctctgcgc	ggcggctctg	acctgccggc	ggccccggg	cccccccacg	240
gcctggcccg	gccggaggcg	ctttactacc	ccggagcctt	actgcctttg	taccccactc	300
gggccatggg	ctccccgttt	cctctggtga	acctgcctac	accctatac	cccatgatgt	360
gccccatgga	acacccccctt	tctgctgaca	tgcctatggc	cacccgtgca	gatgaggacg	420
gagacacgcc	tctccatatt	gctgtggtgc	agggtaacct	gccagctgtg	caccggctgg	480
tcaacctctt	ccagcagggg	ggccggggagc	tgcacatcta	caacaacctc	cggcagacac	540
cgtccacctt	ggctgtgatc	accacattac	ogtctgtggt	ccggtcctctg	gtgacagctg	600
gtgccagccc	catggcgctg	gaccgccatg	gccagacggc	cgtcacctctg	gcgtgcgagc	660
accgcagccc	gacctgcctg	cagaccctgc	tggacagcgc	agctccgggc	acgttggacc	720
tggaggcccg	caattatgac	gggctcaccg	ccctgcacgt	ggcagtgaac	accgagtgcc	780
aagaaaccgt	gcagctcttg	ctagagcgcg	gtgccgacat	cgcgcagctg	gacattaaga	840
gcggccgctc	cccgtctatc	cacgccgtgg	aaaacaacag	ccttagcatg	gtgcagctgc	900
tgctgcagca	cggcgccaac	gtgaacgcgc	aatgtactc	cggcagctcc	gcctgcact	960
cagcgtccgg	ccgcgggctc	ctcccgtctg	tgcgcacgct	ggccccgagc	ggcgtgaca	1020

ES 2 530 022 T3

gcagcctcaa gaactgccac aacgacacgc cgctcatggg ggcgcgcagc cgcaggggtca 1080
 tcgacatcct gagggggaag gccaccggc ctgcttccac ctcccagcca gaccctccc 1140
 ctgaccggag cgccaacacc tccccgaga gcagcagccg cctcagctcc aatggctctc 1200
 tctccgcata accatcctcc tcaccctccc agtctcccc cagggacccc cctggattcc 1260
 ccattggtcc tcccaatttc ttcttctctt ccccatctcc acccgcttc ctgcccttg 1320
 ctggggctct ccgagggcct ggcggcccg tgccccctc cccagctcca ggaggcagct 1380
 gaggggatg gggggcaga tcttgactc atgaggagg gccccctgc ccagaggggt 1440
 caacccttct ggaaactgtg aagatctgac ttgccccc ccccccca tcttcgggac 1500
 caggatttgc acagaagcac atgcacctac ccatacccc cctcttctga gcacagatgt 1560
 tccccatct cgtccctcc caggactctg accccagcat tctcaggcac cagtccctgt 1620
 ccggaatgcc acccacatct tccatttcca tgtccctcc cagagctggg ggaccaggg 1680
 aacagcact cccctccact ctctaccaga taactgagga ggggagagg gggccgtaac 1740
 gggcacgat cacgatgtaa attattaagc attttggtg gatttctttt gtaataaact 1800
 attttgtac cat 1813

<210> 2
 <211> 2746
 <212> ADN
 <213> Homo sapiens

5

<400> 2
 ggctccgact tggactccct gctccgctgc tgccgcttcg gccccgcacg cagccagccg 60
 ccagccgccc gcccgccca gctcccgccg eggcccttg ccgcggtccc tctcctggtc 120
 ccctcccggg tggccgggg gtggcaggg ggcaggcgg gcgcccaggg gaagctcgag 180
 ggacgcgcgc gcgaaggctc ctttgtggac ttcacggccg ccaacatctg ggcgagcgc 240
 gggccaccgc tggcgtctc gccgcccgt cgccttgggg acccgagggg gctcagcccc 300
 aaggacggag acttcgattc gggaccagcc ccccggatg cggtagcggc cgtgtgctgg 360
 aggccgcgaa gcagctgcag ccgcccgcgc gcagatccac gctggctccg tgcgccatgg 420
 tcaccacag caagtttccc gccgcccgga tgagccgccc cctggacacc agcctgcgcc 480
 tcaagacctt cagctccaag agcgagtacc agctgggtgt gaacgcagtg cgcaagctgc 540
 aggagagcgg cttctactgg agcgcagtga ccggcggcga ggcgaacctg ctgctcagtg 600
 ccgagcccgc cggcaccttt ctgatccgcy acagctcgga ccagcggcac ttcttcacgc 660
 tcagcgtcaa gaccagctct gggaccaaga acctgcgcat ccagtgtgag ggggagcgt 720
 tctctctgca gagcgatccc cggagcacgc agcccgtgcc ccgcttcgac tgcgtgctca 780
 agctggtgca ccactacatg ccgccccctg gagccccctc cttccctcg ccacctactg 840

ES 2 530 022 T3

aaccctcctc cgaggtgccc gagcagccgt ctgccagccc actcctggg agtccccca 900
gaagagccta ttacatctac tccggggcg agaagatccc cctggtgttg agcgggcccc 960
tctcctccaa cgtggccact cttcagcadc tctgtcggaa gaccgtcaac ggccacctgg 1020
actcctatga gaaagtcacc cagctgccgg ggccattcgg ggagttcctg gaccagtacg 1080
atgccccgct ttaaggggta aagggcgcaa agggcatggg tcgggagagg ggacgcaggc 1140
ccctctcctc cgtggcacat ggcacaagca caagaagcca accaggagag agtcctgtag 1200
ctctgggggg aaagagggcg gacaggcccc tccctctgcc ctctccctgc agaatgtggc 1260
aggcggacct ggaatgtgtt ggaggggaagg gggagtacca cctgagtctc cagcttctcc 1320
ggaggagcca gctgtcctgg tgggacgata gcaaccacaa gtggattctc cttaattcc 1380
tcagcttccc ctctgcctcc aaacagggga cacttcggga atgctgaact aatgagaact 1440
gccaggggat cttcaaaact tccaacggaa cttgtttgct ctttgatttg gtttaaacct 1500
gagctgggtg tggagcctgg gaaaggtgga agagagagag gtcctgaggg cccagggct 1560
gcgggctggc gaagggaaatg gtcacacccc ccgccaccc caggcgagga tcctggtgac 1620
atgctcctct ccctggctcc ggggagaagg gcttggggtg acctgaaggg aacctcctg 1680
gtaccccaca tctctcctc cgggacagtc accgaaaaca caggttccaa agtctacctg 1740
gtgcctgaga gccagggcc ctctcctcgt ttttaaggggg aagcaacatt tggaggggat 1800
ggatgggctg gtcagctggt ctcttttcc tactcatact ataccttctt gtacctgggt 1860
ggatggagcg ggaggatgga ggagacggga catctttcac ctcaggctcc tggtagagaa 1920
gacaggggat tctactctgt gcctcctgac tatgtctggc taagagattc gccttaaatg 1980
ctcctgtcc catggagagg gaccagcat aggaaagcca catactcagc ctggatgggt 2040
ggagaggctg agggactcac tggagggcac caagccagcc cacagccagg gaagtgggga 2100
ggggggcg aaaccatgc ctcccagctg agcactggga atgtcagccc agtaagtatt 2160
ggccagtcag gcgcctcgtg gtcagagcag agccaccagg tcccactgcc ccgagccctg 2220
cacagccctc cctcctgcct ggggtgggga ggctggaggt cattggagag gctggactgc 2280
tgccaccccg ggtgctcccg ctctgccata gcactgatca gtgacaattt acaggaatgt 2340
agcagcgatg gaattacctg gaacagtttt ttgtttttgt ttttgttttt gttttgtgg 2400
gggggggcaa ctaaacaaac acaaagtatt ctgtgtcagg tattgggctg gacagggcag 2460
ttgtgtgttg ggggtgtttt tttctctatt tttttgttg tttctgttt ttttaataatg 2520
tttacaatct gcctcaatca ctctgtcttt tataaagatt ccacctccag tcctctctcc 2580
tccccctac tcaggccctt gaggtatta ggagatgctt gaagaactca acaaaatccc 2640
aatccaagtc aaactttgca catatttata tttatattca gaaaagaaac atttcagtaa 2700
tttataataa agagcactat tttttaatga aaaaaaaaa aaaaaa 2746

<210> 3
<211> 2684
<212> ADN
<213> Homo sapiens

5

<400> 3

ES 2 530 022 T3

cccgagagga gtcggtggca gcggcggcgg cgggaccggc agcagcagca gcagcagcag 60
 cagcaaccac tagcctcctg ccccgcgggc ctgcccgcacg agccccacga gccgctcacc 120
 ccgccgttct cagcgtgccc cgaccccgtt ggcgcgcctt cccgccgcca gtcccggcag 180
 cgcctcagt tgtcctccga ctgcacctcg gccttccggc ccagccgcag ccacagccgc 240
 aacgccaccc gcagccacag ccacagccac agccccaggc atagccttcg gcacagcccc 300
 ggctccggct cctgcggcag ctctctgagg caccgtccct gcgccgacat cctggagggt 360
 gggatgctct tgtccaaaat caactcgctt gccacactgc gcgccgcgcc ctgcaacgac 420
 ctgcacgcca ccaagctggc gcccggaag gagaaggagc ccctggagtc gcagtaccag 480
 gtgggcccgc tactgggcag cggcggcttc ggctcggctt actcaggcat ccgctctcc 540
 gacaacttgc cggtgccat caaacacgtg gagaaggacc ggatttccga ctggggagag 600
 ctgcctaagt gcaactcagt gcccatgaa gtggctcctgc tgaagaaggc gagctcgggt 660
 ttctccggcg tcattaggct cctggactgg ttcgagaggc ccgacagttt cgtcctgatc 720
 ctggagaggc ccgagccggc gcaagatctc ttgacttca tcacggaaag gggagccctg 780
 caagaggagc tggcccgcag cttcttctgg cagggtctgg aggccgtgcg gcaactgccac 840
 aactgcgggg tgctccaccg cgacatcaag gacgaaaaca tccttatcga cctcaatcgc 900
 ggcgagctca agctcatcga ctccgggtcg ggggcgctgc tcaaggacac cgtctacacg 960
 gacttcgatg ggacccgagt gtatagccct ccagagtgga tccgctacca tcgctaccat 1020
 ggcaggtcgg cggcagctctg gtccctgggg atcctgctgt atgatatggt gtgtggagat 1080
 attcctttcg agcatgacga agagatcctc aggggccagg ttttcttcag gcagagggtc 1140
 tcttcagaat gtcagcatct cattagatgg tgcttgcccc tgagaccatc agataggcca 1200
 acctcgaag aatccagaa ccatccatgg atgcaagatg ttctcctgcc ccaggaaact 1260
 gctgagatcc acctccacag cctgtcggcg gggcccagca aatagcagcc tttctggcag 1320
 gtcctcccct ctctgtcag atgcccgagg gagggaagc ttctgtctcc agcttcccga 1380
 gtaccagtga cacgtctcgc caagcaggac agtgcttgat acaggaacaa catttacaac 1440
 tcattccaga tcccaggccc ctggaggctg cctcccaaca gtggggaaga gtgactctcc 1500
 aggggtccta ggctcaact cctcccatag atactctctt cttctcatag gtgtccagca 1560
 ttgctggact ctgaaatc cgggggtgg ggggtggggg tgggtcagaa ccctgccatg 1620
 gaactgtttt cttcatcatg agttctgctg aatgcccgca tgggtcaggt aggggggaaa 1680

