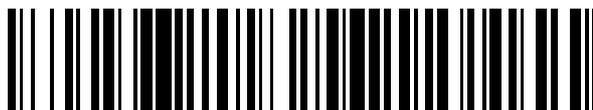


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 666 206**

51 Int. Cl.:

C07K 14/47 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **02.10.2008** E 11182288 (8)

97 Fecha y número de publicación de la concesión europea: **24.01.2018** EP 2420509

54 Título: **Genes marcadores novedosos para linfocitos T reguladores de sangre humana**

30 Prioridad:

02.10.2007 US 976866 P

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

03.05.2018

73 Titular/es:

**UNIVERSITÄT ZU KÖLN (50.0%)
Albertus-Magnus-Platz 1
50923 Köln, DE y
BECTON, DICKINSON AND COMPANY (50.0%)**

72 Inventor/es:

**SCHULTZE, JOACHIM LUDWIG;
BEYER, MARC DANIEL;
WARNER, NOEL y
HINGORANI, RAVI**

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

Observaciones :

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 666 206 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

Genes marcadores novedosos para linfocitos T reguladores de sangre humana

La presente invención proporciona genes marcadores novedosos para la identificación y la caracterización específicas de linfocitos T supresores y/o reguladores humanos, incluyendo linfocitos T CD4⁺CD25⁺FOXP3⁺ naturales, adaptativos y expandidos en individuos sanos, así como en pacientes con tumores o pacientes con enfermedades autoinmunitarias.

Antecedentes de la invención

Ya en 1971, una población de células denominadas "supresoras" fue descrita por primera vez por Gershon y Kondo cuando transfirieron la tolerancia específica de antígeno a animales sin exposición previa mediante la transferencia de linfocitos T con experiencia antigénica (Gershon y Kondo, *Immunology*; 21: 903-914 (1971)). Debido a resultados contradictorios el concepto de supresión de linfocitos T, sin embargo, cayó en el olvido a finales de la década de los 80.

Sakaguchi y col. fueron los primeros en describir los ahora denominados linfocitos "reguladores" T (linfocitos T_{reg} mediante la identificación de una población de linfocitos T CD4⁺ que expresan altamente CD25 y que previenen la autoinmunidad en un modelo murino (Sakaguchi y col., *J. Immunol.* 155: 1151-1164 (1995)). En los años siguientes un número de informes iluminó los aspectos principales de la biología de los linfocitos T_{reg}, caracterizando diferentes subpoblaciones de linfocitos T con propiedades reguladoras incluyendo linfocitos T_{reg} CD4⁺CD25^{alto} de origen natural, linfocitos T_{reg} inducidos, por ejemplo linfocitos Tr1 y TH3, así como linfocitos T_{reg} CD4⁺CD25^{alto} adaptativos que se desarrollan en la periferia mediante la conversión de linfocitos T CD4⁺CD25⁻. Todas estas diferentes poblaciones de linfocitos T con función reguladora coexisten y contribuyen a la supresión inmunitaria (Mills y McGuirk, *Semin Immunol.*; 16: 107-117 (2004); Sakaguchi, *Annu. Rev. Immunol.*; 22: 531-562; Sakaguchi, *Nat. Immunol.*; 6: 345-352 (2005); Vigouroux y col., *Blood*; 104: 26-33 (2004)).

En el ratón, CD25 es un buen marcador para linfocitos T_{reg} ya que los animales se mantienen en condiciones libres de patógenos. Sin embargo, los seres humanos están constantemente expuestos a antígenos extraños que conducen a una fracción significativa de linfocitos T efectoros CD25⁺ recientemente activados. En la búsqueda de marcadores más específicos de linfocitos T_{reg}, se ha identificado que el factor de transcripción FOXP3 se expresa únicamente en linfocitos T_{reg} en el ratón (Fontenot y col., *Nat Immunol.*; 4: 330-336 (2003); Hori y col., *Science*; 299: 1057-1061 (2003); Khattri y col., *Nat Immunol.*; 4: 337-342 (2003)) y la expresión se ha propuesto como un marcador de linaje ya en linfocitos T_{reg} en desarrollo (Bennett y col., *Nat. Genet.*; 27: 20-21 (2001); Brunkow y col., *Nat. Genet.*; 27: 68-73 (2001)). Sin embargo, todavía se recomienda precaución sobre su especificidad ya que informes recientes en seres humanos demostraron la inducción de FOXP3 en linfocitos T activados convencionales sin actividad supresora (Allan y col., *J. Clin. Invest.*; 115: 3276-3284 (2005); Morgan y col., *Hum. Immunol.*; 66: 13-20 (2005); Walker y col., *J. Clin. Invest.*; 112: 1437-1443 (2003)). Modelos de ratones desarrollados recientemente sugieren que FOXP3 es obligatorio para la función supresora de los linfocitos T_{reg} y el establecimiento final de un fenotipo de linfocitos T_{reg}, mientras que la asignación de linaje aparentemente es independiente de la expresión de FOXP3 (Lin y col., *Nat. Immunol.*; 8: 359-368 (2007); Wan y Flavell, *Nature*; 445: 766-770 (2007); Williams y Rudensky, *Nat. Immunol.*; 8: 277-284 (2007)). Son características de los linfocitos T_{reg} CD4⁺CD25^{alto}FOXP3⁺ su estado anérgico, su capacidad para inhibir activamente linfocitos T CD4⁺CD25⁻, linfocitos T CD8⁺, DC (células dendríticas, por sus siglas en inglés), NK, NKT y células B en un contacto célula-célula y de una manera dependiente de la dosis (Azuma y col., *Cancer Res.*; 63: 4516-4520 (2003); Chen, *Front Biosci.*; 11: 1360-1370 (2006); Lim y col., *J. Immunol.*; 175: 4180-4183 (2005); Murakami y col., *Proc. Natl. Acad. Sci. USA.*; 99: 8832-8837 (2002); Romagnani y col., *Eur. J. Immunol.*; 35: 2452-2458 (2005); Trzonkowski y col., *Clin. Immunol.*; 112: 258-267 (2004)). Fenotípicamente los linfocitos T_{reg} CD4⁺CD25^{alto}FOXP3⁺ se caracterizan como linfocitos T de memoria con experiencia antigénica, aunque últimamente algunos informes han descrito linfocitos T CD4⁺CD25^{alto}FOXP3⁺ vírgenes en ratones, así como en seres humanos (Beyer y col., *Blood* (2006); Valmori y col., *J. Clin. Invest.*; 115: 1953-1962 (2005)). Entre los marcadores de superficie celular asociados al fenotipo y la función de linfocitos T_{reg}, la proteína asociada a linfocitos T citotóxicos 4 (CTLA4) y la proteína relacionada con TNFR inducida por glucocorticoides (GITR) son las moléculas más prominentes (McHugh y col., *Immunity*; 16: 311-323 (2002); Lee y col., *J. Exp. Med.*; 192: 295-302 (2000); Shimizu y col., *Nat. Immunol.*; 3: 135-142 (2002); Takahashi y col., *J. Exp. Med.*; 192: 303-310 (2000)). Adicionalmente, IL-10 y TGFβ, aunque rara vez se expresan *in vitro*, podrían tener una importancia funcional para los linfocitos T_{reg} *in vivo*, en particular en el contexto de la enfermedad (Hara y col., *J. Immunol.*; 166: 3789-3796 (2001); Nakamura y col., *J. Exp. Med.*; 194: 629-644 (2001)). Las materias más importantes de la investigación actual son la caracterización de defectos de linfocito T_{reg} en enfermedades autoinmunitarias y su papel en enfermedades infecciosas y en la tolerancia al trasplante, en particular después del trasplante alógeno de médula ósea (Belkaid y Rouse, *Nat. Immunol.*; 6: 353-360 (2005); Hoffmann y col., *Curr. Top Microbiol. Immunol.*; 293: 265-285 (2005); Paust y Cantor, *Immunol. Rev.*; 204: 195-207 (2005); Sakaguchi, *Annu. Rev. Immunol.*; 22: 531-562 (2004); Sakaguchi, *Nat. Immunol.*; 6: 345-352 (2005); Waldmann y col., *Semin. Immunol.*; 16: 119-126 (2004)). Aunque la investigación en la biología de los linfocitos T_{reg} se está intensificando, todavía no está claro si los linfocitos T_{reg} en los seres humanos, en particular en el contexto de las enfermedades humanas, se ceban principalmente en el timo o emergen en la periferia debido a la estimulación específica de antígeno. La falta de marcadores de superficie celular más específicos es una de las principales razones por las que muchos aspectos funcionalmente relevantes de los linfocitos T_{reg} siguen siendo desconocidos.

Los linfocitos T_{reg} protegen al hospedador de la enfermedad autoinmunitaria mediante la supresión de células autorreactivas. Como tales, los linfocitos T_{reg} también pueden bloquear las respuestas inmunitarias antitumorales. En particular, en el contexto del cáncer, las frecuencias y la función de los linfocitos T_{reg} son importantes ya que el aumento del número podría favorecer el desarrollo o crecimiento tumoral e influir en el curso de la enfermedad. En la actualidad, un número de preguntas importantes son objeto de intensa investigación. ¿Es el aumento de las frecuencias de los linfocitos T_{reg} un evento temprano en la aparición de la enfermedad o, más probablemente, una respuesta del sistema inmunitario durante la progresión tumoral? ¿Los linfocitos T_{reg} específicos de órganos y, más importantes, los específicos de tumores desempeñan un papel? ¿Cómo influye la terapia en el número de linfocitos T_{reg}, en particular en tumores ya establecidos? ¿Existe una posibilidad de agotamiento a largo plazo de linfocitos T_{reg} y está esto conectado a la inducción de la autoinmunidad?

Un número de modelos de ratón pudieron establecer el desarrollo de linfocitos T_{reg} durante la progresión tumoral (Peng y col., *J. Immunol.*; 169: 4811-4821 (2002)). La inducción relativamente temprana de linfocitos T_{reg} durante el desarrollo tumoral tiene un impacto significativo en enfermedades humanas ya que el punto temporal de la inducción de linfocitos T_{reg} en pacientes con cáncer sin duda precede al momento del diagnóstico en la mayoría de los pacientes. Además, se estableció un efecto supresor de los linfocitos T_{reg} de origen natural frente a linfocitos T CD8⁺ específicos de tumor en un modelo de melanoma B16 poco inmunógeno (Turk y col., *J. Exp Med.*; 200: 771-782 (2004)). Se establecieron pruebas adicionales de la interferencia de los linfocitos T_{reg} con respuestas inmunitarias antitumorales mediadas por linfocitos T CD8 *in vivo* en un modelo de carcinoma de colon murino transgénico donde los linfocitos T_{reg} anulaban el rechazo del tumor mediado por linfocitos T CD8⁺ mediante la supresión específica de la citotoxicidad de LTC (linfocitos T citotóxicos, CTL por sus siglas en inglés) (Chen y col., *Proc. Natl. Acad. Sci. USA.*; 102: 419-424 (2005)). Se ha sugerido que la supresión de la inmunidad antitumoral por linfocitos T_{reg} se produce predominantemente en el sitio tumoral y que la inversión local de la supresión, incluso tarde durante el desarrollo tumoral, puede ser un tratamiento eficaz (Yu y col., *J. Exp. Med.*; 201: 779-791 (2005)).

El análisis del GL que drena tumor demostró que tanto los linfocitos T efectores antitumorales y los linfocitos T_{reg} FOXP3⁺ se ceban en el mismo GL durante la progresión tumoral. Estos linfocitos T_{reg} específicos de antígenos tumorales poseían las mismas propiedades funcionales que los linfocitos T_{reg} que surgen naturalmente en el timo (Hiura y col., *J. Immunol.*; 175: 5058-5066 (2005)).

Ya antes de la identificación de los linfocitos T_{reg} CD4⁺CD25⁺ los datos iniciales indicaron que el agotamiento no específico de los linfocitos T CD4⁺ puede conducir a la inducción de la inmunidad antitumoral eficiente (Fu y col., *Int. J. Cancer*; 87: 680-687 (2000)). Más específicamente, el direccionamiento a linfocitos T_{reg} mediante la administración de mAb CD25 anuló la falta de respuesta inmunitaria a los tumores e indujo el desarrollo espontáneo de linfocitos T efectores CD8⁺ específicos de tumor y linfocitos NK (Shimizu y col., *J. Immunol.*; 163: 5211-5218 (1999)). Curiosamente, el agotamiento de linfocitos T_{reg} condujo a la inmunidad tumoral con reactividad cruzada contra tumores de diversos orígenes (Golgher y col., *Eur. J. Immunol.*; 32: 3267-3275 (2002)). El momento de la eliminación de los linfocitos T_{reg} también parece ser un aspecto importante. La administración de mAb CD25 más tarde de 2 días después de la inoculación de células de mieloma no provocó ninguna regresión tumoral, independientemente del agotamiento de los linfocitos T_{reg} (Onizuka y col., *Cancer Res*; 59: 3128-3133; (1999)). Como ya se ha esbozado, esto podría deberse a la inducción de tolerancia antitumoral en un punto temporal relativamente temprano del desarrollo tumoral que da como resultado la activación ineficiente de las células efectoras. Además, el número de linfocitos T_{reg} después del agotamiento de CD25 se restablece en el tiempo y la capacidad de montar una respuesta antitumoral disminuye progresivamente (Casares y col., *J. Immunol.*; 171: 5931-5939 (2003)).

El agotamiento de los linfocitos T_{reg} junto con otros enfoques inmunoestimuladores, por ejemplo, el bloqueo de CTLA4, también se ha sometido a ensayo. La combinación del agotamiento de los linfocitos T_{reg} y el bloqueo de CTLA4 fue sinérgica y dio como resultado el máximo rechazo del tumor. La sinergia observada indica que ambas vías representan dos alternativas para la supresión de linfocitos T autorreactivos de manera que la intervención simultánea podría ser un concepto prometedor para la inducción de inmunidad antitumoral terapéutica (Sutmuller y col., *J. Exp. Med.*; 194: 823-832 (2001)). Puesto que las respuestas inmunitarias a los tumores malignos con frecuencia son débiles e ineficaces, solamente agotar los linfocitos T_{reg} podría no siempre dar como resultado la regresión tumoral. Por tanto, podrían ser más beneficiosos enfoques que combinen el agotamiento de los linfocitos T_{reg} con otras intervenciones inmunitarias, por ejemplo, la transferencia de linfocitos T activados o vacunas a base de DC (células dendríticas) (Prasad y col., *J. Immunol.*; 174: 90-98 (2005); Tanaka y col., *J. Immunother.*; 25: 207-217; Van Meirvenne y col., *Mol. Ther.*; 12: 922-932 (2005)).

Durante mucho tiempo se ha reconocido que la ciclofosfamida ejerce un efecto inmunoestimulador (Greenberg y col., *J. Exp. Med.*; 154: 952-963 (1981)). Los datos iniciales indicaron que la ciclofosfamida destruye preferentemente los linfocitos T supresores CD4⁺ provocando la regresión mediada inmunológicamente de linfomas inmunógenos en ratones (Awwad y North, *Cancer Res.*; 49: 1649-1654 (1989)). En un modelo de cáncer de colon en rata la administración de ciclofosfamida agotó los linfocitos T_{reg} y retrasó la excrecencia tumoral (Ghiringhelli y col., *Eur. J. Immunol.*; 34: 336-344 (2004)). La combinación de ciclofosfamida e inmunoterapia incluso curó los ratones, mientras que ambas estrategias aplicadas solas no tuvieron ningún efecto curativo (Ercolini y col., *J. Exp. Med.*; 201: 1591-1602 (2005); Ghiringhelli y col., *Eur. J. Immunol.*; 34: 336-344 (2004)). La ciclofosfamida a dosis bajas no solo disminuye el número de linfocitos T_{reg} sino que también conduce a la disminución de la función, a la potenciación de la apoptosis y a la disminución de la proliferación homeostática (Lutsiak y col., *Blood*; 105: 2862-2868 (2005)). Esto sugiere que la ciclofosfamida puede integrarse satisfactoriamente en la quimioinmunoterapia como se demostró recientemente por Dudley y col. en *Science*; 298: 850-854 (2002)). La combinación de la transferencia adoptiva de linfocitos T específicos de tumor activados *ex vivo* a pacientes de melanoma linfopénicos después de la quimioterapia con ciclofosfamida y fludarabina indujo la regresión del tumor en hasta el 50 % de los pacientes

tratados.

Puesto que los linfocitos T_{reg} son un mecanismo celular importante que suprime que los linfocitos T convencionales específicos de auto-antígeno ataquen a los tejidos propios, el agotamiento no específico de estas células podría ser un enfoque demasiado burdo para utilizarse sin conducir a un daño colateral significativo. El equilibrio fino entre el beneficio y el daño de manipular linfocitos T_{reg} se demostró elegantemente en el siguiente experimento: la transferencia de una mezcla de linfocitos T CD4⁺CD25⁻ y CD4⁺CD25⁺ evitó la inmunoterapia adoptiva eficaz de un melanoma establecido. Por el contrario, la transferencia adoptiva de linfocitos T CD4⁺CD25⁻ junto con linfocitos T CD8⁺ tumorales, así como autorreactivos, dentro de hospedadores deficientes en linfocitos T CD4⁺, seguida de vacunación, indujeron tanto la regresión del melanoma establecido como también la autoinmunidad grave y no deseada (Antony y col., *J. Immunol.*; 174: 2591-2601 (2005)). De forma similar, el agotamiento de linfocitos T_{reg} con mAb CD25 en un modelo de tumor de glándula mamaria dio como resultado la regresión del tumor, pero aumentó significativamente la susceptibilidad a la tiroiditis autoinmune. Esta sensibilización *in vivo* tanto a tumores como a autoantígenos atestigua la presencia de efectores inmunitarios de otro modo indetectables que se encuentran bajo regulación negativa y demuestra que la modulación de linfocitos T_{reg} es una estrategia poderosa en la terapia contra el cáncer, pero también puede aumentar significativamente las complicaciones autoinmunitarias (Wei y col., *Cancer Res.*; 65: 8471-8478 (2005); Wei y col., *Cancer Immunol. Immunother.*; 53: 73-78 (2004)).

Linfocitos T reguladores humanos en el cáncer - conocimiento actual y preguntas abiertas. Ya en la década de 1990 se notificaron linfocitos T con función reguladora en pacientes con cáncer, sin embargo, e estos informes no les siguió la identificación de los linfocitos T_{reg} CD4⁺CD25⁺ a mediados de 1990 (Sakaguchi y col., *J. Immunol.*; 155: 1151-1164 (1995)). Desde entonces, se ha notificado un aumento de linfocitos T_{reg} en pacientes con cáncer por numerosos investigadores. A diferencia del sistema murino, la definición de linfocitos T_{reg} humanos ha sido más difícil y la evaluación del marcador más específico, a saber, FOXP3, no se ha realizado en muchos de los estudios iniciales. Aunque los linfocitos T CD4⁺CD25^{alto} humanos son los más enriquecidos en linfocitos T FOXP3⁺, todavía hay un número significativo de células FOXP3⁺ dentro de la población de linfocitos T CD4⁺CD25^{bajo}. En ausencia de marcadores de superficie celular más específicos, todavía no es posible estudiar los linfocitos T_{reg} FOXP3⁺ humanos, independientemente de su expresión de CD25. Estas limitaciones también explican por qué actualmente es necesario caracterizar los linfocitos T_{reg} en los seres humanos mediante una combinación de la expresión de FOXP3 y de CD25, así como el análisis de la función inhibidora de poblaciones de linfocitos T enriquecidos en células FOXP3⁺, principalmente mediante la clasificación linfocitos T CD25^{alto}.

La comparabilidad de los informes anteriores se ve desafiada aún más mediante el uso de diferentes anticuerpos para detectar CD25 o diferentes estrategias de activación en la evaluación de células CD25⁺/CD25^{alto}. De forma similar, la función de los linfocitos T_{reg} se ha evaluado con numerosos enfoques *in vitro* haciendo bastante difícil comparar los resultados de diferentes estudios.

Woo y col. fueron los primeros en informar de un aumento de porcentajes de linfocitos T_{reg} CD4⁺CD25⁺ en LIT en el cáncer de pulmón no microcítico y el cáncer de ovario (Woo y col., *Cancer Res.*; 61: 4766-4772, (2001)). Se demostró que estos linfocitos T_{reg} secretan TGFβ, proporcionando una primera evidencia de que los linfocitos T_{reg} contribuyen a la disfunción inmunitaria en pacientes con cáncer (Woo y col., *Cancer Res.*; 61: 4766-4772, (2001)). La caracterización adicional de estas células mostró una expresión constitutiva de alto nivel de LTCA-4. Más importante, los linfocitos T_{reg} mediaron una inhibición potente de la proliferación de linfocitos T (Woo y col., *J. Immunol.*; 168: 4272-4276 (2002)). Apoyando este informe inicial, un estudio más amplio concluyó que la prevalencia de linfocitos T_{reg} CD4⁺CD25⁺ aumenta no solo en el microambiente tumoral de pacientes con cáncer de mama o de páncreas invasivos, sino también en SP, lo que sugiere que el aumento de linfocitos T_{reg} es un fenómeno generalizado (Liyanage y col., *J. Immunol.*; 169: 2756-2761 (2002)).

En el melanoma maligno se observó un aumento de linfocitos T_{reg} CD4⁺CD25⁺ funcionales (Javia y Rosenberg, *J. Immunother.*; 26: 85-93 (2003)), que se vinculó adicionalmente con los aumentos en el nivel sérico de H-ferritina (Gray y col., *Clin. Cancer Res.*; 9: 2551-2559 (2003)).

En pacientes con tumores malignos gastrointestinales el aumento relativo de linfocitos T_{reg} podría explicarse en realidad por una reducción significativa de linfocitos T CD4⁺CD25⁻. Curiosamente, en pacientes con carcinoma gástrico, el mal pronóstico y la disminución de las tasas de supervivencia se correlacionaron estrechamente con mayores frecuencias de linfocitos T_{reg} (Ichihara y col., *Clin Cancer Res.*; 9: 4404-4408 (2003); Sasada y col., *Cancer*; 98: 1089-1099 (2003)). Después de resecciones curativas, el número previamente elevado de linfocitos T_{reg} se redujo significativamente. Por el contrario, la prevalencia de linfocitos T_{reg} volvió a aumentar en los pacientes que recaen después de la resección del tumor (Kono y col., *Cancer Immunol. Immunother.*; 1-8 (2005)). Estos hallazgos subrayan la estrecha correlación entre el crecimiento tumoral y las frecuencias de los linfocitos T_{reg}.

Curiel y col. demostraron que los linfocitos T_{reg} CD4⁺CD25⁺FOXP3⁺ suprimen la inmunidad de linfocitos T específicos de tumor en el cáncer de ovario, contribuyen al crecimiento tumoral y se acumulan durante la progresión (Curiel y col., *Nat Med.*; 10: 942-949 (2004)). Además, el aumento de las frecuencias de linfocitos T_{reg} se asocia a una tasa alta de riesgo de muerte y una supervivencia reducida. Los linfocitos T_{reg} preferentemente se trasladaron y se acumularon en los tumores y la ascitis, pero rara vez entraron en el GL drenante en estadios posteriores del cáncer. Las células tumorales y los macrófagos circundantes produjeron la quimiocina CCL22, que medió el tráfico de los linfocitos T_{reg} al tumor a través de CCR4. Este reclutamiento específico de linfocitos T_{reg} podría representar un mecanismo mediante el cual los tumores podrían fomentar el privilegio inmunitario.

Para pacientes con carcinoma de células escamosas de la cabeza y el cuello se demostró una frecuencia significativamente elevada de linfocitos T_{reg} FOXP3⁺GITR⁺ (Schaefer y col., *Br. J. Cancer*; 92: 913-920 (2005)). Estos linfocitos T_{reg} fueron significativamente más sensibles a la apoptosis que los linfocitos no T_{reg} que podrían referirse

indirectamente a una rápida renovación en la circulación periférica (Schaefer y col., *Br. J. Cancer*; 92: 913-920 (2005)). Sin embargo, no se ha abordado todavía cómo influye la mayor sensibilidad a la apoptosis sobre las frecuencias de linfocitos T_{reg}.

También se ha notificado un aumento del número de linfocitos T_{reg} en SP y LIT de pacientes con carcinoma hepatocelular (Ormandy y col., *Cancer Res.*; 65: 2457-2464 (2005)). Aunque el aumento de linfocitos T_{reg} parece ser un tema común en los tumores sólidos, existen claras diferencias, pero aún no explicadas, entre entidades tumorales individuales. En un estudio comparativo se demostraron diferencias en las frecuencias de linfocitos T_{reg} en los derrames pleurales malignos de pacientes con mesotelioma en comparación con los derrames pleurales cancerosos de pacientes con cáncer de pulmón no microcítico o cáncer de mama (DeLong y col., *Cancer Biol. Ther.*; 4: 342-346 (2005)).

En general, el trabajo previo ha establecido claramente que los linfocitos T_{reg} están aumentados en la mayoría de los tumores sólidos humanos. Además, parece que hay un aumento de linfocitos T_{reg} dependiente del estadio, correlacionándose probablemente los linfocitos T_{reg} con la supervivencia global. Sin embargo, poco se sabe acerca de los mecanismos que conducen a este aumento. Un primer estudio de Wolf y col. podría ayudar a comprender los mecanismos moleculares subyacentes (Wolf y col., *Cancer Immunol. Immunother.*; 1-11 (2005)). En este estudio se demostró que el aumento de las frecuencias de linfocitos T_{reg} en SP de pacientes con cáncer se deben a la proliferación activa en lugar de a la redistribución desde otros compartimentos (es decir, órganos linfoides secundarios o médula ósea). Este hallazgo, en combinación con la atracción propuesta de los linfocitos T_{reg} al tumor a través de CCL22/CCR4 y la inducción de linfocitos T_{reg} mediante PGE₂ o H-ferritina, podría ser un posible mecanismo responsable de la expansión de linfocitos T_{reg} en pacientes con cáncer.

Aunque la cuestión de los linfocitos T_{reg} en tumores sólidos despertó el interés relativamente pronto, solo recientemente se han realizado estudios acerca de linfocitos T_{reg} en neoplasias hemáticas.

En pacientes con leucemia linfocítica crónica (LLC) de células B podría establecerse un aumento dependiente del estadio de linfocitos T_{reg} CD4⁺CD25^{alto}FOXP3⁺CTLA4⁺GITR⁺ con capacidad supresora completa. Sin embargo, cuando los pacientes con LLC se trataron con fludarabina las frecuencias de linfocitos T_{reg} disminuyeron y los linfocitos T_{reg} mostraron un deterioro de la función. Los estudios en curso están abordando la cuestión de cómo media la fludarabina este efecto (Beyer y col., *Blood*; 106: 2018-2025 (2005)). El aumento de linfocitos T_{reg} CTLA4⁺ en pacientes con LLC no tratados que se correlacionó con un estadio avanzado de la enfermedad y una citogenética desfavorable fueron confirmados recientemente por otros (Motta y col., *Leukemia*; 19: 1788-1793 (2005)). De forma similar, en pacientes con linfomas no Hodgkin de células B (LNH-B) se han observado aumentos de frecuencias de linfocitos T_{reg} FOXP3⁺CTLA4⁺. La expresión de PD1 fue parcialmente responsable de la actividad supresora de estos linfocitos T_{reg} infiltrantes de GL. Además, como se notificó para el cáncer de ovario, las células tumorales liberaron CCL22 y, de ese modo, atrajeron linfocitos T_{reg} CCR4⁺ al área del linfoma (Yang y col., *Blood* (2006)). Para pacientes con leucemia mieloide aguda (LMA) se han observado frecuencias más altas de T_{reg} CD4⁺CD25^{alto}. De forma similar a las observaciones en tumores sólidos, los linfocitos T_{reg} de pacientes con LMA fueron menos resistentes a la apoptosis, pero mostraron una proliferación mayor en comparación con los individuos sanos (Wang y col., *Eur. J. Haematol.*; 75: 468-476 (2005)).

De forma comparable a otras neoplasias hematológicas, se pudo demostrar el aumento de las frecuencias de linfocitos T_{reg} CD4⁺CD25^{alto}FOXP3⁺ en pacientes con gammapatía monoclonal de significado incierto (GMSI) o mieloma múltiple (MM) (Beyer y col., *Blood* (2006)). Independiente de la terapia previa o del estadio de la enfermedad, los linfocitos T_{reg} presentaron una fuerte capacidad inhibitoria. Además, el aumento de los linfocitos T_{reg} también fue dependiente del estadio y dio como resultado una expansión periférica. Además, pudo establecerse por primera vez una expansión de linfocitos T_{reg} CD4⁺CD25^{alto}FOXP3⁺ vírgenes que coexpresaban CD45RA y CCR7, apoyando adicionalmente el concepto de la expansión periférica de este compartimiento de linfocitos T. La importancia de la identificación de marcadores más específicos, así como de ensayos funcionales más normalizados es apoyada por un informe reciente sobre células FOXP3⁺ en SP de pacientes con MM (Prabhala y col., *Blood*; 107: 301-304 (2006)). Debido a un enfoque experimental alternativo que solamente evalúa FOXP3 en el contexto de linfocitos T CD4⁺ pero no de linfocitos CD25⁺, estos datos son difíciles de comparar con otros estudios acerca de linfocitos T_{reg} CD4⁺CD25^{alto} de origen natural. Aunque este informe llegó a la conclusión de que los linfocitos T_{reg} son disfuncionales en pacientes con MM, los ensayos elegidos para evaluar la función de los linfocitos T_{reg} permitieron explicaciones alternativas de los resultados observados, incluyendo defectos ya descritos en linfocitos T autólogos convencionales en pacientes con MM (Mariani y col., *Br. J. Haematol.*; 113: 1051-1059 (2001)). Para conciliar estos hallazgos recientes sobre linfocitos T_{reg} es más importante identificar marcadores de superficie celular específicos para los linfocitos T_{reg} que permitan aislar estas células y someterlas a ensayo funcionalmente en el contexto de una enfermedad maligna.

En su conjunto, el concepto de linfocitos T_{reg} aumentados se ha establecido para tumores sólidos, así como para neoplasias hemáticas. Sin embargo, se han de desarrollar marcadores más específicos tales como FOXP3, así como ensayos funcionales más sofisticadas y normalizados.

Como ya se ha esbozado, se ha notificado una correlación de linfocitos T_{reg} aumentados con una mayor carga de enfermedad y una supervivencia global peor. En LLC, pudieron observarse frecuencias reducidas de linfocitos T_{reg} funcionalmente deteriorados después del tratamiento con fludarabina, sin embargo, no todos los agentes quimioterápicos parecen inducir este efecto ya que en la LLC o el MM ningún otro tratamiento, incluyendo el trasplante autólogo de células madre, indujo efectos similares. En consonancia con esta observación, la frecuencia y la función supresora de los linfocitos T_{reg} en el GL que drena tumor derivado de pacientes con cáncer de cuello uterino no se vieron influenciadas por la quimioterapia o la quimiorradiación combinada (Fattorossi y col., *Cancer*; 100: 1418-1428 (2004)).

Un trabajo reciente ha demostrado que se requiere la señalización de IL-2 para el desarrollo del timo, la expansión periférica y actividad supresora de los linfocitos T_{reg} (Malek y Bayer, *Nat. Rev. Immunol.*; 4: 665-674 (2004)). Durante la reconstitución inmunitaria después de la quimioterapia, la terapia con IL2 condujo a una expansión periférica homeostática de linfocitos T_{reg} y a un compartimento de linfocitos T_{reg} marcadamente aumentado. La terapia con IL2 indujo la expansión de linfocitos T_{reg} existentes en hospedadores normales y esta expansión se aumentó adicionalmente por la linfopenia. Los linfocitos T_{reg} generados por la terapia con IL2 expresaron FOXP3 en los niveles observados en individuos sanos y estos linfocitos T_{reg} también tuvieron una potencia similar, lo que sugiere que la IL2 y la linfopenia son moduladores de la homeostasia de los linfocitos T_{reg} (Zhang y col., *Nat Med.*; 11: 1238-1243 (2005)). De forma similar, en pacientes con melanoma o carcinoma de células renales (CCR) la frecuencia de linfocitos T_{reg} totalmente funcionales aumentó significativamente después del tratamiento con IL2, que también se acompañó de un aumento de FOXP3 demostrando que la administración de dosis altas de IL-2 aumenta la frecuencia de linfocitos T_{reg} FOXP3⁺ circulantes (Ahmadzadeh y Rosenberg, *Blood* (2005)). Esto también podría explicar por qué la terapia con IL-2 en pacientes con CCR aún no está a la altura de las expectativas, puesto que una inducción significativa de linfocitos T_{reg} podría contrarrestar los posibles efectos antitumorales de la IL-2.

Sorprendentemente, también se demostró que la vacunación de pacientes con melanoma con DC (células dendríticas) ya sea cargadas con péptidos sintéticos o lisados tumorales induce el aumento de frecuencias de linfocitos T_{reg}, concomitante con la expansión de LTC específicos de tumor. Aunque esto potencie la tolerancia antitumoral e influya negativamente en la inducción de respuestas inmunitarias antitumorales clínicamente eficientes, necesita más atención, puesto que los mecanismos de este fenómeno todavía no se comprenden (Chakraborty y col., *Hum. Immunol.*; 65: 794-802 (2004)).

Los modelos murinos han establecido que la eliminación selectiva de linfocitos T_{reg} solos o en combinación con otras opciones de tratamiento podría inducir la regresión de tumores ya establecidos. Se han iniciado estudios de primer piloto en pacientes con cáncer para eliminar selectivamente linfocitos T_{reg}. Un enfoque prometedor y específico podría ser el direccionamiento de CD25 sobre la superficie de linfocitos T_{reg}. Danull y col. usaron conjugado de IL2 y toxina diftérica DAB(389)IL2 (denileukin difitox, ONTAK) para eliminar selectivamente linfocitos T_{reg} que expresan CD25 de las CMSP de pacientes con cáncer sin inducir toxicidad en otras células que solo expresaban CD25 en niveles de intermedios a bajos (Dannull y col., *J. Clin. Invest.*; 115: 3623-3633 (2005)). DAB(389)IL2 redujo significativamente el número de linfocitos T_{reg} presentes en la SP de pacientes con CCR metastásico y anuló la actividad inmunosupresora mediada por linfocitos T_{reg} *in vivo*. Además, la eliminación de linfocitos T_{reg} seguida de la vacunación con DC (células dendríticas) transfectadas con ARN mejoró significativamente la estimulación de respuestas de linfocitos T específicos de tumor en comparación con la vacunación sola.

En resumen, este primer estudio clínico que elimina específicamente linfocitos T_{reg} ha mostrado resultados prometedores que necesitan evaluarse adicionalmente. Un aspecto importante de los futuros estudios será describir claramente la ventana terapéutica de la eliminación de linfocitos T_{reg} como un filtro importante del reconocimiento de autoantígenos.

En los seres humanos, la caracterización de linfocitos T_{reg} se ha centrado principalmente en la coexpresión de CD4 y CD25 mientras que el estado de diferenciación, las frecuencias de subtipos de linfocitos T_{reg}, por ejemplo, linfocitos T_{reg} naturales o inducidos, linfocitos Tr1 o TH3, están menos bien caracterizados. Para comprender y estudiar mejor la biología de los linfocitos T_{reg} en relación con el desarrollo y la progresión tumorales, es crítico identificar marcadores de superficie celular más específicas ya que la evaluación de FOXP3 no permite el ensayo funcional posterior de células FOXP3⁺.

Utilizando el análisis de la transcripción del genoma entero se identificaron 43 genes candidatos, lo que podría servir como dianas diagnósticas, así como terapéuticas, para futuros estudios clínicos. Entre estos genes candidatos, se especificaron adicionalmente seis transcritos que eran hasta ahora solo se habían descrito como productos proteicos expresados supuestamente y que se descubrió que eran específicos para linfocitos T_{reg} humanos.

Sumario de la invención

La presente invención proporciona

- (1) un procedimiento de identificación o caracterización de linfocitos T humanos supresores y reguladores en una muestra del paciente, en el que el procedimiento comprende la detección de la expresión de una secuencia seleccionada entre las SEQ ID NO: 1 o 2;
- (2) una realización preferida de (1) anterior, que comprende (a) poner en contacto una población celular obtenida de sangre humana con uno o más ligandos que se unen específicamente a una o más proteínas marcadoras que comprenden una secuencia de la SEQ ID NO: 2 (denominándose abreviadamente dicha proteína marcadora en lo sucesivo en el presente documento "Colonia 1" o "Col 1") y (b) identificar o caracterizar los linfocitos T humanos supresores y reguladores de la población celular;
- (3) un procedimiento para aislar linfocitos T humanos supresores y reguladores de una muestra del paciente, en el que el procedimiento comprende (a) poner en contacto una población celular obtenida de sangre humana con uno o más ligandos que se unen específicamente a una o más proteínas marcadoras que comprenden una secuencia de la SEQ ID NO: 2 y (b) aislar los linfocitos T humanos supresores y reguladores de la población celular;
- (4) una realización preferida de (2) o (3) anteriores, en la que los ligandos son anticuerpos, preferentemente son anticuerpos monoclonales o fragmentos de los mismos que se unen a una o más proteínas marcadoras que comprenden una secuencia de la SEQ ID NO: 2;

(5) el uso de un kit que comprende uno o más ligandos que se unen específicamente a una o más proteínas marcadoras como se han definido en (2) o (3) anteriormente para el aislamiento, la identificación o la caracterización de linfocitos T humanos supresores y reguladores de una muestra del paciente, siendo dicho kit en particular útil para la realización del procedimiento de (2), (3) o (4) anteriores; y

5 (6) el uso del anticuerpo o fragmento de anticuerpo como se ha definido en (4) anteriormente para aislar o caracterizar los linfocitos T humanos supresores y reguladores de una muestra del paciente.

Breve descripción de las figuras

Figura 1: Visión general del estudio para identificar los genes del Ejemplo 1

10 Figura 2: Gráfico de contorno de FOXP3 que representa los valores de expresión sobre todo el conjunto de datos. Los linfocitos T convencionales se muestran en color azul, los linfocitos T_{reg} se muestran en color rojo.

Figura 3: Gráfico de contorno tridimensional de genes incluidos en el transcriptoma núcleo de T_{reg}

15 Figura 4: Potencial inhibidor de linfocitos T reguladores. Se estimulan linfocitos T CD4⁺ convencionales con CMSP alógenas solas o en combinación con diferentes concentraciones de linfocitos T_{reg}. La proliferación de linfocitos T se midió mediante la incorporación de BrdU.

Figura 5: Estrategia de clasificación para CD4⁺CD25^{alto} frente a linfocitos T CD4⁺CD25⁻.

20 Figura 6: Estrategia de clasificación para linfocitos T_{reg} frente a linfocitos T CD4⁺ convencionales que integran CD127. En esta figura la estrategia de clasificación para linfocitos T_{reg} frente a linfocitos T CD4⁺ convencionales se muestra integrando CD127 como marcador de linaje.

Figura 7: El análisis de agrupamiento se realizó en genes variables.

25 Figura 8: Perfiles de expresión de 3 genes marcadores de linfocitos T_{reg} conocidos. Con fines de selección se generaron gráficos de contorno de perfiles de expresión de genes seleccionados. Se representaron valores de expresión en todo el conjunto de datos. Los linfocitos T convencionales se muestran en color azul, los linfocitos T_{reg} se muestran en color rojo. Se muestran tres genes marcadores conocidos para linfocitos T_{reg}: CD25, FOXP3 y CD127.

30 Figura 9: Perfil de expresión de los 6 genes marcadores de linfocitos T_{reg} identificados. Se representan los valores de expresión relativos en las condiciones dadas. A: FANK1, B: PLEKHK1, C: HPGD, D: DNAPTP6, E: Col5, F: Col6.

Figura 10: Descripción de los 6 genes marcadores. La localización de los 6 genes en el cromosoma, la longitud del ARNm, la secuencia codificante, así como la longitud de la proteína se describen en las Figuras 10A-F (A: FANK1, B: PLEKHK1, C: HPGD, D: DNAPTP6, E: Col5, F: Col6).

35 Figura 11: Mapas del vector pTrcHis-TOPO que incluye los 6 genes con codón de parada. En la Figura 11A se muestra un mapa del vector pTrcHis-TOPO vacío. Este vector (4390 pb) contiene un promotor trc, un ATG de iniciación, un marcador 6xHis y un epítipo Xpress y un gen de resistencia a ampicilina. Por tanto, todas las proteínas clonadas en este vector están en el extremo N terminal marcado con His y Xpress. Las figuras 11B-G representan el vector que incluye las secuencias de ADNc, incluyendo el codón de parada de los 6 genes (B: FANK1, C: PLEKHK1, D: HPGD, E: DNAPTP6, F: Col5, G: col6; las secuencias exactas de 11B-G se proporcionan en las SEQ ID NO: 141-146).

40 Figura 12: Mapas del vector pTrcHis-TOPO que incluye los 6 genes sin codón de parada. Las Figuras 12A-F representan el vector pTrcHis-TOPO con las secuencias de ADNc que excluyen el codón de parada de los 6 genes (A: FANK1, B: PLEKHK1, C: HPGD, D: DNAPTP6, E: Col5, F: Col6). Estos vectores son adecuados para la clonación del ADNc sin codón de parada en diferentes vectores de expresión de mamíferos.

45 Figura 13: Mapas del vector pENTR4-eGFP que incluye los 6 genes sin codón de parada. En la Figura 13A se muestra un mapa del vector pENTR4 modificado con una molécula de eGFP. Este vector (3441 pb) contiene un origen pUC, un marcador eGFP, un gen de resistencia a la kanamicina y dos sitios attL para la recombinación específica de sitio en vectores de destino. Por tanto, todas las proteínas clonadas en este vector están en el extremo C marcado con eGFP. Las figuras 13B-G representan el vector que incluye las secuencias de ADNc que excluyen el codón de parada de los 6 genes (B: FANK1, C: PLEKHK1, D: HPGD, E: DNAPTP6, F: Col5, G: Col6).

50 Figura 14: Mapas del vector pLenti4/V5-DEST que incluye los 6 genes sin codón de parada. En la Figura 14A se muestra un mapa del vector pLenti4/V5-DEST. Este vector (8634 pb) contiene un promotor CMV, una señal de empaquetado psi (ψ) de VIH-1, un elemento de respuesta Rev de HIV-1 (RRE, por sus siglas en inglés), un gen de resistencia a ampicilina para la selección bacteriana, un gen de resistencia a blasticidina para la selección de células de mamíferos infectadas y dos sitios attR para la recombinación específica de sitio en el vector de destino. Adicionalmente, en el extremo C terminal el vector incluye un epítipo V5. Por tanto, todas las proteínas clonadas en este vector están en el extremo C marcado con eGFP y V5. Las figuras 14B-G representan el vector que incluye las secuencias de ADNc excluyendo el codón de parada de los 6 genes (B: FANK1, C: PLEKHK1, D: HPGD, E: DNAPTP6, F: Col5, G: Col6).

55 Figura 15: Mapas del vector pcDNA6-HisA incluyendo los 6 genes sin codón de parada. En la Figura 15A se muestra un mapa del vector pcDNA6. Este vector (5100 pb) contiene un promotor CMV, un gen de resistencia a ampicilina para la selección bacteriana y un gen de resistencia a blasticidina para la selección de células de mamífero infectadas. Adicionalmente, en el extremo C terminal el vector incluye un epítipo V5, así como una cola de His 6x. Por tanto, todas las proteínas clonadas en este vector están en el extremo C marcado con V5 y His. Las Figuras 15B-G muestran vectores pcDNA6-HisA incluyendo los 6 genes (B: FANK1, C: PLEKHK1, D: HPGD, E: DNAPTP6, F: Col5, G: Col6).

Descripción detallada de la invención

Los aspectos (1) y (2) de la invención se refieren a un procedimiento para la identificación o caracterización de linfocitos T supresores y reguladores que expresan una proteína marcador específico. A continuación, se describen procedimientos y etapas necesarios para identificar genes específicos de linfocitos T_{reg}. Comienza con la descripción de los donantes y las diferentes estrategias de aislamiento para purificar linfocitos T reguladores CD4⁺CD25^{alto}, linfocitos T CD4⁺CD25⁻ convencionales, así como la inclusión de CD127 en una segunda estrategia para aislar linfocitos T_{reg} y linfocitos T CD4⁺CD127⁺CD25^{bajo} convencionales. Además, se representan las condiciones de cultivo *in vitro* para la estimulación de la población celular diferente, seguidas de un protocolo para preparar muestras de ARNc para la hibridación de micromatrices, que incluye una breve descripción de los procedimientos de análisis utilizados.

El uso de este enfoque bioinformático condujo a la identificación de un total de 43 genes (de los cuales 6 genes se eligieron y se sometieron a ensayo adicionalmente) que se expresa específicamente en linfocitos T_{reg} humanos. Estos genes son los siguientes (en lo sucesivo en el presente documento dichos genes y las proteínas codificadas de este modo se denominan "genes de la invención" y "proteínas de la invención"):

FANK1: FANK1 humano (dominios de repetición 1 de fibronectina tipo III y anquirina; en lo sucesivo en el presente documento también denominados "Colonia 1" o "Col1") se ubica en la región genómica 127575098-127688151 en el cromosoma 10 en la región 10q26.2. Consiste en 11 exones resultantes de la transcripción en un ARNm que tiene 1395 pares de bases (véase la SEQ ID NO: 1). La secuencia codificante en sí consiste en 1038 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 346 aminoácidos (SEQ ID NO: 2).

La proteína se exploró para determinar los dominios funcionales y Colonia 1 tiene un dominio de fibronectina de tipo 3 en la región de 9-94, que es uno de los tres tipos de repeticiones internas descubiertos en la proteína fibronectina plasmática posiblemente relacionados con la secreción extracelular y la localización de la fibronectina, así como 5 dominios de repetición de anquirina en las regiones 109-139, 143-172, 176-205, 209-238, 243-273, 277-306, que median las interacciones proteína-proteína en muy diversas familias de proteínas. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 3 regiones transmembrana potenciales (206-208, 222, 322-324). La proteína supuesta tiene una alta homología (> 88 %) a las secuencias de proteínas supuestas de FANK1 descritas para ratón y rata. El análisis de la expresión en diferentes tejidos demostró solamente la expresión a niveles bajos o ausentes de Colonia 1 en todos los tejidos analizados, incluyendo médula ósea, hígado, corazón, bazo, pulmón, riñón, timo, músculo esquelético, cerebro, médula espinal, próstata y páncreas.

Pueden determinarse fragmentos/ligandos del CMH de polipéptidos adecuados de acuerdo con Rammensee y col. (Rammensee y col., *Immunogenetics*; 41: 178-228 (1995)) (Rammensee y col., *Immunogenetics*; 50: 213-219 (1999)) y la base de datos SYFPEITHI (acceso a través de: www.syfpeithi.de).

En vista del número continuamente creciente de motivos, era necesario configurar una base de datos que facilite la búsqueda de péptidos y permita la predicción de epítomos de linfocitos T. La predicción se basa en motivos publicados (secuenciación por combinación, ligandos naturales) y toma en consideración los aminoácidos en el anclaje y las posiciones de anclaje auxiliares, así como otros aminoácidos frecuentes. La puntuación se calcula de acuerdo con las siguientes reglas: A los aminoácidos de un determinado péptido se les proporciona un valor específico en función de si son de anclaje, anclaje auxiliar o resto preferido. A los anclajes ideales se les proporcionarán 10 puntos, a los anclajes inusuales 6-8 puntos, a los anclajes auxiliares 4-6 puntos a los restos preferidos 1-4 puntos. A los aminoácidos que se considera que tienen un efecto negativo sobre la capacidad de unión se les proporcionan valores entre -1 y -3. En lo que se refiere a los epítomos de linfocitos T, solo se han seleccionado aquellos que son propensos a procesarse naturalmente.

Se muestran ligandos del CMH particulares de Colonia 1 en la SEQ ID NO: 152-255. PLEKHK1: PLEKHK1 Humano (miembro 1 de la familia K que contiene dominio de homología a pleckstrina, también conocido como rhotecina 2; en lo sucesivo en el presente documento también denominado "Colonia 2" o "Col2"), se ubica en la región genómica 63698472-63622959 en el cromosoma 10 en la región 10q21.2. Consiste en 12 exones resultantes de la transcripción en un ARNm que tiene 6659 pares de bases (véase la SEQ ID NO: 3, teniendo una variante adicional del mismo 2123 pares de bases que se muestran en la SEQ ID NO: 5). La secuencia codificante en sí consiste en 1830 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 609 aminoácidos (SEQ ID NO: 4).

Colonia 2 codifica una proteína altamente homóloga a la proteína efectora Rho-GTPasa humana, rhotecina. La rhotecina pertenece a la Clase 1 de proteínas efectoras Rho, así como PKN y rhofilina (Collier y col., *Biochem. Biophys. Res. Commun.*; 324: 1360-1369 (2004)).

La proteína se exploró para determinar los dominios funcionales y Colonia 2 tiene un dominio de homología a Pleckstrina (PH, por sus siglas en inglés) en la región 295-391, que con frecuencia está implicado en el direccionamiento de las proteínas a la membrana plasmática, pero pocos muestran una fuerte especificidad en la

unión a lípidos. Cualquier especificidad por lo general se determina mediante regiones de bucle o inserciones en el extremo N del dominio, que no están conservados en todos los dominios pH. Los dominios PH se encuentran en proteínas de señalización celular, tales como serina/treonina cinasa, tirosina cinasas, reguladores de proteínas G, GTPasas endocíticas, adaptadores, así como moléculas asociadas al citoesqueleto y en enzimas asociadas a lípidos. Colonia 2, adicionalmente tiene un dominio efector Rho o de homólogos de la región 1 de homología a cinasa relacionada con la proteína cinasa C en la región 19-95, que se describió por primera vez como una región de homología repetida tres veces de la parte no catalítica del extremo N de la proteína cinasa PRK1 (PKN). Más tarde se demostró que las primeras dos de estas repeticiones se unen a la pequeña G proteína rho que se sabe que activa PKN en su forma unida a GTP. También aparecen dominios de unión a rho similares en un número de otras proteínas cinasas y en las proteínas de unión a rho rhotecina y rhotecina. Recientemente, se ha determinado por cristalografía de rayos X la estructura de la repetición REM N terminal que forma complejo con RhoA. Forma un pliegue superenrollado antiparalelo denominado un dedo ACC. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 4 regiones transmembrana potenciales (41-44, 197-199, 287-297, 300-301), mientras que la región de 289-294 tenía la probabilidad más alta. La proteína supuesta tiene una alta homología (> 73 %) a las secuencias de proteínas supuestas de PLEKHK1 descritas para ratón y rata. Varias variantes de corte y empalme se caracterizan por Collier y col. y, además, muestran una expresión selectiva de este gen en linfocitos T CD4⁺ en reposo. Esta proteína se encuentra principalmente en el citoplasma. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en el tejido pulmonar, mientras que solo se observó la expresión a niveles bajos o ausentes de Colonia 2 en todos los demás tejidos analizados, incluyendo médula ósea, hígado, corazón, bazo, riñón, timo, músculo esquelético, cerebro, la médula espinal, próstata y páncreas. Se sugiere que Colonia 2 puede ser un marcador de madurez con un papel importante en el desarrollo de linfocitos ((Collier y col., *Biochem. Biophys. Res. Commun.*; 324: 1360-1369 (2004))) y la regulación de este gen puede asociarse a la proliferación celular y/o la señalización de activación de linfocitos T, la división y la diferenciación de linfocitos.

Se pueden identificar ligandos/fragmentos del CMH de Colonia 2 adecuados mediante el procedimiento descrito en relación con Colonia 1 anterior.

Se muestran fragmentos particulares del CMH de Colonia 2 en la SEQ ID NO: 256-435.

HPGD: HPGD Humano (hidroxiprostaglandina deshidrogenasa 15-(NAD); en lo sucesivo en el presente documento también denominado "Colonia 3" o "Col3"), se ubica en la región genómica 175680186-175647955 en el cromosoma 4 en la región 4q34-35. Consiste en 7 exones resultantes de la transcripción en un ARNm que tiene 2592 pares de bases. La secuencia codificante en sí consiste en 801 pares de bases (véase la SEQ ID NO: 7). La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 266 aminoácidos (SEQ ID NO: 8).

Las prostaglandinas están implicadas en muchos procesos fisiológicos y celulares, tales como la inflamación (Pichaud y col., *Hum Genet.*; 99: 279-281 (1997)) y HPGD es la principal enzima de degradación de las prostaglandinas. Mediante la catálisis de la conversión del grupo 15-hidroxilo de las prostaglandinas en un grupo ceto, esta enzima ubicua reduce fuertemente la actividad biológica de estas moléculas.

La proteína se exploró para determinar los dominios funcionales y Colonia 3 tiene un dominio deshidrogenasa de cadena corta en la región 8-250. Esta familia contiene una amplia diversidad de deshidrogenasas. Colonia 3 es un homodímero situado principalmente en el citoplasma con la capacidad de catalizar la reacción de (5Z,13E)-(15S)-11-alfa,15-dihidroxi-9-oxoprost-13-enoato + NAD⁺ a (5Z,13E)-11-alfa-hidroxi-9,15-dioxoprost-13-enoato + NADH y da como resultado la inactivación de las prostaglandinas. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 5 regiones transmembrana potenciales (80, 162-170, 188-195, 238-240, 278-286), mientras que la región 164-167 tenían la probabilidad más alta. La proteína supuesta tiene una alta homología (> 88 %) a las secuencias de proteínas supuestas de Colonia 3 descritas para ratón y rata. El análisis de la expresión en diferentes tejidos demostró solamente la expresión a niveles bajos o ausentes de Colonia 3 en todos los tejidos analizados, incluyendo médula ósea, hígado, corazón, bazo, pulmón, riñón, timo, músculo esquelético, cerebro, médula espinal, próstata y páncreas.

Pueden identificarse ligandos/fragmentos adecuados del CMH de Colonia 3 mediante el procedimiento descrito en relación con Colonia 1 anterior.

Se muestran fragmentos particulares del CMH de Colonia 3 en la SEQ ID NO: 436-536.

DNAPTP6: DNAPTP6 humano (proteína 6 transactivada por ADN polimerasa; en lo sucesivo en el presente documento también denominada "Colonia 4" o "Col4"), se ubica en la región genómica 200879041-201051498 en el cromosoma 2 en la región 2q33.1. Consiste en 13 exones resultantes de la transcripción en un ARNm que tiene 2355 pares de bases (véase la SEQ ID NO: 9). La secuencia codificante en sí consiste en 1677 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 558 aminoácidos (SEQ ID NO: 10).

La proteína se exploró para determinar los dominios funcionales y Colonia 4 tiene un dominio DUF1387 en la región

59-368, que representa un dominio de función desconocida. Esta familia representa una región conservada de aproximadamente 300 restos de longitud dentro de un número de proteínas hipotéticas de función desconocida que parecen estar restringidas a mamíferos. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 4 regiones transmembrana potenciales (41-44, 72, 88, 409-416, 585-593), mientras que la región 587-591 tenía la probabilidad más alta. La proteína supuesta tiene una alta homología (> 89 %) a las secuencias de proteínas supuestas de DNAPTP6 descritas para ratón y rata. El análisis de la expresión en diferentes tejidos demostró solamente la expresión a niveles bajos o ausentes de Colonia 4 en todos los tejidos analizados, incluyendo médula ósea, hígado, corazón, bazo, pulmón, riñón, timo, músculo esquelético, cerebro, médula espinal, próstata y páncreas.

10 Pueden identificarse ligandos/fragmentos del CMH adecuados de Colonia 4 mediante el procedimiento descrito en relación con Colonia 1 anterior.

Se muestran fragmentos particulares del CMH de Colonia 4 en la SEQ ID NO: 537-660.

15 C1orf78: C1orf78 humano (marco de lectura abierto 78 del cromosoma 1; en lo sucesivo en el presente documento también denominado "Colonia 5" o "Col5"), se ubica en la región genómica 36562342-36560219 en el cromosoma 1 en la región 1p34.3. Consiste en 3 exones resultantes de la transcripción en un ARNm que tiene 1054 pares de bases (véase la SEQ ID NO: 11). La secuencia codificante en sí se compone de 495 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 164 aminoácidos (SEQ ID NO: 12).

20 La proteína se exploró para determinar los dominios funcionales y no pudieron identificarse dominios conocidos anteriormente para Colonia 5. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 1 región transmembrana potencial (68-93) en la que la región 69-92 tenía la probabilidad más alta. La proteína supuesta tiene una alta homología (> 89 %) a las secuencias de proteínas supuestas de Colonia 5 descritas para ratón y rata. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en tejidos de pulmón, bazo, timo y cerebro, mientras que solo se observó una expresión a niveles bajos o ausentes de Colonia 5 en todos los demás tejidos analizados, incluyendo médula ósea, hígado, corazón, riñón, músculo esquelético, médula espinal, próstata y páncreas.

25 Pueden identificarse ligandos/fragmentos adecuados del CMH de Colonia 5 mediante el procedimiento descrito en relación con Colonia 1 anterior. Se muestran fragmentos particulares del CMH de Colonia 5 en la SEQ ID NO: 661-708.

30 FLJ45983: FLJ45983 humano (proteína FLJ45983; en lo sucesivo en el presente documento también denominada "Colonia 6" o "Col6"), se ubica en la región genómica 8135453-8132419 en el cromosoma 10 en la región 10p14. Consiste en 2 exones resultantes de la transcripción en un ARNm que tiene 2213 pares de bases (véase la SEQ ID NO: 13). La secuencia codificante en sí se compone de 378 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 125 aminoácidos (SEQ ID NO: 14).

35 La proteína se exploró para determinar los dominios funcionales y no pudieron identificarse dominios conocidos anteriormente para Colonia 6. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 1 región transmembrana potencial (100-108). El producto de proteína supuesta de Colonia 6 no tiene una homología obvia a las secuencias de proteínas conocidas que se describen para ratón y rata. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en el tejido del riñón mientras que solo se observó una expresión a niveles bajos o ausentes de Colonia 6 en todos los demás tejidos analizados, incluyendo médula ósea, hígado, corazón, bazo, pulmón, timo, músculo esquelético, cerebro, médula espinal, próstata y páncreas.

40 Pueden identificarse ligandos/fragmentos del CMH adecuados de Colonia 6 mediante el procedimiento descrito en relación con Colonia 1 anterior.

45 Se muestran fragmentos particulares del CMH de Colonia 6 en la SEQ ID NO: 709-734.

50 ACTA2: ACTA2 Humano (actina, alfa 2, músculo liso, aorta) se ubica en la región genómica 90702490-90684810 en el cromosoma 10 en la región 10q23.3. Consiste en 9 exones resultantes de la transcripción en un ARNm que tiene 1330 pares de bases (véase la SEQ ID NO: 15). La secuencia codificante en sí consiste en 1134 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 377 aminoácidos (SEQ ID NO: 16).

Actina alfa 2, el gen de actina de músculo liso aórtico humano es una de las seis isoformas de actina diferentes que se han identificado. Las actinas son proteínas altamente conservadas que están implicadas en la motilidad, la estructura y la integridad celulares. Las actinas alfa son un constituyente principal del aparato contráctil.

55 La proteína se exploró para determinar los dominios funcionales y ACTA2 tiene un dominio de actina en la región 1 377, que pertenece a la subfamilia de actina de la superfamilia de actina/MreB/azúcarcinasa/Hsp70. Cualquier especificidad por lo general se determina mediante regiones bucle o inserciones en el extremo N del dominio, que no

están conservadas en todos los dominios PH. Los dominios PH se encuentran en proteínas de señalización celular, tales como serina/treonina cinasa, tirosina cinasas, reguladores de las proteínas G, GTPasas endocíticas, adaptadores, así como moléculas asociadas al citoesqueleto y en enzimas asociadas a lípidos. La actina es una proteína ubicua implicada en la formación de filamentos que son los componentes principales del citoesqueleto. Estos filamentos interactúan con miosina para producir un efecto de deslizamiento, que es la base de la contracción muscular y muchos aspectos de la motilidad celular, incluyendo la citocinesis. Cada promotor de actina se une a una molécula de ATP y tiene un sitio de alta afinidad para los iones de calcio o magnesio, así como varios sitios de baja afinidad. La actina existe como un monómero en bajas concentraciones de sal, pero se forman filamentos rápidamente a medida que la concentración de sal aumenta, con la consecuente hidrólisis de ATP. La actina de muchas fuentes forma un complejo estrecho con desoxirribonucleasa (DNasa I), aunque la importancia de esto aún se desconoce. La formación de este complejo da como resultado la inhibición de la actividad de DNasa I y la actina pierde su capacidad de polimerizarse. Se ha demostrado que un dominio de ATPasa de actina comparte similitud con dominios ATPasa de las proteínas hexocinasa y hsp70. En los vertebrados existen tres grupos de isoformas de actina: alfa, beta y gamma. Las actinas alfa se encuentran en los tejidos musculares y son un constituyente principal del aparato contráctil. Las actinas beta y gamma coexisten en la mayoría de tipos celulares como componentes del citoesqueleto y como mediadores de la motilidad celular interna. En las plantas existen muchas isoformas que probablemente están implicadas en una diversidad de funciones tales como la corriente citoplasmática, la determinación de la forma de la célula, el crecimiento de la punta, graviperception, deposición de la pared celular, etc. Recientemente se han identificado algunas proteínas similares a actina divergentes en varias especies. Estas proteínas incluyen contractina (actina-RPV) de mamíferos, hongos de levadura ACT5, ro-4 de *Neurospora crassa* y *Pneumocystis carinii*, que parece ser un componente de un complejo centrosómico de múltiples subunidades implicado en la motilidad de vesículas a base de microtúbulos (esta subfamilia se conoce como ARP1); subfamilia ARP2, que incluye ACTL de pollo, ACT2 de *Saccharomyces cerevisiae*, 14D de *Drosophila melanogaster* y ACTC de *Caenorhabditis elegans*; subfamilia ARP3, que incluye actina 2 de mamíferos, *Drosophila* 66B, ACT4 de levadura y ACT2 de *Schizosaccharomyces pombe*; y subfamilia ARP4, que incluye ACT3 de levadura y 13E de *Drosophila*. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 2 regiones transmembrana potenciales (133-145, 343-351), mientras que las regiones de 137-142 y 346-348 tenían la probabilidad más alta. La proteína tiene una alta homología (> 99 %) a las secuencias de proteínas de ACTA2 descritas para ratón y rata. Esta proteína se ubica principalmente en el citoplasma. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en el tejido del corazón mientras que solo se observó la expresión a niveles bajos o ausentes de ACTA2 en todos los demás tejidos analizados, incluyendo médula ósea, hígado, pulmón, bazo, riñón, timo, músculo esquelético, cerebro, médula espinal, próstata y páncreas.

Pueden identificarse ligandos/fragmentos adecuados del CMH de ACTA2 humano mediante el procedimiento descrito en relación con Colonia 1 anterior.

BFSP2: BFSP2 humana (proteína estructural de filamento en cuentas 2, facinina) se sitúa en la región genómica 134601479-134676745 en el cromosoma 3 en la región 3q21-q25. Consiste en 7 exones resultantes de la transcripción en un ARNm que tiene 1644 pares de bases (véase la SEQ ID NO: 17). La secuencia codificante en sí consiste en 1248 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 415 aminoácidos (SEQ ID NO: 18).

Más del 99 % de la lente ocular de los vertebrados está compuesta de células de fibra óptica diferenciadas terminalmente. Se expresan dos proteínas similares a filamentos intermedios específicas de lentes, el producto proteico de este gen (BFSP2) y filensina (también conocida como CP115), solo después de que haya comenzado la diferenciación celular de la fibra. Ambas proteínas se encuentran en un elemento citoesquelético estructuralmente único que se denomina filamento de cuentas (BF, del inglés *beaded filament*). Las mutaciones en este gen se han asociado a las cataratas progresivas de aparición juvenil y a la epidermólisis bullosa simple de Dowling-Meara.

La proteína se exploró para determinar los dominios funcionales y BFSP2 tiene 2 dominios superenrollados en la región 199-248 y 356-395. El superenrollamiento es un dominio proteico que forma un haz de dos o tres hélices alfa. Los dominios superenrollados cortos están implicados en las interacciones proteicas pero los dominios superenrollados largos que forman varillas largas se producen en proteínas estructurales o motoras. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 7 regiones transmembrana potenciales (10-77, 80-203, 213-249, 254-320, 322-339, 344-1093, 1098-1240), mientras que las regiones 34-65, 110-185, 355-1085 y 1124-1220 tenían la probabilidad más alta. La proteína tiene una alta homología (> 85 %) a las secuencias proteicas de BFSP2 descritas para ratón. Esta proteína se ubica principalmente en el citoplasma y la membrana plasmática. El análisis de la expresión en diferentes tejidos demostró una expresión de alto nivel en el tejido del ojo mientras que solo se observaron niveles bajos o expresión ausente de BFSP2 en todos los demás tejidos analizados, incluyendo médula ósea, hígado, pulmón, corazón, bazo, riñón, timo, músculo esquelético, cerebro, médula espinal, próstata y páncreas. El filamento de cuentas es una estructura citoesquelética que es abundante en las células de las fibras de la lente ocular. BFSP2 y CP115 son los principales componentes del filamento de cuentas. Alizadeh y col. (Alizadeh y col., *Invest. Ophthalmol. Vis. Sci.*; 45: 884-891 (2004)), caracterizaron el genotipo y el fenotipo de la cepa de ratón 129 (129/SvJ), que lleva una mutación en el gen BFSP2. La mutación consiste en una delección de 6303-pb desde el extremo del intrón B y el comienzo del exón 2, dando como resultado la pérdida del sitio aceptor de corte y empalme

del exón 2, la ausencia del exón 2 del ARNm de BFSP2 y niveles drásticamente reducidos de ARNm de BFSP2. Los niveles de transcripción de filensina (CP115), compañero de montaje de BFSP2, son normales, pero los niveles de proteína se reducen drásticamente. La microscopía óptica estableció que la diferenciación inicial y el alargamiento de las células de la fibra transcurrieron normalmente. La microscopía electrónica mostró la ausencia de filamentos de cuentas, mientras que la microscopía de lámpara de rendija mostró una pérdida progresiva y que emerge lentamente de claridad óptica. Por tanto, Alizadeh y col. (Alizadeh y col., *Invest. Ophthalmol. Vis. Sci.*; 45: 884-891 (2004)) llegaron a la conclusión de que la cepa 129 de ratón se comporta como un ratón nuligénico BFSP2 funcional.

Pueden identificarse ligandos/fragmentos adecuados del CMH de BFSP2 humano mediante el procedimiento descrito en relación con Colonia 1 anterior.

10 CCL5: CCL5 Humano (ligando de quimiocina (motivo C-C) 5), también conocido como RANTES, se ubica en la región genómica 31231489-31222607 en el cromosoma 17 en la región 17q11.2-q12. Consiste en 3 exones resultantes de la transcripción en un ARNm que tiene 1237 pares de bases (véase la SEQ ID NO: 19). La secuencia codificante en sí consiste en 276 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 91 aminoácidos (SEQ ID NO: 20).