ES 2 530 022 T3

caggttggga tgggatagga ctagcaccat ttttaagtccc tgtcacctct tccgactctt 1740
 tctgagtgcc ttctgtgggg actccggctg tgctgggaga aatacttgaa cttgcctctt 1800
 ttacctgetg cttctccaaa aatctgcctg ggttttgttc cctatttttc tctcctgtcc 1860
 tccctcacc cctccttcat atgaaagtg ccatggaaga ggctacaggg ccaaacgctg 1920
 agccacctgc ccttttttct gcctccttta gtaaaactcc gagtgaactg gtcttccttt 1980
 ttggttttta cttactgtt tcaaagccaa gacctcacac acacaaaaaa tgcacaaaca 2040
 atgcaatcaa cagaaaagct gtaaagtgt gtacagttgg catggtagta tacaaaaaga 2100
 ttgtagtggga tctaattttt aagaaatttt gcctttaagt tattttacct gtttttgttt 2160
 cttgttttga aagatgcgca ttctaacctg gaggtcaatg ttatgtattt atttatttat 2220
 ttatttggtt cccttctat tccaagcttc catagctgct gccctagttt tctttcctcc 2280
 tttcctcctc tgacttgggg accttttggg ggaggctgc gacgcttgc ctgtttgtgg 2340
 ggtgacggga ctcaggggg acagtgtgc agctccctgg cttctgtggg gccctcacc 2400
 tacttaccca ggtgggtccc ggctctgtgg gtgatgggga ggggcattgc tgactgtgta 2460
 tataggataa ttatgaaaag cagttctgga tgggtgtcct tccagatcct ctctggggct 2520
 gtgttttgag cagcaggtag cctgctggtt ttatctgagt gaaatactgt acaggggaat 2580
 aaaagagatc ttattttttt ttttatactt ggcgtttttt gaataaaaac cttttgtctt 2640
 aaaaaaaaa aaaaaaaaa aaaaaaaaa aaaaaaaaa aaaa 2684

<210> 4
 <211> 4948
 <212> ADN
 <213> Homo sapiens

5

<400> 4
 gtgcgcgggg cccccaagat ggcgccgaa acccggaagt gagcggcggc agctgcgagg 60
 ctcgagaaa caggcgcgc gggctccgc cccggccga cccgggccg agatcatgat 120
 gctgccgcca ccgccccac cacggagcga gaagcccaga tagacgccc gccggccccg 180
 ggtcctggag tcccgccgc tgctgcccgg ccgagagacc caccctcct gccgcccgat 240
 gcttgacagt gggcccgcca tggacagga ttaccgcag catgaacccc cgcggcggg 300
 cagcctcctg tacagcccgc gccctgca gagcgcctg ctgcactgcc cctactggaa 360
 cacctctctg ctgccgcat accctgcct ctccagcag agccgcccgt tcatgagctc 420
 cgcctccttc ctggcagcc agccctgcc agacaccagc tatgccccg tggccaccgc 480
 ctccagcttg ccaccaaga cctgcgactt tgctcaggac tcctcctatt ttgaggactt 540
 ctccaacatc tccatcttct cctcgtccgt ggactccctg tcggacatcg tggacacgcc 600
 cgacttcctg ccggctgaca gcctcaacca ggtgtccacc atctgggacg ataaccctgc 660

ES 2 530 022 T3

cccctccacc cacgataagc tgttccagct cagcaggccg tttgcaggct togaggactt 720
 tctgccctcc cacagcacc cgttctcgt cagctaccag gagcagagtg tgcagagcca 780
 gccagaggag gaggacgagg ctgaggagga ggaggcggag gagctggggc acacagagac 840
 ctacgccgac tacgtgccgt ccaagtccaa gatcgggaag cagcaccag accgcgtggt 900
 ggagaccagc aactgtcca gcgtcccacc cccagacatc acctacacc tggcctgcc 960
 ctcgacagc gggccctgt ctgccctgca gctagaggcc atcaacctac cctgccagca 1020
 acacgaggtc ctgctcccca gcgggcagcg cgcgggcttt ctcatcggcg atggggccgg 1080
 cgtgggcaaa ggcgggagc tggcggagt catcctggag aaccacctgc ggggcccga 1140
 gaaagcattg tggttcagcg tctccaacga cctcaagtac gatgaggagc ggcacctgcg 1200
 ggacatcgaa gccacgggca tcgagggtca cgcgctcagc aagatcaagt acgggtgacac 1260
 cactacctca gagggcgtcc tcttcgccac ctactccgcc ctgattgggg agagccaggc 1320
 cggcggccag caccgcactc gcctccgca gatcctggac tgggtgtgggg aggccttcga 1380
 gggcgtcatc gtgttcgacg agtgtcacia agccaagaat gccggctcca ccaagatggg 1440
 caaggctgtg ctagacctgc agaacaagct gccctggcc cgcgtggtct acgccagcgc 1500
 cacaggtgcc tctgagcctc ggaacatgat ctacatgagc cgcttgggtc tctggggcga 1560
 gggcacacc ttccggaact ttgaggagt cctgcacgcc atcgagaaga gggcgttgg 1620
 cgccatggag atcgtggcca tggacatgaa ggtcagcggc atgtacatcg caccgagct 1680
 cagcttctcc ggcgtcacct tccgcatcga ggagatcccg ctggccccag ccttcgagtg 1740
 cgtctacaac cgcggggccc tgcgtgggc cgaggccctg aacgtgttcc agcagcggc 1800
 cgactggatc ggcctggagt cgcgcaagtc cctgtggggc cagttctggt cggcacacca 1860
 gcgcttcttc aagtatctgt gcacgcagc caagggtgcg cggtggtgg agctggcccg 1920
 agaggagctg gcgcgagaca agtgcgtggt catcgggctg cagtccacgg gcgagcgcg 1980
 caccggggag gtgctggggg agaacgatgg gcacctcaac tgcttcgtct cggccgctga 2040
 aggcgtgttc ctgtcgctaa ttcagaagca cttccgctcc accaagagaa agcgggacag 2100
 aggagcgggc agcaagcggg aacggcgacc tcggggacgc ggggcaaaag cccccggct 2160
 ggcgtgagag acagcgggag tcatccgcat cagtgcagc agcagcacgg agtcggacc 2220
 tggcctggac agcgaactca actcctccc cgagtcctg gtggatgacg acgttgcac 2280
 cgttgatgca gtcgggctcc ccagtgcagc ccggggaccc ctgtgcctcc tgcagagaga 2340
 cccgcatggc cccggggtcc tggagcgggt ggagcggctg aagcaggatc tgctggacaa 2400
 agtgcggcgg ctgggcccgg aactgccagt caacacctg gacgagctca tcgaccagct 2460
 gggcggcccc cagcgggtgg cggagatgac cggcaggaaa ggcgcgtgg tgtccaggcc 2520

ES 2 530 022 T3

cgacgggacg gtggccttcg agtcgcgggc agagcagggc ctgtccatcg accacgtgaa 2580
 cctcagggag aagcagcgct tcatgagcgg cgagaagctc gtggccatca tctcggaggc 2640
 ctccagctcg ggtgtctccc tccaagccga ccgccgtgtc cagaaccagc ggcgcccgct 2700
 gcacatgacc ttggagctgc cgtggagcgc cgaccgcgcc atccagcagt tcggccgcac 2760
 ccaccgggtc aaccaggtct ccgcgccaga gtatgtcttc ctcatctcgg agctggccgg 2820
 ggagcgcggg ttgcctcca tcgtggccaa gcgcctggag agtctggggg ccctgaccca 2880
 cggagaccgc cgcccaagg agtcccgtga cctcagcaag tacaactttg agaacaagta 2940
 tggcaccocg gccctgcact gtgtcctcac caccatcctg agccagactg agaacaagt 3000
 gcctgtgccc cagggatacc ctggaggggt ccccacctc ttccgggaca tgaagcaggg 3060
 cctgtgtct gtgggcattg gtggccggga gtcccgaat ggctgcctgg acgtggagaa 3120
 ggactgttcc atcaccaagt tcctgaaccg catcctgggg ctggaggtgc acaagcagaa 3180
 cgcctgttc cagtacttct cagacacctt cgaccacctc atcgagatgg acaagcggga 3240
 gggcaatac gacatgggca tcctggacct tgctcccgtt atcgaggaga tctacgagga 3300
 gagccagcag gtgttctctg ctcccgggca cccgcaggac gggcaggtgg tcttctaaa 3360
 gatcagcgtg gaccgcggcc tgaagtggga ggacgcctt gcccaagtcgc tggcctgac 3420
 gggcccctat gacggcttct acctctcta caaggtccgc ggtaacaagc ccagctgcct 3480
 gctggcggag cagaaccgcg gccagttctt cacggtgtac aagcccaaca tcggccggca 3540
 gagccagctg gaggccctgg acagcctccg ccgcaagttc caccgggtca ccgcggagga 3600
 ggccaaggag ccctgggaga gtggctacgc tttgtcgtg acgcactgca gccacagcgc 3660
 ctggaaccgg cactgcgggc tggcgcagga gggtaaggac tgctcagag ggctgcggct 3720
 gcggcaaccac tacatgctgt gcggcgcgct gctgcgcgtg tggggccgca tcgcccgct 3780
 catggccgac gtcagcagca gcagctacct gcagatcgtg cggctgaaga ccaaggacag 3840
 gaagaagcaa gtgggcatca agatccccga gggctgcgtg cgcggggtgc tgcaggagct 3900
 gcggctgatg gatgcggacg tgaagcgcag gcaggcgcgc gccctgggct gcccccccc 3960
 gcccccccc cgcccgcctg cgctgcctt cgccccggga gaggtgctgg acctcaacta 4020
 cagccccccg gccgaggcct tcccgcgcc cccgcacttc tcttcccgg cgcctgtgc 4080
 cctgacgcc ggccccggcg tcgtgccgtt gggcaccccc gacgccagg ccgacctgc 4140
 ggccctcgcg caccagggct gcgacatcaa cttcaaggag gtgctggagg acatgctgcg 4200
 ctcctgcac gcggggccgc cctccgaggc cgcctgggg gagggcgcgg gggcgggggg 4260
 cgcggcgggc ggtggtccc agcggcagag cgtgatccag ttcagcccac ccttccccgg 4320
 cgcccaggct cctctctgac acgccttag gcgaaacatg cccaagaca cagggaccgt 4380
 ttctccccta ggagcagcgg tggggagcag ggccaaggtc ccctgaccac tgctcagagg 4440