15 Este gen es uno de varios genes de citocinas CC agrupados en el brazo q del cromosoma 17. Las citocinas son una familia de proteínas secretadas implicadas en procedimientos inmunorreguladores e inflamatorios. Las citocinas CC son proteínas caracterizadas por dos cisteínas adyacentes. La citocina codificada por este gen funciona como un quimioatrayente para los monocitos de la sangre, los linfocitos T auxiliares de memoria y los eosinófilos. Produce la liberación de histamina de los basófilos y activa los eosinófilos. Esta citocina es uno de los principales factores supresores de VIH producidos por los linfocitos CD8⁺. Funciona como uno de los ligandos naturales para el receptor de quimiocinas CCR5 y suprime la replicación in vitro de las cepas R5 del VIH-1, que usa CCR5 como correceptor.

La proteína se exploró para determinar los dominios funcionales y CCL5 tiene un dominio de familia alfa de intercrinas (citocina pequeña C-X-C) (quimiocina CXC) en la región 30-88, que con frecuencia están implicadas en la quimiotaxis de células específicas, la mediación del crecimiento celular y la respuesta inflamatoria. Muchos factores de bajo peso molecular secretados por células incluyendo fibroblastos, macrófagos y células endoteliales, en respuesta a una diversidad de estímulos tales como factores de crecimiento, interferones, transformación vírica y productos bacterianos, se relacionan estructuralmente. La mayoría de los miembros de esta familia de proteínas parecen tener actividades mitógenas, quimiotácticas o inflamatorias. Estas pequeñas citocinas también se denominan intercrinas o quimiocinas. Son proteínas catiónicas de 70 a 100 restos de aminoácidos que comparten cuatro restos de cisteína conservados implicados en dos enlaces disulfuro. Estas proteínas pueden clasificarse en dos grupos basados en el espaciado de las dos cisteínas aminoterminales. En el primer grupo, las dos cisteínas se separan por un único resto (C-X-C), mientras que, en el segundo grupo, son adyacentes (C-C). CLL5, adicionalmente, tiene un dominio de péptido señal N terminal en la región 1-23. Por lo general se ubica en el extremo N y normalmente está ausente en la proteína madura. Normalmente se refiere a la secuencia (aproximadamente 20 aminoácidos) que interactúa con la partícula de reconocimiento de señal y dirige el ribosoma al retículo endoplasmático donde tiene lugar la inserción cotraduccional. Los péptidos señal son altamente hidrófobos, pero con algunos restos de carga positiva. La secuencia señal normalmente es retirada de la cadena peptídica en crecimiento por la peptidasa de señal, una proteasa específica se ubica en la cara cisternal del retículo endoplasmático. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 1 región transmembrana potencial (6-19), mientras que la región de 6-18 tenía la probabilidad más alta. La proteína tiene una alta homología (> 80 %) a las secuencias de proteínas de CCL5 descritas para ratón y rata. Esta proteína se ubica principalmente en el citoplasma y el núcleo. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en la médula ósea, el hígado, el bazo, el timo y el tejido pulmonar, mientras que solo se observó la expresión a niveles bajos o ausentes de CCL5 en todos los demás tejidos analizados, incluyendo corazón, riñón, músculo esquelético, cerebro, médula espinal, próstata y páncreas. Cocchi y col. (Cocchi y col., *Science*; 270: 1811-1815 (1995)) demostraron que RANTES es uno de los principales factores supresores del VIH, producidos por linfocitos T CD8-positivos. Cocchi y col. (Cocchi y col., *Science*; 270: 1811-1815 (1995)) especularon que el control mediado por RANTES del VIH puede producirse ya sea directamente, a través de su actividad antienterovírica inherente, o indirectamente, a través de su capacidad para quimioatraer linfocitos T y monocitos a la proximidad de los focos de infección. Sin embargo, este último mecanismo también puede tener el efecto opuesto de proporcionar nuevas dianas no infectadas para la infección por el VIH. El uso de astrocitos obtenidos a partir de prosencéfalos fetales de 5-10 semanas de edad, Bakhiet y col. (Bakhiet y col., *Nat. Cell. Biol.*; 3: 150-157 (2001)) demostraron que la expresión de ARNm de RANTES y de proteína aumentan con la edad. Los receptores de RANTES son CCR1, CCR3 y CCR5. El análisis inmunohistoquímico y por hibridación in situ demostraron que las células que expresan CCR5 no expresan el ARNm de RANTES, mientras que las células que carecen de CCR5 sí expresan el ARNm de RANTES, lo que sugiere un modo paracrina de acción. Bakhiet y col. (Bakhiet y col., *Nat. Cell. Biol.*; 3: 150-157 (2001)) demostraron que RANTES inhibió la proliferación y la supervivencia prolongadas de las células de tipo silvestre pero no tuvo efecto sobre las células Ccr5^{-/-}. En respuesta a la estimulación de RANTES, los astrocitos humanos de 5 y 10 semanas de edad potenciaron su producción de gamma-interferón, pero solo las células más viejas expresaron el receptor IFNG. La estimulación de RANTES de células 5 semanas de edad aumentó rápidamente la actividad tirosina cinasa y la fosforilación de proteínas. La estimulación de RANTES indujo la translocación de STAT1 desde el citoplasma al núcleo. Se descubrieron tres elementos supuestos de respuesta a PPAR (PPRE, por sus siglas en inglés) en el

promotor de RANTES humano.

Una vez que las células infectadas por virus son eliminadas por los linfocitos citotóxicos, la retirada de estas células muertas requiere el aclaramiento de los macrófagos sin que los macrófagos sean destruidos por el virus. Tyner y col. (Tyner y col., *Nat Med.*; 11: 1180-1187 (2005)) demostraron que los ratones deficientes en CCL5 tenían un aclaramiento vírico retrasado, la inflamación excesiva de las vías respiratorias y la muerte respiratoria después de la infección ya sea con el virus paragripal murino o el virus de la gripe humano. Se requería CCL5 para mantener la apoptosis y la disfunción mitocondrial en macrófagos de ratón infectados por virus in vivo y macrófagos de ratón y humanos ex vivo y el efecto protector de CCL5 requería la activación de CCR5 y las vías de señalización ERK1/ERK2 y AKT corriente abajo.

10 RANTES es uno de los ligandos naturales para el receptor de quimiocina CCR5 y suprime potentemente la replicación in vitro de las cepas R5 del VIH-1, que usan CCR5 como correceptor.

Pueden identificarse ligandos/fragmentos adecuados del CMH de CCL5 humano mediante el procedimiento descrito en relación con Colonia 1 anterior.

15 CCR7: CCR7 humano (receptor de quimiocina (motivo C-C) 7) se ubica en la región genómica 35975249-35963546 en el cromosoma 17 en la región 17q12-q21.2. Consiste en 3 exones resultantes de la transcripción en un ARNm que tiene 2188 pares de bases (véase la SEQ ID NO: 21). La secuencia codificante en sí consiste en 1137 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 378 aminoácidos (SEQ ID NO: 22).

20 La proteína codificada por este gen es un miembro de la familia de receptores acoplados a proteína G. Este receptor se identificó como un gen inducido por el virus de Epstein-Barr (VEB) y se piensa que es un mediador de los efectos de VEB sobre los linfocitos B. Este receptor se expresa en diversos tejidos linfoides y activa los linfocitos B y T. Se ha demostrado que controla la migración de linfocitos T de memoria a tejidos inflamados, así como estimula la maduración de células dendríticas. Se ha notificado que el ligando 19 de quimiocina (motivo C-C) (CCL19/ECL) es un ligando específico de este receptor. La proteína se exploró para determinar los dominios funcionales y CCR7
25 tiene 7 dominios transmembrana en la región 63-85, 97-116, 131-150, 171-193, 229-251, 264-286 y 306-328. Los dominios transmembrana se componen de 15-30 restos hidrófobos y se encuentran en proteínas que abarcan una membrana en una célula. Una proteína puede tener un único dominio TM (por ejemplo, receptores transmembrana de Tipo I tales como el receptor de EGF) o múltiples dominios TM (por ejemplo, receptores acoplados a proteína G). Para abarcar el núcleo hidrocarbonado de la membrana de ~ 3 nm requiere una hélice alfa de ~20 restos
30 predominantemente apolares sin carga. La membrana separa dos compartimentos acuosos mediante una fase lipídica bidimensional delgada. Las proteínas de membrana generalmente abarcan esta fase lipídica y, por tanto, necesitan acomodarse al tramo hidrófilo a ambos lados de la membrana, así como al entorno hidrófobo en el núcleo de la bicapa. La estructura de las porciones embebidas en la membrana consiste ya sea en hélices alfa transmembrana montadas con frecuencia en haces de hélices, o en láminas beta antiparalelas que forman poros en
35 forma de barril. La longitud de la hélice transmembrana puede incluso correlacionarse con el espesor de su membrana. Las proteínas que abarcan una sola vez la membrana del retículo endoplasmático (RE) y Golgi tienen generalmente dominios transmembrana más cortos que las proteínas de la membrana plasmática. Puesto que el contenido de colesterol y, por tanto, el espesor de la bicapa lipídica también aumenta junto con la vía secretora, esto podría reflejar un papel de los segmentos y proteínas transmembrana en la clasificación de los lípidos. En proteínas
40 que abarcan la membrana múltiples veces las hélices transmembrana se unen estrechamente a la estructura compacta, globular, de la que se excluyen los lípidos. La proteína tiene una alta homología (> 86 %) a las secuencias de proteínas de CCR7 descritas para ratón y rata. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en el tejido de la médula ósea, el bazo y el timo, así como en los linfocitos, mientras que solo se observaron niveles de expresión bajos o ausentes de CCR7 en todos los demás tejidos
45 analizados, incluyendo pulmón, hígado, corazón, riñón, músculo esquelético, cerebro, médula espinal, próstata y páncreas. Los linfocitos T vírgenes y auxiliares 1 (TH1, por sus siglas en inglés) expresan CCR7 y retornan a la vaina linfoide periarteriolar en el bazo, mientras que los linfocitos TH2 activados carecen de CCR7 y forman anillos en la periferia de las zonas de linfocitos T cerca de los folículos de células B. Randolph y col. (Randolph y col., *Science*; 286: 2159-2162 (1999)) descubrieron que la transducción retroviral de los linfocitos TH2 con CCR7 les obligó a localizarse en un patrón similar a TH1 e inhibieron su participación en la ayuda de las células B in vivo, pero
50 no in vitro. La expresión diferencial de los receptores de quimiocinas da como resultado patrones de migración celulares únicos que son importantes para respuestas inmunitarias eficaces. Sallusto y col. (Sallusto y col., *Nature*; 401: 708-712 (1999)) demostraron que la expresión de CCR7 divide los linfocitos T de memoria humanos en 2 subconjuntos funcionalmente distintos. Las células de memoria CCR7 expresan receptores para la migración a tejidos inflamados y presentan una función efectora inmediata. Por el contrario, las células de memoria CCR7⁺
55 expresan receptores de retorno a ganglios linfáticos y carecen de función efectora inmediata, pero estimulan eficientemente las células dendríticas y se diferencian en células efectoras CCR7 tras la estimulación secundaria. Los linfocitos T CCR7⁺ y CCR7⁻, que Sallusto y col. (Sallusto y col., *Nature*; 401: 708-712 (1999)) denominaron de memoria central (T-CM, por sus siglas en inglés) y de memoria efectora (T-EM, por sus siglas en inglés), se diferencian en una forma etapa a etapa a partir de linfocitos T vírgenes, que persisten durante años después de la
60 inmunización, y permiten una división del trabajo en la respuesta de memoria. Los análisis por citometría de flujo de la expresión de marcadores y de la división celular identificaron 4 subconjuntos de linfocitos T CD8⁺, lo que

representa un patrón de diferenciación de linaje: CD45RA+CCR7⁺ (doble positivo); CD45RA-CCR7⁺; CD45RA-CCR7⁻ (doble negativo); CD45RA+CCR7⁻. La capacidad de división celular, medida mediante diacetato de 5- (y 6-) carboxil-fluoresceína, éster de succinimidilo y la tinción intracelular para el antígeno nuclear Ki67, se limita en gran medida a los subconjuntos CCR7⁺ y se produce más rápidamente en las células que también son CD45RA⁺. Aunque las células doble negativo no se dividieron ni se expandieron después de la estimulación, sí invirtieron la positividad, ya sea para CD45RA o CCR7 o ambos. Los linfocitos CD45RA⁺CCR7⁻, que se consideran diferenciados de forma terminal, dejan de dividirse, pero no producen interferón-gamma y expresan altos niveles de perforina. Los linfocitos B recirculan entre compartimentos ricos en células B (foliculos o zonas B) en órganos linfoides secundarios, sondeando el antígeno. Después de la unión del antígeno, las células B se mueven hasta el límite de las zonas B y T para interactuar con linfocitos T auxiliares. Reif y col. (Reif y col., *Nature*; 416: 94-99) demostraron que las células B comprometidas a antígeno tienen una mayor expresión de CCR7, el receptor para las quimiocinas de la zona T CCL19 (también conocida como ELC) y CCL21, y que presentan una sensibilidad aumentada a ambos quimioatrayentes. En ratones que carecen de quimiocinas CCL19 y CCL21 linfoides, o con células B que carecen de CCR7, el compromiso de antígeno no consigue provocar el movimiento a la zona T. Usando la transferencia de genes mediada por retrovirus, los autores demostraron que el aumento de la expresión de CCR7 es suficiente para dirigir las células B a la zona T. Recíprocamente, la sobreexpresión de CXCR5, el receptor para la quimiocina de la zona B CXCL13, es suficiente para superar el movimiento de células B inducido por antígenos a la zona T. Reif y col. (Reif y col., *Nature*; 416: 94-99 (2002)) llegaron a la conclusión de que sus resultados definieron el mecanismo de relocalización de células B en respuesta al antígeno y establecieron que la posición de la célula in vivo puede determinarse mediante el equilibrio de sensibilidad a quimioatrayentes realizado en zonas separadas pero adyacentes.

Pueden identificarse ligandos/fragmentos adecuados del CMH de CCR7 humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

CD40LG: CD40LG humano (ligando CD40; superfamilia TNF, miembro 5, síndrome de hiper IgM), también conocido como CD154, se ubica en la región genómica 135558001-135570214 en el cromosoma 10 en la región Xq26. Consiste en 5 exones resultantes de la transcripción en un ARNm que tiene 1834 pares de bases (véase la SEQ ID NO: 23). La secuencia codificante en sí consiste en 786 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 261 aminoácidos (SEQ ID NO: 24).

La proteína codificada por este gen se expresa en la superficie de los linfocitos T. Regula la función de las células B mediante el compromiso de CD40 sobre la superficie de células B. Un defecto en este gen da como resultado una incapacidad de experimentar cambio de clase de inmunoglobulinas y se asocia al síndrome de hiper-IgM.

La proteína se exploró para determinar los dominios funcionales y CD40LG tiene un dominio transmembrana en la región 23-45. Los dominios transmembrana se componen de 15-30 restos hidrófobos y se encuentran en proteínas que abarcan una membrana en una célula. Una proteína puede tener un único dominio TM (por ejemplo, receptores transmembrana de Tipo I tales como el receptor de EGF) o múltiples dominios TM (por ejemplo, receptores acoplados a proteína G). Para abarcar el núcleo hidrocarbonado de la membrana de ~ 3 nm requiere una hélice alfa de ~20 restos predominantemente apolares sin carga. La membrana separa dos compartimentos acuosos mediante una fase lipídica bidimensional delgada. Las proteínas de membrana generalmente abarcan esta fase lipídica y, por tanto, necesitan acomodarse al tramo hidrófilo a ambos lados de la membrana, así como al entorno hidrófobo en el núcleo de la bicapa. La estructura de las porciones embebidas en la membrana consiste ya sea en hélices alfa transmembrana montadas con frecuencia en haces de hélices, o en láminas beta antiparalelas que forman poros en forma de barril. La longitud de la hélice transmembrana puede incluso correlacionarse con el espesor de su membrana. Las proteínas que abarcan una sola vez la membrana del retículo endoplasmático (RE) y Golgi tienen generalmente dominios transmembrana más cortos que las proteínas de la membrana plasmática. Puesto que el contenido de colesterol y, por tanto, el espesor de la bicapa lipídica también aumenta junto con la vía secretora, esto podría reflejar un papel de los segmentos y proteínas transmembrana en la clasificación de los lípidos. En proteínas que abarcan la membrana múltiples veces las hélices transmembrana se unen estrechamente a la estructura compacta, globular, de la que se excluyen los lípidos. CD40LG, adicionalmente, tiene un dominio de la familia del factor de necrosis tumoral en la región de 122-261, que pertenece a una familia de citocinas que forman complejos homotriméricos o heterotriméricos. CD40L es una citocina que parece ser importante en el desarrollo y la activación de células B. Todas estas citocinas parecen formar complejos homotriméricos (o heterotriméricos en el caso de LT-alfa/beta) que son reconocidos por sus receptores específicos. El patrón PROSITE para esta familia se ubica en una cadena beta en la sección central de la proteína que se conserva en todos los miembros. La proteína tiene una alta homología (> 77 %) a las secuencias proteicas de CD40LG descritas para ratón y rata. Esta proteína se ubica principalmente en la membrana plasmática. El análisis de la expresión en diferentes tejidos demostró un alto nivel de expresión en linfocitos T y plaquetas, mientras que solo se observaron niveles de expresión bajos o ausentes de CD40L en todos los demás tejidos analizados. La molécula de ligando de CD40 ayuda en la estimulación de las células B en la respuesta inmunitaria. La molécula CD40 es una glicoproteína expresada en linfocitos B, células epiteliales y algunas células de carcinoma. El entrecruzamiento de CD40 mediante anticuerpos monoclonales anti-CD40 media la proliferación, la adhesión y la diferenciación de células B. Gauchat y col. (Gauchat y col., *FEBS Lett.*; 315: 259-266 (1993)) demostraron que el ligando CD40 humano puede detectarse en linfocitos T, pero está ausente en células B y monocitos. Se expresa tanto en linfocitos T CD4-positivos como CD8-positivos. El ligando CD40 también se expresa en plaquetas y se libera de ellas en la activación. Se han identificado mutaciones múltiples en el

gen CD40LG que se asocian al síndrome de inmunodeficiencia hiper-IgM de tipo 1.

Pueden identificarse ligandos/fragmentos adecuados del CMH de CD40LG humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

5 **CTLA4:** CTLA4 humano (proteína 4 asociada a linfocitos T citotóxicos), también conocida como CD152, se ubica en la región genómica 204440753-204446927, en el cromosoma en la región 2q33. La isoforma A consiste en 4 exones resultantes de la transcripción en un ARNm que tiene 1988 pares de bases (véase la SEQ ID NO: 25). La secuencia codificante en sí consiste en 672 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 223 aminoácidos (SEQ ID NO: 26). La isoforma B consiste en 3 exones resultantes de la transcripción en un ARNm que tiene 1878 pares de bases (véase la SEQ ID NO: 27). La secuencia codificante en sí consiste en 525 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 174 aminoácidos (SEQ ID NO: 28).

15 Este gen es un miembro de la superfamilia de las inmunoglobulinas y codifica una proteína que transmite una señal inhibitoria a linfocitos T. La proteína contiene un dominio V, un dominio transmembrana y una cola citoplasmática. Se han caracterizado variantes de corte y empalme transcripcionales alternativos, que codifican diferentes isoformas. La isoforma unida a la membrana funciona como un homodímero interconectado mediante un enlace disulfuro, mientras que la isoforma soluble funciona como un monómero. Las mutaciones en este gen se han asociado a la diabetes mellitus dependiente de insulina, la enfermedad de Graves, la tiroiditis de Hashimoto, la enfermedad celíaca, el lupus eritematoso sistémico, la orbitopatía asociada a la tiroides y otras enfermedades autoinmunitarias.

20 La proteína se exploró para determinar los dominios funcionales y CTLA4 tiene un dominio similar a la inmunoglobulina en la región 53-131. Las moléculas de Ig son proteínas altamente modulares, en las que los dominios variables y constantes tienen patrones de secuencia claras y conservadas. Los dominios en las moléculas Ig y similares a Ig se agrupan en cuatro tipos: conjunto V (variable), conjunto C1 (constante 1), conjunto C2 (constante 2) y conjunto I-set (intermedio). Estudios estructurales han demostrado que estos dominios comparten una estructura sándwich beta de llave griega núcleo común, difiriendo los tipos en el número de cadenas en las láminas beta, así como en sus patrones de secuencia. Pueden encontrarse dominios similares a Ig que se relacionan tanto en secuencia como en estructura en varias familias de proteínas diversas. Los dominios similares a Ig están implicados en una diversidad de funciones, incluyendo el reconocimiento célula-célula, los receptores de superficie celular, la estructura muscular y el sistema inmunitario. Además, CTLA4 tiene un dominio transmembrana en la región 162-184. Los dominios transmembrana se componen de 15-30 restos hidrófobos y se encuentran en proteínas que abarcan una membrana en una célula. Una proteína puede tener un único dominio TM (por ejemplo, receptores transmembrana de Tipo I tales como el receptor de EGF) o múltiples dominios TM (por ejemplo, receptores acoplados a proteína G). Para abarcar el núcleo hidrocarbonado de la membrana de ~ 3 nm requiere una hélice alfa de ~20 restos predominantemente apolares sin carga. La membrana separa dos compartimentos acuosos mediante una fase lipídica bidimensional delgada. Las proteínas de membrana generalmente abarcan esta fase lipídica y, por tanto, necesitan acomodarse al tramo hidrófilo a ambos lados de la membrana, así como al entorno hidrófobo en el núcleo de la bicapa. La estructura de las porciones embebidas en la membrana consiste ya sea en hélices alfa transmembrana montadas con frecuencia en haces de hélices, o en láminas beta antiparalelas que forman poros en forma de barril. La longitud de la hélice transmembrana puede incluso correlacionarse con el espesor de su membrana. Las proteínas que abarcan una sola vez la membrana del retículo endoplasmático (RE) y Golgi tienen generalmente dominios transmembrana más cortos que las proteínas de la membrana plasmática. Puesto que el contenido de colesterol y, por tanto, el espesor de la bicapa lipídica también aumenta junto con la vía secretora, esto podría reflejar un papel de los segmentos y proteínas transmembrana en la clasificación de los lípidos. En proteínas que abarcan la membrana múltiples veces las hélices transmembrana se unen estrechamente a la estructura compacta, globular, de la que se excluyen los lípidos. En proteínas que abarcan la membrana múltiples veces las hélices transmembrana se unen estrechamente a la estructura compacta, globular, de la que se excluyen los lípidos. CTLA4, adicionalmente, tiene un dominio de péptido señal N terminal en la región 1-36. Por lo general se encuentra en el extremo N y normalmente está ausente en la proteína madura. Normalmente se refiere a la secuencia (aproximadamente 20 aminoácidos) que interactúa con la partícula de reconocimiento de señal y dirige el ribosoma al retículo endoplasmático donde tiene lugar la inserción cotraslacional. Los péptidos señal son altamente hidrófobos, pero con algunos restos de carga positiva. La secuencia señal normalmente es retirada de la cadena peptídica en crecimiento por la peptidasa señal, una proteasa específica ubicada en la cara cisternal del retículo endoplasmático. La proteína tiene una alta homología (> 74 %) a las secuencias proteicas de CTLA4 descritas para ratón y rata. Esta proteína se ubica principalmente en la membrana plasmática. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en linfocitos T activos, el bazo, la médula ósea, el timo y el tejido pulmonar, mientras que solo se observaron niveles de expresión bajos o ausentes de CTLA4 en todos los demás tejidos analizados, incluyendo hígado, corazón, riñón, músculo esquelético, cerebro, médula espinal, próstata y páncreas. CTLA4 es un miembro de la superfamilia de las inmunoglobulinas y es una molécula coestimuladora expresada por linfocitos T activados. CTLA4 es similar al linfocito T coestimulador CD28 y ambas moléculas se unen a B7-1 (CD80) y B7-2 (CD86) sobre las células presentadoras de antígeno. CTLA4 transmite una señal inhibitoria a los linfocitos T, mientras que CD28 transmite una señal estimuladora. CTLA4 contiene una secuencia líder, un dominio V, un dominio transmembrana y una cola citoplasmática codificada por 4 exones, respectivamente. La cola citoplasmática contiene 2 sitios de fosforilación potenciales. El análisis por transferencia

Northern detectó 2 transcritos de ARNm de CTLA4 1,8 y 0,8 kb en linfocitos T de sangre periférica activados. Linsley y col. (Linsley y col., *J. Biol. Chem.*; 270: 15417-15424 (1995)) demostraron que la proteína CTLA4 es un homodímero interconectado por un enlace disulfuro en el dominio extracelular en el resto de cisteína 120. Cada polipéptido monomérico contiene un sitio de unión de alta afinidad para las moléculas coestimuladoras B7-1 y B7-2.

5 Usando RT-PCR, Magistrelli y col. (Magistrelli y col., *Eur. J. Immunol.*; 29: 3596-3602 (1999)) descubrieron que los linfocitos T de sangre periférica humana no activados expresaban una forma cortada y empalmada alternativa de CTLA4. El corte y empalme induce la delección de la región transmembrana codificada por el exón 2, dando como resultado la producción de una forma soluble de la proteína con una masa molecular de 23 kD; CTLA4 de membrana tiene una masa molecular aparente de 45 kD, lo que sugiere que la forma soluble es una proteína monomérica.

10 Estudios funcionales demostraron que la forma soluble de CTLA4 se regula negativamente mediante la activación de linfocitos T, a diferencia de CTLA4 de membrana, que se regula positivamente mediante la activación de linfocitos T. CTLA4 se une a las isoformas B7 con una afinidad que es de 10 a 100 veces la de CD28. Ostrov y col. (Ostrov y col., *Science*; 290: 816-819 (2000)) determinaron la estructura cristalina de la porción extracelular de Ctl4 murino a una resolución de 2,0 angstrom. De acuerdo con su pertenencia a la superfamilia de Ig, Ctl4 muestra una topología de cadena similar a los dominios V-alfa. Ctl4 tiene un modo de dimerización inusual que coloca los sitios de unión a B7 distales a la interfaz de dimerización, permitiendo que cada dímero Ctl4 se una a 2 moléculas de B7 divalentes. Los autores sugirieron que la reordenación periódica de estos componentes podría explicar el papel de CTLA4 en la regulación de sensibilidad de los linfocitos T. Oaks y col. (Oaks y col., *Cell Immunol.*; 201: 144-153 (2000)) descubrieron que la activación de linfocitos humanos dio como resultado la desaparición de sCTLA4 a las 48 horas seguida de su reaparición gradual, mientras que CTLA4-TM aumentó rápidamente y se mantuvo presente al mismo nivel después de la activación. CTLA4Ig es una proteína química soluble que consiste en el dominio extracelular de CTLA4 humano y un fragmento de la porción Fc de la IgG1 humana. Se une a moléculas de B7-1 y B7-2 sobre células presentadoras de antígeno (CPA) y de ese modo bloquea la señal coestimuladora mediada por CD28 para la activación de los linfocitos T. Fallarino y col. (Fallarino y col., *Nat. Immunol.*; 4: 1206-1212 (2003)) demostraron que las células Tr CD4⁺/CD25⁺ de ratón que expresan Ctl4 condicionaron que DC (células dendríticas) que expresan B7 expresen Ido y produzcan IFN γ . Las DC (células dendríticas) condicionadas in vitro por los linfocitos Tr mediaron efectos supresores in vivo que fueron dependientes del catabolismo eficaz del triptófano. El requisito para la expresión de Ctl4 se superó en linfocitos Tr estimulados con lipopolisacárido para producir cantidades sustanciales de Ifn γ e interleucina 10 (IL-10). Fallarino y col. (Fallarino y col., *Nat. Immunol.*; 4: 1206-1212 (2003)) concluyeron que los linfocitos Tr ceban las DC (células dendríticas) para la inducción de tolerancia a través de la inmunorregulación a base de IDO. Usando ensayos de migración in vitro y microscopía de barrido láser de 2 fotones in vivo, Schneider y col. en. (Schneider y col., *Science*; 313: 1972-1975 (2006)) demostraron que CTLA4 aumenta la motilidad de los linfocitos T y anula la señal de parada inducida por el receptor de linfocitos T (TCR, por sus siglas en inglés) necesaria para la formación de conjugados estables entre linfocitos T y células presentadoras de antígeno. Este evento condujo a periodos de contacto reducidos entre linfocitos T y células presentadoras de antígeno que, a su vez, disminuyeron la producción y proliferación de citocinas. Schneider y col. (Schneider y col., *Science*; 313: 1972-1975 (2006)) llegaron a la conclusión de que sus resultados sugieren un modelo fundamentalmente diferente de la señalización de parada inversa, mediante el cual CTLA4 modula el umbral para la activación de linfocitos T y protege frente a la autoinmunidad.

40 Pueden identificarse ligandos/fragmentos adecuados del CMH de CTLA4 humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

CTSL1: CTSL1 Humana (catepsina L1) se ubica en la región genómica 89530799-89536127 en el cromosoma 9 en la región 9q21-q22. La isoforma a consiste en 6 exones resultantes de la transcripción en un ARNm que tiene 1731 pares de bases. La isoforma b consiste en 6 exones resultantes de la transcripción en un ARNm que tiene 1587 pares de bases (véase la SEQ ID NO: 29). Ambas isoformas codifican la misma proteína cuya secuencia codificante real consiste en sí en 1002 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 333 aminoácidos (SEQ ID NO: 30).

La proteína codificada por este gen es una cisteína proteinasa lisosómica que desempeña un papel fundamental en el catabolismo intracelular de proteínas. Sus sustratos incluyen el colágeno y la elastina, así como el inhibidor de proteasa alfa-1, un elemento de control fundamental de la actividad elastasa de neutrófilos. La proteína codificada se ha implicado en varios procesos patológicos, incluyendo la necrosis de miofibrillas en las miopatías y en la isquemia miocárdica y en la respuesta tubular renal a proteinuria. Esta proteína, que es un miembro de la familia C1 de peptidasas, es un dímero compuesto de cadenas pesadas y ligeras unidas por puentes disulfuro, producidas ambas a partir de un único precursor proteico. Se han descubierto para este gen al menos dos variantes de transcrito que codifican la misma proteína. La proteína se exploró para determinar los dominios funcionales y CTSL1 tiene un dominio cisteína proteasa de la familia de la papaína en la región 114-332. La familia de la papaína tiene una amplia diversidad de actividades, incluyendo endopeptidasas de gama amplia (papaína) y de gama estrecha, aminopeptidasas, dipeptidil peptidasas y enzimas con actividad tanto exo como endopeptidasa. Los miembros de la familia de la papaína están extendidos, se encuentran en baculovirus, eubacterias, levadura y prácticamente todos los protozoos, plantas y mamíferos. Las proteínas son normalmente lisosómicas o secretadas, y se requiere la escisión proteolítica del propéptido para la activación enzimática, aunque la bleomicina hidrolasa es citosólica en hongos y mamíferos. Las cisteína proteinasas similares a papaína se sintetizan esencialmente como proenzimas inactivas (zimógenos) con regiones propeptídicas N terminales. El proceso de activación de estas enzimas incluye la

eliminación de las regiones propeptídicas. Las regiones propeptídicas sirven para una diversidad de funciones in vivo e in vitro. Se requiere que la pro-región para el plegamiento apropiado de la enzima recién sintetizada, la inactivación del dominio peptidasa y la estabilización de la enzima frente a la desnaturalización en condiciones neutras a alcalinas. Los restos de aminoácidos dentro de la pro-región median su asociación a la membrana y desempeñan un papel en el transporte de la proenzima a los lisosomas. Entre las características más notables de los propéptidos está su capacidad para inhibir la actividad de sus enzimas afines y que ciertos propéptidos presentan una alta selectividad para la inhibición de las peptidasas a partir de las cuales se originan. Los restos catalíticos de la papaína son Cys-25 y Sus-159, siendo otros restos importantes Gln-19, que ayuda a formar el 'agujero oxianión' y Asn-175, que orienta el anillo de imidazol de His-159. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 2 regiones transmembrana potenciales (6-19, 239-249, 278-283), mientras que las regiones de 7-17 y 242-243 tenían la probabilidad más alta. La secuencia proteica de CTSL1 no tiene homología a genes descritos para ratón y rata. La catepsina L es una cisteína proteinasa lisosómica con un papel importante en el catabolismo de la proteína intracelular. También muestra la actividad in vitro colagenolítica y elastinolítica más potente de cualquiera de las catepsinas. La catepsina L inactiva proteolíticamente el inhibidor de proteasa alfa-1, un elemento de control fundamental de la actividad elastasa de neutrófilos humanos in vivo. La catepsina L se ha implicado en procesos patológicos, incluyendo la necrosis de miofibrillas en las miopatías y en la isquemia miocárdica y en la respuesta tubular renal a proteinuria (Joseph y col., *J. Clin. Invest.*; 81: 1621-1629). Urbich y col. (Urbich y col., *Nat Med.*; 11: 206-213 (2005)) descubrieron que la catepsina L se expresa altamente en células progenitoras en comparación con células endoteliales maduras y es esencial para la degradación de la matriz y la invasión in vitro. Ratones nulos para Ctsl mostraron una recuperación funcional alterada después de la isquemia del miembro posterior y las células progenitoras deficientes en Ctsl no retornaron a los sitios de isquemia ni aumentaron la neovascularización. La expresión forzada de la catepsina L en células endoteliales maduras potenció la actividad invasiva y la neovascularización in vivo. Urbich y col. (Urbich y col., *Nat. Med.*; 11: 206-213 (2005)) concluyeron que la catepsina L tiene un papel crítico en la integración de células progenitoras endoteliales circulantes en el tejido isquémico y se requiere para la neovascularización mediada por células progenitoras endoteliales. Las proteasas lisosómicas generan péptidos presentados por moléculas del CMH de clase II a linfocitos T CD4⁺. Para determinar si las proteasas lisosómicas específicas influyen en el resultado de una respuesta autoinmunitaria dependiente de linfocitos T CD4⁺, Maehr y col. (Maehr y col., *J. Clin. Invest.*; 115: 2934-2943) generaron ratones que carecían de Ctsl en el fondo de la misma estirpe NOD propensa a la diabetes autoinmunitaria. La ausencia de CTSL proporcionó una fuerte protección frente a la enfermedad en la etapa de infiltración pancreática. Dentro de los compartimentos de linfocitos T CD4⁺ de ratones deficientes en Ctsl, hubo una mayor proporción de linfocitos T reguladores en comparación con la de los compañeros de camada suficientes en Ctsl. Maehr y col. (Maehr y col., *J. Clin. Invest.*; 115: 2934-2943 (2005)) sugirieron que es este equilibrio desplazado de linfocitos T CD4⁺ reguladores frente a agresivos que protege a los ratones deficientes en Ctsl de la enfermedad autoinmunitaria. Los resultados identificaron Ctsl como una enzima cuya actividad es esencial para el desarrollo de diabetes de tipo I en el ratón NOD.