ES 2 530 022 T3

agccctaggg cctggccgca gtgccttcag cgccccaccc gggccccccac ctggtcagcc 4500
ctggcggggc ccactoagga cagctggggg ccggggcggtg gcagggccct ctctgtgcct 4560
ctcctcccaa gtaggaaggg gctccgggtg gctgctctgg gactgggcac ccacaagggc 4620
tcagtgggoc caaaccttg aatccgtga aaccgggtgg tcccaagagc tagaaactca 4680
ggaaacccca ggtgctcagg gccccgcgtc tcgggggctc cgtggggcag acccctgcta 4740
atatatgcaa ttctccctcc cccagccctt ccctgacccc taagttattg cccgctcacc 4800
tctcccaggc cccaggccgc ggagctggca gggtgggccc tgcggtttct atgtatttat 4860
agcaagttct gatgtacata tgtaaaggac ttttttaaat atatgtgcct tttgcctact 4920
tccccaaaaa aaaaaaaaaa aaaaaaaaaa 4948

<210> 5
<211> 1661
<212> ADN
<213> Homo sapiens

5

<400> 5
tgctgcagcc gctgccgccc attccggatc tcattgccac gcgccccga cgaccgcccg 60
acgtgcattc ccgattcctt ttggttccaa gtccaatag gcaactctaa aggatcagct 120
gatttataat cttctaaagg aagaacagac cccccagaat aagattacag ttgttggggg 180
tggtgctggt ggcatggcct gtgccatcag tatcttaatg aaggacttgg cagatgaact 240
tgctcttgtt gatgtcatcg aagacaaatt gaaggagag atgatggatc tccaacatgg 300
cagccttttc cttagaacac caaagattgt ctctggcaaa gactataatg taactgcaaa 360
ctccaagctg gtcattatca cggctggggc acgtcagcaa gagggagaaa gccgtcttaa 420
tttggccag cgtaacgtga acatatttaa attcatcatt cctaagtgtg taaaatacag 480
cccgaactgc aagttgctta ttgtttcaaa tccagtggat atcttgacct acgtggcttg 540
gaagataagt ggttttccca aaaaccgtgt tattggaagt ggttgcaatc tggattcagc 600
ccgattccgt tacctgatgg gggaaaggct gggagttcac ccattaagct gtcatgggtg 660
ggtccttggg gaacatggag attccagtgt gcctgatgg agtggaatga atgttgctgg 720
tgtctctctg aagactctgc acccagattt agggactgat aaagataagg aacagtggaa 780
agaggttcac aagcaggtgg ttgagagtgc ttatgaggtg atcaaactca aaggctacac 840
atcctgggct attggactct ctgtagcaga tttggcagag agtataatga agaatttag 900
gcgggtgcac ccagtttcca ccatgattaa gggcttttac ggaataaagg atgatgtctt 960
ccttagtggt ccttgcatth tgggacagaa tggaatctca gaccttgtga aggtgactct 1020
gacttctgag gaagaggccc gtttgaagaa gagtgcagat acactttggg ggatccaaaa 1080
ggagctgcaa ttttaaagtc ttctgatgic atatcatttc actgtctagg ctacaacagg 1140

ES 2 530 022 T3

attctaggtg gaggttgtgc atggtgtcct ttttatctga tctgtgatta aagcagtaat 1200
 attttaagat ggactgggaa aaacatcaac tcctgaagtt agaaataaga atggtttgtg 1260
 aatccacag ctatatcctg atgctggatg gtattaatct tgtgtagtct tcaactggtt 1320
 agtgtgaaat agttctgcca cctctgacgc accactgccca atgctgtacg tactgcattt 1380
 gccccttgag ccaggtggat gtttaccgtg tgttatataa cttcctggct ccttactga 1440
 acatgcctag tccaacattt ttcccgatg agtcacatcc tgggatccag tgtataaatc 1500
 caatatcatg tcttgtgcat aattcttcca aaggatctta ttttgtgaac tatatcagta 1560
 gtgtacatta ccataaatg taaaaagatc tacatacaaa caatgcaacc aactatccaa 1620
 gtgttatacc aactaaaacc cccaataaac cttgaacagt g 1661

<210> 6
 <211> 2484
 <212> ADN
 <213> Homo sapiens

5

<400> 6
 catcgaacaa acaactgaga tcctgttgtg cctatcacct gttgaagttg ccagtcttaa 60
 ggaaggaatc aatttcttcc gcaataagag cactggcaaa gactacgtct tgtacaagaa 120
 taagagccga ctgagggcat gcaagaatat gtgcaagcac caaggaggcc tgttcataaa 180
 agatatcgag gatttagccg gaagttgttg aaatggatga aaacaacgga cttttgcttt 240
 tagaactgaa tcctcctaac ccttgggact tacagcccag atctcctgaa gagtggctt 300
 ttggagaagt acagataaca tatctcactc atgcctgcat ggacctcaag ttagggagaca 360
 agagaatggt gtttgaccct tggttaatcg gtcctgcttt tgcccgtgga tgggtggtgc 420
 tccatgagcc tccatctgat tggctggaga ggctgtgccca ggcagacctc atttacatca 480
 gtcactctgca ctgagaccac ctgagttacc ccacactgaa aaagcttgcct gggagaagac 540
 cagatattcc catttatggt ggaacacag aaaggcctgt attttggaaat ctgaatcaga 600
 ggggtgtcca gttgactaat atcaatgtcg tgccatttgg aatatggcag caggtggaca 660
 aaaatcttcc attcatgac ttgatggacg gtgttcatcc tgagatggac acttgcatca 720
 ttgtggagta caaaggtcat aaaatactca atatagtaga ctgcaccaga cccaatgggg 780
 gaaggctgcc tatgaaggtt gctctaataga tgagtgattt tgctggagga gcatcaggct 840
 ttccaatgac tttcagtggt ggaaaattta cggaggaatg gaaagcccaa ttcattaata 900
 cagaaaggaa gaagctcctg aactacaagg ctcggtcgtg gaagaacctg caaccccgaa 960
 tttattgtcc ctttgctggg tattttgtgg aatctcacc ctcagacaag tacatcaagg 1020
 aaacaaacac caaaaatgac ccaaatgaac tcaacaatct tatcaagaaa aactctgatg 1080
 tgataacatg gacccctcga ccgggagcca cccttgatct gggagaatg ctgaaggatc 1140

ES 2 530 022 T3

gaacagacag caagggcatc atagagcctc cagaggggac aaaaatttac aaggattcct 1200
 gggattttga accttatttg gaaatcttga atgctgctct aggagatgaa atatttcttc 1260
 actcatcctg gataaaagaa tacttcactt gggctggatt taaagattac aaccttgtgg 1320
 tcaggatgat tgagacagat gaggacttca atccttttcc tggaggatat gactatttgg 1380
 ttgacttttt agatttatcc ttcccaaaag aaagaccaca acgagaacat ccctatgagg 1440
 aatccatag cgggtggat gtcacagac acgtggtgaa gaatggtcta ctctgggatg 1500
 agttgtatat aggattcaa acacggctcc agcgggatcc tgacatatac catcacctgt 1560
 tttggaatca ttttcaaata aaactcccc tcacaccacc caactggaag tcattcctga 1620
 tgtgctgtga gcagaatggg cctgtgattt tgcaagaatg caaaaccaca tgaaaatttc 1680
 aagaattcac tgatctgatg caaaataaaa atttatcatt acatcttgaa cccaggaagc 1740
 ttacagcaa gagactatgc tttatgacgt cagcaataga taattccacg ttgcctttgt 1800
 gatttgata tatagcttac atttgtggat cactacatag ccagattcaa aatattttta 1860
 cttgttccat ccacagttct ctacagaaag aaccaatgaa cccaatagga acaaattctc 1920
 tgtggaaaac aaagcatagc tgtagtagat acgaatcaa tcacagagga aacaggaaga 1980
 gaaaaacatc caagactaca gtgaaaactg gaaatggtct gttttcgtga tattcgtatg 2040
 attaagatgc aaatttttct ttaggaaaat gtgattgtta actagcattc tgttttacat 2100
 gttgacattt ctaacacaca caccactgat ttgaacttca aaatttattt tctgattata 2160
 tatgctaggt ctgattctga agatacaaga attcaatggt ggaatttgtc tcctgaaatt 2220
 tctagattca taattagaag tgtttattac caacttctca cttagattaa catcagtgat 2280
 gtttgattat tctggttaac cgctcacatg cataacaata atgctagaaa ttcaggaatt 2340
 ataattttgt gatttatact gtgtggtaaa ttattcagtt ctcactgcaa ataaatataa 2400
 tgtatatcaa tgtaatgtag gtcagaatgt ctatggtgag caaataaaaag attattttta 2460
 aatatgaaaa aaaaaaaaaa aaaa 2484

<210> 7
 <211> 1279
 <212> ADN
 <213> Homo sapiens

5

<400> 7
 cagcgaggag cggcgcctc ccgcgccccg cggcgcctc ggagtaattt cggatgccca 60
 gccgcggccg ccttcccag tagaccggg agaggagtg cggccaactt gtgtgccttt 120
 cttccgccc ggtgggagcc ggcgctgcgc gaagggctct cccgcggtc catgctgccg 180
 gccctgcgcc tccccagcct cgggtgagcc gcctccggag agacggggga gcgcgggcgc 240
 gccgcgggct cggcgtgctc tcctccggg acgcgggacg aagcagcagc cccgggcgcg 300

ES 2 530 022 T3

cgccagagggc atggagcgcct gccccagcct aggggtcacc ctctacgccc tggtaggtggt 360
cctggggctg cgggcgacac cggccggcgg ccagcactat ctccacatcc gcccggcacc 420
cagcgacaac ctgcccctgg tggacctcat cgaacacca gaccctatct ttgaccccaa 480
ggaaaaggat ctgaacgaga cgctgctgcg ctcgctgctc gggggccact acgaccaggg 540
cttcatggcc acctcgcccc ccgaggaccg gcccggcggg ggcgggggtg cagctggggg 600
cgcgaggagc ctggcggagc tggaccagct gctgcggcag cggccgtcgg gggccatgcc 660
gagcgagatc aaagggctag agttctccga gggcttgccc cagggaaga agcagcgcct 720
aagcaagaag ctgcgaggga agttacagat gtggctgtgg tcgcagacat tctgccccgt 780
gctgtacgcg tggaacgacc tgggcagccg cttttggccg cgctacgtga aggtgggcag 840
ctgcttcagt aagcgcctgt gctccgtgcc cgagggcatg gtgtgcaagc cgtccaagtc 900
cgtgcacctc acggtgctgc ggtggcgtg tcagcggcgc gggggccagc gctgcccgtg 960
gattcccatac cagtacccca tcatttccga gtgcaagtgc tcgtgctaga actcgggggc 1020
cccctgccg caccgggaca cttgatgat ccccaccgac gccccctgca ccgcctcaa 1080
ccagttccac caccctctag cgagggtttt caatgaactt ttttttttt ttttttttt 1140
ttctgggcta cagagaccta gctttctggt tcctgtaatg cactgtttaa ctgtgtagga 1200
atgtatatgt gtgtgtatat acgggtcccag ttttaattta cttattaataa ggtcagtatt 1260
atacgtaa aaaaaaaaa 1279

<210> 8
<211> 2106
<212> ADN
<213> Homo sapiens

5

<400> 8
agtttccctt ccgctcacct ccgcctgagc agtggagaag gcggcactct ggtggggctg 60
ctccaggcat gcagatccca caggcgcctt ggccagtcgt ctgggcgggtg ctacaactgg 120
gctggcggcc aggatggttc ttagactccc cagacaggcc ctggaacccc cccaccttct 180
tcccagccct gctcgtggtg accgaagggg acaacgccac cttcacctgc agcttctcca 240
acacatcgga gagcttcgtg ctaaactggt accgcatgag ccccagcaac cagacggaca 300
agctggccgc cttccccgag gaccgcagcc agcccggcca ggactgccgc ttccgtgtca 360
cacaactgcc caacgggcgt gacttcaca tgagcgtggt cagggcccgg cgcaatgaca 420
gcggcacctc cctctgtggg gccatctccc tggcccccaa ggcgcagatc aaagagagcc 480
tgcgggcaga gctcaggggt acagagagaa gggcagaagt gccacagcc caccocagcc 540
cctcaccag gccagccggc cagttccaaa ccctggtggt tgggtgctgtg ggcggcctgc 600
tgggcagcct ggtgctgcta gtctgggtcc tggccgtcat ctgctcccgg gccgcacgag 660

ES 2 530 022 T3

ggacaatagg agccaggcgc accggccagc ccctgaagga ggacccctca gccgtgcctg 720
 tgttctctgt ggactatggg gagctggatt tccagtggcg agagaagacc cccggagcccc 780
 ccgtgccctg tgtccctgag cagacggagt atgccaccat tgtctttcct agcggaatgg 840
 gcacctcatc ccccgcccgcc aggggctcag ccgacggccc tcggagtgcc cagccactga 900
 ggctgagga tggacactgc tcttgcccc tctgaccggc ttccttgccc accagtgttc 960
 tgcagaccct ccaccatgag cccgggtcag cgcatttcct caggagaagc aggcagggtg 1020
 caggccattg caggccgtcc aggggctgag ctgcctgggg gcgaccgggg ctccagcctg 1080
 cacctgcacc aggcacagcc ccaccacagg actcatgtct caatgccac agtgagccca 1140
 ggcagcaggt gtcaccgtcc cctacagga gggccagat cagtactgc ttcaggtcct 1200
 gccagcacag agctgcctgc gtccagctcc ctgaatctct gctgctgctg ctgctgctgc 1260
 tgctgctgcc tgcggcccg ggctgaaggc gccgtggccc tgctgacgc cccggagcct 1320
 cctgcctgaa cttgggggct ggttgagat ggccttgag cagccaaggt gccctggca 1380
 gtggcatccc gaaacgccct ggacgcaggg cccaagactg ggcacaggag tgggaggtac 1440
 atggggctgg ggactcccc ggagtatatct gctccctgca ggcctagaga agtttcaggg 1500
 aaggtcagaa gagctcctgg ctgtggtggg cagggcagga aaccctccc acctttacac 1560
 atgcccaggc agcacctcag gccctttgtg gggcagggaa gctgaggcag taagcgggca 1620
 ggcagagctg gaggcctttc aggccagcca gcaactctggc ctctgcccgc cgcattccac 1680
 cccagccctc cacaccactc gggagagggc catcctacgg tcccaaggtc aggagggcag 1740
 ggctgggggtt gactcaggcc cctcccagct gtggccacct ggggtgtggg agggcagaag 1800
 tgcaggcacc tagggcccc catgtgcccc ccctgggagc tctccttggg acccattcct 1860
 gaaattattt aaaggggttg gccgggctcc caccagggcc tgggtgggaa ggtacaggcg 1920
 ttccccggg gcctagtacc cccgctggc ctatccactc ctacatcca cactgcac 1980
 cccactcct ggggcagggc caccagcatc cagggcgcca gcaggcacct gagtggctgg 2040
 gacaagggat ccccttccc tgtggttcta ttatattata attataatta aatatgagag 2100
 catgct 2106

<210> 9
 <211> 894
 <212> ADN
 <213> Homo sapiens

<400> 9
 tttgggtttg gacgggtgcg tttgttttc atagtgcagg tctggagccg gactgagatg 60
 gggagtagag ggaccccttg aaacttcagg gtggatgaat atgaagttc atcaccagct 120
 gtgatgtttg atattggact gtggttgtgc cagttgacac tcaggagtgg agactagatc 180

5

ES 2 530 022 T3

tgcaatctgt tgggactgtg atgaggggat catcctgccc tcgaaccaga cccagccctg 240
tagcccaggc actattgcct ggagttcctg ggctctcagc gccaggaaat catgaggttc 300
agtgtctcag gcatgaggac cgactacccc aggagtgtgc tggctcctgc ttatgtgtca 360
gtctgtctcc tcctcttgtg tccaagggaa gtcacgctc ccgctggctc agaaccatgg 420
ctgtgccagc cggcaccocag gtgtggagac aagatctaca accccttggg gcagtgtctgt 480
tacaatgacg ccacgtgtc cctgagcgag acccgccaat gtggtcccc ctgcaccttc 540
tggccctgct ttgagctctg ctgtcttgat tcctttggcc tcacaaacga ttttgttgtg 600
aagctgaagg ttcaggggtg gaattccag tgccactcat ctccatctc cagtaaatgt 660
gaaagcagaa gacgttttcc ctgagaagac atagaaagaa aatcaacttt cactaaggca 720
tctcagaaac ataggctagg gtaatatgtg taccagtaga gaagcctgag gaatttacia 780
aatgatgcag ctccaagcca ttgtatggcc catgtgggag actgatggga catggagaat 840
gacagtagat tatcaggaaa taaataaagt ggtttttcca atgtaaaaaa aaaa 894

<210> 10
<211> 1095
<212> ADN
<213> Homo sapiens

5

<400> 10
atgggttcag cccaaggcct ggcatattgt ggatgcttcg tagaagttat tgtgggctta 60
ctgtcagctg tgcacaaggc accaggagaa caggggctgg aggtaggaga gatgttgcca 120
gactgtcagt tacttcttcc agtttagctt tcagttgtgt tgcagtatgc ctctggagga 180
ggggaagtgt ctttgatgct ctggaataca accgggcccg cccccggcgc accggccccg 240
gcgggcacct ggcggtcgca gcggacacag caggaggagg ataggccgcg ctgggcccggc 300
gtcgcggggc cgcgcttggg gttcctgagc cgcggccgaa ctttttccca agaagtcca 360
cgtttggagg cccgaggagc gggacgggga ggagagactg gagttcctgg aaggagcagc 420
agcagcagcc ccacaacatg gcctcctcag cctcctcctc tcccggggc accgaggact 480
ccgctcctgc ccagggtggc tttggatcag attattcacg aagctctaga aaggtaagtg 540
aacttggcat tggccattgc ttattatgct tcctgtctgg ctcttactca tgttgccatt 600
ttgtggacta aaactcaagg cagttcaaaa atactgatat gcaggtgcca tggatgctgt 660
tggagtgtgt gtcctaagag agagggtaag tgctggtttt ccattttaa catagctcgt 720
gatatcaagt tgccccctcc cttcattccc tgactcccgt cttcctcaa aagcggctct 780
cttggtttag taagcatctc taccctgcaa agcggaaata atgaaaaatg agattcactt 840
tgcaagcttg ttgggtgagc ttctggagaa taatttttac caaaagcctg aatcatttg 900
gcattgttg tgttttcaat gcatctgcat cacagcact tgcatttcac atctggccag 960
tttggaatga aacttagaaa atcaccagac tatagtgatg taattgtcat gaagtggttt 1020
ttgtagctta gagaaaagaa aaacacttgg ctgaaagaat ggtctgaagt cccacccaaa 1080
gtaaaacatt ctaaa 1095

<210> 11
<211> 846
<212> ADN

10

ES 2 530 022 T3

<213> Homo sapiens

<400> 11

```

acctctcaag cagccagcgc ctgcctgaat ctgttctgcc ccctccccac ccatttcacc      60
accacatga caccgggcac ccagtctcct ttcttcctgc tgctgctcct cacagtgttt      120
acagttgtta cgggttctgg tcatgcaagc tctaccccag gtggagaaaa ggagacttcg      180
gctaccacaga gaagttcagt gccagctct actgagaaga atgcttttaa ttcctctctg      240
gaagatccca gcaccgacta ctaccaagag ctgcagagag acatttctga aatggctgtc      300
tgtcagtgcc gccgaaagaa ctacgggcag ctggacatct ttccagcccg ggatacctac      360
cctcctatga gcgagtaccc cacctaccac acccatgggc gctatgtgcc cctagcagt      420
accgatcgtg gccctatga gaaggtttct gcaggtaatg gtggcagcag cctctcttac      480
acaaaccacag cagtggcagc cacttctgcc aacttgtagg ggcacgtcgc ccgctgagct      540
gagtggccag ccagtgccat tccactccac tcaggttctt cagggccaga gccctgcac      600
cctgtttggg ctggtgagct gggagttcag gtgggctgct cacagcctcc ttcagaggcc      660
ccaccaatth ctgcgacact tctcagtgtg tggaagctca tgtgggcccc tgagggtctca      720
tgcctgggaa gtgttgtggt gggggctccc aggaggactg gccagagag ccctgagata      780
gctgggatcc tgaactggac tgaataaaac gtggtctccc actgcgcaa aaaaaaaaaa      840
aaaaaa                                             846

```

5

<210> 12

<211> 730

<212> ADN

<213> Homo sapiens

10

<220>

<221> rasgo_misc

<222> (728)..(728)

<223> n es a, c, g, t o u

15

<400> 12

```

atthtaacag caatgagatg acatgagaat atcttcaatt catcaaggat ctgtaaagaa      60
ggagaggtac aagatatatg aaacccaaat ctcaaaacaa tgatttagtg aatttcccat      120
gaactttaa cagtgattgc ttcaaaatth ccaagagcca tactctccct ccagctgctg      180
tgtgtgtgtg tgtgtataaa tgcacactat thtaacctaa aatggtgcc ttgtgctgcc      240
attctctaac tcttgacata ttaaacatth attcttggtc aaattaaaac ctcatgcatt      300
tccaaagata taaatgcctt gcctggagaa gttagatctt gcaagtctca ggagggccga      360
gatggtttgt cttatgccta tagctgttha tgtcccacca gtgggtgttt gtttcattag      420
gtgccgtttc cagccaaatg ttctcattct tcacatcttc aatggtgagt agcaaacaga      480
agaacatcct tcttagcata atattgcttc actggactga ttgtgaactc aaaatacctc      540
ttgtttcttg tgaagggttt gccttttgta aacaatataa gatcacttht ggtcaaccac      600
cctgtctgaa thtatctggg ctgctataat aaagtatcat aaactgggag gcttgtaaat      660
aaatttattt cttacagtht tggaaagctg ggaaccccaa aatcaaggcc aatttggtgt      720
ctggtgangg                                             730