Pueden identificarse ligandos/fragmentos adecuados del CMH de CTSL1 humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

EOMES: EOMES Humano (homólogo de eomesodermina (*Xenopus laevis*)) se ubica en la región genómica 27738788-27732871 en el cromosoma 3 en la región 3p21.3-p21.2. Consiste en 6 exones resultantes de la transcripción en un ARNm que tiene 2756 pares de bases (véase la SEQ ID NO: 33). La secuencia codificante en sí consiste en 2061 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 686 aminoácidos (SEQ ID NO: 34).

Este gen codifica un miembro de una familia de proteínas conservadas que comparte un dominio de unión a ADN común, la caja T. Los genes de la caja T codifican factores de transcripción implicados en la regulación de procedimientos de desarrollo. Se demuestra que un gen similar interrumpido en ratones es esencial durante el desarrollo del trofoblasto y la gastrulación. La proteína se exploró para determinar los dominios funcionales y EOMES tiene un dominio CAJAT descubierto primero en el locus T de ratones (*Brachyury*) en la región 266-461. Los factores de transcripción de la familia caja T se requieren tanto para las decisiones de destino celular tempranas como para las necesarias para la formación del plan corporal de vertebrados básico y para la diferenciación y organogénesis. La caja T se define como la región mínima dentro de la proteína caja T que es tanto necesaria como suficiente para la unión a ADN específica de secuencia, todos los miembros de la familia examinados hasta ahora se unen a la secuencia de consenso de ADN TCACACCT. La caja T es un dominio de unión a ADN relativamente grande, que comprende generalmente aproximadamente un tercio de la proteína completa (17-26 kDa). Estos genes se descubrieron basándose en la similitud al dominio de unión a ADN del producto génico *Brachyury* (T) murino, cuya similitud es la característica definitoria de la familia. El gen *Brachyury* se denomina así por su fenotipo, que se identificó hace 70 años como una cepa mutante de ratón con una cola corta truncada. El gen y sus parálogos, se han convertido en un modelo bien estudiado para la familia y, por tanto, gran parte de lo que se conoce acerca de la familia Caja T deriva del gen *Brachyury* murino. De acuerdo con su localización nuclear, la proteína *Brachyury* tiene una actividad de unión a ADN específica de secuencia y puede actuar como un regulador transcripcional. Los mutantes homocigotos para el gen experimentan extensas anomalías del desarrollo, convirtiendo la mutación en letal. El papel postulado de *Brachyury* es como un factor de transcripción, que regula la especificación y la

diferenciación del mesodermo posterior durante la gastrulación en una forma dependiente de la dosis. Las proteínas caja T tienden a expresarse en órganos específicos o tipos celulares, especialmente durante el desarrollo, y por lo general se requieren para el desarrollo de esos tejidos, por ejemplo, Brachyury se expresa en el mesodermo posterior y en la notocorda en desarrollo, y se requiere para la formación de estas células en ratones. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 4 regiones transmembrana potenciales (10-14, 118-130, 250-261, 316-317), mientras que la región 121-127 tenía la probabilidad más alta. La proteína tiene una alta homología (> 86 %) a las secuencias proteicas de EOMES descritas para ratón y rata. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en el tejido del bazo, mientras que solo se observaron niveles de expresión bajos o ausentes de EOMES en todos los demás tejidos analizados, incluyendo médula ósea, hígado, pulmón, corazón, bazo, riñón, timo, músculo esquelético, cerebro, médula espinal, próstata y páncreas. Las mutaciones en el gen T (Brachyury) de ratón provocan defectos en la formación del mesodermo. Varios genes relacionados, miembros de la familia de genes caja T, codifican un dominio de unión a ADN N terminal similar, la caja T, y desempeñan papeles críticos en el desarrollo embrionario humano. Las mutaciones en TBX5 y TBX3 humanos, por ejemplo, provocan los trastornos del desarrollo síndrome de Holt-Oram y síndrome ulnar-mamario, respectivamente. Pearce y col. (Pearce y col., *Science*; 302: 1041-1043 (2003)) mostraron que eomesodermina, un parólogo de Tbet, se induce en linfocitos T CD8(+) efectora in vitro e in vivo. La expresión ectópica de Eomes era suficiente para recurrir a atributos de efector linfocitos T CD8(+), incluyendo interferón-gamma, perforina y granzima B. El análisis de pérdida de función sugirió que Eomes también puede ser necesario para la diferenciación efectora completa de los linfocitos T CD8(+). Pearce y col. (Pearce y col., *Science*; 302: 1041-1043 (2003)) sugirieron que la eomesodermina probablemente complementa las acciones de Tbet y actúa como un gen regulador clave en el desarrollo de la inmunidad mediada por células.

Pueden identificarse ligandos/fragmentos adecuados del CMH de EOMES humanos mediante el procedimiento descrito en relación con la Colonia 1 anterior.

EPB41L3: EPB41L3 Humano (3 similar a banda 4.1 de proteína de membrana de eritrocitos) se ubica en la región genómica 5533985-5382387 en el cromosoma 18 en la región 18p11.32. Consiste en 16 exones resultantes de la transcripción en un ARNm que tiene 4446 pares de bases (véase la SEQ ID NO: 35). La secuencia codificante en sí consiste en 3264 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 1087 aminoácidos (SEQ ID NO: 36).

La proteína se exploró para determinar los dominios funcionales y EPB41L3 tiene un dominio banda de homólogos de Banda 4.1 en la región 106-301, que también se conoce como dominio de proteína ezrina/radixina/moesina (ERM) presente en miosinas, ezrina, radixina, moesina, proteína tirosina fosfatasas. Estas proteínas desempeñan papeles estructurales y reguladores en el ensamblaje y la estabilización de dominios plasmamembrana especializados. Algunas proteínas que contienen dominio PDZ se unen una o más de esta familia. Ahora incluye las JAK. Este dominio se encuentra en un número de proteínas asociadas al citoesqueleto que se asocian a diversas proteínas en la interfaz entre la membrana plasmática y el citoesqueleto. Es un dominio N terminal conservado de aproximadamente 150 restos implicado en la unión de proteínas citoplasmáticas a la membrana. EPB41L3, adicionalmente, tiene un dominio superenrollado en la región 708-742. El superenrollamiento es un dominio proteico que forma un haz de dos o tres hélices alfa. Los dominios superenrollados cortos están implicados en las interacciones proteicas pero los dominios superenrollados que forman varillas largas se producen en proteínas estructurales o motoras. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 4 regiones transmembrana potenciales (218-231, 311-322, 586-590, 630-667), mientras que las regiones 220-230, 313-320 y 632-664 tenían la probabilidad más alta. La proteína tiene una alta homología (> 88 %) a las secuencias proteicas de EPB41L3 descritas para ratón y rata. Esta proteína se ubica principalmente en la membrana plasmática. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en cerebro, riñón, testículos y tejido intestinal mientras que solo se observaron niveles de expresión bajos o ausentes de EPB41L3 en todos los demás tejidos analizados, incluyendo médula ósea, pulmón, hígado, corazón, bazo, timo, músculo esquelético, médula espinal, próstata y páncreas.

Pueden identificarse ligandos/fragmentos adecuados del CMH de EPB41L3 humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

FCGBP: FCGBP humano (fragmento Fc de la proteína de unión a IgG) se sitúa en la región genómica 45132372-45045810 en el cromosoma 19 en la región 19q13.1. FCGBP consiste en 31 exones resultantes de la transcripción en un ARNm que tiene 9378 pares de bases (véase la SEQ ID NO: 39). La secuencia codificante en sí consiste en 9213 pares de bases. La traducción de esta secuencia da como resultado una proteína con una longitud de 3070 aminoácidos (SEQ ID NO: 40).

En algunos tejidos no linfoides, se expresa el locus (TRGy) gamma del receptor de linfocitos T no reordenado. El transcrito resultante contiene un subconjunto de los segmentos génicos TRGy y es más corto que los transcritos TRGy expresados en tejidos linfoides. Este registro RefSeq representa el transcrito de locus TRGy reordenado; el locus de TRGy completo se representa mediante el RefSeq genómico NG_001336. El transcrito representado por este RefSeq tiene dos marcos de lectura abiertos (ORF, por sus siglas en inglés) que codifican diferentes proteínas.

El ORF corriente abajo está en el mismo marco que TRGy y su producto proteico es similar a las proteínas TRGy. El ORF corriente arriba usa un marco de lectura diferente y codifica una proteína nueva.

La proteína se exploró para determinar los dominios funcionales y FCGBP tiene 13 dominios D de tipo factor de von Willebrand (vWF) en la región 463-627, 862-1019, 1249-1406, 1662-1831, 2064-2222, 2450-2607, 2863-3032, 3265-3423, 3651-3808, 4064-4233, 4463-4624, 4848-5002 y 5228-5382. El factor de Von Willebrand contiene varios dominios de tipo D: D1 y D2 están presentes dentro del propéptido N terminal, mientras que los dominios D restantes se requieren para la multimerización. Una familia de reguladores de crecimiento (denominados originalmente cef10, factor de crecimiento de tejido conjuntivo, fisp-12, cyr61, o, como alternativa, IG-M1 beta y IG-M2 beta), pertenecen todos a genes tempranos inmediatos expresados después de la inducción por factores de crecimiento o ciertos oncogenes. El análisis de secuencia de esta familia reveló la presencia de cuatro módulos distintos. Cada módulo tiene homólogos en otras proteínas de mosaico extracelulares tales como factor de Von Willebrand, rendija, tromboespondinas, colágenos fibrilares, proteínas de unión a IGF y mucinas. La clasificación y el análisis de estos módulos sugieren la ubicación de regiones de unión y, por analogía a módulos mejor caracterizados en otras proteínas, arroja algo de luz sobre la estructura de esta nueva familia. El dominio de vWF se encuentra en diversas proteínas plasmáticas: factores del complemento B, C2, CR3 y CR4; las integrinas (dominios I); tipos de colágeno VI, VII, XII y XIV; y otras proteínas extracelulares. Aunque la mayoría de las proteínas que contienen VWA son extracelulares, las más antiguas presentes en todos los eucariotas son todas proteínas intracelulares implicadas en funciones tales como la transcripción, la reparación del ADN, el transporte ribosómico y de membrana y el proteosoma. Una característica común parece ser la implicación en complejos multiproteicos. Las proteínas que incorporan dominios de vWF participan en numerosos eventos biológicos (por ejemplo, adhesión celular, la migración, retorno, formación de patrones y transducción de señales), implicando la interacción con una gran diversidad de ligandos. Un número de enfermedades humanas surgen de mutaciones en los dominios VWA. La predicción de estructura secundaria de 75 secuencias de vWF alineadas ha revelado una secuencia en gran parte alternante de hélices alfa y cadenas beta. Una de las funciones del factor de von Willebrand (vWF) es servir como un transportador del factor de coagulación VIII (FVIII). La conformación nativa del dominio D' de vWF no solo se requiere para la unión al factor VIII (FVIII), sino también para la multimerización normal y la secreción óptima. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó para 8 regiones transmembrana potenciales FCGBP (11-15, 312-318, 580-586, 416-419, 1801-1807, 1827-1830, 2578-2581, 2901-2904). No se ha notificado ninguna homología a secuencias proteicas conocidas de ratón o rata para la secuencia de la proteína humana de FCGBP. Esta proteína se ubica principalmente extracelularmente. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en intestino, bazo, riñón, pulmón, timo, próstata y tejido páncreas mientras que solo se observaron niveles de expresión bajos o ausentes de FCGBP en todos los demás tejidos analizados, incluyendo médula ósea, hígado, corazón, cerebro, músculo esquelético y médula espinal.

Pueden identificarse ligandos/fragmentos adecuados del CMH de FCGBP humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

FHIT: FHIT Humano (gen de la tríada de histidina frágil), también conocido como FRA3B o AP3Aasa, se ubica en la región genómica 59.712.992-60.497.735 en el cromosoma 3 en la región 3p14.2. Consiste en 5 exones resultantes de la transcripción en un ARNm que tiene 1095 pares de bases (véase la SEQ ID NO: 41). La secuencia codificante en sí consiste en 444 pares de bases. La traducción de esta secuencia da como resultado una proteína con una longitud de 147 aminoácidos (SEQ ID NO: 42).

Este gen, un miembro de la familia de genes de la tríada de histidina, codifica una diadenosina 5',5"-P1,P3-trifosfato hidrolasa implicada en el metabolismo de purinas. El gen abarca el sitio frágil común FRA3B en el cromosoma 3, donde el daño inducido por carcinógenos puede conducir a translocaciones y transcritos aberrantes de este gen. De hecho, se han descubierto transcritos aberrantes de este gen en aproximadamente la mitad de todos los carcinomas de esófago, de estómago y de colon.

La proteína se exploró para determinar los dominios funcionales, FHIT tiene un motivo tríada de histidina (HIT, por sus siglas en inglés) que abarca toda la proteína. El motivo tríada de histidina (HIT), His-phi-His-phi-His-phi-phi (phi, un aminoácido hidrófobo) se identificó como que está altamente conservado en una diversidad de organismos. La estructura cristalina de Hint de conejo, purificado como una proteína de unión a adenosina y AMP, mostró que las proteínas en la superfamilia HIT se conservan como proteínas de unión a nucleótidos y que los homólogos de Hint, que se encuentran en todas las formas de vida, se relacionan estructuralmente con homólogos de Fhit y enzimas relacionadas con GalT, que tienen perfiles filogénicos más restringidos. Los homólogos de Hint incluyendo Hint de conejo y Hnt1 de levadura hidrolizan sustratos de adenosina 5' monofosforamida tales como AMP-NH2 y AMP-lisina a AMP más el producto de amina y funcionan como reguladores positivos de Cdk7/Kin28 in vivo. Los homólogos de Fhit son diadenosina polifosfato hidrolasas y funcionan como supresores tumorales en ser humano y ratón, aunque la función de supresión tumoral de Fhit no depende de la hidrólisis de ApppA. La tercera rama de la superfamilia HIT, que incluye homólogos de GALT, contiene un motivo His-X-His-X-Gln relacionado y transfiere restos de monofosfatos de nucleósidos a segundos sustratos fosforilados en lugar de hidrolizarlos.

La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> no identificó regiones transmembrana. FHIT tiene una alta

homología (> 88 %) a las secuencias proteicas de FHIT para ratón y rata. Huebner y Croce ((Huebner y Croce, *Br. J. Cancer*; 88: 1501-1506)) revisaron las mutaciones en FHIT en tumores primarios y las características clínicas asociadas. Señalaron que, desde que el gen FHIT se encontró en 1996 se habían publicado más de 350 informes de estudios. Los datos que resumieron indicaron que FHIT está alterado en muchos tumores humanos, en particular en los provocados por agentes carcinógenos ambientales, tales como los presentes en el humo del tabaco. En muchos de estos tumores, en particular en los inducidos por el tabaco u otros carcinógenos ambientales, las alteraciones de FHIT se producen muy pronto durante el proceso de múltiples etapas de la carcinogénesis. Huebner y Croce (Huebner y Croce, *Nat. Rev. Cancer*; 1: 214-221 (2001)) demostraron que las células cancerosas negativas para FHIT son muy sensibles a la expresión de FHIT; por ejemplo, la infección con virus recombinantes de FHIT puede provocar la regresión y la prevención de tumores en animales de experimentación. Por tanto, es lógico predecir el desarrollo de un enfoque de terapia génica para el tratamiento y la prevención de cánceres humanos negativos para FHIT.

Pueden identificarse ligandos/fragmentos adecuados del CMH de FHIT humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

FLOT1: FLOT1 Humana (flotilina 1), también conocida como componente de membrana integral de caveolas, se ubica en la región genómica 30803465-30818443 en el cromosoma 6 en la región 6p21.3. Consiste en 10 exones resultantes de la transcripción en un ARNm que tiene 1462 pares de bases (véase la SEQ ID NO: 43). La secuencia codificante en sí consiste en 1284 pares de bases. La traducción de esta secuencia da como resultado una proteína con una longitud de 305 aminoácidos (SEQ ID NO: 44).

Las caveolas son dominios pequeños en la membrana celular interna implicada en el tráfico vesicular y la transducción de señales. FLOT1 codifica una proteína integral de membrana asociada a caveolas.

FLOT1 se exploró para determinar los dominios funcionales y tiene un dominio Band 7 en la región 1-185 y un dominio flotilina en la región 190-363. Las flotilinas son proteínas integrales de membrana que se ha demostrado que están presentes en varios componentes subcelulares, incluyendo caveolas (microdominios de membrana plasmática invaginados), balsas de lípidos (microdominios de membrana plasmática resistentes a detergente y ricos en esfingolípidos y colesterol) y el aparato de Golgi. La función molecular de las flotilinas sigue siendo desconocida. Probablemente están implicadas en la organización de la estructura de caveolas y balsas de lípidos, y otros dominios de membrana resistentes a detergente. También pueden estar implicadas en la transducción de señales. Se ha demostrado que las flotilinas se acumulan en células cerebrales con el desarrollo de la patología de Alzheimer. También se incluyen en esta familia las proteínas Reggie, que se expresan en dominios de la membrana plasmática neuronal no caveolar. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> no identificó regiones transmembrana potenciales. FLOT1 tiene una homología muy alta (> 98 %) a las secuencias proteicas de FLOT1 descritas para ratón y rata. Zhang y col. (Zhang y col., *Genome Res.*; 10: 1546-1560 (2000)) purificaron FLOT1 humana mediante RT-PCR a partir de sangre de cordón y médula ósea de adultos CD34+. Mediante el análisis de micromatrices, encontraron una expresión débil de FLOT1 en 4 de 5 estirpes celulares hematopoyéticas y ninguna expresión en una estirpe celular promielocítica. El análisis de la transferencia Northern de tejidos de ratón reveló la expresión en tejido adiposo, corazón, músculo esquelético y pulmón. Mediante análisis de transferencia Western, también se encontró Flot1 como una proteína de 47 kD en los dominios de membrana ricos en caveolina aislados de pulmón humano y de adipocitos de ratón cultivados.

Pueden identificarse ligandos/fragmentos adecuados del CMH de FLOT1 humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

FOXP3: FOXP3 humana (caja forkhead P3), también conocida como IPEX o escurfina, se ubica en la región genómica 48993841-49008232 en el cromosoma X en la región Xp11.23. El transcrito de longitud completa consiste en 12 exones resultantes de la transcripción en un ARNm que tiene 6659 pares de bases. La secuencia codificante en sí consiste en 1830 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 609 aminoácidos.

Chatila y col. (Chatila y col., *J. Clin. Invest.*; 106: R75-81 (2000)) indicaron que el marco de lectura abierto de FOXP3 humana de 1146 pb codifica una proteína de 381 aminoácidos deducida. Mediante PCR de una biblioteca de ADNc de células mononucleares de sangre periférica humana, Smith y col. (Smith y col., *Immunology*; 119: 203-211 (2006)) clonaron FOXP3 de longitud completa y variantes de corte y empalme que carecen del exón 2 y ambos exones 2 y 7. La proteína de longitud completa contiene 431 aminoácidos. La variante que carece del exón 2 codifica una proteína de 396 aminoácidos que carece de parte del dominio rico en prolina y la variante que carece de los exones 2 y 7 codifica una proteína de 369 aminoácidos que también carece de una gran parte de la supuesta cremallera de leucina. Smith y col. (Smith y col., *Immunology*; 119: 203-211 (2006)) observaron que los ratones parecen carecer de variantes de Foxp3. FOXP3 se exploró para determinar los dominios funcionales y tiene un dominio Forkhead (FH) en la región 295-391. Forkhead (FH), también conocido como "hélice alada" se denomina así por la proteína forkhead de *Drosophila*, un factor de transcripción que promueve el desarrollo terminal más que segmentario. Esta familia de los dominios de factor de transcripción, que se unen a B-ADN como monómeros, también se encuentran en proteínas de factor nuclear de hepatocitos (HNF, por sus siglas en inglés), que

proporcionan regulación génica específica de tejido. La estructura contiene de 2 bucles flexibles o "alas" en la región C terminal, por tanto, de ahí el término hélice alada. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> no identificó regiones transmembrana potenciales. FOXP3 tiene una alta homología (> 86 %) a las secuencias de proteínas de FOXP3 descritas para ratón y rata. Stock y col. (Stock y col., *Nat. Immunol.*; 5: 1149-1156 (2004)) describieron un tipo de linfocito Tr específico de antígeno que se desarrolló in vivo a partir de linfocitos T vírgenes CD4-positivos/CD25-negativos durante una respuesta inmunitaria polarizada Th1. Estos linfocitos Tr fueron inducidos por células dendríticas CD8A-positivas, produjeron tanto IFNG como IL10 y expresaron el 'factor de transcripción principal' de Th1, TBET (TBX21), así como ICOS y FOXP3. Las células inhibieron la hiperreactividad de las vías respiratorias inducida por alérgenos de una manera dependiente de IL10 en ratones. Stock y col. (Stock y col., *Nat. Immunol.*; 5: 1149-1156 (2004)) llegaron a la conclusión de que estos linfocitos Tr adaptativos están relacionados con, pero son distintos de, los linfocitos Th1. Propusieron que existe un espectro de tipos de linfocitos Tr adaptativos, que comprende linfocitos similares a Th1 y los linfocitos Th2 notificados anteriormente que son inducidos por células dendríticas CD8A-negativas y expresan el 'factor de transcripción principal' de Th2, GATA3 y FOXP3. Pueden identificarse ligandos/fragmentos adecuados del CMH de FOXP3 humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

GPLY: GPLY Humana (granulinsina), también conocida como LAG2, se ubica en la región genómica 85774925-85779380 en el cromosoma 2 en la región 2p11.2. La granulinsina se corta y empalma de forma alternativa, dando como resultado los transcritos NKG5 y 519. El transcrito NKG5 consiste en 5 exones resultantes de la transcripción en un ARNm que tiene 738 pares de bases (véase la SEQ ID NO: 47). La secuencia codificante en sí consiste en 438 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 145 aminoácidos (SEQ ID NO: 48). El transcrito 519 consiste en 5 exones resultantes de la transcripción en un ARNm que tiene 853 pares de bases (véase la SEQ ID NO: 49). La secuencia codificante en sí consiste en 390 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 129 aminoácidos (SEQ ID NO: 50). La granulinsina es una proteína presente en gránulos citotóxicos de linfocitos T citotóxicos y linfocitos citolíticos naturales. La granulinsina es un miembro de la familia de proteínas similares a saposina (SAPLIP) y se ubica en los gránulos citotóxicos de los linfocitos T, que se liberan tras la estimulación por el antígeno. La granulinsina tiene actividad antimicrobiana frente a M. tuberculosis y otros organismos.

GPLY se exploró para determinar los dominios funcionales y tiene un dominio Saposina B (SapB) en la región 46-126. Las saposinas son proteínas lisosómicas pequeñas que sirven como activadores de diversas enzimas lisosómicas de degradación de lípidos. Probablemente actúan aislando el sustrato lipídico de los alrededores de la membrana, haciéndolo de este modo más accesible a las enzimas de degradación solubles. Todas las saposinas de mamíferos se sintetizan como una sola molécula precursora (prosaposina) que contiene cuatro dominios Saposina B, produciendo las saposinas activas después de la escisión proteolítica, y dos dominios Saposina-A que se retiran en la reacción de activación. Los dominios Saposina B también aparecen en otras proteínas, muchas de ellas activas en la lisis de las membranas. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 1 región transmembrana potencial (5-19) para el transcrito NKG5, mientras que para el transcrito 519 no pudo identificarse ninguna región transmembrana. La proteína GPLY no tiene homología descrita a las secuencias proteicas GPLY descritas para ratón y rata.

Pueden identificarse ligandos/fragmentos adecuados del CMH de GPLY humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

ICA1: ICA1 humano (autoantígeno 1 de las células de los islotes), también conocido como ICA69, se ubica en la región genómica 8119940-8268710 en el cromosoma 7 en la región 7p22. Se han descrito variantes cortadas y empalmadas de forma alternativa que codifican diferentes isoformas de la proteína; sin embargo, no todas las variantes se han caracterizado completamente. Hasta la fecha, se han descrito 2 transcritos para ICA1. El primer transcrito consiste en 14 exones resultantes de la transcripción en un ARNm que tiene 2396 pares de bases (véase la SEQ ID NO: 51). El segundo transcrito consiste en 14 exones resultantes de la transcripción en un ARNm que tiene 2268 pares de bases (véase la SEQ ID NO: 53). La secuencia codificante en sí de ambas transcripciones consiste en 1452 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 483 aminoácidos (SEQ ID NO: 54).

ICA1 se encuentra en el citosol y como forma unida a la membrana en el complejo de Golgi y gránulos secretores inmaduros. Se cree que esta proteína es un autoantígeno en la diabetes mellitus dependiente de insulina y en el síndrome de Sjögren primario.

ICA1 se exploró para determinar los dominios funcionales y tiene un dominio Arfaptina en la región 22 a 249. Arfaptina es una proteína expresada de forma ubicua implicada en mediar la interacción entre Rac, un miembro de la familia Rho y GTPasas pequeñas Arf; la Arfaptina se une a Arf6 y Arf1 unidos a GTP, pero se une Rac.GTP y Rac.GDP con afinidades similares. Las estructuras de Arfaptina con Rac unido ya sea a GDP o al análogo lentamente hidrolizable GMPPNP, muestran que las regiones de cambio adoptan conformaciones similares en ambos complejos. Se cree que Arf1 y Arf6 se unen a la misma superficie que Rac. ICA1, adicionalmente, tiene un dominio ICA69 (ICA69) autoantigénico de células de los islotes en la región 261-483. Esta familia incluye una

5 proteína de 69 kD que se ha identificado como un autoantígeno de células de los islotes en la diabetes mellitus de tipo I. Su función precisa es desconocida. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> no identificó regiones transmembrana potenciales. ICA1 tiene una alta homología (> 89 %) a las secuencias proteicas de ICA1 descritas para ratón y rata. ICA1 se expresa en glándulas lagrimales y salivales. En ratones NOD, en los que la pérdida de la función secretora salival se desarrolla espontáneamente (como en el síndrome de Sjogren humano). Winer y col. (Winer y col., *Lancet*; 360: 1063-1069 (2002)) descubrieron que la alteración del gen Ica1 evitó la enfermedad de la glándula lagrimal y redujo en gran medida la enfermedad de la glándula salival. Estos animales desarrollan diabetes de tipo 1 con un ligero retraso, pero con la misma incidencia que en animales de tipo silvestre, asignando un papel facultativo más que obligado a ICA1 en el desarrollo de la diabetes.

10 Pueden identificarse ligandos/fragmentos adecuados del CMH de ICA1 humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

15 IL10RA: IL10RA humano (receptor de interleucina 10, alfa 1) se ubica en la región genómica 117362319-117377404 en el cromosoma 11 en la región 11q23. Consiste en 7 exones resultantes de la transcripción en un ARNm que tiene 3649 pares de bases (véase la SEQ ID NO: 57). La secuencia codificante en sí consiste en 1737 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 578 aminoácidos (SEQ ID NO: 58).

20 La proteína codificada por este gen es un receptor para la interleucina 10. Esta proteína está relacionada estructuralmente con receptores de interferón. Se ha demostrado que media la señal inmunosupresora de la interleucina 10 y, por tanto, inhibe la síntesis de citocinas proinflamatorias. Se ha notificado que este receptor promueve la supervivencia de las células progenitoras mieloides a través de la vía del sustrato del receptor de insulina-2/PI 3-kinasa/AKT. La activación de este receptor conduce a la fosforilación de tirosinas de las cinasas JAK1 y TYK2.

25 IL10RA se exploró para determinar los dominios funcionales y tiene 2 dominios de pliegue similares a Fibronectina de tipo 3 en las regiones 22-124 y 125-235.

30 La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 2 regiones transmembrana potenciales (5-16 y 232-256). La proteína supuesta muestra una homología intermedia (~ 56 %) a las secuencias proteicas de IL10RA descritas para ratón y rata. La interleucina-10 (IL-10) es una citocina producida por las células B, los linfocitos T auxiliares y las células del linaje de monocitos/macrófagos, que presenta diversas actividades en diferentes estirpes celulares. Tan y col. (Tan y col., *J. Biol. Chem.*; 268: 21053-21059 (1993)) demostraron que la proteína puede yodarse enzimáticamente para una alta radiactividad específica con retención de la actividad biológica. Se descubrió que el ligando radiomarcado se une específicamente a su receptor en varias estirpes celulares de ratón y humanas. Para las dos estirpes celulares de ratón y humana examinadas, parecía que había, a lo sumo, solo unos pocos cientos de receptores de IL-10 por célula. La IL10 de ratón fue capaz de bloquear la unión de IL-10 humana a células de ratón, pero no a células humanas. Ho y col. (Ho y col., *Proc. Natl. Acad. Sci. USA.*; 90: 11267-11271 (1993)) descubrió que el IL10R de ratón está estructuralmente relacionado con receptores de interferón. Puesto que IL10 inhibe la activación de los macrófagos por el interferón-gamma, una posible implicación de esta relación es la interacción de IL10R y IFN-gamma-R o sus vías de señalización.

40 Pueden identificarse ligandos/fragmentos adecuados del CMH de IL10RA humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

45 IL7R: IL7R humano (receptor de interleucina 7), también conocido como CD127, se ubica en la región genómica 35892748-35915462 en el cromosoma 5 en la región 5p13. Consiste en 8 exones resultantes de la transcripción en un ARNm que tiene 1802 pares de bases (véase la SEQ ID NO: 55). La secuencia codificante en sí consiste en 1380 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 459 aminoácidos (SEQ ID NO: 56).

50 La proteína codificada por este gen es un receptor de interleucina 7 (IL7). La función de este receptor requiere la cadena gamma del receptor de interleucina 2 (IL2RG), que es una cadena gamma común compartida por los receptores de diversas citocinas, incluyendo interleucina 2, 4, 7, 9 y 15. Se ha demostrado que esta proteína desempeña un papel crítico en la recombinación V(D)J durante el desarrollo de los linfocitos. También se ha descubierto que esta proteína controla la accesibilidad del locus gamma de TCR por acetilación de STAT5 e histonas. Los estudios de inactivación en ratones sugirieron que el bloqueo de la apoptosis es una función esencial de esta proteína durante la diferenciación y la activación de los linfocitos T. Los defectos funcionales en esta proteína pueden estar asociados a la patogenia de la inmunodeficiencia combinada grave (IDCG).

55 IL7R se exploró para determinar los dominios funcionales y tiene un dominios de pliegue similares a Fibronectina de tipo 3 en la región 125-228.

La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 2 regiones transmembrana potenciales (8-18 y 241-264).

La proteína supuesta muestra una homología intermedia (> 64 %) a las secuencias proteicas de IL7R descritas para ratón y rata. De esta relación es la interacción de IL10R y IFN-gamma-R o sus vías de señalización. Reche y col. (Reche y col., *J. Immunol.*; 167: 336-343 (2001)) demostraron que la expresión del receptor de linfopoyetina estromal tímico (TSLPR) e IL7R, juntos, pero no solos, indujo una respuesta proliferativa a TSLP, pero no a IL7, lo que indica que el receptor de TSLP funcional consiste en estas 2 subunidades. El análisis por PCR de bibliotecas de ADNc sugirió que las DC (células dendríticas) y los monocitos coexpresan IL7R y TSLPR.

Pueden identificarse ligandos/fragmentos adecuados del CMH de IL7R humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

LASS6: LASS6 humana (homólogo de LAG1, ceramida sintasa 6 (*S. cerevisiae*)) se ubica en la región genómica 169021081-169339398 en el cromosoma 2 en la región 2q24.3. Consiste en 10 exones resultantes de la transcripción en un ARNm que tiene 6259 pares de bases (véase la SEQ ID NO: 61). La secuencia codificante en sí consiste en 1155 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 384 aminoácidos (SEQ ID NO: 62).

LASS6 se exploró para determinar los dominios funcionales y tiene un dominio Homeobox en la región 81-126. Los dominios Homeobox son dominios de unión a ADN implicados en la regulación transcripcional de procesos de desarrollo eucariotas clave y pueden unirse al ADN como monómeros o como homo y/o heterodímeros, de una forma específica de la secuencia. LASS6, adicionalmente, tiene un dominio LAG1 en la región 150-348. Los miembros de esta familia están implicados en la determinación de la esperanza de vida. Los mecanismos moleculares mediante los cuales LAG1 determina la longevidad no están claros, aunque ciertas pruebas sugieren una participación en la síntesis de ceramida.

La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó de 5 a 7 regiones transmembrana potenciales (40-55, 140-149, 182-192 (?), 207-224, 230-237, 261-286, 301-325). Por tanto, LASS6 es una proteína de membrana de múltiples pasadas potencial. La proteína LASS6 tiene una homología muy alta (> 95 %) a las secuencias proteicas de LASS6 descritas para ratón y rata. LASS6 puede estar implicada en la síntesis de esfingolípidos o su regulación (determinada por similitud) y se ubica en el núcleo, la membrana nuclear, el retículo endoplasmático y la membrana del retículo endoplasmático.