```

20

<210> 13

ES 2 530 022 T3

<211> 1834
 <212> ADN
 <213> Homo sapiens

5

<400> 13
 actttgacag tcttctcatg ctgcctctgc caccttctct gccagaagat accatttcaa 60
 ctttaacaca gcatgatcga aacatacaac caaacttctc cccgatctgc ggccactgga 120
 ctgcccataca gcatgaaaat ttttatgtat ttacttactg tttttcttat caccagatg 180
 attgggtcag cactttttgc tgtgtatcct catagaaggt tggacaagat agaagatgaa 240
 aggaatcttc atgaagattt tgtattcatg aaaacgatac agagatgcaa cacaggagaa 300
 agatccttat ccttactgaa ctgtgaggag attaaaagcc agtttgaagg ctttgtgaag 360
 gatataatgt taaacaaaga ggagacgaag aaagaaaaca gctttgaaat gcaaaaaggt 420
 gatcagaatc ctcaaattgc ggcacatgtc ataagtgagg ccagcagtaa aacaacatct 480
 gtgttacagt gggctgaaaa aggatactac accatgagca acaacttggg aaccctgga 540
 aatgggaaac agctgaccgt taaaagacaa ggactctatt atatctatgc ccaagtcacc 600
 ttctgttcca atcgggaagc ttogagtcaa gctccattta tagccagcct ctgcctaaag 660
 tccccggta gattcgagag aatcttactc agagctgcaa ataccacag ttccgcaaaa 720
 ccttgcgggc aacaatccat tcacttggga ggagtatttg aattgcaacc aggtgcttcg 780
 gtgtttgtca atgtgactga tccaagccaa gtgagccatg gcaactggct cacgtccttt 840
 ggcttactca aactctgaac agtgtcacct tgcaggctgt ggtggagctg acgctgggag 900
 tcttcataat acagcacagc ggttaagccc acccctgtt aactgcctat ttataaccct 960
 aggatcctcc ttatggagaa ctatttatta tacactccaa ggcatgtaga actgtaataa 1020
 gtgaattaca ggtcacatga aacaaaacg ggcctgctc cataagagct tatatatctg 1080
 aagcagcaac cccactgatg cagacatcca gagagtctta tgaaaagaca aggcattat 1140
 gcacagggtg aattctgagt aaacagcaga taacttgcca agttcagttt tgtttctttg 1200
 cgtgcagtgt ctttccatgg ataatgcatt tgatttatca gtgaagatgc agaagggaaa 1260
 tggggagcct cagctcacat tcagttatgg ttgactctgg gttcctatgg ccttgttggg 1320
 gggggccagg ctctagaacg tctaacacag tggagaaccg aaaccccccc ccccccccg 1380
 ccaccctctc ggacagttat tcattctctt tcaatctctc tctctccatc tctctctttc 1440
 agtctctctc tctcaacctc tttcttccaa tctctctttc tcaatctctc tgtttccctt 1500
 tgtcagtctc ttcctcccc cagtctctct tctcaatccc cctttctaac acacacacac 1560
 acacacacac acacacacac acacacacac acacacacac agagtcaggc cgttgcctagt 1620
 cagttctctt ctttccacct tgtocctatc tctaccacta tagatgaggg tgaggagtag 1680
 ggagtgcagc cctgagcctg cccactctc attacgaaat gactgtattt aaaggaaatc 1740
 tattgtatct acctgcagtc tccattgttt ccagagtgaa cttgtaatta tcttgttatt 1800
 tattttttga ataataaaga cctcttaaca ttaa 1834

REIVINDICACIONES

1. Un método para diagnosticar Artritis reumatoide en un paciente, comprendiendo el método:

obtener una muestra de células T CD4+ del paciente; y

determinar los niveles de expresión de todos los genes del grupo que consiste en

5 BCL3

SOCS3

PIM1

SBNO2

LDHA

10 CMAH

NOG

PDCD1

IGFL2

LOC731186

15 MUC1

GPRIN3; y

comparar dichos niveles de expresión con niveles de expresión de referencia, en donde

una diferencia en la expresión de dichos genes indica una mayor probabilidad de que el paciente tenga Artritis reumatoide (AR).

20 2. Un método como en la Reivindicación 1, en donde el grupo comprende adicionalmente el gen CD40LG.

3. Un método como en las Reivindicaciones 1 o 2, en donde los niveles de expresión de referencia son representativos de los niveles encontrados en muestras que comprenden células de un paciente que no tiene Artritis reumatoide (AR).

4. Un método para tipificar una muestra de un individuo, comprendiendo el método:

25 obtener una muestra de células T CD4+ del individuo; y

determinar los niveles de expresión de todos los genes del grupo que consiste en

BCL3

SOCS3

PIM1

30 SBNO2

LDHA

CMAH

NOG

PDCD1

35 IGFL2

LOC731186

MUC1

GPRIN3; y

tipificar dicha muestra basándose en los niveles de expresión determinados;

en donde dicha tipificación proporciona información pronóstica relacionada con el riesgo de que el individuo tenga artritis reumatoide (AR).

5. Un método como en la Reivindicación 4, en donde el grupo comprende adicionalmente el gen CD40LG.
- 5 6. Un método como en cualquiera de las Reivindicaciones anteriores, en donde los niveles de expresión se determinan determinando los niveles de ARN.
7. Un método como en cualquiera de las Reivindicaciones anteriores, en donde la muestra de células T CD4+ se obtiene a partir de sangre completa periférica.
- 10 8. Un método como en la Reivindicación 7, que incluye la etapa de separar las células T CD4+ de sangre completa periférica.
9. Un método como en la Reivindicación 7, que incluye la etapa de extraer ARN de las células T CD4+.
10. Métodos como en cualquiera de las Reivindicaciones anteriores, que comprenden adicionalmente la etapa de combinar los resultados con los resultados de análisis de predicción conocidos.
- 15 11. Un método como en la Reivindicación 10, en donde el análisis de predicción conocido es la regla de predicción de Leiden.
12. Un método para diagnosticar artritis reumatoide en un paciente, comprendiendo el método:
- obtener una muestra de sangre del paciente; y
- determinar los niveles de expresión/ ARNm de todos los genes del grupo que consiste en
- 20 BCL3
- SOCS3
- PIM1
- SBNO2
- LDHA
- CMAH
- 25 NOG
- PDCD1
- IGFL2
- LOC731186
- MUC1
- 30 GPRIN3; y
- comparar dichos niveles de expresión/ARNm con un conjunto de niveles de expresión/ARNm de referencia, en donde una diferencia en la expresión de dichos genes indica una mayor probabilidad de que el paciente tenga Artritis reumatoide.
13. El uso de un conjunto de sondas que consiste solamente en polinucleótidos específicos para los genes
- 35 BCL3
- SOCS3
- PIM1
- SBNO2
- LDHA
- 40 CMAH

NOG

PDCD1

IGFL2

LOC731186

5 MUC1

GPRIN3

para determinar el riesgo de un individuo de padecer artritis reumatoide.

14. El uso de un conjunto de sondas como en la reivindicación 13, comprendiendo adicionalmente dichas sondas un polinucleótido específico para CD40LG.

10

Figura 1.

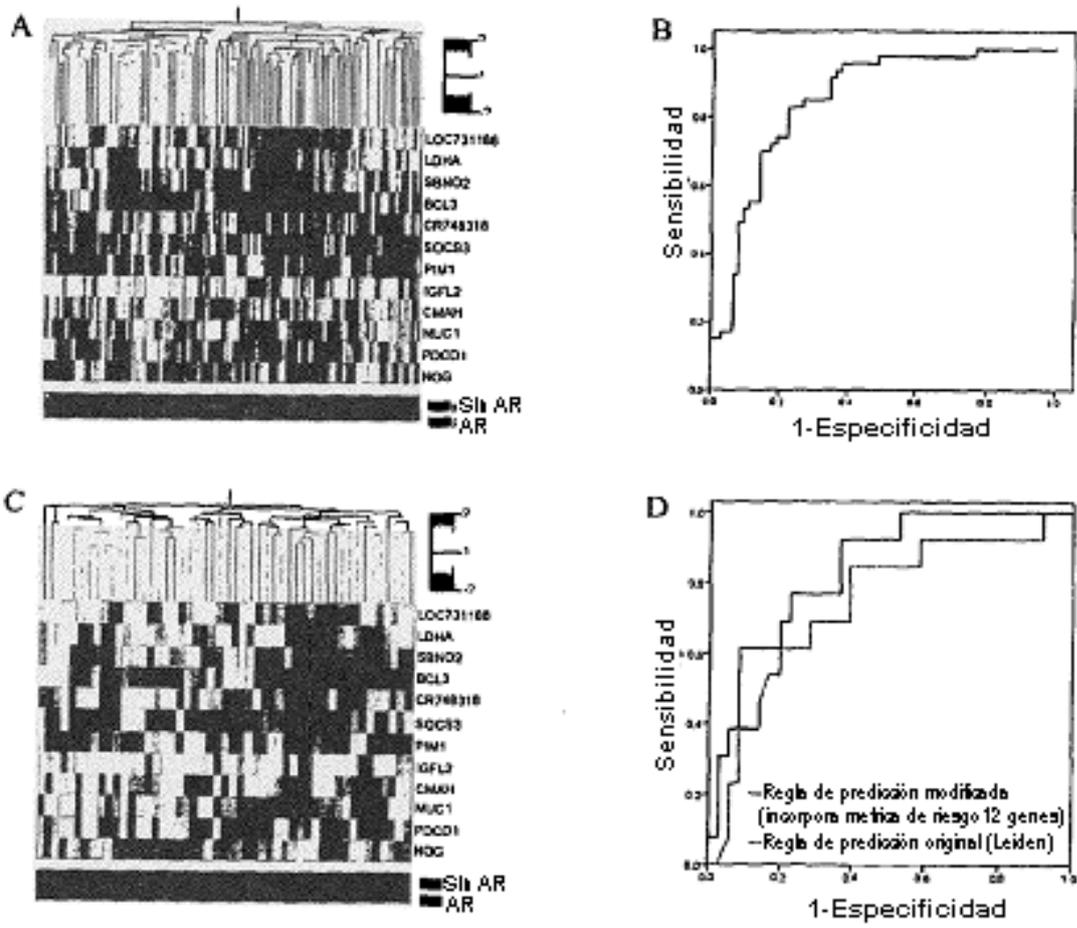


Figura 2.

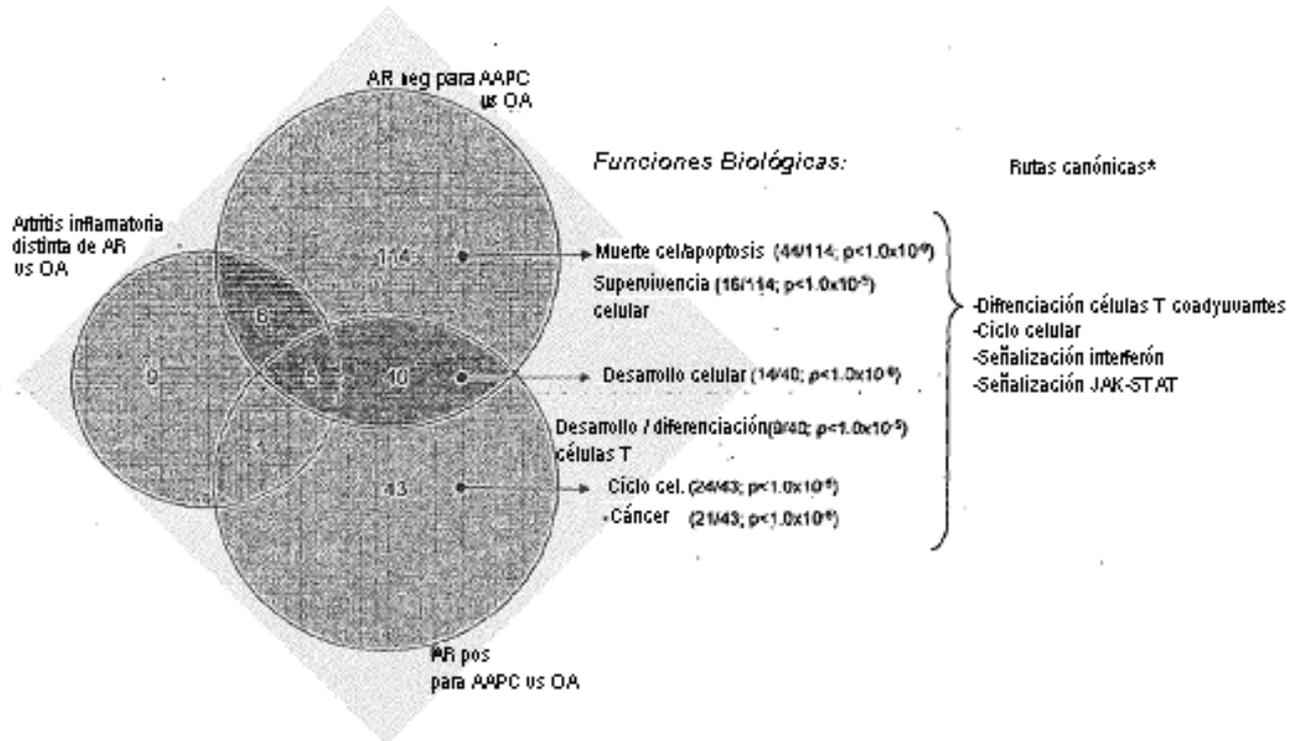


Figura 3.