Pueden identificarse ligandos/fragmentos adecuados del CMH de LASS6 humana mediante el procedimiento descrito en relación con la Colonia 1 anterior.

MGST2: MGST2 humana (glutación S-transferasa microsómica 2), también conocida como GST2, se ubica en la región genómica 140806372-140844857 en el cromosoma 4 en la región 4q28.3. Consiste en 5 exones resultantes de la transcripción en un ARNm que tiene 801 pares de bases (véase la SEQ ID NO: 65). La secuencia codificante en sí consiste en 444 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 147 aminoácidos (SEQ ID NO: 66).

Este gen codifica una proteína que cataliza la conjugación de leucotrieno A4 y glutatión reducido para producir leucotrieno C4.

MGST2 se exploró para determinar los dominios funcionales y tiene un dominio Proteínas Asociadas a la Membrana en el Metabolismo de Eicosanoide y Glutación (MAPEG) en la región 3-130. El MAPEG (Proteínas Asociadas a la Membrana en el Metabolismo de Eicosanoide y Glutación) consiste en seis proteínas humanas, varias de las cuales están implicadas en la producción de leucotrienos y prostaglandina E, mediadores importantes de la inflamación.

La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 4 regiones transmembrana potenciales (7-15, 60-71, 78-88, 110-127). Por tanto, LASS6 es una proteína potencial de membrana de múltiples pasadas. La proteína MGST2 tiene una homología muy alta (> 95 %) a las secuencias proteicas de MGST2 descritas para ratón y rata. El análisis por transferencia Northern reveló que el gen se expresa como un transcrito de 0,6 kb en una diversidad de tejidos humanos, pero solo se expresa débilmente (si es que lo hace) en pulmón, cerebro, placenta y médula ósea. Por el contrario, se expresa ARNm de MGST2 en pulmón, diversos órganos del sistema inmunitario y leucocitos de sangre periférica. Pueden identificarse ligandos/fragmentos adecuados del CMH de MGST2 humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

TARP: TARP humana (proteína de marco de lectura alternativo de TCR gamma), también conocida como cadena gamma del receptor de linfocitos T, se ubica en la región genómica 38279772-38265768 en el cromosoma 7 en la región 7p15-p14. La isoforma a consiste en 4 exones resultantes de la transcripción en un ARNm que tiene 1027 pares de bases (véase la SEQ ID NO: 103). La secuencia codificante en sí consiste en 177 pares de bases. La traducción de esta secuencia da como resultado una proteína con una longitud de 58 aminoácidos (SEQ ID NO: 104). La isoforma b consiste en 4 exones resultantes de la transcripción en un ARNm que tiene 1027 pares de bases. La secuencia codificante en sí consiste en 336 pares de bases. La traducción de esta secuencia da como resultado una proteína con una longitud de 111 aminoácidos.

En algunos tejidos no linfoides, se expresa el locus no reordenado del receptor de linfocitos T gamma (TRGγ). El transcrito resultante contiene un subconjunto de los segmentos génicos TRGγ y es más corto que los transcritos TRGγ expresados en los tejidos linfoides. Este registro RefSeq representa el transcrito de locus TRGγ no reordenado; el locus TRGγ completo está representado por el RefSeq genómico NG_001336. El transcrito representado por este RefSeq tiene dos marcos de lectura abiertos (ORF, por sus siglas en inglés) que codifican proteínas diferentes. El ORF corriente abajo está en el mismo marco que TRGγ y su producto proteico es similar a las proteínas TRGγ. El ORF corriente arriba usa un marco de lectura diferente y codifica una proteína nueva.

La proteína se exploró para determinar los dominios funcionales y TARP no tiene dominios funcionales conocidos. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó para la isoforma a 2 regiones transmembrana potenciales (8-16, 28-39), mientras las regiones de 10-13 y 30-37 tenían la probabilidad más alta, y para la isoforma b una región transmembrana potencial (76-100), mientras que la región de 77-99 tenía la probabilidad más alta. La proteína supuesta tiene una alta homología (> 68 %) a las secuencias proteicas supuestas de TARP descritas para ratón. Esta proteína se ubica principalmente en el núcleo y las mitocondrias. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en el tejido de la próstata, mientras que solo se observaron niveles de expresión bajos o ausentes de TARP en todos los demás tejidos analizados, incluyendo médula ósea, hígado, corazón, bazo, riñón, timo, músculo esquelético, cerebro, médula espinal, pulmón y páncreas. El gen TARP está embebido dentro de un intrón del locus del receptor gamma de linfocitos T (TCRG), que codifica un receptor de linfocitos T alternativo que siempre se coexpresa con el receptor delta de linfocitos T (TCRD). Aunque las proteínas TARP y TCRG son distintas, el ARNm de TARP se detecta mediante sondas de ADNc que detectan TCRG. A diferencia de TCRG, sin embargo, TARP se detecta en las bibliotecas de cáncer de próstata en ausencia de TCRD. TARP está regulada positivamente por andrógenos (Maeda y col., *J. Biol. Chem.*; 279: 24.561-24.568 (2004)). Usando análisis por fraccionamiento subcelular, inmunocitoquímica y transferencia Western, Maeda y col. (Maeda y col., *J. Biol. Chem.*; 279: 24561-24568 (2004)) descubrieron que TARP se regulaba positivamente por andrógenos en la mitocondria de las estirpes celulares de próstata. El fraccionamiento de las mitocondrias y la inmunohistoquímica indicaron que TARP se ubicaba en la membrana mitocondrial externa. Maeda y col. (Maeda y col., *J. Biol. Chem.*; 279: 24561-24568 (2004)) propusieron que TARP puede actuar sobre las mitocondrias para realizar su función biológica. Maeda y col. (Maeda y col., *J. Biol. Chem.*; 279: 24561-24568 (2004)) descubrieron que TARP se regulaba positivamente por andrógenos en la mitocondria de estirpes celulares de próstata. El fraccionamiento de las mitocondrias y la inmunohistoquímica indicaron que TARP se ubicaba en la membrana mitocondrial externa. Maeda y col. (Maeda y col., *J. Biol. Chem.*; 279: 24561-24568 (2004)) propusieron que TARP puede actuar sobre las mitocondrias para realizar su función biológica.

Pueden identificarse ligandos/fragmentos adecuados del CMH de TARP humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

RBMS1: RBMS1 humana (motivo de unión a ARN, proteína de interacción monocatenaria 1) se ubica en la región genómica 161058550-160838757 en el cromosoma 2 en la región 2q24.2. La isoforma a consiste en 11 exones resultantes de la transcripción en un ARNm que tiene 2438 pares de bases (véase la SEQ ID NO: 81). La secuencia codificante en sí consiste en 1221 pares de bases. La traducción de esta secuencia da como resultado una proteína con una longitud de 406 aminoácidos (SEQ ID NO: 82). La isoforma b consiste en 12 exones resultantes de la transcripción en un ARNm que tiene 2596 pares de bases (véase la SEQ ID NO: 83). La secuencia codificante en sí consiste en 1170 pares de bases. La traducción de esta secuencia da como resultado una proteína con una longitud de 389 aminoácidos (SEQ ID NO: 84). La isoforma c consiste en 11 exones resultantes de la transcripción en un ARNm que tiene 2429 pares de bases (véase la SEQ ID NO: 85). La secuencia codificante en sí consiste en 1212 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 403 aminoácidos (SEQ ID NO: 86).

RBMS codifica un número de una pequeña familia de proteínas que se unen a ADN/ARN monocatenario. Estas proteínas se caracterizan por la presencia de dos grupos de secuencia de consenso de ribonucleoproteína (RNP-CS, por sus siglas en inglés) que contienen motivos conservados, RNP1 y RNP2, descritos originalmente en proteínas de unión a ARN y necesarias para la unión al ADN. Estas proteínas se han implicado en diversas funciones tales como la replicación del ADN, la transcripción de genes, la progresión del ciclo celular y la apoptosis. Se han descrito múltiples variantes de transcripción, como resultado del corte y empalme alternativo y que codifican diferentes isoformas. Varios de estas se aislaron en virtud de su unión a cualquier cadena de un elemento corriente arriba de c-myc (MSSP) o por complementación fenotípica de mutantes *cdc2* y *cdc13* de levadura (*scr2*) o como un represor humano potencial del promotor de la transcripción alfa de VIH-1 y ILR-2 (YC1). Se encuentra un pseudogen para este locus en el cromosoma 12.

La proteína se exploró para determinar los dominios funcionales y RBMS1 tiene 2 motivos de reconocimiento de ARN en la región 63-131 y 142-213. Muchas proteínas eucarióticas que se sabe o se supone que se unen a ARN monocatenario contienen una o más copias de un dominio de unión a ARN putativo de aproximadamente 90 aminoácidos. Esto se conoce como el distintivo RNP-1 de región de unión al ARN supuesto eucariota o motivo de reconocimiento del ARN (RRM, por sus siglas en inglés). Los RRM se encuentran en una diversidad de proteínas de unión al ARN, incluyendo ribonucleoproteínas nucleares heterogéneas (hnRNPs), proteínas implicadas en la regulación del corte y empalme alternativo y los componentes proteicos de ribonucleoproteínas nucleares pequeñas

(snRNPs). El motivo también aparece en unas pocas proteínas de unión al ADN monocatenario. La estructura del RRM consiste en cuatro cadenas y dos hélices dispuestas en un sándwich alfa/beta, con una tercera hélice presente durante la unión al ARN en algunos casos. Se construyeron dos modelos individuales que identifican subtipos de este dominio, pero no existe ninguna diferencia funcional entre los subtipos. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> no identificó regiones transmembrana potenciales. La proteína tiene una alta homología (> 96 %) a las secuencias proteicas de RBMS1 descritas para ratón y rata. Esta proteína se ubica principalmente en el núcleo. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en el pulmón y el tejido renal, mientras que solo se observaron niveles de expresión bajos o ausentes de Colonia 2 en todos los demás tejidos analizados, incluyendo médula ósea, hígado, corazón, bazo, timo, músculo esquelético, cerebro, médula espinal, próstata y páncreas.

Pueden identificarse ligandos/fragmentos adecuados del CMH de la proteína RBMS1 humana mediante el procedimiento descrito en relación con la Colonia 1 anterior.

NCF4: NCF4 humano (factor citosólico de neutrófilos 4, 40 kDa) se ubica en la región genómica 35586990-35604004 en el cromosoma 22 en la región 22q13.1. La isoforma a consiste en 9 exones resultantes de la transcripción en un ARNm que tiene 1386 pares de bases (véase la SEQ ID NO: 67). La secuencia codificante en sí consiste en 1020 pares de bases. La traducción de esta secuencia da como resultado una proteína con una longitud de 339 aminoácidos (SEQ ID NO: 68). La isoforma b consiste en exones resultantes de la transcripción en un ARNm que tiene 1631 pares de bases (véase la SEQ ID NO: 69). La secuencia codificante en sí consiste en 1047 pares de bases. La traducción de esta secuencia da como resultado una proteína con una longitud de 348 aminoácidos (SEQ ID NO: 70).

La proteína se exploró para determinar los dominios funcionales y NCF4 tiene un dominio homólogo de PhOx en la región 19-140, que es un dominio eucariótico de función desconocida presente en las proteínas phox. El dominio PX (phox) aparece en una diversidad de proteínas eucarióticas y se ha implicado en funciones muy diversas tales como la señalización celular, el tráfico vesicular, la clasificación de proteínas y la modificación de lípidos. Los dominios PX son módulos de unión a fosfoinosítidos importantes, que tienen diferentes especificidades de unión a lípidos. El dominio PX tiene aproximadamente 120 restos de longitud y se pliega en una lámina beta de tres cadenas seguida de tres hélices y una región rica en prolina que precede inmediatamente a un bucle de interacción con la membrana y abarca aproximadamente ocho restos hidrófobos y polares. El dominio PX de p47phox se une al dominio SH3 de la misma proteína. La fosforilación de p47 (phox), un activador citoplasmático de la oxidasa de fagocito microbicida (phox), desencadena la interacción de p47 (phox) con los fosfoinosítidos. El cambio conformacional impulsado por la fosforilación de la proteína de p47 (phox) permite a su dominio PX unirse a fosfoinosítidos, cuya interacción desempeña un papel crucial en el reclutamiento de p47 (phox) desde el citoplasma a las membranas y la posterior activación de la oxidasa de fagocitos. La actividad de unión a lípidos de esta proteína normalmente está suprimida mediante la interacción intramolecular del dominio PX con el dominio 3 de homología a Src C terminal (SH3). El dominio PX se conserva desde la levadura hasta los seres humanos. Puede encontrarse una alineación múltiple reciente de secuencias de dominio PX representativas y, aunque muestra relativamente poca conservación de la secuencia, su estructura parece estar altamente conservada. Aunque el fosfatidilinositol-3-fosfato (PtdIns(3)P) es la diana principal de los dominios PX, también se ha notificado la unión a ácido fosfatídico, fosfatidilinositol-3,4-bisfosfato (PtdIns(3,4)P2), fosfatidilinositol-3,5-bisfosfato (PtdIns(3,5)P2), fosfatidilinositol-4,5-bisfosfato (PtdIns(4,5)P2) y fosfatidilinositol-3,4,5-trisfosfato (PtdIns(3,4,5)P3). El PX-dominio es también un dominio de interacción proteína-proteína. NCF4, adicionalmente, tiene un dominio 3 de homología a Src en la región 175-226, que se une a proteínas diana a través de secuencias que contienen prolina y aminoácidos hidrófobos. Los polipéptidos que contienen Pro pueden unirse a dominios SH3 en 2 orientaciones de unión diferentes. Los dominios SH3 (homología a src-3) son módulos de proteínas pequeñas que contienen aproximadamente 50 restos de aminoácidos. Se encuentran en una gran diversidad de proteínas intracelulares o asociadas a la membrana, por ejemplo, en una diversidad de proteínas con actividad enzimática, en proteínas adaptadoras que carecen de secuencias catalíticas y en proteínas del citoesqueleto, tales como fodrina y la proteína de unión a actina de levadura ABP-1. El dominio SH3 tiene un pliegue característico que consiste en cinco o seis cadenas beta dispuestas como dos láminas beta anti-paralelas empaquetadas estrechamente. Las regiones de engarce pueden contener hélices cortas. La superficie del dominio SH2 lleva un bolsillo de unión al ligando hidrófobo, plano, que consiste en tres surcos poco profundos definidos por restos aromáticos conservadores en los que el ligando adopta una disposición helicoidal hacia la izquierda extendida. El ligando se une con baja afinidad, pero esto puede potenciarse mediante múltiples interacciones. La región unida por el dominio SH3 es, en todos los casos, rica en prolina y contiene PXXP como un motivo de unión conservado en el núcleo. La función del dominio SH3 no se comprende bien, pero puede mediar en muchos procedimientos diversos, tales como el aumento de la concentración local de proteínas, alterando su localización subcelular y la mediación en el ensamblaje de grandes complejos multiproteicos. Además, NCF4 tiene una señal de localización nuclear en la región 148-158. Las secuencias de señal de localización nuclear (NLS, por sus siglas en inglés) clásicas incorporan regiones enriquecidas en aminoácidos básicos y, en general, se ajustan a uno de tres motivos. La NLS "pat4" consiste en un tramo continuo de cuatro aminoácidos básicos (lisina o arginina) o tres aminoácidos básicos asociados a histidina o prolina. La NLS "pat7" comienza con una prolina y va seguida, después de tres restos, de una secuencia de aminoácidos que contiene tres restos básicos de cada cuatro. El tercer tipo de NLS, conocido como un motivo "bipartito", consiste en

dos aminoácidos básicos, un espaciador de 10 aminoácidos y una secuencia de cinco aminoácidos que contiene al menos tres restos básicos. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 4 regiones transmembrana potenciales (39-42, 115-124, 196-200, 218-224), mientras que la región 117-119 tenía la probabilidad más alta. La proteína tiene una alta homología (> 75 %) a las secuencias proteicas de NCF4 descritas para ratón y rata. Esta proteína se ubica principalmente en el citoplasma. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en la médula ósea, el bazo y el tejido pulmonar, mientras que solo se observaron niveles de expresión bajos o ausentes de tNCF4 en todos los demás tejidos analizados.

Pueden identificarse ligandos/fragmentos adecuados del CMH de las proteínas NCF4 humanas mediante el procedimiento descrito en relación con la Colonia 1 anterior.

PTTG3: PTTG3 humano (transformante de tumor hipofisario 3) se ubica en la región genómica 67842186-67842794 en el cromosoma 8 en la región 8q13.1. Consiste en 1 exón resultante de la transcripción en un ARNm que tiene 609 pares de bases (véase la SEQ ID NO: 77). La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 202 aminoácidos (SEQ ID NO: 78).

PTTG3 se exploró para determinar los dominios funcionales y tiene un dominio Segurina inhibidor de la separación de cromátidas hermanas en la región 1-191. La segurina, también conocida como producto del gen transformante de tumor hipofisario, es una proteína reguladora que desempeña un papel fundamental en la estabilidad de los cromosomas en la vía p53/TP53 y en la reparación del ADN. Probablemente actúa bloqueando la acción de las proteínas clave, por ejemplo, durante la mitosis bloquea la función Separasa/ESPL1 evitando la proteólisis del complejo cohesina y la posterior segregación de los cromosomas. Al inicio de la anafase, se ubiquitina, conduciendo a su destrucción y a la liberación de ESPL1. Su función, sin embargo, no se limita a una actividad inhibidora, puesto que es necesaria para activar ESPL1. La regulación negativa de la actividad transcripcional y la actividad apoptótica relacionada de TP53 pueden explicar la fuerte capacidad transformante de la proteína cuando se sobreexpresa. La sobreexpresión de segurina se asocia a una serie de tumores y se ha propuesto que esto puede deberse a la separación de cromátidas erróneas conduciendo la ganancia o pérdida de cromosomas. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 1 región transmembrana potencial (181-191). No se han descrito homologías para secuencias proteicas de PTTG3 descritas para ratón y rata.

Pueden identificarse ligandos/fragmentos adecuados del CMH de PTTG3 humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

FAM129A: FAM129A humano (familia con similitud de secuencia 129, miembro A), también conocido como NIBAN o C1orf24, se ubica en la región genómica 183026787-183210305 en el cromosoma 1 en la región 8q13.1. Consiste en 14 exones resultantes de la transcripción en un ARNm que tiene 6928 pares de bases (véase la SEQ ID NO: 37). La secuencia codificante en sí consiste en 2787 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 928 aminoácidos (SEQ ID NO: 38).

FAM129A se exploró para determinar los dominios funcionales. No se identificaron dominios funcionales. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> no identificó regiones transmembrana potenciales. La proteína supuesta tiene una homología intermedia (> 71 %) a las secuencias proteicas de FAM129A descritas para ratón y rata. La expresión de FAM129A se regula positivamente en diversos tipos de tumores tiroideos. Pueden identificarse ligandos/fragmentos adecuados del CMH de FAM129A humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

KIAA1600: KIAA1600 humana, también conocida como proteína hipotética LOC57700, se ubica en la región genómica 116571493-116649581 en el cromosoma 10 en la región 1q25. Consiste en 17 exones resultantes de la transcripción en un ARNm que tiene 5806 pares de bases (véase la SEQ ID NO: 59). La secuencia codificante en sí consiste en 2298 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 765 aminoácidos (SEQ ID NO: 60).

La proteína se exploró para determinar los dominios funcionales y para KIAA1600 no se detectaron dominios funcionales. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 1 región transmembrana potencial (689-701). La proteína supuesta tiene una alta homología (> 92 %) a las secuencias de proteínas supuestas de KIAA1600 descritas para ratón y rata.

Pueden identificarse ligandos/fragmentos adecuados del CMH de KIAA1600 humana mediante el procedimiento descrito en relación con la Colonia 1 anterior.

C6ORF19Q: C6ORF190 humano (marco de lectura abierto 190 del cromosoma 6), también conocido como C6orf207, se ubica en la región genómica 128065879-128281435 en el cromosoma 6 en la región 6q22.33. C6ORF190 existe en 3 variantes de transcrito. La primera variante consiste en 7 exones resultantes de la transcripción en un ARNm que tiene 2248 pares de bases (SEQ ID NO: 735). La traducción de esta secuencia da

como resultado una proteína supuesta con una longitud de 643 aminoácidos (SEQ ID NO: 736). La segunda variante consiste en 7 exones resultantes de la transcripción en un ARNm que tiene 4029 pares de bases. La secuencia codificante en sí consiste en 1926 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 641 aminoácidos. La tercera transcripción consiste en 7 exones resultantes de la transcripción en un ARNm que tiene 4101 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 562 aminoácidos.

Todos los transcritos de C6ORF190 se exploraron para determinar los dominios funcionales. Ninguno de los tres transcritos contiene ningún dominio funcional. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> no identificó regiones transmembrana potenciales. La proteína supuesta tiene una alta homología (> 81 %) a las secuencias de la proteína supuesta de E430004N04Rik descritas para ratón. No se describen homologías para rata.

Pueden identificarse ligandos/fragmentos adecuados de las proteínas C6ORF190 humanas mediante el procedimiento descrito en relación con la Colonia 1 anterior.

MGC33556: MGC33556 humana, también conocida como proteína LOC339541 hipotética, se ubica en la región genómica 44912981-44963848 en el cromosoma 1 en la región 1p34.1. MGC33556 consiste en 10 exones resultantes de la transcripción en un ARNm que tiene 2113 pares de bases (véase la SEQ ID NO: 63). La secuencia codificante en sí consiste en 924 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 307 aminoácidos (SEQ ID NO: 64). MGC33556 se exploró para determinar los dominios funcionales, pero no se detectaron dominios funcionales. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 2 regiones transmembrana potenciales (140-152, 298-308). La proteína supuesta tiene una alta homología (> 82 %) a las secuencias de proteínas supuestas de Gm1661 descritas para ratón y RGD1563714_prevista para rata.

Pueden identificarse ligandos/fragmentos adecuados de MGC33556 humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

NELL2: NELL2 humana (proteína similar a NEL 2 (pollo)), también conocida como proteína 2 similar al factor de crecimiento epidérmico neural, se ubica en la región genómica 43556404-43188331 en el cromosoma 12 en la región 12q13.11-q13.12. Consiste en 13 exones resultantes de la transcripción en un ARNm que tiene 3198 pares de bases (véase la SEQ ID NO: 71). La secuencia codificante en sí consiste en 2451 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 816 aminoácidos (SEQ ID NO: 72).

Este gen codifica una proteína citoplasmática que contiene repeticiones similares al factor de crecimiento epidérmico (EGF). La proteína heterotrimérica codificada puede estar implicada en la regulación del crecimiento y la diferenciación celulares. Una proteína similar en roedores está implicada en la craneosinostosis. Se ha descrito una variante de corte y empalme alternativa, pero su secuencia de longitud completa no se ha determinado.

La proteína se exploró para determinar los dominios funcionales y NELL2 tiene 2 dominios C de tipo factor de von Willebrand (vWF) en la región 274-330 y 700-755. El dominio vWF se encuentra en diversas proteínas plasmáticas: factores del complemento B, C2, CR3 y CR4; las integrinas (dominios I); tipos de colágeno VI, VII, XII y XIV; y otras proteínas extracelulares. Aunque la mayoría de las proteínas que contienen VWA son extracelulares, las más antiguas presentes en todos los eucariotas son todas proteínas intracelulares implicadas en funciones tales como la transcripción, la reparación del ADN, el transporte ribosómico y de membrana y el proteosoma. Una característica común parece ser la implicación en complejos de múltiples proteínas. Las proteínas que incorporan dominios de vWF participan en numerosos eventos biológicos (por ejemplo, adhesión celular, migración, retorno, formación de patrones y transducción de señales), implicando la interacción con una gran diversidad de ligandos. Un número de enfermedades humanas surgen de mutaciones en los dominios VWA. La predicción de la estructura secundaria de 75 secuencias de vWF alineadas ha revelado una secuencia en gran parte alternante de hélices alfa y cadenas beta. El dominio se denomina así por la repetición de tipo C del factor de von Willebrand (VWF) que se encuentra en proteínas de múltiples dominios/proteínas multifuncionales implicadas en el mantenimiento de la homeostasia. Para el factor de von Willebrand se cree que el dominio VWF duplicado participa en la oligomerización, pero no en la etapa inicial de dimerización. La presencia de esta región en un número de otras proteínas formadoras de complejos apunta a la posible implicación del dominio VWF en la formación de complejos. Además, NELL 2 tiene 6 dominios similares al EGF en la región 395-438, 440-480, 486-521, 525-552, 555-591 y 602-640. Además, NELL2 tiene un dominio superenrollado en la región 237-368. El superenrollamiento es un dominio de proteína que forma un haz de dos o tres hélices alfa. Los dominios superenrollados cortos están implicados en las interacciones proteicas pero los dominios superenrollados largos que forman varillas largas se producen en proteínas estructurales o motoras. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 4 regiones transmembrana potenciales (7-24, 100-106, 155-164, 210-213), mientras que las regiones 9-21 y 158-162 tenían la probabilidad más alta. La proteína supuesta tiene una alta homología (> 93 %) a las secuencias de proteínas supuestas de NELL descritas para ratón y rata. Esta proteína se ubica principalmente en el citoplasma. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en el cerebro, el timo, el bazo y el tejido pulmonar, mientras que solo se observaron

niveles de expresión bajos o ausentes de NELL2 en todos los demás tejidos analizados. Las proteínas que contienen repeticiones similares al factor de crecimiento epidérmico (EGF) con frecuencia están implicadas en la regulación del crecimiento y la diferenciación.

5 Pueden identificarse ligandos/fragmentos adecuados de NELL2 humana mediante el procedimiento descrito en relación con la Colonia 1 anterior.

10 PTPLA: PTPLA humana (similar a proteína tirosina fosfatasa) se ubica en la región genómica 17699378-17671963 en el cromosoma 10 en la región 10p14-p13. Consiste en 6 exones resultantes de la transcripción en un ARNm que tiene 1323 pares de bases (véase la SEQ ID NO: 73). La secuencia codificante en sí consiste en 867 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 288 aminoácidos (SEQ ID NO: 74).

15 La proteína codificada por este gen contiene un motivo catalítico característico de la familia de las proteína tirosina fosfatasas (PTP). El motivo PTP de esta proteína tiene el resto de arginina altamente conservado reemplazado por un resto de prolina; por tanto, puede representar una clase distinta de PTP. Se sabe que los miembros de la familia PTP son moléculas que regulan una diversidad de procesos celulares de señalización. Este gen se expresa preferentemente en el corazón tanto adulto como fetal. Se detectó un nivel de expresión mucho más bajo en los tejidos del músculo esquelético y liso y no se observó ninguna expresión en otros tejidos. La expresión específica de tejido en el corazón en desarrollo y adulto sugiere un papel en la regulación del desarrollo y la diferenciación del corazón. La proteína se exploró para determinar los dominios funcionales y PTPLA tiene 4 dominios transmembrana en la región 68-90, 121-143, 206-228 y 248-270. Los dominios transmembrana se componen de 15-30 restos hidrófobos y se encuentran en proteínas que abarcan una membrana en una célula. Una proteína puede tener un único dominio TM (por ejemplo, receptores transmembrana de Tipo I tales como el receptor EGF) o múltiples dominios TM (por ejemplo, receptores acoplados a proteína G). Para abarcar el núcleo hidrocarbonado de la membrana de ~ 3 nm requiere una hélice alfa de ~20 restos predominantemente apolares sin carga. La membrana separa dos compartimentos acuosos mediante una fase lipídica bidimensional delgada. Las proteínas de membrana generalmente abarcan esta fase lipídica y, por tanto, necesitan acomodarse al tramo hidrófilo a ambos lados de la membrana, así como al entorno hidrófobo en el núcleo de la bicapa. La estructura de las porciones embebidas en la membrana consiste ya sea en hélices alfa transmembrana montadas con frecuencia en haces de hélices, o en láminas beta antiparalelas que forman poros en forma de barril. La longitud de la hélice transmembrana puede incluir correlacionarse con el espesor de su membrana. Las proteínas que abarcan una sola vez la membrana del retículo endoplasmático (RE) y Golgi tienen generalmente dominios transmembrana más cortos que las proteínas de la membrana plasmática. Puesto que el contenido de colesterol y, por tanto, el espesor de la bicapa lipídica también aumenta junto con la vía secretora, esto podría reflejar un papel de los segmentos y proteínas transmembrana en la clasificación de los lípidos. En proteínas que abarcan la membrana múltiples veces las hélices transmembrana se unen estrechamente a la estructura compacta, globular, de la que se excluyen los lípidos. La proteína tiene una alta homología (> 86 %) a las secuencias proteicas de PTPLA descritas para ratón y rata. Esta proteína se ubica principalmente en las membranas. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en el tejido cardíaco y en células del músculo liso mientras que solo se observaron niveles de expresión bajos o ausentes de PTPLA en todos los demás tejidos analizados. Las proteína tirosina fosfatasas (PTP) median la defosforilación de fosfotirosina. Se sabe que los PTP están implicados en muchas vías de transducción de señales que conducen al crecimiento celular, la diferenciación y la transformación oncogénica. PTPLA es una proteína similar a PTP que contiene el sitio catalítico conservado de las proteínas PTP pero con un resto de prolina en lugar de un resto de arginina conservado (Uwanogho y col., *Genomics*; 62: 406-416). Se sugiere que la sustitución de una arginina esencial por prolina dentro del presunto sitio activo de PTPLA puede volver la proteína inactiva, haciendo de la PTPLA una antifosfatasa que antagoniza competitivamente las PTP.

45 Pueden identificarse ligandos/fragmentos adecuados del CMH de la PTPLA humana mediante el procedimiento descrito en relación con la Colonia 1 anterior.

50 PTTG1: PTTG1 humano (transformante de tumor hipofisario 1) se ubica en la región genómica 159781442-159788323 en el cromosoma 5 en la región 5q35.1. Consiste en 6 exones resultantes de la transcripción en un ARNm que tiene 728 pares de bases (véase la SEQ ID NO: 75). La secuencia codificante en sí consiste en 609 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 202 aminoácidos (SEQ ID NO: 76).

55 PTTG1 es un homólogo de proteínas segurina de levadura, que evita que las separinas promuevan la separación de cromátidas hermanas. Es un sustrato del complejo promotor de la anafase (APC, por sus siglas en inglés) que se asocia a un separina hasta la activación del APC. El producto génico tiene actividad transformante in vitro y actividad oncogénica in vivo y el gen se expresa altamente en diversos tumores. El producto génico contiene 2 motivos PXXP, que son necesarios para sus actividades transformantes y oncogénicas, así como por su estimulación de la expresión del factor de crecimiento básico de fibroblastos. También contiene una caja de destrucción (caja D) que se necesita para su degradación por el APC. La región ácida C terminal de la proteína codificada puede actuar como un dominio de transactivación. El producto génico es principalmente una proteína citosólica, aunque se localiza parcialmente en el núcleo.

60

La proteína se exploró para determinar los dominios funcionales y PTTG1 tiene un dominio básico N terminal y un dominio ácido C terminal y es una proteína rica en prolina con varios sitios de unión a SH3 supuestos. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 2 regiones transmembrana potenciales (140-146, 185-187). La proteína tiene una alta homología (> 72 %) a las secuencias proteicas de PTTG1 descritas para ratón y rata. Esta proteína se ubica principalmente en el citoplasma y el núcleo. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en el tejido de la médula ósea, los testículos y el timo, mientras que solo se observaron niveles de expresión bajos o ausentes de PTTG1 en todos los demás tejidos analizados, incluyendo pulmón, hígado, corazón, bazo, riñón, músculo esquelético, cerebro, médula espinal, próstata y páncreas. Kakar (Kakar, *Cytogenet. Cell Genet.*; 83: 93-95 (1998)) descubrieron que la sobreexpresión de PTTG1 en células NIH3T3 de fibroblasto de ratón dio como resultado un aumento en la proliferación celular y la transformación celular. La inyección de células NIH3T3 transfectadas en ratones desnudos dio como resultado la formación de tumores, indicando que PTTG1 es oncogénico. Zhang y col. (Zhang y col., *Mol. Endocrinol.*; 13: 156-166 (1999)) mutaron los restos de prolina de los sitios de unión al dominio SH3 de PTTG1 y observaron que la actividad de transformación in vitro y la de inducción de tumores in vivo, así como la estimulación del factor de crecimiento de fibroblastos básico (FGFb), se anularon. Zhang y col. (Zhang y col., *Mol. Endocrinol.*; 13: 156-166 (1999)) llegaron a la conclusión de que el PTTG1 humano funciona a través de las vías de transducción de señales mediada por SH3 y la activación de factores de crecimiento. Como PTTG1 parecía ser un sustrato del complejo promotor de la anafase (APC, por sus siglas en inglés) que asociaba a ESP1 hasta la activación del APC, Zou y col. (Zou y col., *Science*; 285: 418-422 (1999)) identificaron a PTTG1 como la segurina humana. Zou y col. (Zou y col., *Science*; 285: 418-422 (1999)) observaron la similitud de secuencia entre las segurinas de vertebrados a lo largo de toda la secuencia. Un motivo conservado en la proteína PTTG1 coincide con la caja de destrucción (caja D) compartida por muchos sustratos de APC. Al igual que otras segurinas, PTTG1 contiene agrupaciones de dominios ácidos y básicos. La mitad N terminal es rica en restos de lisina que alrededor de la caja D. Esto es común para sustratos de APC que contienen caja D, presumiblemente porque la lisina es el resto que forma un enlace covalente isoapéptídico con ubiquitina. Zou y col. (Zou y col., *Science*; 285: 418-422 (1999)) descubrieron que la segurina humana comienza a acumularse en el inicio de la fase S y alcanza su máximo en las fases G2-M en paralelo con la ciclina B1 (123836). Su nivel cae precipitadamente cuando se activa APC, como se indica por la disminución de la ciclina B1. Zou y col. (Zou y col., *Science*; 285: 418-422 (1999)) mutaron los restos RKAL de la caja D de segurina a AKAA y descubrieron que la segurina mutante fue estable en extractos mitóticos, lo que confirma que la secuencia RKAL es necesaria para la degradación. Zou y col. (Zou y col., *Science*; 285: 418-422 (1999)) llegaron a la conclusión de que la identificación de la segurina humana como un oncogén sugiere que la regulación errónea de la separación de cromátidas puede contribuir a la generación de tumores malignos.

Pueden identificarse ligandos/fragmentos adecuados del CMH de PTTG1 humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

RAB6IP1: RAB6IP1 humana (proteína 1 de interacción con RAB6) se ubica en la región genómica 9243446-9116947 en el cromosoma 11 en la región 11p15.4. Consiste en 17 exones resultantes de la transcripción en un ARNm que tiene 4965 pares de bases (véase la SEQ ID NO: 79). La secuencia codificante en sí consiste en 3864 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 1287 aminoácidos (SEQ ID NO: 80).

La proteína se exploró para determinar los dominios funcionales y RAB6IP1 tiene un dominio 2 de homología a la lipoxigenasa (barril beta) en la región 957-1059. Este dominio se encuentra en una diversidad de proteínas asociadas a la membrana o a lípidos. Está presente en lipogenasas, enzimas implicadas en diversas etapas de la biosíntesis de leucotrienos con hierro como cofactor. La estructura conocida de la lipasa pancreática demuestra que este dominio se une a procolipasa que media la asociación a la membrana. Este dominio puede mediar la unión a la membrana a través de otros compañeros de unión a proteínas. La estructura de este dominio es conocida por muchos miembros de la familia y consiste en un sándwich beta. RAB6IP1, adicionalmente, tiene un dominio de péptido señal N terminal en la región 1-48. Por lo general, se ubica en el extremo N y normalmente está ausente en la proteína madura. Normalmente se refiere a la secuencia (de aproximadamente 20 aminoácidos) que interactúa con una partícula de reconocimiento de señal y dirige el ribosoma al retículo endoplasmático donde tiene lugar la inserción cotraduccional. Los péptidos señal son altamente hidrófobos, pero con algunos restos de carga positiva. La secuencia señal normalmente es retirada de la cadena peptídica en crecimiento mediante la peptidasa señal, una proteasa específica ubicada en la cara cisternal del retículo endoplasmático. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 12 regiones transmembrana potenciales (21, 205-212, 298-311, 327-331, 342-354, 795-797, 799-799, 936-939, 943-961, 1135-1143, 1234-1239, 1260-1266), mientras que las regiones 207-209, 305-307, 345-350 y 948-959 tenían la probabilidad más alta. La proteína tiene una alta homología (> 97 %) a las secuencias proteicas de RAB6IP1 descritas para ratón y rata. La ubicación de esta proteína es desconocida. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en casi todos los tejidos analizados, incluyendo médula ósea, hígado, pulmón, corazón, bazo, riñón, timo, músculo esquelético, cerebro, médula espinal, próstata y páncreas.

Pueden identificarse ligandos/fragmentos adecuados del CMH de RAB6IP1 humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

RHOU: RHOU humano (familia de genes homólogos a ras, miembro de U) se ubica en la región genómica 226937491-226949033 en el cromosoma 1 en la región 1q42.11-q42.3. Consiste en 3 exones resultantes de la transcripción en un ARNm que tiene 4369 pares de bases (véase la SEQ ID NO: 87). La secuencia codificante en sí consiste en 777 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 258 aminoácidos (SEQ ID NO: 88).

RHOU codifica un miembro de la familia Rho de GTPasas. Esta proteína puede activar PAK1 y JNK1 y puede inducir la formación de filopodios y la disolución de fibras de estrés. También puede mediar los efectos de la señalización de WNT1 en la regulación de la morfología celular, la organización del citoesqueleto y la proliferación celular.

La proteína se exploró para determinar los dominios funcionales y RHOU tiene la subfamilia Rho (de homología a Ras) de dominio de GTPasas pequeñas similares a Ras en la región 52-225. Las GTPasas pequeñas están implicadas en los procedimientos de señalización celular intracelulares. La familia Ras incluye un gran número de GTPasas pequeñas. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 7 regiones transmembrana potenciales (53-56, 58-61, 63-67, 88-90, 121-133, 156-159, 217-222), mientras que la región 122-132 tenía la probabilidad más alta. La proteína tiene una alta homología (> 93 %) a las secuencias proteicas de RHOU descritas para ratón. Esta proteína se ubica principalmente en el citoplasma. RHOU tiene características conservadas entre los miembros de la familia Rho de GTPasas incluyendo dominios de unión a GTP y GDP, el dominio efector y una señal de modificación lipídica CAAX en el extremo C terminal. Tiene un dominio N terminal único que contiene varios motivos de unión a SH3 PXXP supuestos. RHOU se expresa en el cerebro, el músculo esquelético y la placenta, con una expresión moderada en el hígado, el pulmón y el corazón y una baja expresión en el colon, el bazo, el riñón y el intestino delgado. Tao y col. (Tao y col., *Genes Dev.*; 15: 1796-1807 (2001)) demostraron que RHOU, como CDC42, puede activar PAK1 y JNK1 (MAPK8) y puede inducir la formación de filopodios y la disolución de fibras de estrés. RHOU activo estimuló las células quiescentes para que entraran de nuevo en el ciclo celular. La sobreexpresión de RHOU en células epiteliales mamarias de ratón indujo la transformación morfológica que imita la transformación inducida por Wnt1. Tao y col. (Tao y col., *Genes Dev.*; 15: 1796-1807 (2001)) formularon la hipótesis de que RHOU podría mediar los efectos de la señalización de WNT1 en la regulación de la morfología celular, la organización del citoesqueleto y la proliferación celular. Pueden identificarse ligandos/fragmentos adecuados del CMH de RHOU humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

SATB1: SATB1 humana (homeobox SATB 1), también conocida como proteína especial 1 de unión a secuencias ricas en AT, se ubica en la región genómica 18440343-18364437 en el cromosoma 3 en la región 3p23. Consiste en 11 exones resultantes de la transcripción en un ARNm que tiene 6659 pares de bases (véase la SEQ ID NO: 89). La secuencia codificante en sí consiste en 2292 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 763 aminoácidos (SEQ ID NO: 90).

La proteína se exploró para determinar los dominios funcionales y SATB1 tiene un dominio homeobox (HOX) en la región 644-707, que son con frecuencia factores de unión al ADN que están implicados en la regulación transcripcional de procesos de desarrollo clave. El dominio homeobox se identificó por primera vez en una serie de proteínas homeóticas y de segmentación de *Drosophila*, pero ahora se sabe que está bien conservado en muchos otros animales, incluyendo los vertebrados. Los genes HOX codifican reguladores transcripcionales que contienen homeodominio que operan programas genéticos diferenciales a lo largo del eje anteroposterior de cuerpos animales. El dominio se une al ADN a través de una estructura de hélice-giro-hélice (HTH, por sus siglas en inglés). El motivo HTH se caracteriza por dos hélices alfa, que hacen contactos íntimos con el ADN y se unen mediante un giro corto. La segunda hélice se une al ADN a través de un número de enlaces de hidrógeno e interacciones hidrófobas, que se producen entre cadenas laterales específicas y las bases expuestas y grupos metilo de timina en el surco mayor del ADN. La primera hélice ayuda a estabilizar la estructura. El motivo es muy similar en secuencia y estructura en una amplia gama de proteínas de unión al ADN (por ejemplo, proteínas cro y represoras, proteínas homeóticas, etc.). Una de las principales diferencias entre los motivos HTH en estas proteínas diferentes surge del requisito estereoquímico de glicina en el giro que se necesita para evitar la interferencia estérica del carbono beta con la cadena principal: para proteínas cro y represoras la glicina parece ser obligatoria, mientras que para muchos de las proteínas homeóticas y de unión al ADN el requisito se relaja. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 6 regiones transmembrana potenciales (74-79, 110-119, 163-165, 217-220, 392-392, 655-663), mientras que la región 114-114 tenía la probabilidad más alta. La proteína tiene una alta homología (> 98 %) a las secuencias proteicas de SATB1 descritas para ratón y rata. Se caracterizan varias variantes de corte y empalme y, adicionalmente, muestran una expresión selectiva de este gen en linfocitos T CD4⁺ en reposo. Esta proteína se ubica principalmente en el núcleo. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en el timo, los testículos y los linfocitos Th2, mientras que solo se observaron niveles de expresión bajos o ausentes de SATB1 en todos los demás tejidos analizados. SATB1 es una proteína matricial que se une específicamente a la matriz nuclear/ADN de asociación de armazón (MARS/SAR), elementos de control que mantienen esferas independientes de actividad génica. Dickinson y col. (Dickinson y col., *Cell*; 70: 631-645 (1992)) clonaron un ADNc humano que codifica SATB1. Se expresa predominantemente en el timo. SATB1 se une selectivamente a secuencias especiales de ADN ricas en AT, bicatenarias, en las que la cadena 1 consiste exclusivamente en nucleótidos A, T y C bien mezclados (secuencias de ATC). Yasui y col. (Yasui y col., *Nature*; 419:

641-645 (2002)) señalaron que la ablación de Satb1 en ratones mediante direccionamiento génico dio como resultado la expresión errónea temporal y espacial de numerosos genes y detuvo el desarrollo de linfocitos T, lo que sugiere que Satb1 es un regulador génico global específico de tipo celular. Descubrieron que Satb1 se coimmunoprecipitó con varios factores de remodelación de la cromatina en un complejo proteico grande que no se observó en extractos de timocitos nulos para Satb1. Yasui y col. (Yasui y col., *Nature*; 419: 641-645 (2002)) examinaron el papel de Satb1 en la expresión del gen del receptor alfa de interleucina 2 (IL2RA), que se transcribe ectópicamente en timocitos nulos para Satb1. Satb1 reclutó la histona desacetilasa (HDAC1) contenida dentro del complejo de remodelación de la cromatina NURD a un sitio unido a SATB1 en el locus Il2ra y medió la desacetilación específica de las histonas en un dominio grande dentro del locus. Satb1 también se dirigió a las subunidades Acf1 (BAZ1A) y Iswi (SMARCA5) de los complejos de movilización del nucleosoma CADHC y ACF a este sitio específico y reguló el posicionamiento del nucleosoma a lo largo de varias kilobases. Los cromosomas eucarióticos se empaquetan en núcleos en muchos órdenes de plegado. El empaquetamiento de cromatina de orden superior puede afectar a la expresión génica. SATB1 es una proteína nuclear específica de tipo celular que recluta factores de remodelación de la cromatina y regula numerosos genes durante la diferenciación de timocitos. Cai y col. (Cai y col., *Nat. Genet.*; 34: 42-51 (2003)) demostraron que, en los núcleos de los timocitos, SATB1 tiene una distribución de 'red' similar a una jaula que circunscribe la heterocromatina y ata selectivamente secuencias de ADN especializadas en su red. Esto se demostró mediante hibridación in situ por fluorescencia en timocitos de tipo silvestre y nulos para Satb1 usando secuencias unidas a SATB1 in vivo como sondas. Muchos loci de genes, incluyendo el de MYC y un gen específico del cerebro, están anclados por la red SATB1 en sitios genómicos específicos y este fenómeno se correlaciona de forma precisa con la regulación apropiada de genes distantes. El análisis de modificación de histonas a través de una región genómica enriquecida en genes de 70 kb demostró la acetilación de la histona H3 a los picos lys9 y lys14 en el sitio de unión a SATB1 y se extiende sobre una región de aproximadamente 10 kb que abarca los genes regulados por SATB1. Por el contrario, en timocitos nulos para Satb1, este sitio está marcado por la metilación en lys9 de H3. Cai y col. (Cai y col., *Nat. Genet.*; 34: 42-51 (2003)) propusieron que SATB1 es un nuevo tipo de regulador génico con una arquitectura nuclear novedosa, que proporciona sitios para la organización específica de tejido de secuencias de ADN y que regula la modificación de histonas específica de región. Kumar y col. (Kumar y col., *Nat. Cell Biol.*; 9: 45-56 (2007)) demostraron que la fosforilación de SATB1 en ser185 por PKC (PRKCA) actuó como un interruptor para determinar si SATB1 interactuaba con HDAC1 o PCAF en estirpes celulares humanas. La fosforilación y la desfosforilación de SATB1 ejercían efectos opuestos sobre la actividad de genes indicadores. SATB1 interactuó tanto con el dímero CBP (CREBBP) - p300 (EP300) como con PCAF. La interacción SATB1-PCAF permitió que PCAF acetilara la lys136 dentro del dominio similar a PDZ de SATB1, lo que condujo a la pérdida de la actividad de unión al ADN de SATB1 y alteró la expresión génica en un nivel global. Cai y col. (Cai y col., *Nat. Genet.*; 38: 1278-1288 (2006)) descubrieron que tras la activación de los linfocitos T auxiliares 2 (Th2) en ratones, la expresión SATB1 se indujo rápidamente para formar una estructura de cromatina transcripcionalmente activa en el locus de la citocina Th2 de 200 kb en el cromosoma 11. En esta estructura, la cromatina se plegó en numerosos bucles pequeños, todos anclados a Satb1 en su base. Antes de la activación, el locus de la citocina Th2 ya se había asociado a Gata3 y Stat6, mostrando algo de formación de bucles, pero éstos eran insuficientes para inducir la expresión de genes de citocina. Usando interferencia de ARN, Cai y col. (Cai y col., *Nat. Genet.*; 38: 1278-1288 (2006)) demostraron que, en la activación de células, Satb1 era necesario no solo para la compactación de la cromatina en bucles densos en el locus de la citocina, sino también para inducir la expresión de Il4, Il5, Il13 y Maf.

Pueden identificarse ligandos/fragmentos adecuados del CMH de SATB1 humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

SELP: SELP humana (selectina P (proteína de membrana de gránulos de 140 kDa, antígeno CD62)) se ubica en la región genómica 167866030-167824713 en el cromosoma 1 en la región 1q22-q25. Consiste en 17 exones resultantes de la transcripción en un ARNm que tiene 3199 pares de bases (véase la SEQ ID NO: 91). La secuencia codificante en sí consiste en 3493 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 830 aminoácidos (SEQ ID NO: 92).

SELP es una proteína de membrana de gránulos alfa de plaquetas de peso molecular 140.000 que se redistribuye a la membrana plasmática durante la activación y desgranulación de plaquetas. Es un miembro de una familia de receptores de adhesión/retorno. Pueden aparecer variantes de corte y empalme alternativas, pero no están bien documentadas.

La proteína se exploró para determinar los dominios funcionales y SELP tiene un dominio de lectina de tipo C (CTL) en la región 30-159, que con frecuencia funciona como módulos de unión a carbohidratos dependiente de calcio. Las lectinas de animales muestran una amplia diversidad de arquitecturas. Se clasifican de acuerdo con el dominio de reconocimiento de carbohidratos (CRD) del cual hay dos tipos principales, el tipo S y el tipo C. Las lectinas de tipo C muestran una amplia gama de especificidades. Requieren Ca²⁺ para su actividad. Se encuentran predominantemente, pero no exclusivamente, en vertebrados. Además, SELP tiene un dominio similar al factor de crecimiento epidérmico en la región 162-195. Los factores de crecimiento epidérmico y los factores de crecimiento transformantes pertenecen a una clase general de proteínas que comparten un patrón de repetición que implica un número de restos de Cys conservados. Los factores de crecimiento están implicados en el reconocimiento y la división celulares. Se piensa que el patrón de repetición, especialmente de las cisteínas (la denominada repetición EGF), es importante para la estructura 3D de las proteínas y, por tanto, para su reconocimiento por receptores y

5 otras moléculas. El distintivo EGF de tipo 1 incluye seis cisteínas conservadas que se cree que están implicadas en la formación de enlaces disulfuro. El motivo de EGF se encuentra frecuentemente en la naturaleza, en particular en proteínas extracelulares. Además, SELP tiene 9 dominios sushi en las regiones 200-257, 262-319, 324-381, 386-443, 448-505, 510-567, 572-629, 642-699, 704-761. Los dominios sushi, también conocidos como módulos de proteína de control del complemento (CCP, por sus siglas en inglés) o repeticiones de consenso cortas (SCR, por sus siglas en inglés), existen en una amplia diversidad de proteínas del complemento y de adhesión. La estructura es conocida para este dominio, se basa en una disposición de sándwich beta; una cara compuesta de tres cadenas beta unidas mediante enlaces de hidrógeno para formar una región de triple cadena en su centro y la otra cara formada a partir de dos cadenas beta separadas. SELP, adicionalmente, tiene un dominio transmembrana en la región 773-795. Los dominios transmembrana se componen de 15-30 restos hidrófobos y se encuentran en proteínas que abarcan una membrana en una célula. Una proteína puede tener un único dominio TM (por ejemplo, receptores transmembrana de Tipo I tales como el receptor EGF) o múltiples dominios TM (por ejemplo, receptores acoplados a proteína G). Para abarcar el núcleo hidrocarbonado de la membrana de ~ 3 nm requiere una hélice alfa de ~20 restos predominantemente apolares sin carga. La membrana separa dos compartimentos acuosos mediante una fase lipídica bidimensional delgada. Las proteínas de membrana generalmente abarcan esta fase lipídica y, por tanto, necesitan acomodarse al tramo hidrófilo a ambos lados de la membrana, así como al entorno hidrófobo en el núcleo de la bicapa. La estructura de las porciones embebidas en la membrana consiste ya sea en hélices alfa transmembrana montadas con frecuencia en haces de hélices, o en láminas beta antiparalelas que forman poros en forma de barril. La longitud de la hélice transmembrana puede incluso correlacionarse con el espesor de su membrana. Las proteínas que abarcan una sola vez la membrana del retículo endoplasmático (RE) y Golgi tienen generalmente dominios transmembrana más cortos que las proteínas de la membrana plasmática. Puesto que el contenido de colesterol y, por tanto, el espesor de la bicapa lipídica también aumenta junto con la vía secretora, esto podría reflejar un papel de los segmentos y proteínas transmembrana en la clasificación de los lípidos. En proteínas que abarcan la membrana múltiples veces las hélices transmembrana se unen estrechamente a la estructura compacta, globular, de la que se excluyen los lípidos. La proteína tiene una alta homología (> 69 %) a las secuencias proteicas de SELP descritas para ratón y rata. Esta proteína se ubica principalmente en la membrana plasmática. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en casi todos los tejidos analizados. SELP es una molécula de adhesión de 140 kD, expresada en la superficie de las células activadas, que media en la interacción de células endoteliales o plaquetas activadas con leucocitos. SELP se detectó en megacariocitos y plaquetas, así como en células endoteliales vasculares, pero no se encontró en una diversidad de otros tipos celulares examinados. En las células endoteliales, la proteína se ubicó en las membranas de los cuerpos de Weibel-Palade, los gránulos de almacenamiento intracelular para el factor de von Willebrand. CD24 es un ligando para P-selectina. Florian y col. (Florian y col., *Biochem. Biophys. Res. Commun.*; 281: 1045-1050 (2001)) demostraron que la nexina-17 (SNX17) de clasificación interactúa con el dominio citosólico de P-selectina y sugirió que SNX17 puede actuar en el tráfico intracelular de P-selectina. Mayadas y col. (Mayadas y col., *Cell*; 74: 541-554 (1993)) generaron ratones deficientes en P-selectina mediante el direccionamiento de genes en células madre embrionarias y se descubrió que presentaban una serie de defectos en el comportamiento de los leucocitos, incluyendo los números elevados de neutrófilos circulantes, la ausencia prácticamente total de leucocitos rodando en las vénulas mesentéricas y el retraso en el reclutamiento de neutrófilos a la cavidad peritoneal tras la inflamación inducida experimentalmente.

Pueden identificarse ligandos/fragmentos adecuados del CMH de SELP humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

45 **SEMA3G:** SEMA3G humano (dominio sema, dominio de inmunoglobulina (Ig), dominio básico corto, 3G secretado (semaforina)) se ubica en la región genómica 52454082-52442307 en el cromosoma 3 en la región 3p21.1. Consiste en 16 exones resultantes de la transcripción en un ARNm que tiene 4700 pares de bases (véase la SEQ ID NO: 93). La secuencia codificante en sí consiste en 2349 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 782 aminoácidos (SEQ ID NO: 94).

50 La proteína se exploró para determinar los dominios funcionales y SEMA3G tiene un dominio semaforina en la región 58-503, que aparece en semaforinas, una gran familia de proteínas secretadas y transmembrana, algunas de las cuales actúan como señales repelentes durante la orientación de axones. Los dominios Sema también aparecen en un receptor del factor de crecimiento de hepatocitos, en proteínas SEX y en proteínas víricas. El dominio Sema se caracteriza por un conjunto conservado de restos de cisteína, que forman cuatro enlaces disulfuro para estabilizar la estructura. El pliegue del dominio Sema es una variación de la topología de la hélice beta, con siete hojas dispuestas radialmente alrededor de un eje central. Cada hoja contiene una lámina beta antiparalela de cuatro cadenas (cadenas A a D). La cadena interior de cada hoja (A) alinea el canal en el centro de la hélice, con las cadenas B y C de la misma repetición irradiándose hacia el exterior y la cadena D de la siguiente repetición formando el borde exterior de la hoja. El gran tamaño del dominio Sema no se debe a un único dominio insertado, sino que es resultado de la presencia de elementos de estructura secundaria adicionales insertados en la mayoría de las hojas. El dominio Sema usa un sistema de 'bucle y gancho' para cerrar el círculo entre las hojas primera y última. Las hojas se construyen secuencialmente con una cadena beta N terminal cerrando el círculo al proporcionar la cadena más exterior (D) de la séptima hoja (extremo C). La hélice beta se estabiliza adicionalmente mediante una extensión del extremo N terminal, proporcionando una quinta cadena beta adicional en el borde exterior de la hoja 6. SEMA3G, adicionalmente, tiene un dominio PSI en la región 521-574, que se ha descubierto en plexinas,

semaforinas e integrinas. Las semaforinas inducen el colapso y la parálisis de los conos de crecimiento neuronales. Además, SEMA3G tiene un dominio similar a la inmunoglobulina en la región 588-674. Las moléculas de Ig son proteínas altamente modulares, en las que los dominios variables y constantes tienen patrones de secuencia conservados claros. Los dominios en las moléculas de Ig y similares a Ig se agrupan en cuatro tipos: conjunto V (variable), conjunto C1 (constante 1), conjunto C2 (constante 2) y conjunto I-set (intermedio). Estudios estructurales han demostrado que estos dominios comparten una estructura sándwich beta de llave griega núcleo común, difiriendo los tipos en el número de cadenas en las láminas beta, así como en sus patrones de secuencia. Pueden encontrarse dominios similares a inmunoglobulina que se relacionan tanto en secuencia como en estructura en varias familias de proteínas diversas. Los dominios similares a Ig están implicados en una diversidad de funciones, incluyendo el reconocimiento célula-célula, los receptores de superficie celular, la estructura muscular y el sistema inmunitario. Además, SEMA3G tiene un dominio de péptido señal N terminal en la región 1-23. Por lo general se encuentra en el extremo N y normalmente está ausente en la proteína madura. Normalmente se refiere a la secuencia (de aproximadamente 20 aminoácidos) que interactúa con la partícula de reconocimiento de señal y dirige el ribosoma al retículo endoplasmático donde tiene lugar la inserción cotraduccional. Los péptidos señal son altamente hidrófobos, pero con algunos restos de carga positiva. La secuencia señal normalmente es retirada de la cadena peptídica en crecimiento por la peptidasa de señal, una proteasa específica ubicada en la cara cisternal del retículo endoplasmático. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 8 regiones transmembrana potenciales (7-18, 74-78, 145-145, 326-344, 461-469, 584-591, 665-678, 713-715), mientras que las 3 regiones 8-17, 328-341 y 666-677 tenían la probabilidad más alta. La proteína tiene una alta homología (> 86 %) a las secuencias proteicas de SEMA3G descritas para ratón y rata. Esta proteína es una proteína de membrana integral. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en el corazón, el bazo y el tejido pulmonar, mientras que solo se observaron de expresión bajos o ausentes de SEMA3G en todos los demás tejidos analizados, incluyendo médula ósea, hígado, riñón, timo, músculo esquelético, cerebro, médula espinal, próstata y páncreas. Pueden identificarse ligandos/fragmentos adecuados del CMH de SEMA3G humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

SHMT2: SHMT2 humana (serina hidroximetiltransferasa 2 (mitocondrial)) se ubica en la región genómica 55909818-55914980 en el cromosoma 12 en la región 12q12-q14. Consiste en 12 exones resultantes de la transcripción en un ARNm que tiene 2113 pares de bases (véase la SEQ ID NO: 95). La secuencia codificante en sí consiste en 1515 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 504 aminoácidos (SEQ ID NO: 96).

La proteína se exploró para determinar los dominios funcionales y SHMT2 tiene un dominio serina hidroximetiltransferasa en la región 49-448. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 2 regiones transmembrana potenciales (255-261, 333-338). La proteína tiene una alta homología (> 94 %) a las secuencias proteicas de SHMT2 descritas para ratón y rata. Esta proteína se ubica principalmente en el núcleo y las mitocondrias. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en casi todos los tejidos analizados, incluyendo hígado, corazón, bazo, riñón, timo, músculo esquelético, médula espinal, próstata y páncreas.

Pueden identificarse ligandos/fragmentos adecuados del CMH de SHMT2 humana mediante el procedimiento descrito en relación con la Colonia 1 anterior.

STAM: STAM humana (molécula adaptadora de la transducción de señales (dominio SH3 y motivo ITAM) 1) se ubica en la región genómica 17726129-17797912 en el cromosoma 10 en la región 10p14-p13. Consiste en 13 exones resultantes de la transcripción en un ARNm que tiene 2967 pares de bases (véase la SEQ ID NO: 97). La secuencia codificante en sí consiste en 1623 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 540 aminoácidos (SEQ ID NO: 98).

STAM se identificó por la rápida fosforilación de tirosina de su producto en respuesta a la estimulación por citocinas. La proteína codificada contiene un dominio SH3 y el motivo inmunoreceptor de activación a base de tirosina (ITAM). Esta proteína se asocia a cinasas JAK3 y JAK2 a través de su región ITAM y es fosforilada por las cinasas JAK tras la estimulación por citocinas, lo que sugiere que la función de esta proteína es como una molécula adaptadora implicada en la señalización corriente abajo de los receptores de citocinas. Se ha descubierto que HGS/HRS (sustrato de tirosina cinasa regulada por el factor de crecimiento de hepatocitos) se une y contrarresta la función de esta proteína. La proteína se exploró para determinar los dominios funcionales y STAM tiene un dominio VHS en la región 9-139. El dominio VHS es un dominio de ~140 restos de largo, cuyo nombre deriva de su aparición en VPS-27, Hrs y STAM. El dominio VHS se encuentra siempre en el extremo N de las proteínas lo que sugieren que dicha topología es importante para la función. Se considera que el dominio tiene un papel carga de direccionamiento a la membrana/reconocimiento en el tráfico vesicular. La resolución de la estructura cristalina del dominio VHS de Hrs y Tom1 humana reveló que consiste en ocho hélices dispuestas en una superhélice de doble capa. La existencia de parches conservados de restos sobre la superficie del dominio sugiere que los dominios VHS pueden estar implicados en el reconocimiento y el acoplamiento proteína-proteína. En general, la similitud de secuencias es baja (de aproximadamente el 25 %) entre los miembros de la familia del dominio. STAM, adicionalmente, tiene un dominio de motivo de interacción de ubiquitina (UIM, por sus siglas en inglés) en la región 171-190, que se describió

por primera vez en la subunidad del proteosoma 26S, PSD4/RPN-10. Se sabe que se une a múltiples ubiquitinas y también se encontró en muchas proteínas implicadas en la vía endocítica. Además, STAM tiene un dominio 3 de homología a Src (SH3) en la región 213-268, que se une a proteínas diana a través de secuencias que contienen prolina y aminoácidos hidrófobos. Los polipéptidos que contienen Pro pueden unirse a dominios SH3 en 2 orientaciones de unión diferentes. Los dominios SH3 (homología a src-3) son módulos de proteínas pequeñas que contienen aproximadamente 50 restos de aminoácidos. Se encuentran en una gran diversidad de proteínas intracelulares o asociadas a la membrana, por ejemplo, en una diversidad de proteínas con actividad enzimática, en proteínas adaptadoras que carecen de secuencias catalíticas y en proteínas del citoesqueleto. El dominio SH3 tiene un pliegue característico que consiste en cinco o seis cadenas beta dispuestas como dos láminas beta anti-paralelas empaquetadas estrechamente. Las regiones de engarce pueden contener hélices cortas. La superficie del dominio SH2 lleva un bolsillo de unión al ligando hidrófobo, plano, que consiste en tres surcos poco profundos definidos por restos aromáticos conservadores en los que el ligando adopta una disposición helicoidal hacia la izquierda extendida. El ligando se une con baja afinidad, pero esto puede potenciarse mediante múltiples interacciones. La región unida por el dominio SH3 es, en todos los casos, rica en prolina y contiene PXXP como un motivo de unión conservado en el núcleo. La función del dominio SH3 no se comprende bien, pero puede mediar en muchos procedimientos diversos, tales como el aumento de la concentración local de proteínas, alterando su localización subcelular y la mediación en el ensamblaje de grandes complejos multiproteicos. STAM también tiene un dominio de motivo inmunoreceptor de activación a base de tirosina en la región 359-387, que es un motivo que puede fosforilarse doblemente en la tirosina y que une receptores de antígenos a la maquinaria de señalización corriente abajo. Los motivos de señalización inmunorreceptores fosforilados (ITAM) presentan habilidades únicas para unirse y activar las tirosina cinasas Lyn y Syk. El motivo puede fosforilarse doblemente en la tirosina que une receptores de antígenos a la maquinaria de señalización corriente abajo. Además, STAM, adicionalmente, tiene un dominio superenrollado en la región 343-377. El superenrollamiento es un dominio proteico que forma un haz de dos o tres hélices alfa. Los dominios superenrollados cortos están implicados en las interacciones proteicas pero los dominios superenrollados largos que forman varillas largas aparecen proteínas estructurales o motoras. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 1 región transmembrana potencial (64-74), mientras que la región 66-71 tenía la probabilidad más alta. La proteína tiene una alta homología (> 91 %) a las secuencias proteicas de STAM descritas para ratón y rata. Esta proteína se ubica principalmente en el citoplasma. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en casi todos los tejidos analizados. Takeshita y col. (Takeshita y col., *Biochem. Biophys. Res. Commun.*; 225: 1035-1039 (1996)) sugirieron que STAM actúa como una molécula adaptadora en las vías de transducción de señales de receptores de citocinas. Asao y col. (Asao y col., *J. Biol. Chem.*; 272: 32785-32791 (1997)) demostraron que HGS se une a STAM a través de secuencias superenrolladas y parece regular la proliferación en respuesta a citocinas.

35 Pueden identificarse ligandos/fragmentos adecuados del CMH de STAM humana mediante el procedimiento descrito en relación con la Colonia 1 anterior.

STOM: STOM humana (estomatina), se ubica en la región genómica 123172365-123141173 en el cromosoma 9 en la región 9q34.1. La isoforma de STOM 1 consiste en 7 exones resultantes de la transcripción en un ARNm que tiene 3108 pares de bases (véase la SEQ ID NO: 99). La secuencia codificante en sí consiste en 867 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 288 aminoácidos (SEQ ID NO: 100). La isoforma de STOM 2 consiste en 3 exones resultantes de la transcripción en un ARNm que tiene 2593 pares de bases (véase la SEQ ID NO: 101). La secuencia codificante en sí consiste en 372 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 123 aminoácidos (SEQ ID NO: 102).

45 La proteína se exploró para determinar los dominios funcionales y la isoforma de STOM 1 tiene un dominio de homólogos a pro-hibitina (PHB) en la región 52-211, que es un dominio de proteína integral de membrana que se piensa que regula la conductancia de cationes. Una diversidad de proteínas pertenece a esta familia. Éstas proteínas incluyen las prohibitinas, proteínas antiproliferativas citoplasmáticas y estomatina, una proteína de membrana de eritrocitos. La proteína HflC bacteriana también pertenece a esta familia. La isoforma de STOM 2 tiene un dominio transmembrana en la región 32-54. Los dominios transmembrana se componen de 15-30 restos hidrófobos y se encuentran en las proteínas que abarcan una membrana en una célula. Una proteína puede tener un único dominio TM (por ejemplo, receptores transmembrana de Tipo I tales como el receptor de EGF) o múltiples dominios TM (por ejemplo, receptores acoplados a proteína G). Para abarcar el núcleo hidrocarbonado de la membrana de ~ 3 nm requiere una hélice alfa de ~20 restos predominantemente apolares sin carga. La membrana separa dos compartimentos acuosos mediante una fase lipídica bidimensional delgada. Las proteínas de membrana generalmente abarcan esta fase lipídica y, por tanto, necesitan acomodarse al tramo hidrófilo a ambos lados de la membrana, así como al entorno hidrófobo en el núcleo de la bicapa. La estructura de las porciones embebidas en la membrana consiste ya sea en hélices alfa transmembrana montadas con frecuencia en haces de hélices, o en láminas beta antiparalelas que forman poros en forma de barril. La longitud de la hélice transmembrana puede incluso correlacionarse con el espesor de su membrana. Las proteínas que abarcan una sola vez la membrana del retículo endoplasmático (RE) y Golgi tienen generalmente dominios transmembrana más cortos que las proteínas de la membrana plasmática. Puesto que el contenido de colesterol y, por tanto, el espesor de la bicapa lipídica también aumenta junto con la vía secretora, esto podría reflejar un papel de los segmentos y proteínas transmembrana en la

clasificación de los lípidos. En proteínas que abarcan la membrana múltiples veces las hélices transmembrana se unen estrechamente a la estructura compacta, globular, de la que se excluyen los lípidos. La proteína supuesta tiene una alta homología (> 87 %) a las secuencias de proteínas supuestas de STOM descritas para ratón y rata. STOM es una proteína integral de membrana 29000 kD que se expone en la superficie citoplasmática de la membrana y es susceptible a la fosforilación por una proteína cinasa dependiente de AMPc. La misma proteína puede demostrarse en estirpes celulares humanas de origen epitelial y linfóide, en particular, en células HeLa. El análisis estructural asignó STOM a las proteínas transmembrana de tipo Ib. STOM tiene un amplio patrón de expresión, con altos niveles de ARNm en el corazón, el hígado, el músculo esquelético y los testículos, pero niveles bajos en los pulmones, el cerebro y el bazo. Los modelos de la estructura prevista de la proteína mostraron una cabeza corta NH₂-terminal, un tramo de 28 aminoácidos fuertemente hidrófobos que presumiblemente codifica un único dominio que abarca la membrana y un gran dominio compuesto de lámina beta y hélice alfa. Pueden identificarse ligandos/fragmentos adecuados del CMH de STOM humana mediante el procedimiento descrito en relación con la Colonia 1 anterior.