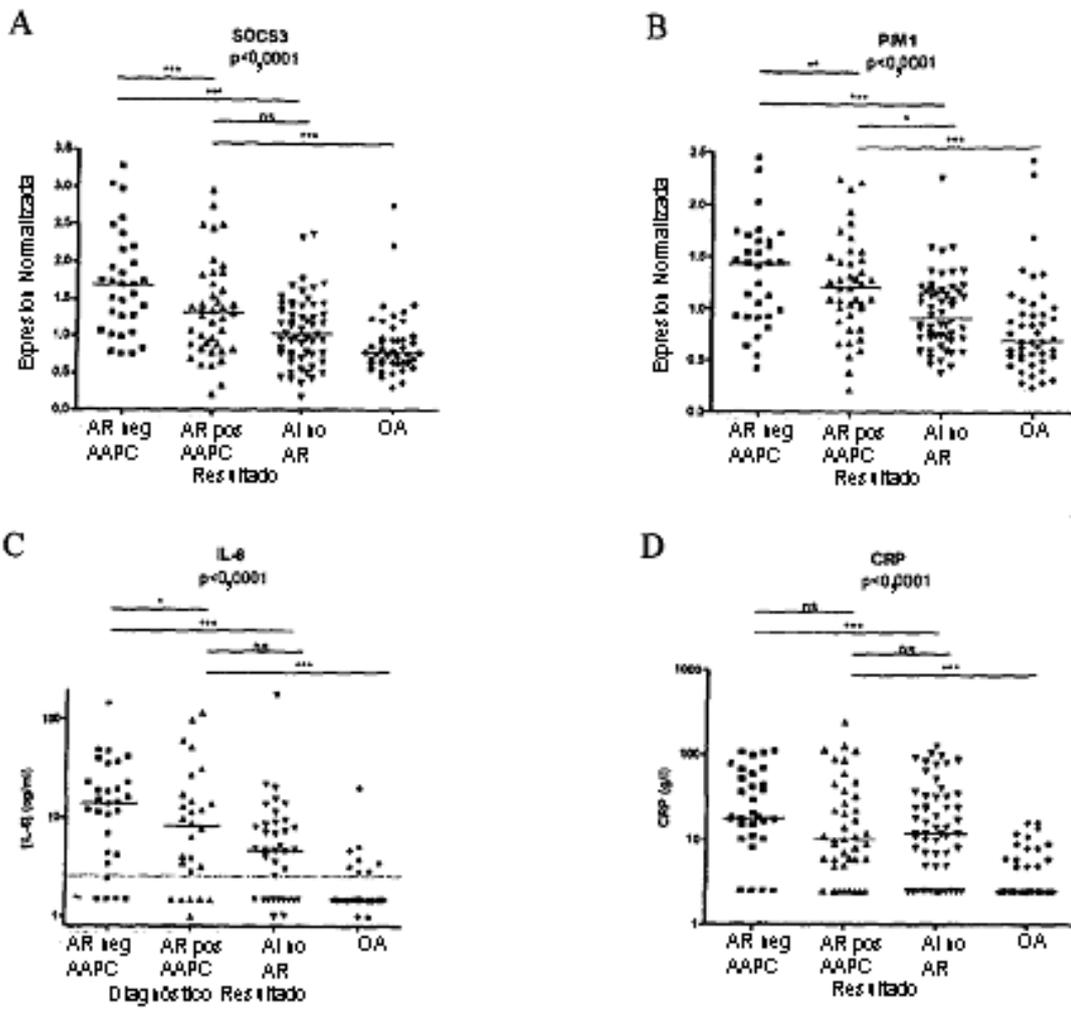


Figura 4.

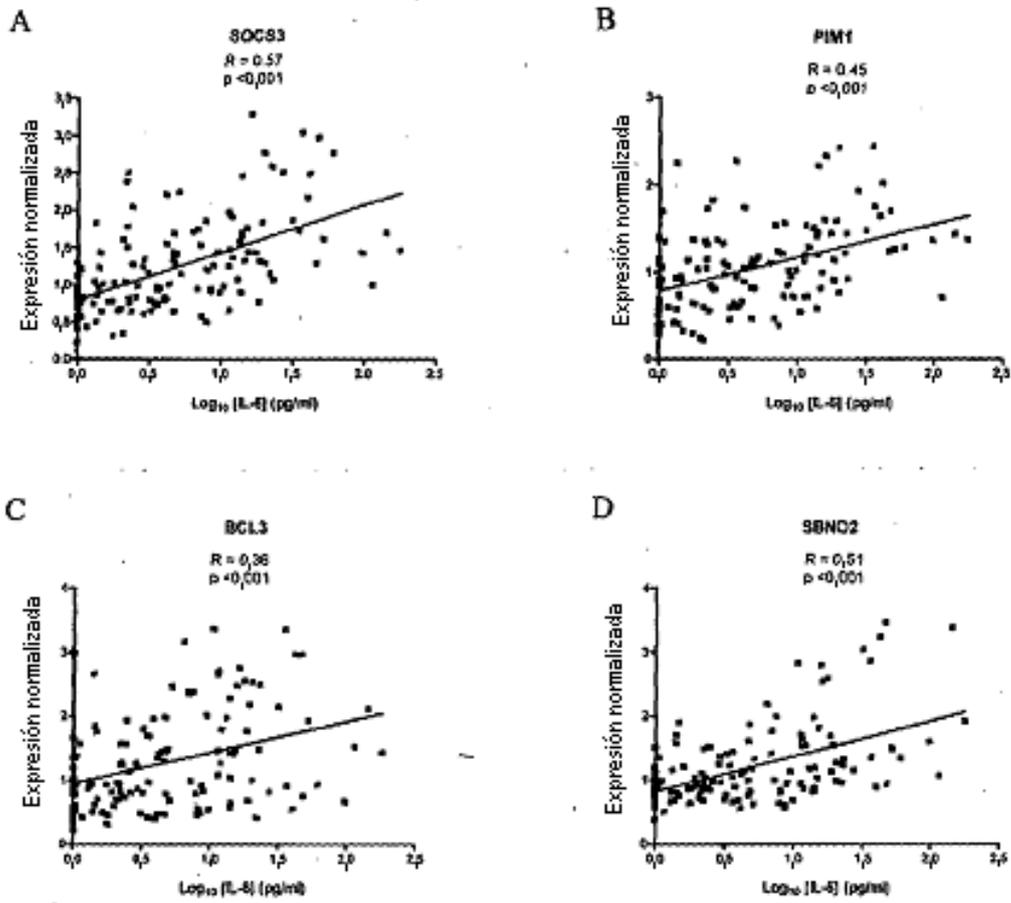


Figura 5.

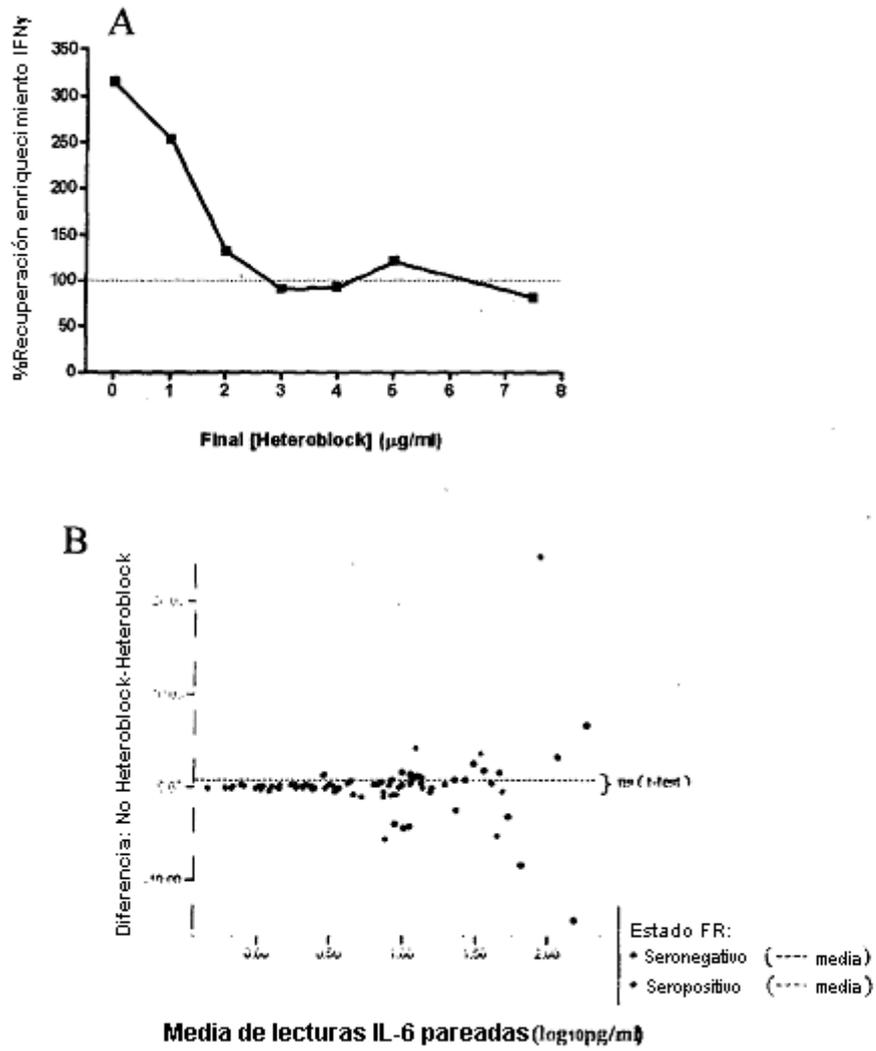


Figura 6

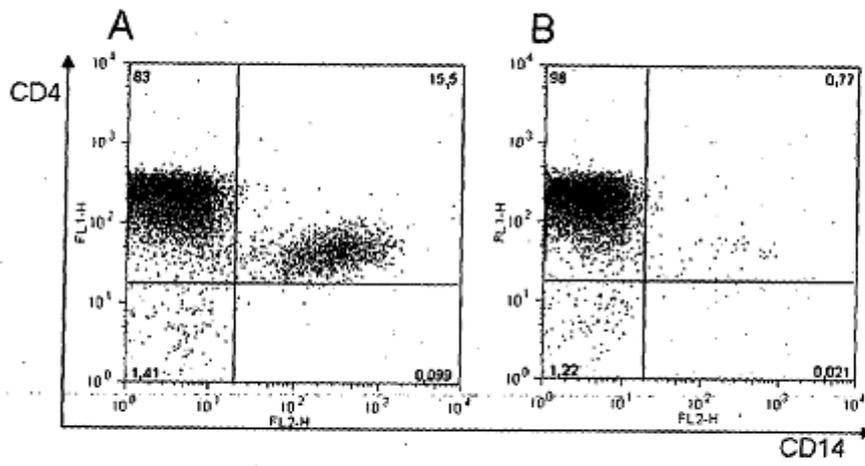


Figura 7

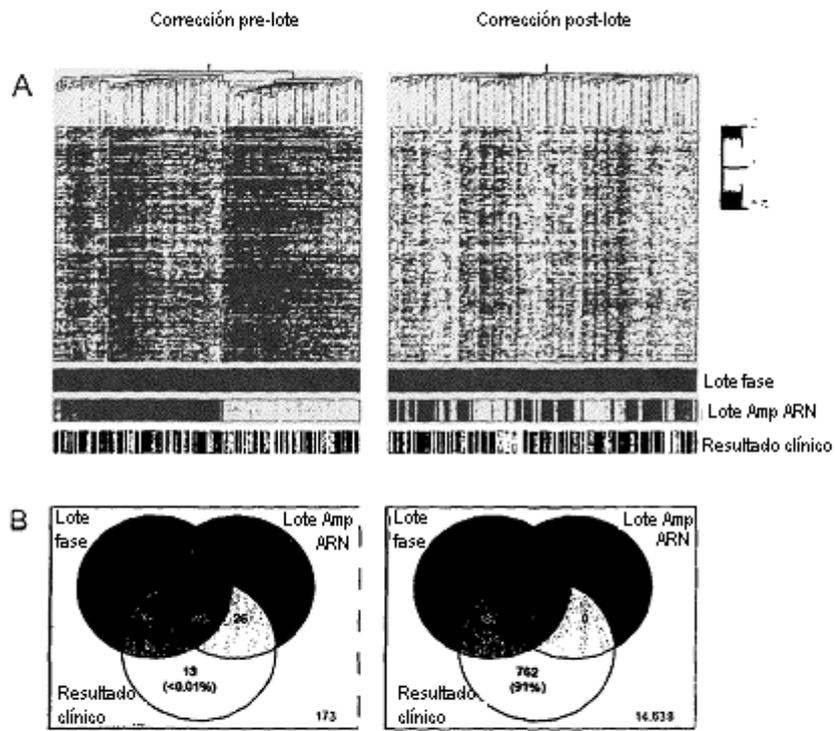


Figura 8

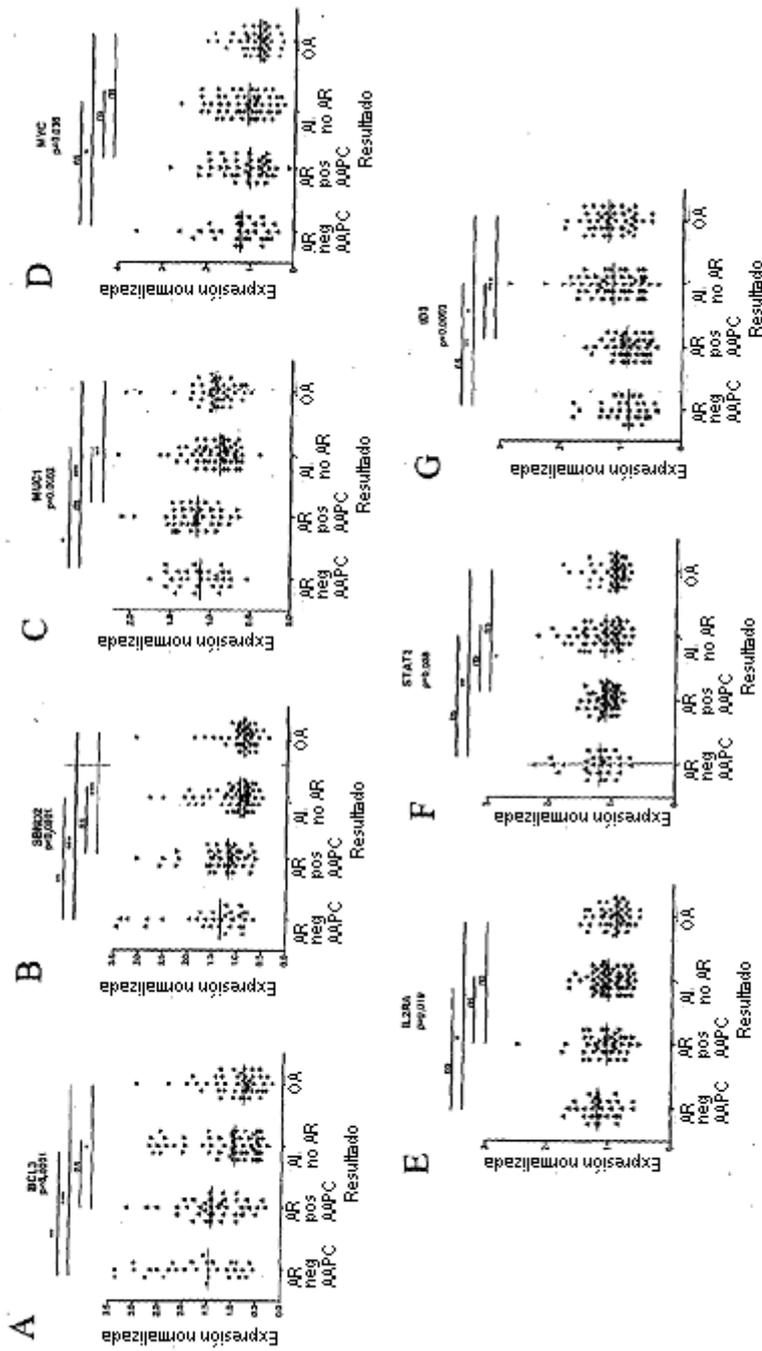


Figura 9.