TCF7: TCF7 humano (factor de transcripción 7 (específico de linfocitos T, caja HMG)), se ubica en la región genómica 133478300-133511818 en el cromosoma 5 en la región 5q31.1. La isoforma 1 de TCF 7 consiste en 9 exones resultantes de la transcripción en un ARNm que tiene 3283 pares de bases (véase la SEQ ID NO: 107). La secuencia codificante en sí consiste en 1155 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 384 aminoácidos (SEQ ID NO: 108). La isoforma 2 de TCF 7 consiste en 7 exones resultantes de la transcripción en un ARNm que tiene 2821 pares de bases (véase la SEQ ID NO: 109). La secuencia codificante en sí consiste en 810 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 269 aminoácidos (SEQ ID NO: 110). La isoforma 3 de TCF 7 consiste en 7 exones resultantes de la transcripción en un ARNm que tiene 1254 pares de bases (véase la SEQ ID NO: 111). La secuencia codificante en sí consiste en 810 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 269 aminoácidos (SEQ ID NO: 112). La isoforma 4 de TCF 7 consiste en 8 exones resultantes de la transcripción en un ARNm que tiene 2917 pares de bases (véase la SEQ ID NO: 113). La secuencia codificante en sí consiste en 807 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 268 aminoácidos (SEQ ID NO: 114).

La proteína se exploró para determinar los dominios funcionales y TCF7 tiene un dominio de grupo de alta movilidad (HMG, por sus siglas en inglés) en la región 268-338. Las proteínas HMG son una familia de componentes no histonas de peso molecular relativamente bajo en la cromatina. HMG1 (también denominada HMG-T en el pescado) y HMG2 son dos proteínas relacionadas muy estrechamente que se unen a ADN monocatenario preferentemente y desenrollan el ADN bicatenario. Aunque no tienen especificidad de secuencia, tienen una alta afinidad por el ADN doblado o distorsionado y doblan el ADN lineal. HMG1 y HMG2 contienen dos dominios caja HMG de unión al ADN (A y B) que muestran diferencias estructurales y funcionales, y que tienen un dominio C terminal ácido largo y rico en restos de ácido aspártico y glutámico. La cola ácida modula la afinidad de las cajas HMG en tándem en HMG1 y 2 para una diversidad de dianas de ADN. HMG1 y 2 parecen desempeñar papeles arquitectónicos importantes en el ensamblaje de complejos de nucleoproteínas en una diversidad de procesos biológicos, por ejemplo, la recombinación V(D)J, el inicio de la transcripción y la reparación del ADN. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 1 región transmembrana potencial (217-233), 3 regiones transmembrana potenciales (11-13, 82-84, 100-109), 3 regiones transmembrana potenciales (10-13, 80-85, 98-109), respectivamente 3 regiones transmembrana potenciales (11-13, 82-84, 99-109). La proteína supuesta tiene una alta homología (> 93 %) a las secuencias de proteínas supuestas de TCF7 descritas para ratón y rata. Esta proteína se ubica principalmente en el núcleo. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en casi todos los tejidos analizados.

La expresión del gen CD3E parece estar controlada por un elemento potenciador específico de linfocitos T corriente abajo. Van de Wetering y col. (van de Wetering y col., *J. Biol. Chem.*; 267: 8530-8536 (1992)) identificaron un factor de transcripción específico de linfocitos T, que denominaron TCF1/TCF7, que demostró unión a este elemento. Kingsmore y col. (Kingsmore y col., *Mamm. Genome*; 6: 378 (1995)) cartografiaron el gen homólogo al cromosoma de ratón 11. Van de Wetering y col. (van de Wetering y col., *EMBO J.*; 10: 123-132 (1991)) cartografiaron el gen a 5q31.1 mediante una combinación de estudio de híbridos de células somáticas e hibridación in situ por fluorescencia. Aunque la bibliografía continúa usando el símbolo TCF1 para este gen, su designación oficial es TCF7. Verbeek y col. (Verbeek y col., *Nature*; 374: 70-74 (1995)) generaron 2 mutaciones independientes de la estirpe germinal en TCF7 mediante alteración dirigida y descubrieron que el desarrollo de timocitos en ratones mutantes, por lo demás normales, se bloquea en la transición desde la etapa de CD8+ inmaduro positivo simple a la etapa CD4+/CD8+ doble positivo. A diferencia de los ratones de tipo silvestre, la mayoría de las células positivo simple inmaduras en los mutantes no están en el ciclo celular y el número de linfocitos T inmunocompetentes en los órganos linfoides periféricos está reducido. Verbeek y col. (Verbeek y col., *Nature*; 374: 70-74 (1995)) concluyeron que TCF7 controla una etapa esencial en la diferenciación de los timocitos. Roose y col. (Roose y col., *Science*; 285: 1923-1926 (1999)) demostraron que uno de los genes diana de TCF4, o TCF7L2, en las células epiteliales es TCF7. Las isoformas de TCF7 más abundantes carecen de un dominio de interacción catenina beta. Los ratones TCF7 -/- desarrollan adenomas en el intestino y las glándulas mamarias. La introducción de un alelo APC mutante en estos ratones aumenta sustancialmente el número de estos adenomas. TCF7 puede actuar como un represor de

retroalimentación de genes diana TCF4 (TCF7L2) de catenina beta y, por tanto, puede cooperar con APC para suprimir la transformación maligna de las células epiteliales. Roose y col. (Roose y col., *Science*; 285: 1923-1926 (1999)) detectaron proteína TCF7 nuclear en tejidos humanos normales, en las células epiteliales intestinales en proliferación y en células epiteliales basales de epitelio de la glándula mamaria. Roose y col. (Roose y col., *Science*; 285: 1923-1926 (1999)) formularon la hipótesis de que una posible explicación para el fenotipo tumoral en ratones TCF7 ^{-/-} es que TCF7 actúa como un represor de retroalimentación transcripcional de la catenina beta en genes diana TCF4 y que la alteración de este bucle de realimentación negativa permite la formación de tumores epiteliales muy similar a la pérdida de APC. Esta noción predice la sinergia entre la pérdida de TCF7 y de APC. Roose y col. (Roose y col., *Science*; 285: 1923-1926 (1999)) propusieron un modelo en el que la activación transcripcional de genes diana tales como c-myc y ciclina D1 por catenina beta/TCF4 es contrarrestada por las isoformas represoras de TCF7.

Pueden identificarse ligandos/fragmentos adecuados del CMH de las proteínas TCF7 humanas mediante el procedimiento descrito en relación con la Colonia 1 anterior.

TNFRSF1B: TNFRSF1B humano (superfamilia del receptor del factor de necrosis tumoral, miembro 1B), también conocido como CD120b o TNF-R-II, se ubica en la región genómica 12149646-12191863 en el cromosoma 1 en la región 1p36.3-p36.2. Consiste en 10 exones resultantes de la transcripción en un ARNm que tiene 3682 pares de bases (véase la SEQ ID NO: 115). La secuencia codificante en sí consiste en 1386 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 461 aminoácidos (SEQ ID NO: 116).

TNFRSF1B es un miembro de la superfamilia de receptores de TNF. Esta proteína y el receptor de TNF 1 forman un heterocomplejo que media el reclutamiento de dos proteínas antiapoptóticas, c-IAP1 y c-IAP2, que poseen actividad ubiquitina ligasa E3. La función de las IAP en la señalización de los receptores del TNF es desconocida, sin embargo, se piensa que c-IAP1 potencia la apoptosis inducida por TNF mediante la ubiquitinación y degradación del factor 2 asociado a receptores del TNF, que media señales antiapoptóticas. Estudios de inactivación en ratones también sugieren un papel de esta proteína en la protección de las neuronas frente a la apoptosis mediante la estimulación de las vías antioxidativas.

La proteína se exploró para determinar los dominios funcionales y TNFRSF1B tiene 4 dominios de repetición de receptor de factor de necrosis tumoral/receptor de factor de crecimiento nervioso en las regiones 40-75, 78-118, 120-161 y 164-200. TNFRSF1B, adicionalmente, tiene un dominio transmembrana en la región 258-280. Los dominios transmembrana se componen de 15-30 restos hidrófobos y se encuentran en proteínas que abarcan una membrana en una célula. Una proteína puede tener un único dominio TM (por ejemplo, receptores transmembrana de Tipo I tales como el receptor de EGF) o múltiples dominios TM (por ejemplo, receptores acoplados a proteína G). Para abarcar el núcleo hidrocarbonado de la membrana de ~ 3 nm requiere una hélice alfa de ~20 restos predominantemente apolares sin carga. La membrana separa dos compartimentos acuosos mediante una fase lipídica bidimensional delgada. Las proteínas de membrana generalmente abarcan esta fase lipídica y, por tanto, necesitan acomodarse al tramo hidrófilo a ambos lados de la membrana, así como al entorno hidrófobo en el núcleo de la bicapa. La estructura de las porciones embebidas en la membrana consiste ya sea en hélices alfa transmembrana montadas con frecuencia en haces de hélices, o en láminas beta antiparalelas que forman poros en forma de barril. La longitud de la hélice transmembrana puede incluso correlacionarse con el espesor de su membrana. Las proteínas que abarcan una sola vez la membrana del retículo endoplasmático (RE) y Golgi tienen generalmente dominios transmembrana más cortos que las proteínas de la membrana plasmática. Puesto que el contenido de colesterol y, por tanto, el espesor de la bicapa lipídica también aumenta junto con la vía secretora, esto podría reflejar un papel de los segmentos y proteínas transmembrana en la clasificación de los lípidos. En proteínas que abarcan la membrana múltiples veces las hélices transmembrana se unen estrechamente a la estructura compacta, globular, de la que se excluyen los lípidos. Además, TNFRSF1B tiene un dominio de péptido señal N terminal en la región 1-23. Por lo general se encuentra en el extremo N y normalmente está ausente en la proteína madura. Normalmente se refiere a la secuencia (de aproximadamente 20 aminoácidos) que interactúa con la partícula de reconocimiento de señal y dirige el ribosoma al retículo endoplasmático donde tiene lugar la inserción cotraduccional. Los péptidos señal son altamente hidrófobos, pero con algunos restos de carga positiva. La secuencia señal es retirada normalmente de la cadena peptídica en crecimiento por la peptidasa de señal, una proteasa específica ubicada en la cara cisternal del retículo endoplasmático. La proteína tiene una homología modesta (> 60 %) a las secuencias de proteínas supuestas de TNFRSF1B descritas para ratón y rata. Esta proteína se ubica principalmente en la membrana celular. TNFRSF1B está presente en muchos tipos celulares, especialmente los de origen mieloide, y se expresa fuertemente en los linfocitos T y B estimulados. Beltinger y col. (Beltinger y col., *Genomics*; 35: 94-100 (1996)) señalaron que TNFR2 es el principal receptor de TNF que se encuentra en linfocitos T circulantes y es el principal mediador de la apoptosis autorreguladora en linfocitos CD8+. TNFR2 puede actuar con TNFR1 para destruir células no linfocoides. El preensamblaje o la autoasociación de dímeros de receptores de citocinas se produce a través de los mismos contactos de aminoácidos que son críticos para la unión al ligando. Chan y col. (Chan y col., *Science*; 288: 2351-2354) descubrieron que, por el contrario, los receptores TNFA p60 (TNFRSF1A) y p80 (TNFRSF1B) se autoensamblan a través de un dominio funcional distinto en el dominio extracelular del TNFR, denominado dominio de ensamblaje del preligando (PLAD, por sus siglas en inglés), en ausencia de ligando. La supresión del PLAD da como resultado la presentación monomérica de p60 o p80. El análisis por citometría de flujo demostró que la unión eficiente a TNFA depende del autoensamblaje del receptor. También descubrieron que otros miembros de la superfamilia de receptores de TNF, incluyendo los

dominios extracelulares de TRAIL, CD40 y FAS, se autoasocian todos ellos, pero no interactúan con receptores heterólogos. Usando linfocitos T Jurkat, que expresan TNFR1, pero poco TNFR2, y células Jurkat transfectadas de forma estable con TNFR2, Li y col. (Li y col., *Nature*; 416: 345-347 (2002)) confirmaron que la estimulación por TNF o la estimulación con un agonista de TNFR2, pero no de TNFR1, provoca una pérdida de TRAF2 en las células que expresan TNFR2, pero no en la estirpe celular parental, a través de un proceso dependiente de la ubiquitinación y el proteosoma. El análisis de unión indicó que TRAF2 interactúa con CIAP1 y CIAP2, que posee actividad ubiquitina ligasa E3. Ensayos de ubiquitinación y análisis por DSS-EGPA demostraron que en presencia de una enzima de conjugación a E2, CIAP1, pero no CIAP2, induce ubiquitinación de TRAF2 fuera de su dominio RING. Ambos CIAP se unen, pero ninguno ubiquitina TRAF1. La expresión de CIAP1 no consigue proteger a las células que expresan TNFR2 de la apoptosis inducida por TNF, mientras que un CIAP1 mutante inactivo para E3 y CIAP2 de tipo silvestre sí protegen a las células de la regulación positiva de TRAF2 y provocan un retraso en la muerte celular. Li y col. (Li y col., *Nature*; 416: 345-347 (2002)) llegaron a la conclusión de que la estimulación de TNFR2 provoca la ubiquitinación de TRAF2 por CIAP1, que puede desempeñar un papel proapoptótico en la señalización de TNF.

Pueden identificarse ligandos/fragmentos adecuados del CMH de TNFRSF1B humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

TRIM 16: TRIM16 Humano (16 que contiene motivo tripartito), también conocida como proteína caja B sensible a estrógeno, se ubica en la región genómica 15526917-15472004 en el cromosoma 17 en la región 17p11.2. Consiste en 9 exones resultantes de la transcripción en un ARNm que tiene 2920 pares de bases (véase la SEQ ID NO: 117). La secuencia codificante en sí consiste en 1695 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 564 aminoácidos (SEQ ID NO: 118).

Este gen se identificó como un gen regulado por estrógenos y antiestrógenos en células epiteliales que expresan de forma estable receptores de estrógenos. La proteína codificada por este gen contiene dos dominios caja B y una región superenrollada que son característicos de la familia de proteínas de dedos de cinc caja B. Se ha notificado que las proteínas de esta familia participan en una diversidad de procesos biológicos, incluyendo el crecimiento celular, la diferenciación y la patogenicidad. Se detectó la expresión de este gen en la mayoría de los tejidos. Su función, sin embargo, aún no se ha determinado.

La proteína se exploró para determinar los dominios funcionales y TRIM16 tiene un dominio de dedos de cinc caja B en la región 126-651. El dedo de cinc caja B es un dominio de aproximadamente 40 aminoácidos. Una o dos copias de este motivo generalmente se asocian a un dedo RING y un motivo superenrollado para formar el denominado motivo tripartito. Se encuentra esencialmente en factores de transcripción, ribonucleoproteínas y protooncoproteínas, pero no hay ninguna función asignada claramente a este dominio. Se ha demostrado que es esencial pero no suficiente para ubicar la proteína PML en un patrón puntiforme en núcleos en interfase. Entre los 7 posibles ligandos para el átomo de cinc contenido en una caja B, solo se usan 4 y se unen a un átomo de cinc en una disposición tetraédrica Cys2-His2. El análisis por RMN revela que la estructura de la caja B comprende dos cadenas beta, dos vueltas helicoidales y tres regiones de bucle extendidas diferentes de cualquier otro motivo de unión a cinc. TRIM16, adicionalmente, tiene un dominio PRY en la región 372-425. PRY es un dominio asociado a dominios SPRY. El dominio SPRY es de función desconocida, sin embargo, homólogos distantes son dominios en butirofilina/marenostrina/pirina. La liberación de Ca²⁺ desde el retículo sarcoplásmico o endoplasmático, el almacenamiento intracelular de Ca²⁺, está mediada por el receptor de rianodina (RyR) y/o el receptor de inositol trifosfato (IP3R). Contiene un dominio de dedo RING N terminal con respecto al dominio PRY. El dedo RING es un tipo especializado de dedo de cinc de 40 a 60 restos que se une a dos átomos de cinc y, probablemente, está implicado en la mediación de las interacciones proteína-proteína. Existen dos variantes diferentes, la de tipo C3HC4 y una de tipo C3H2C3, que está claramente relacionada a pesar del patrón de cisteína/histidina diferente. Este conjunto de proteínas se describe como miembros de la familia TRIM (Motivo TRIPartito, por sus siglas en inglés) y están implicados en la compartimentación celular. Las secuencias de la familia TRIM se definen por un dominio de dedo RING, una caja B de tipo 1 (B1) y una caja B de tipo 2 (B2) seguidos de una región superenrollada (CC, por sus siglas en inglés). Los genes que pertenecen a esta familia están implicados en una diversidad de procesos tales como el desarrollo y el crecimiento celulares y están implicados en la enfermedad humana. Muchas de estas proteínas, si no todas la que tienen el dominio PRY, tienen un número de distintivos C terminales, dominio SPRY, similar a RFP (B30.2) y butirofilina. El dominio similar a B30.2 es un dominio bien conservado C terminal de 160-170 aminoácidos que se encuentra en las proteínas nucleares y citoplasmáticas, así como en proteínas transmembrana y secretadas. La función del dominio similar a B30.2 no se conoce, pero se ha demostrado que el dominio similar a B30.2a de butirofilina interactúa con la xantina oxidasa. TRIM16, adicionalmente, tiene un dominio SPRY en la región 426-551. El dominio SPRY es de función desconocida. Homólogos distantes son dominios en butirofilina/marenostrina/pirina. La liberación de Ca²⁺ desde el retículo sarcoplásmico o endoplasmático, el almacenamiento intracelular de Ca²⁺, está mediada por el receptor de rianodina (RyR) y/o el receptor de inositol trifosfato (IP3R). Además, TRIM16, adicionalmente, tiene 3 dominios superenrollados en la región 166-203, 251-271 y 320-340. El superenrollamiento es un dominio de proteína que forma un haz de dos o tres hélices alfa. Los dominios superenrollados cortos están implicados en las interacciones de proteínas, pero los dominios superenrollados largos que forman varillas largas se producen en proteínas estructurales o motoras. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó 2 regiones transmembrana potenciales (440-442, 501-507). La proteína supuesta tiene una alta homología (> 75 %) a las secuencias de proteínas supuestas de TRIM16 descritas

para ratón y rata. Esta proteína se ubica principalmente en el citoplasma. El análisis de la expresión de TRIM16 en diferentes tejidos demostró una expresión de nivel medio en casi todos los tejidos analizados. Liu y col. (Liu y col., *Mol. Endocrinol.*; 12: 1733-1748 (1998)) descubrieron que el tratamiento con estrógenos de una estirpe celular epitelial mamaria humana que expresa RE aumentó los niveles en estado estacionario de ARNm de TRIM16 3 veces. Demostraron que el aumento dependiente de estrógenos en el ARNm de TRIM16 fue un evento temprano que no requirió la síntesis de nuevas proteínas. Beer y col. (Beer y col., *J. Biol. Chem.*; 277: 20740-20749 (2002)) descubrieron que la estimulación por KGF de queratinocitos humanos quiescentes provocó una expresión bifásica de ARNm de TRIM16. Después de un aumento inicial menor 1 hora después de la estimulación, los niveles de ARNm se redujeron fuertemente a las 5 a 8 horas. La expresión de TRIM16 se reguló negativamente en el epitelio engrosado de heridas cutáneas en ratón. La sobreexpresión de TRIM16 en los queratinocitos no afectó a su tasa de proliferación, pero potenció el proceso de diferenciación temprano y aumentó la expresión de genes asociados a la diferenciación de queratinocitos.

Pueden identificarse ligandos/fragmentos adecuados del CMH de TRIM16 humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

UTS2: UTS2 humana (urotensina 2), se ubica en la región genómica 7836158-7830261 en el cromosoma 1 en la región 1p36. Se han descrito para este gen dos variantes de transcrito que codifican diferentes isoformas de preproteína. La isoforma a consiste en 5 exones resultantes de la transcripción en un ARNm que tiene 558 pares de bases (véase la SEQ ID NO: 119). La secuencia codificante en sí consiste en 420 pares de bases. La traducción de esta secuencia da como resultado una preproteína con una longitud de 139 aminoácidos (SEQ ID NO: 120). La isoforma b consiste en 4 exones resultantes de la transcripción en un ARNm que tiene 652 pares de bases (véase la SEQ ID NO: 121). La secuencia codificante en sí consiste en 375 pares de bases. La traducción de esta secuencia da como resultado una preproteína con una longitud de 124 aminoácidos (SEQ ID NO: 122).

Este gen codifica un péptido maduro que es un heptapéptido cíclico activo absolutamente conservado desde la lamprea hasta el ser humano. El péptido activo actúa como un vasoconstrictor y se expresa solo en el tejido cerebral. A pesar de la similitud del nombre de la familia de genes, este gen no es homólogo a la urocortina, un miembro de la familia de salvajina/factor liberador de corticotropina/urotensina I. La mayor parte de la preproteína se escinde para fabricar el péptido maduro.

La proteína se exploró para determinar los dominios funcionales y UTS2 tiene un dominio de péptido señal N terminal en la región 1-18 respectivamente 1-21. Por lo general se ubica en el extremo N y normalmente está ausente en la proteína madura. Normalmente se refiere a la secuencia (de aproximadamente 20 aminoácidos) que interactúa con la partícula de reconocimiento de señal y dirige el ribosoma al retículo endoplasmático donde tiene lugar la inserción cotraduccional. Los péptidos señal son altamente hidrófobos, pero con algunos restos de carga positiva. La secuencia señal es retirada normalmente de la cadena peptídica en crecimiento por la peptidasa de señal, una proteasa específica ubicada en la cara cisternal del retículo endoplasmático. La predicción del dominio transmembrana con el servidor de predicción de dominio transmembrana en línea "DAS" ubicado en <http://www.sbc.su.se/%7Emiklos/DAS/> identificó para la isoforma a 3 regiones transmembrana potenciales (8-13, 35-44, 113-119), mientras que la región 40-41 tiene la probabilidad más alta, y para la isoforma b 2 regiones transmembrana potenciales (6-22, 98-103), mientras que la región 8-19 tiene la mayor probabilidad. Las proteínas supuestas tienen una homología modesta (50 %) a las secuencias proteicas de UTS2 descritas para ratón y rata. Esta proteína se ubica principalmente en la membrana plasmática o secretada. El análisis de la expresión en diferentes tejidos demostró una expresión de nivel medio en la médula ósea, el bazo y el tejido nervioso craneal, así como el tejido cardíaco y los vasos sanguíneos, mientras que solo se observaron niveles de expresión bajos o ausentes de UTS2 en todos los demás tejidos analizados.

UTS2 es un péptido cíclico de 12 aminoácidos que tiene alguna similitud de secuencia, pero no es homólogo, a la somatostatina-14. La caracterización de UTS2 de la urófisis de diversas especies de teleosteos ha demostrado que la estructura del hexapéptido cíclico C terminal se ha conservado completamente, mientras que se han producido varias sustituciones en la región N terminal de la molécula. Posteriormente se descubrió que las moléculas similares a UTS2 están presentes en taxones distantes, desde moluscos hasta anfibios, lo que sugiere que también pueden producirse péptidos relacionados con la urotensina en mamíferos. Además, se demostraron sitios de unión específicos para UTS2 en los vasos sanguíneos de rata y se descubrió que el UTS2 de pescado ejerce efectos cardiovasculares en rata y conejo. También se descubrió que el músculo anococcígeo de ratón tiene una sensibilidad especial a UTS2. Coulouarn y col. (Coulouarn y col., *Proc. Natl. Acad. Sci. USA.*, 95: 15803-15808 (1998)) establecieron claramente la existencia de UTS2 en mamíferos mediante la caracterización de ADNc que codifican precursores de UTS2. La secuencia UTS2 se ubica en la posición C terminal del precursor. UTS2 humano consiste solo en 11 restos de aminoácidos. La región cíclica de UTS2, que es responsable de la actividad biológica del péptido, se ha conservado completamente desde los peces hasta los seres humanos. El análisis por transferencia Northern y transferencia puntual demostró que los ARNm de precursor de UTS2 se encuentran predominantemente en la médula espinal humana. Ames y col. (Ames y col., *Nature*; 401: 282-286 (1999)) identificaron un receptor humano acoplado a proteína G, GPR14, que funciona como un receptor de UTS2. El UTS2 humano se une a GPR14 recombinante humano con una alta afinidad y la unión está acoplada funcionalmente a la movilización de calcio. El UTS2 humano se encuentra dentro del tejido tanto vascular como cardíaco, incluyendo el ateroma coronario, y constriñe eficazmente arterias aisladas de primates no humanos. La potencia de la

vasoconstricción de UTS2 fue un orden de magnitud superior a la de la endotelina-1, convirtiendo al UTS2 humano en el vasoconstrictor de mamífero más potente identificado hasta ese momento. In vivo, el UTS2 humano aumentó notablemente la resistencia periférica total en primates no humanos anestesiados, una respuesta asociada a la disfunción contráctil cardíaca profunda. Por otra parte, se encontró inmunorreactividad a UTS2 dentro de los tejidos del sistema nervioso central y endocrino, lo que sugiere que puede tener actividades adicionales. El UTS2 tiene varias acciones cardiovasculares, incluyendo una actividad vasoactiva potente, y propiedades inotrópicas e hipertróficas cardíacas. Douglas y col. (Douglas y col., *Lancet*; 359: 1990-1997 (2002)) investigaron el grado de expresión de UTS2 y su receptor GPR14 en el miocardio de pacientes con insuficiencia cardíaca congestiva. Obtuvieron muestras de ensayo de miocardio de los corazones de 19 pacientes con insuficiencia cardíaca congestiva en etapa terminal, 5 pacientes con insuficiencia cardíaca congestiva en etapa temprana y 8 controles sanos. Descubrieron una fuerte expresión de UTS2 en los cardiomiocitos y, en menor medida, en las células de músculo liso vascular, las células endoteliales y las células inflamatorias de los pacientes con insuficiencia cardíaca congestiva en etapa terminal. Hubo una expresión de UTS2 significativamente menos en el miocardio de pacientes con insuficiencia cardíaca congestiva en fase temprana y poca o ninguna expresión de UTS2 en el miocardio de los controles sanos. El análisis por RT-PCR mostró concentraciones aumentadas de UTS2 y la presencia de ARNm del receptor de urotensina en el miocardio de pacientes con insuficiencia cardíaca congestiva. El análisis por microscopía confocal mostró un aumento significativo en los sitios de unión para UTS2 en el miocardio de pacientes con insuficiencia cardíaca congestiva en etapa terminal. Pueden identificarse ligandos/fragmentos adecuados del CMH del UTS2 humano mediante el procedimiento descrito en relación con la Colonia 1 anterior.

Las muteínas, derivados y fragmentos de las proteínas de los aspectos (1), (2) y (3) de la invención incluyen

- (i) fragmentos que tienen al menos 8 restos de aminoácidos, preferentemente al menos 9 restos de aminoácidos, incluyendo fragmentos de péptidos que se unen a moléculas del complejo mayor de histocompatibilidad;
- (ii) derivados que incluyen proteínas de fusión con al menos un dominio proteico funcional adicional y proteínas modificadas químicamente (por ejemplo, proteínas PEGiladas) (en los que el tipo del dominio proteico adicional y de la modificación química depende de la utilidad pretendida de la proteína)
- (iii) muteínas que incluyen secuencias en las que hasta 5, preferentemente hasta 3, restos de aminoácidos consecutivos o separados se han suprimido, sustituido o añadido; y
- (iv) fragmentos de proteína que tiene una secuencia seleccionada entre las SEQ ID NO: 152-255, 256-435, 436-536, 537-660, 661-708 o 709-734.

A continuación, la invención se explica más estrechamente por referencia a las proteínas del aspecto ((2) de la invención), a saber, las 6 proteínas Colonia 1 a Colonia 6 y muteínas, derivados y fragmentos de las mismas, que, sin embargo, no se han de considerar una limitación de la invención. La ubicación genómica, su ADN, ARN y estructura proteica de Colonia 1-6, así como de otros genes que codifican las proteínas que se desvelan en el presente documento, así como los dominios proteicos, se han explicado anteriormente en el presente documento. Adicionalmente, la función supuesta y el panel de expresión en otros tejidos se ha demostrado anteriormente en el presente documento.

En la parte experimental se demuestra una estrategia de clonación adecuada mediante una expresión de las 6 proteínas en bacterias (*E. coli*), así como en células de mamífero (células Jurkat, linfocitos T primarios). Se usaron clones de ADNc de longitud completa para clonar el ADNc de los 6 genes en el vector de expresión bacteriano pTrcHis-TOPO (Invitrogen) y los dos vectores de expresión de mamífero pLenti6/V5-DEST y pCDNA6-His (Invitrogen). Se muestran adicionalmente procedimientos para la generación de anticuerpos (monoclonales o policlonales) frente a Colonia 1-6 o un fragmento peptídico de Colonia 1-6, y se describen estrategias para la inactivación funcional de los 6 genes, lo que potenciará el conocimiento acerca de la función y la importancia potenciales de los genes en linfocitos T_{reg}. Por último, se desvelan estrategias de activación para los 6 genes en linfocitos T CD4⁺CD25⁻ convencionales humanos primarios para determinar el potencial de los genes para inducir la función y el fenotipo específicos de linfocitos T_{reg} en linfocitos no T_{reg} se describen. El aspecto (3) de la invención se refiere al aislamiento de linfocitos T humanos supresores y reguladores a partir de sangre periférica humana, preferentemente mediante anticuerpos monoclonales adecuados y usando procedimientos de separación magnética o inmunoadsorción. Además, se prefiere que dichos linfocitos T humanos sean positivos para FOXP3 y posean propiedades reguladoras.

Se desvela adicionalmente un procedimiento para expresar los linfocitos T de los aspectos (1), (2) y (3) anteriores. Dicho procedimiento puede efectuarse *ex vivo* o *in vivo*. El agente estimulador de linfocitos T utilizado en dicho procedimiento es preferentemente una composición que comprende

- (a) ligandos anti-CD3 y/o anti-CD28/anticuerpos monoclonales, incluyendo anticuerpos super-agonistas,
- (b) un ligando/anticuerpo frente al receptor de linfocitos T en la superficie de las células que se han definido anteriormente o frente a los componentes del receptor de linfocitos T; o
- (c) complejos CMH-péptido que se unen a los receptores de linfocitos T expresados en la superficie de las que se han definido anteriormente; o
- (d) un éster de forbol y un ionóforo de calcio.

Los linfocitos T humanos expandidos que se pueden obtener mediante el procedimiento mencionado anteriormente

en el presente documento pueden fijarse. Dicha fijación puede, por ejemplo, conseguirse mediante tratamiento *ex vivo* con paraformaldehído.

En un procedimiento para la preparación de los linfocitos T del aspecto (1), (2) o (3) los linfocitos T se activan/estimulado/expanden con antígenos incluyendo autoantígenos (incluyendo des-mogleína 3 en el caso del péncigo vulgar, MelanA o tirosinasa en el caso del vitiligo; tiroglobulina en caso de la tiroiditis), antígenos extraños (incluyendo antígenos derivados de patógenos tales como la hepatitis C) o aloantígenos/antígenos de trasplante. Los antígenos indefinidos son preferentemente antígenos derivados de tejido o células (incluyendo antígenos que están en forma de ARN derivado de células o tejido necróticos o apoptóticos o híbridos entre células de interés y células dendríticas/células presentadoras de antígeno, otras formas de entrega de antígenos indefinidos dentro de células dendríticas u otras células presentadoras de antígeno) o antígenos derivados de patógenos.