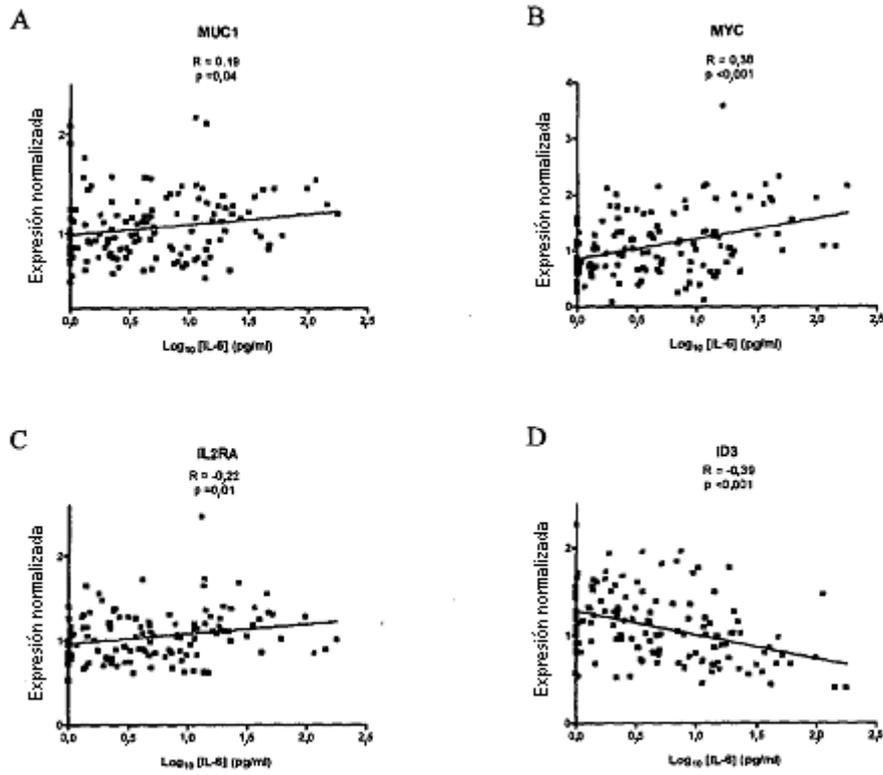


Figura 10

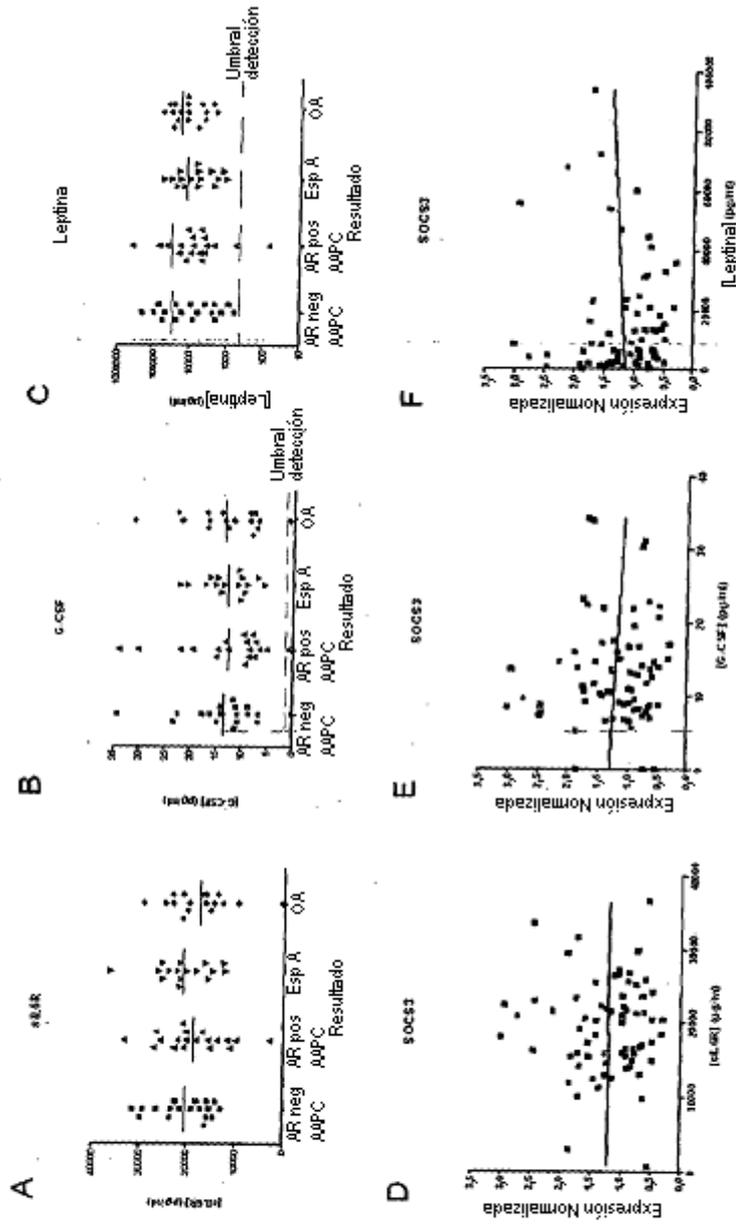
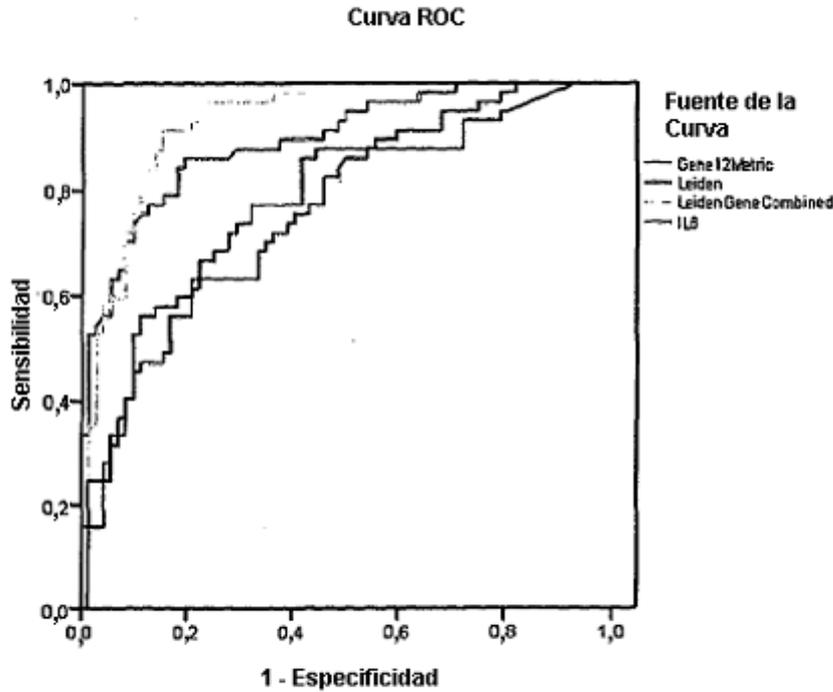


Figura 11



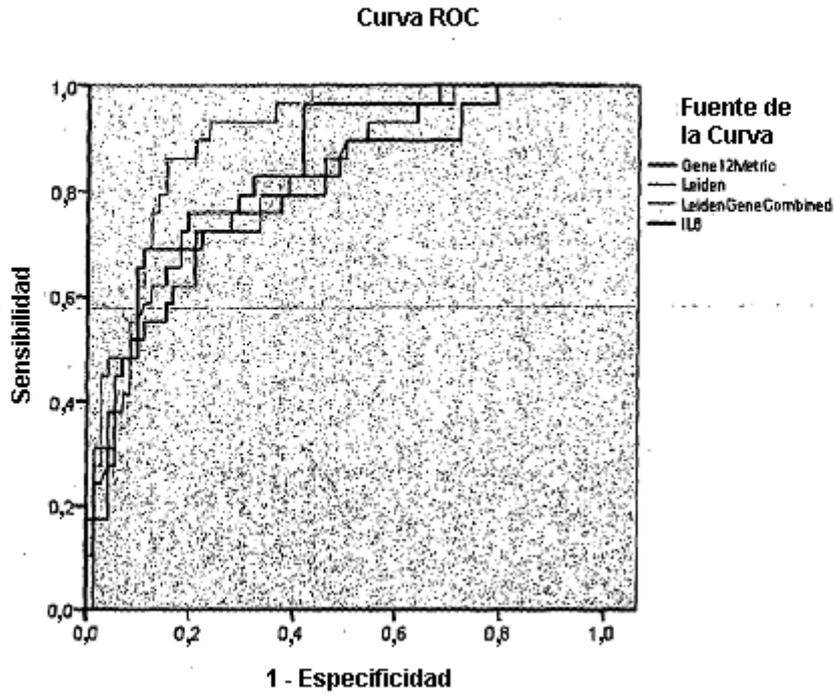
Los segmentos diagonales son producidos por los empates.

Área Bajo la Curva

Variables del Resultado de la Prueba		Área	Error Típico ^a	Sig. Asintótica ^b Límite Inferior	Intervalo de Confianza Asintótico al 95% Límite Superior
Gene12Metric		0,782	0,040	0,000 0,703	0,861
Leiden	0,894	0,028	0,000	0,839 0,949	
LeidenGeneCombined		0,929	0,022	0,000 0,886	0,972
IL6	0,758	0,043	0,000	0,673 0,842	

- 5 Variables del Resultado de la prueba: Leiden, IL6 tiene al menos un empate entre el grupo de estado actual positivo y el grupo de estado actual negativo.
 La estadística puede estar sesgada
 a. Bajo la suposición no paramétrica
 b. Hipótesis nula: área verdadera = 0,5

Figura 12



Los segmentos diagonales son producidos por los empates.

5 **Área Bajo la Curva**

Variables del Resultado de la Prueba	Área	Error Típico ^a	Sig. Asintótica ^b Límite Inferior	Intervalo de Confianza Asintótico al 95% Límite Superior
Gene12Metric	0,833	0,041	0,000 0,752	0,914
Leiden	0,816	0,046	0,000 0,726 0,907	
LeidenGeneCombined	0,905	0,029	0,000 0,848	0,963
IL6	0,812	0,049	0,000 0,715 0,909	

Variables del Resultado de la prueba: Leiden, IL6 tiene al menos un empate entre el grupo de estado actual positivo y el grupo de estado actual negativo.

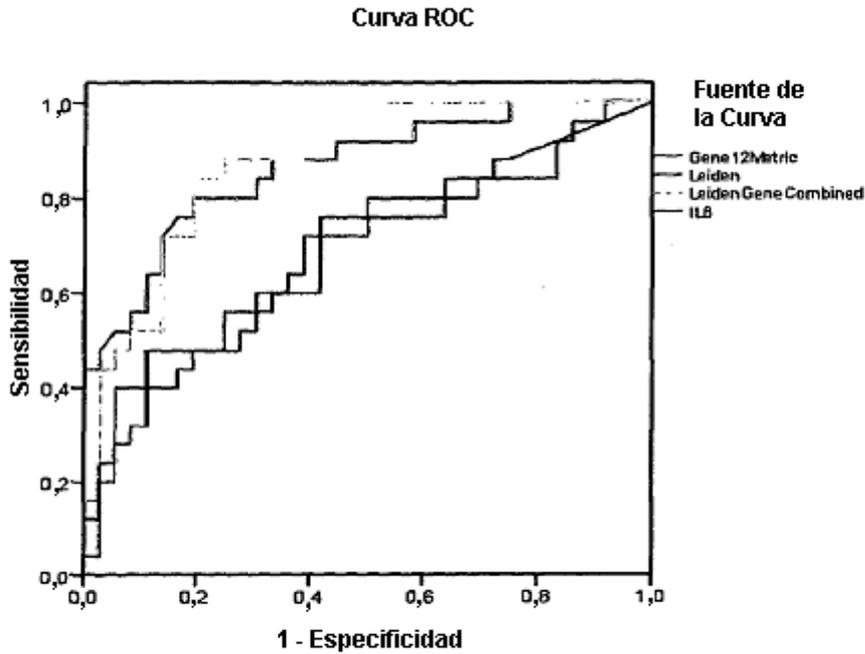
La estadística puede estar sesgada

a. Bajo la suposición no paramétrica

b. Hipótesis nula: área verdadera = 0,5

10

Figura 13



Los segmentos diagonales son producidos por los empates.

Área Bajo la Curva

Variables del Resultado de la Prueba		Área	Error Típico ^a	Sig. Asintótica ^b Límite Inferior	Intervalo de Confianza Asintótico al 95% Límite Superior
Gene12Metric		0,680	0,071	0,018 0,540	0,820
Leiden	0,863	0,048	0,000	0,770 0,957	
LeidenGeneCombined		0,873	0,044	0,000 0,787	0,959
IL6	0,692	0,071	0,011	0,552 0,831	

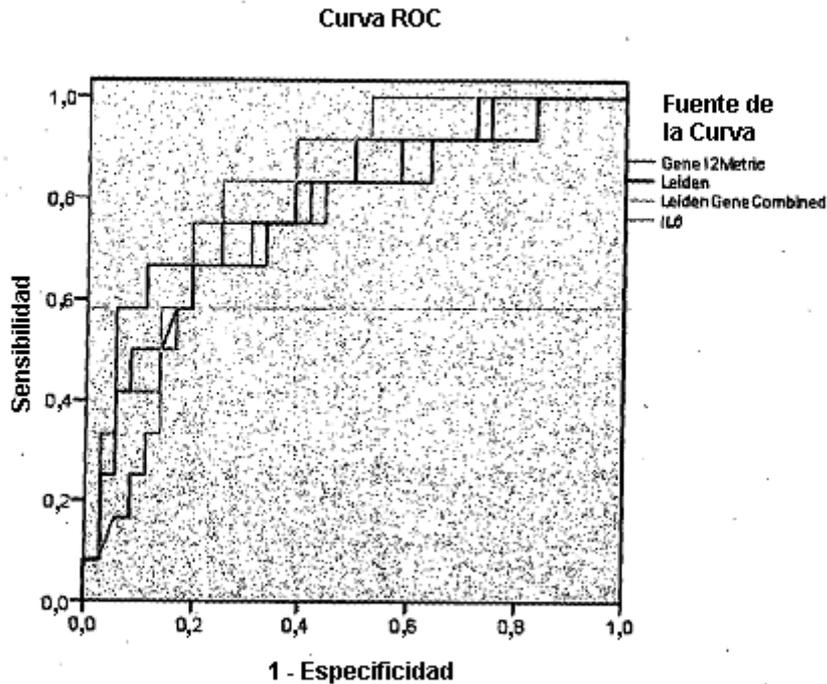
5 Variables del Resultado de la prueba: Leiden, IL6 tiene al menos un empate entre el grupo de estado actual positivo y el grupo de estado actual negativo.

La estadística puede estar sesgada

a. Bajo la suposición no paramétrica

b. Hipótesis nula: área verdadera = 0,5

Figura 14



Los segmentos diagonales son producidos por los empates.

Área Bajo la Curva

Variables del Resultado de la Prueba		Área	Error Típico ^a	Sig. Asintótica ^b Límite Inferior	Intervalo de Confianza Asintótico al 95% Límite Superior
Gene12Metric		0,780	0,079	0,004 0,625	0,936
Leiden	0,752	0,079	0,009	0,598 0,907	
LeidenGeneCombined		0,836	0,061	0,001 0,717	0,954
IL6	0,801	0,079	0,002	0,647 0,955	

- 5 Variables del Resultado de la prueba: Leiden, IL6 tiene al menos un empate entre el grupo de estado actual positivo y el grupo de estado actual negativo.
La estadística puede estar sesgada
a. Bajo la suposición no paramétrica
b. Hipótesis nula: área verdadera = 0,5

10