La composición farmacéutica o medicamento que contiene los linfocitos T de los aspectos (1), (2) y (3) de la invención puede comprender adicionalmente vehículos farmacéuticamente aceptables conocidos para un experto en la materia. La composición farmacéutica o medicamento es preferentemente adecuado para el tratamiento de enfermedades con inmunidad potenciada, incluyendo, pero no limitadas a, enfermedades autoinmunitarias, enfermedad de injerto contra huésped y rechazos de injertos. Los anticuerpos contra las proteínas de los aspectos (2) y (3) de la invención, en concreto los anticuerpos contra las proteínas Colonia 1-6, pueden prepararse mediante procedimientos bien establecidos en la técnica incluyendo la administración de

(I) célula presentadora de antígeno (CPA) pulsada con una proteína como se ha definido en los aspectos (1), (2) y (3) o con Colonia 1-6, o una proteína como se ha definido en los aspectos (1), (2) y (3) o un péptido de Colonia 1-6 que activa las células B para que produzcan anticuerpos que reconozcan linfocitos T_{reg} de una manera específica de proteína (específica de Colonia 1-6),

(II) una proteína como se ha definido en los aspectos (1), (2) y (3) o Colonia 1-6, o un péptido de una proteína como se ha definido en los aspectos (1), (2) y (3) o de Colonia 1-6, no necesariamente en combinación con un adyuvante, que es procesada por una célula presentadora de antígeno, que, a su vez, activa células B para que produzcan anticuerpos que reconozcan linfocitos T_{reg} de una manera específica de proteína (específica de Colonia 1-6),

(III) una molécula de ácido nucleico que codifica una proteína como se ha definido en los aspectos (1), (2) y (3) o Colonia 1-6, o un péptido de una proteína como se ha definido en los aspectos (1) y (2) o de Colonia 1-6, no necesariamente en un vector de expresión que posteriormente se expresa de modo que pueda ser procesado por una célula presentadora de antígeno, que activa las células B para que produzcan anticuerpos que reconozcan la proteína (Colonia 1-6) de una manera específica de proteína (específica de Colonia 1-6), o

(IV) el uso de una proteína como se ha definido en los aspectos (1), (2) y (3) o de Colonia 1-6, o un péptido de una proteína como se ha definido en los aspectos (1), (2) y (3) o de Colonia 1-6, para que se una a anticuerpos expresados por una biblioteca de fagos.

Se expresan numerosos anticuerpos en la biblioteca como fusiones con la proteína de cubierta de un bacteriófago, de modo que se presentan en la superficie de la partícula vírica. El ADN extraído de fagos interactuantes contiene las secuencias de los anticuerpos específicos que reconocen una proteína como se ha definido en los aspectos (1), (2) y (3) o de Colonia 1-6 de una manera específica de proteína (específica de Colonia 1-6).

Los aspectos (1), (2) y (3) de la invención incluyen la detección diagnóstica de linfocitos T_{reg} $CD4^+CD25^+FOXP3^+$ humanos con anticuerpos específicos para una proteína que se desvela en el presente documento, en particular con anticuerpos monoclonal o policlonales específicos para Colonia 1-6 (sin marcar o directamente conjugados) de una manera específica de proteína (específica de Colonia 1-6) mediante inmunohistoquímica, inmunofluorescencia, citometría de flujo, transferencia Western, inmunoprecipitación, cromatografía de afinidad y otras técnicas para la identificación específica de la expresión de proteínas. El anticuerpo puede generarse, por ejemplo, mediante inmunización de mamíferos, por ejemplo, ratón, rata, cabra, burro, conejo, con polipéptidos o péptidos derivados de la proteína que se desvela en el presente documento o de Colonia 1-6 con la adición de adyuvantes específicos, o puede identificarse mediante presentación en fagos como se ha especificado anteriormente.

Los aspectos (1), (2) y (3) de la invención también incluyen la detección diagnóstica de la expresión génica de los péptidos que se desvelan en el presente documento o de la expresión génica de Colonia 1-6 en linfocitos T_{reg} $CD4^+CD25^+FOXP3^+$ humanos con RT-PCR, transferencia Northern, matrices de genes y otras técnicas para la identificación específica de la expresión génica, lo que puede conducir a una identificación más específica de los linfocitos T_{reg}

Adicionalmente, los péptidos de la invención, incluyendo los genes Colonia 1-6, pueden expresarse diferencialmente en linfocitos T_{reg} $CD4^+CD25^+FOXP3^+$ naturales y adaptativos y, por tanto, pueden ayudar a discriminar estas dos poblaciones distintas. Si es así, estos genes pueden usarse como un marcador de linaje.

Además, la presente invención incluye la administración al paciente de un anticuerpo específico para una proteína de la invención, en concreto un anticuerpo específico monoclonal o policlonal Colonia 1-6 (humanizado, quimérico o no humanizado) que conduce a la muerte (en lo sucesivo en el presente documento denominada destrucción) de la célula de una manera específica de Colonia 1-6. El anticuerpo puede generarse como se ha indicado anteriormente.

Se desvela adicionalmente la administración a un paciente de linfocitos T citotóxicos (LTC) (autólogos o alogénos) que eliminan los linfocitos T_{reg} de una manera específica de proteína (específica de Colonia 1-6), restringida al complejo mayor de histocompatibilidad. El LTC puede generarse, por ejemplo, mediante activación con células presentadoras de antígenos que se han pulsado con las proteínas que se desvelan en el presente documento o Colonia 1-6, o un péptido de la proteína que se desvela en el presente documento o de Colonia 1-6 que se une a una molécula de complejo mayor de histocompatibilidad. La invención incluye también un procedimiento alternativo de tratamiento de un paciente para eliminar linfocitos T_{reg} que expresan la proteína que se desvela en el presente documento (Colonia 1-6). Este procedimiento implica la administración al paciente de una célula presentadora de antígeno (CPA) que activa en el paciente un linfocito T citotóxico que destruye la célula de una manera específica de proteína (específica de Colonia 1-6), restringida al complejo mayor de histocompatibilidad. La CPA puede pulsarse con las proteínas que se desvelan en el presente documento o Colonia 1-6, o un péptido de la proteína que se desvela en el presente documento o de Colonia 1-6 que se une a una molécula del complejo mayor de histocompatibilidad.

Otro procedimiento que se desvela es un procedimiento de eliminación de linfocitos T_{reg} en un paciente administrando al paciente una proteína que se desvela en el presente documento o Colonia 1-6, o un péptido de la proteína que se desvela en el presente documento o de Colonia 1-6 que se une a una molécula del complejo mayor de histocompatibilidad, que es procesada por una célula presentadora de antígeno en el paciente, lo que, a su vez, activa un linfocito T citotóxico en el paciente para inducir la muerte celular de la célula que expresa la proteína que se desvela en el presente documento (Colonia 1-6) de una manera específica de proteína (específica de Colonia 1-6), restringida al complejo mayor de histocompatibilidad. La proteína que se desvela en el presente documento o polipéptido o péptido Colonia 1-6 de la proteína que se desvela en el presente documento o de Colonia 1-6 utilizada en este procedimiento puede administrarse al paciente en asociación con un adyuvante.

Se desvela adicionalmente un procedimiento para eliminar específicamente linfocitos T_{reg} que expresan la proteína que se desvela en el presente documento o Colonia 1-6. Este procedimiento implica administrar al paciente una molécula de ácido nucleico que codifica la proteína que se desvela en el presente documento o Colonia 1-6, o un péptido de la proteína que se desvela en el presente documento o de Colonia 1-6 que se une a una molécula del complejo mayor de histocompatibilidad. La molécula de ácido nucleico se expresa en el paciente de manera que pueda ser procesada por una célula presentadora de antígeno en el paciente, que activa un linfocito T citotóxico en el paciente para inducir la muerte celular de la célula que expresa la proteína que se desvela en el presente documento o Colonia 1-6, de una manera específica de proteína (específica de Colonia 1-6), restringida al complejo mayor de histocompatibilidad. La molécula de ácido nucleico que codifica la proteína que se desvela en el presente documento o Colonia 1-6, o un péptido de la proteína que se desvela en el presente documento o de Colonia 1-6 puede estar presente en un vector de expresión.

Cada uno de los procedimientos descritos anteriormente también pueden incluir el tratamiento basado en torno a un segundo (o más) antígeno asociado a linfocito T_{reg} o un péptido del mismo que se une al CMH. En cualquiera de los procedimientos descritos anteriormente, el paciente puede tener linfocitos T_{reg} que expresan la proteína que se desvela en el presente documento o Colonia 1-6. Las CPA utilizadas en estos procedimientos pueden ser, por ejemplo, una célula dendrítica o una célula B activada por CD40. El péptido que se desvela en el presente documento o Colonia 1-6 en estos procedimientos puede unirse a una molécula del complejo mayor de histocompatibilidad (CMH) de clase I o de clase II.

También se desvela un procedimiento de evaluación del nivel de inmunidad de un paciente a la proteína que se desvela en el presente documento o Colonia 1-6, o un péptido de la proteína que se desvela en el presente documento o de Colonia 1-6 que se une a una molécula del complejo mayor de histocompatibilidad. En este procedimiento, el nivel de linfocitos T citotóxicos específicos para la proteína que se desvela en el presente documento o para Colonia 1-6 o para un péptido de la proteína que se desvela en el presente documento o de Colonia 1-6 se mide en una muestra de un paciente. La muestra puede obtenerse del paciente antes, durante o después de que se administre un tratamiento al paciente. Una muestra también puede obtenerse, por ejemplo, antes y después del tratamiento.

También se desvelan péptidos descritos anteriormente en el presente documento de la invención o péptidos Colonia 1-6 (en concreto, péptidos de las proteínas que se unen a moléculas del complejo mayor de histocompatibilidad).

Se desvela adicionalmente un linfocitos T citotóxico generado *ex vivo* que destruye específicamente una célula que expresa el péptido que se desvela en el presente documento o Colonia 1-6 de una manera específica, restringida al complejo mayor de histocompatibilidad, y una célula presentadora de antígeno generada *ex vivo* (por ejemplo, una célula dendrítica o una célula B activada por CD40) que presenta un péptido del péptido que se desvela en el presente documento o de Colonia 1-6 en el contexto de una molécula de complejo mayor de histocompatibilidad.

La composición farmacéutica que se desvela en el presente documento comprende la proteína que se desvela en el presente documento o Colonia 1-6 o un péptido de unión al CMH de la misma que puede utilizarse para la vacunación. Estos incluyen, sin limitación, la proteína de longitud completa que se desvela en el presente documento o la proteína Colonia 1-6 de longitud completa, fragmentos de unión al CMH de la proteína que se desvela en el presente documento o de Colonia 1-6, así como proteínas de fusión que incluyen la proteína que se

desvela en el presente documento o Colonia 1-6 y fragmentos de unión al CMH de la misma. Los péptidos o polipéptidos que incluyen la proteína que se desvela en el presente documento o péptidos y polipéptidos Colonia 1-6 pueden incluir 8, 9, 10, 11, 12 o más tramos de aminoácidos que tienen identidad de secuencia con una región de la proteína que se desvela en el presente documento o de Colonia 1-6. Por ejemplo, los péptidos pueden incluir nueve tramos de aminoácidos, en los que siete, ocho o los nueve aminoácidos en la proteína que se desvela en el presente documento o de la secuencia de nueve aminoácidos del péptido Colonia 1-6 son idénticos a una región de nueve aminoácidos en la proteína que se desvela en el presente documento o Colonia 1-6. Además, un péptido o polipéptido de la proteína que se desvela en el presente documento o un péptido o polipéptido Colonia 1-6 puede incluir hasta la longitud completa de los aminoácidos que son idénticos a una secuencia de aminoácidos que se encuentra en la proteína que se desvela en el presente documento o en Colonia 1-6, respectivamente, por ejemplo, 9-20, 20-40, 40-80, 80-a longitud completa, 80-200 o 200-a longitud completa aminoácidos que son idénticos a una secuencia de aminoácidos que se encuentra en Colonia 1-6, respectivamente. Los polipéptidos que contienen los péptidos de la proteína que se desvela en el presente documento o péptidos Colonia 1-6 pueden contener tramos de aminoácidos adicionales que no corresponden a la secuencia de aminoácidos de la proteína que se desvela en el presente documento o de Colonia 1-6, respectivamente.

Para vacunar para desencadenar una respuesta inmunitaria específica para la proteína que se desvela en el presente documento o una respuesta inmunitaria específica para Colonia 1-6 en pacientes, así como animales, es necesario obtener grandes cantidades del péptido de la proteína que se desvela en el presente documento o de la proteína o péptido Colonia 1-6 y esto puede conseguirse mediante numerosos procedimientos convencionales, por ejemplo, síntesis química (por ejemplo, procedimientos de Fmoc (Sigma Genosys), véase anteriormente) o la expresión en células eucarióticas o procarióticas. Pueden sobreexpresarse péptidos recombinantes de la proteína que se desvela en el presente documento o péptidos Colonia 1-6 *in vivo* mediante la introducción de secuencias codificantes de los péptidos en diversos tipos celulares, o *in vitro*, usando sistemas de expresión sin células que se muestran en la *técnica*. Los productos peptídicos después pueden purificarse para la generación de anticuerpos y LTC específicos para las proteínas que se desvelan en el presente documento o anticuerpos y LTC específicos para Colonia 1-6 (en concreto, los anticuerpos del aspecto (5) de la invención) *ex vivo* y para la producción de vacunas.

De acuerdo con aspectos (3), (4) y (5) de la invención, los anticuerpos contra los péptidos de la proteína que se desvela en el presente documento o contra péptidos Colonia 1-6 también son útiles para ensayos diagnósticos que miden la presencia de linfocitos T_{reg} que expresan la proteína que se desvela en el presente documento o de linfocitos T_{reg} que expresan Colonia 1-6 en una muestra de ensayo. Por ejemplo, la presencia (o el aumento de los niveles) de linfocitos T_{reg} que expresan Colonia 1-6 en una muestra de un sujeto que ha recibido una terapia, en relación con el nivel de linfocitos T_{reg} que expresan Colonia 1-6 en una muestra de referencia (tal como una muestra de pretratamiento del paciente), indica que el paciente ha respondido a la terapia proporcionada.

También pueden ser útiles péptidos purificados de la proteína que se desvela en el presente documento o péptidos Colonia 1-6 para ensayos de diagnóstico que miden la presencia de LTC específicos de Colonia 1-6 en una muestra de ensayo. Por ejemplo, la presencia (o el aumento de los niveles) de LTC específicos de Colonia 1-6 en una muestra de un sujeto que ha recibido una vacunación anti-Colonia 1-6, con respecto al nivel de LTC específicos de Colonia 1-6 en una muestra de referencia (tal como una muestra prevacunación del paciente), indica que el paciente ha montado una respuesta inmunitaria específica de Colonia 1-6.

Los péptidos de la proteína que se desvela en el presente documento o péptidos Colonia 1-6 pueden producirse mediante síntesis química (por ejemplo, mediante los procedimientos descritos en *Solid Phase Peptide Synthesis*, 2^a ed., 1984, The Pierce Chemical Co., Rockford, IL, o mediante otros procedimientos conocidos por los expertos en la materia de la síntesis peptídica).

Puede usarse una amplia diversidad de sistemas de expresión para producir de manera recombinante las proteínas que se desvelan en el presente documento y péptidos de las mismas, en concreto, los péptidos, polipéptidos, fragmentos, proteínas de fusión y variantes de secuencia de aminoácidos de Colonia 1-6. Las proteínas que se desvelan en el presente documento y los péptidos Colonia 1-6 pueden producirse en hospedadores procarióticos (por ejemplo, *E. coli*) o en hospedadores eucarióticos (por ejemplo, *S. cerevisiae*, células de insectos, tales como células Sf9 o células de mamíferos, tales como células COS-1, NIH 3T3, Jurkat, 293, 293T o HeLa). Estas células están disponibles en el mercado, por ejemplo, de American Type Culture Collection, Rockville, Maryland (véase también, por ejemplo, Ausubel y col., *Current Protocols in Molecular Biology*, John Wiley & Sons, Nueva York, Nueva York, 1998). El procedimiento de transformación y la elección del vehículo de expresión (por ejemplo, vector de expresión) dependen del sistema hospedador seleccionado. Se describen procedimientos de transformación y transfección, por ejemplo, por Ausubel y col., citado anteriormente, y pueden elegirse vehículos de expresión entre los numerosos ejemplos que son conocidos en este campo.

En primer lugar, una molécula de ácido nucleico que codifica la proteína que se desvela en el presente documento o el péptido Colonia 1-6 se introduce en un plásmido u otro vector, que después se usa para transformar las células vivas. Pueden usarse para la expresión de proteínas construcciones en las que un ADNc que contiene la secuencia codificante entera de la proteína que se desvela en el presente documento o la secuencia codificante de Colonia 1-6, un fragmento de dicha secuencia codificante, una variación de aminoácidos de dicha secuencia codificante o proteínas de fusión de los mencionados anteriormente, insertadas en la correcta orientación en un plásmido de

expresión. Los sistemas de expresión procarióticos y eucarióticos permiten que diversos dominios inmunogénicos de la proteína que se desvela en el presente documento o péptidos o polipéptidos Colonia 1-6 se recuperen como proteínas de fusión y después se usen para la generación de anticuerpos especificados para la proteína que se desvela en el presente documento o de anticuerpos específicos de Colonia 1-6, pero también LTC. En algunos casos, por ejemplo, cuando una proteína que se desvela en el presente documento o un péptido Colonia 1-6 se ha de expresar directamente dentro de una célula, puede ser deseable expresar la proteína que se desvela en el presente documento o el péptido Colonia 1-6 bajo el control de un promotor inducible o específico de tejido.

Los vectores de expresión típicos contienen promotores que dirigen la síntesis de grandes cantidades de ARNm correspondiente a la proteína insertada desvelada en el presente documento (péptido Colonia 1-6) que codifica una molécula de ácido nucleico en las células que llevan plásmidos. También pueden incluir secuencias de "origen de replicación" eucarióticas o procarióticas, que permitan su replicación autónoma dentro del organismo hospedador, secuencias que codifican rasgos genéticos que permiten que las células que contienen el vector se seleccionen en presencia de fármacos de otro modo tóxicos (tales como antibióticos), y secuencias que aumenten la eficiencia con la que se traduce el ARNm sintetizado. Pueden mantenerse vectores estables, a largo plazo, como entidades que se replican libremente dentro de las células mediante el uso de elementos reguladores de, por ejemplo, virus (por ejemplo, las secuencias OriP del genoma del virus de Epstein Barr). También pueden producirse estirpes celulares que tengan el vector integrado en el ADN genómico, y, de esta manera, el producto génico se produce de forma continua.

La expresión de secuencias extrañas en bacterias tales como *Escherichia coli* requiere la inserción de una molécula de ácido nucleico que codifique un polipéptido en un vector de expresión bacteriano. Los vectores plasmídicos de esta categoría contienen varios elementos necesarios para la propagación del plásmido en bacterias y la expresión del ADN insertado del plásmido por las bacterias que llevan el plásmido. La propagación solamente de las bacterias que llevan plásmidos se consigue mediante la introducción en el plásmido de secuencias que codifican marcadores seleccionables que permiten que las bacterias que llevan plásmidos crezcan en presencia de fármacos de otro modo tóxicos (por ejemplo, antibióticos). El plásmido también incluye un promotor de la transcripción que es capaz de producir grandes cantidades de ARNm del gen clonado. Dichos promotores pueden o pueden no ser promotores inducibles. El plásmido también, preferentemente, contiene un poliengarce para simplificar la inserción del gen en la orientación correcta dentro del vector. Por ejemplo, en un vector de expresión simple de *E. coli* que utiliza el promotor lac, el plásmido vector de expresión contiene un fragmento del cromosoma de *E. coli* que contiene el promotor lac y el gen *lacZ* vecino. En presencia del análogo de lactosa IPTG, la ARN polimerasa normalmente transcribe el gen *lacZ*, produciendo ARNm de *lacZ*, que se traduce en la proteína codificada, β -galactosidasa. El gen *lacZ* se puede cortar y retirar del vector de expresión con endonucleasas de restricción y puede reemplazarse por una secuencia génica de la proteína que se desvela en el presente documento o una secuencia génica del péptido Colonia 1-6 o un fragmento, fusión o mutante del mismo. Cuando el plásmido resultante se transfecta en *E. coli*, la adición de IPTG y posterior transcripción a partir del promotor *lac* produce ARNm que codifica la proteína que se desvela en el presente documento o el polipéptido Colonia 1-6 de interés, que después se traduce en un polipéptido.

Una vez que se construye el vector de expresión apropiado que contiene un gen de la proteína que se desvela en el presente documento o un gen Colonia 1-6, se introduce en una célula hospedadora apropiada mediante técnicas de transformación, transfección o transducción que son conocidas en la técnica, incluyendo la transformación con cloruro de calcio, la transfección con fosfato de calcio, la transfección con DEAE-dextrano, la electroporación, la microinyección, la fusión de protoplastos y la transfección mediada por liposomas. Las células hospedadoras que se transforman con los vectores pueden incluir *E. coli* u otras bacterias, levaduras, hongos, células de insecto (utilizando, por ejemplo, vectores de baculovirus para la expresión), humanas, de ratón u otras células animales. También pueden usarse células de mamíferos para expresar la proteína que se desvela en el presente documento de los péptidos Colonia 1-6 usando un sistema de expresión de virus de la variolovacuna, como se describe por Ausubel y col., citado anteriormente.

La expresión *in vitro* de la proteína que se desvela en el presente documento o de péptidos Colonia 1-6, proteínas, fusiones, fragmentos de polipéptido o versiones mutadas de los mismos codificados por ADN clonado también es posible usando el sistema de expresión del promotor tardío T7. También pueden usarse vectores plasmídicos que contienen promotores tardíos y las polimerasas de ARN correspondientes de bacteriófagos relacionados, tales como T3, T5 y SP6 para la producción *in vitro* de proteínas a partir de ADN clonado. También puede usarse *E. coli* para la expresión usando un fago M13 tal como mGPI-2. Además, para la expresión en *E. coli* también pueden usarse vectores que contienen secuencias reguladoras del fago lambda o vectores que dirigen la expresión de proteínas de fusión, por ejemplo, una proteína de fusión de proteína de unión a maltosa o una proteína de fusión de glutatión S-transferasa. Los sistemas de expresión eucarióticos permiten modificaciones postraduccionales apropiadas de las proteínas expresadas. La transfección transitoria de un plásmido de expresión eucariótico permite la producción transitoria de las proteínas que se desvelan en el presente documento o de péptidos Colonia 1-6 por una célula hospedadora transfectada. Las proteínas que se desvelan en el presente documento de péptidos Colonia 1-6 también pueden ser producidas por una estirpe celular de mamífero transfectada de forma estable. Un número de vectores adecuados para la transfección estable de células de mamífero está disponible al público (por ejemplo, véase Pouwels y col., *Cloning Vectors: A Laboratory Manual*, 1985, Sup. 2, 987), como también lo están los procedimientos para construir dichas estirpes celulares (véase, por ejemplo, Ausubel y col., citado anteriormente). En un ejemplo, el ADNc que codifica una proteína que se desvela en el presente documento o un péptido, proteína, fragmento,

mutante o proteína de fusión Colonia 1-6 se clonó en un vector de expresión que incluía el gen de la dihidrofolato reductasa (DHFR). La integración del plásmido y, por tanto, la integración del gen que codifica la proteína que se desvela en el presente documento o los genes que codifican péptidos Colonia 1-6 en el cromosoma de la célula hospedadora se selecciona mediante la inclusión de metotrexato 0,01-300 μM en el medio de cultivo celular (como se describe por Ausubel y col., citado anteriormente). Esta selección dominante puede lograrse en la mayoría de tipos celulares. La expresión de la proteína recombinante puede aumentarse mediante amplificación mediada por DHFR del gen transfectado. Los procedimientos para seleccionar estirpes celulares que llevan amplificaciones génicas se describen por Ausubel y col., citado anteriormente. Estos procedimientos generalmente implican el cultivo extendido en medio que contiene niveles gradualmente crecientes de metotrexato. Los vectores de expresión que contienen DHFR más habitualmente utilizados son pCVSEII-DHFR y pAdD26SV (A) (descritos por Ausubel y col., citado anteriormente). Las células hospedadoras descritas anteriormente o, preferentemente, una estirpe celular CHO deficiente en (por ejemplo, células CHO DHFR-, N.º de referencia de ATCC CRL 9096) están entre las más preferidas para la selección de DHFR de una estirpe celular transfectada de forma estable o la amplificación génica mediada por DHFR. Pueden usarse de forma análoga otros marcadores de fármacos.

La expresión de proteínas que se desvelan en el presente documento, tales como las que contienen péptidos Colonia 1-6, en células eucarióticas permite la producción de grandes cantidades de proteínas normales o mutantes para el aislamiento y la purificación, y el uso de células que expresan una proteína que contiene péptido Colonia 1-6 proporciona un sistema de ensayo funcional para anticuerpos generados contra un péptido Colonia 1-6 de interés.

Otro sistema de expresión eucariótico preferido es el sistema de vector de expresión trc usando, por ejemplo, el vector pTrcHis, que está disponible de Invitrogen (Karlsruhe, Alemania). Si se desea, este sistema puede usarse para expresar las proteínas que se desvelan en el presente documento o Colonia 1-6 junto con otros marcadores de proteínas, por ejemplo, el marcador myc, His o XPRESS como se conoce en la técnica.

Una vez que una proteína recombinante que se desvela en el presente documento o proteína Colonia 1-6 se expresa, puede aislarse de las células de expresión mediante lisis celular, seguida de técnicas de purificación de proteínas, tales como cromatografía de afinidad. En este ejemplo, un anticuerpo anti-XPRESS puede unirse a una columna y usarse para aislar la proteína recombinante que se desvela en el presente documento o las proteínas que contienen péptidos Colonia 1-6 recombinantes. La lisis y el fraccionamiento de células que albergan péptido Colonia 1-6 anteriores a la cromatografía de afinidad pueden realizarse mediante procedimientos convencionales (véase, por ejemplo, Ausubel y col., citado anteriormente). Una vez aislada, la proteína recombinante puede, si se desea, purificarse adicionalmente, por ejemplo, mediante cromatografía líquida de alto rendimiento (HPLC; por ejemplo, véase Fisher, *Laboratory Techniques in Biochemistry and Molecular Biology*, Work y Burdon, Editores, Elsevier, 1980).

Para la producción de anticuerpos, la proteína que se desvela en el presente documento o Colonia 1-6 o un péptido de unión a CMH de la misma se administra a un animal en asociación con un adyuvante. Por ejemplo, puede usarse un antígeno químico (por ejemplo, adyuvante incompleto de Freund; citoxano; un compuesto de aluminio, tal como hidróxido de aluminio, fosfato de aluminio o hidroxifosfato de aluminio; liposomas; ISCOMS; microesferas; cocleatos de proteínas; vesículas que consisten en tensioactivos no iónicos; dispersiones anfífilas catiónicas en agua; emulsiones aceite/agua; muramididipéptido (MDP) y sus derivados tales como glucosil muramid dipéptido (GMDP), treonil-MDP, mu-rametida y murapalmitina; y QuilA y sus subfracciones, así como otros compuestos diversos tales como monofosforil-lípido A (MIPLA); gamma-inulina; calcitriol; y loxoribina).

Preferentemente, se administra Colonia 1-6 o un péptido de unión al CMH de la misma a un paciente en asociación con un adyuvante. Por ejemplo, un antígeno químico tal como se ha descrito anteriormente.

Un modificador de la respuesta biológica, que es un mediador soluble que afecta a la inducción de una respuesta inmunitaria, también puede usarse como adyuvante. Por ejemplo, pueden usarse citocinas (por ejemplo, IL2 y GM-CSF), quimiocinas, moléculas coestimuladoras (por ejemplo, B7, ICAM, anticuerpos monoclonales de clase I, factor de células madre y linfocitos T estimulados). Además, productos bacterianos, tales como toxinas o, preferentemente, subunidades o fragmentos de las mismas que tienen toxicidad reducida (si es que tienen), pero una actividad adyuvante mantenida. Los tipos adicionales de moléculas adyuvantes que pueden usarse en la invención incluyen, por ejemplo, modificadores biológicos de la respuesta de muerte (por ejemplo, sensibilizadores a la apoptosis) y compuestos o tratamiento que aumentan la susceptibilidad de la célula diana para el tratamiento, tales como la radiación y la quimioterapia. Además, el aumento de expresión de la proteína que se desvela en el presente documento o Colonia 1-6 en la célula puede aumentar la susceptibilidad de la célula a un tratamiento como se desvela en el presente documento.

Por último, como se ha descrito anteriormente, pueden usarse adyuvantes celulares en los procedimientos de inmunización que se desvelan en el presente documento. Por ejemplo, un péptido de la proteína que se desvela en el presente documento o un péptido Colonia 1-6 puede administrarse a un paciente o un animal en la superficie de una célula presentadora de antígeno, en el contexto del CMH. Además de las células presentadoras de antígeno profesionales pueden usarse, por ejemplo, células dendríticas, células B activadas por CD40, linfocitos T_{reg} irradiados, células presentadoras de antígeno alternativas, células presentadoras de antígenos sintéticas (por ejemplo, micelas lipídicas y armazones artificiales similares a CPA) y fusiones de cualquiera de las células

enumeradas anteriormente.

Como una alternativa a la vacunación con una proteína que se desvela en el presente documento o un péptido de la misma o una proteína o péptido Colonia 1-6, la vacunación con una molécula de ácido nucleico que codifica una proteína o péptido de este tipo puede usarse para la vacunación. Dichas moléculas de ácido nucleico pueden administrarse en forma de moléculas de ADN "desnudas", presentes en un plásmido o vector vírico, o empaquetadas en un liposoma o una célula, tal como una célula eucariótica, antes de la administración. Las moléculas de ácido nucleico pueden administrarse *in vivo* o pueden usarse para tratar una célula *ex vivo* (por ejemplo, una célula presentadora de antígeno, tal como una célula dendrítica o una célula B activada por CD40), que después se administra. Como alternativa, puede usarse ARN, por ejemplo, ARNm, en estos procedimientos (Boczkowski y col., *J. Exp. Med.*; 184: 465-472 (1996)).

Para la expresión *in vivo*, un gen que codifica un polipéptido que incluye una proteína que se desvela en el presente documento o Colonia 1-6 o un péptido de unión al CMH de las mismas deben entregarse a las células en una forma que pueda ser captada por las células, en las que se expresa un suficiente nivel de proteína para inducir una respuesta inmunitaria eficaz. Los vectores retrovíricos, adenovíricos, lentivíricos, poxvíricos y otros vectores víricos son adecuados como vectores de expresión de ácido nucleico para la entrega *in vivo*, debido a que muestran una infección y/o una integración y una expresión eficientes; (Cayouette y Gravel, *Hum. Gene Ther.*; 8: 423-430 (1997); Kido y col., *Curr. Eye Res.*; 15: 833-844 (1996); Miyoshi y col., *Proc. Natl. Acad. Sci. USA*, 94: 10319-10323 (1997); Naldini y col., *Science*; 272: 263-267 (1996)). Por ejemplo, cualquier fragmento de ADN que codifica un polipéptido que contiene un péptido de una proteína que se desvela en el presente documento o un péptido Colonia 1-6 puede clonarse en un vector retrovírico o lentivírico y transcribirse a través de su promotor endógeno, a través de un promotor exógeno, a través de un promotor específico para el tipo de célula diana de interés, o, en el caso de vectores retrovíricos o lentivíricos, a través de la repetición terminal larga vírica. Otros vectores víricos que pueden usarse incluyen adenovirus, virus adenoasociados, poxvirus, tales como virus de la variolovacuna o virus del papiloma bovino, o un virus del herpes, tal como el virus de Epstein-Barr.

La transferencia génica *in vivo* también puede conseguirse por medios no víricos. Por ejemplo, un vector plasmídico que codifica un polipéptido que contiene un péptido de una proteína que se desvela en el presente documento o un péptido Colonia 1-6 puede inyectarse directamente en el músculo esquelético o músculo cardíaco mediante procedimientos descritos anteriormente (Wolff y col., *Science*; 247: 1465-1468 (1990)). Los vectores de expresión inyectados en el músculo esquelético *in situ* son captados en núcleos de las células musculares y se usan como moldes para la expresión de sus proteínas codificadas. Se secretan genes que codifican péptidos Colonia 1-6 que se modifican mediante ingeniería genética para que contengan un péptido señal, de células musculares que expresan péptidos Colonia 1-6, después de lo cual inducen una respuesta inmunitaria. La transferencia génica en células dentro de los tejidos de un animal vivo puede conseguirse también mediante lipofección (Brigham y col., *Am. J. Med. Sci.*; 298: 278-281 (1989); Felgner y col., *Proc. Natl. Acad. Sci. USA*, 84: 7413-7417 (1987); Ono y col., *Neurosci. Lett.*; 117: 259-263 (1990)) o conjugación asialoorosomucoide-pollisina (Wu y col., *J. Biol. Chem.*; 264: 16985-16987; Wu y Wu, *J. Biol. Chem.*; 263: 14621-14624 (1988)) y procedimientos análogos.

También pueden usarse vectores retrovíricos, vectores adenovíricos, vectores víricos asociados a adenovirus u otros vectores víricos para entregar genes que codifican péptidos o polipéptidos Colonia 1-6 a células *ex vivo*. Se conocen numerosos vectores útiles para este fin (Anderson, *Science*; 224: 340 (1984); Cornetta y col., *Prog. Nucleic Acid. Res. Mol. Biol.*; 36: 311-322 (1989); Eglitis y col., *Adv. Exp. Med. Biol.*; 241: 19-27 (1988); Friedmann, *Science*; 244: 1275-1281 (1989); Johnson, *Chest*; 107: 77S-83S (1995); Le Gal La Salle y col., *Science*; 259: 988-990 (1993); Miller, *Hum. Gene Ther.*; 1: 5-14 (1990); Moen, *Blood Cells*, 17: 407-416 (1991); Tolstoshev y Anderson, *Curr. Opin. Biotechnol.*; 1: 55-61 (1990)). Los vectores retrovíricos en concreto están bien desarrollados y se han usado en entornos clínicos (Rosenberg y col., *N. Engl. J. Med.*; 323: 570-578 (1990)); Anderson y col., Patente de los EE.UU. N.º 5.399.346).

La transferencia de genes en células *ex vivo* también puede conseguirse mediante la entrega de vectores no víricos, tales como plásmidos de expresión, usando procedimientos tales como transfección fosfato de calcio o DEAE dextrano, electroporación y fusión de protoplastos. Los liposomas también pueden ser potencialmente beneficiosos para la entrega de ADN en una célula.

Las células que han de transducirse o transfectarse *ex vivo* pueden obtenerse de un animal o un paciente (por ejemplo, células de sangre periférica, tales como linfocitos T_{reg}, células B o células dendríticas, células madre de médula ósea o células de una biopsia tumoral) antes de la transfección y pueden reintroducirse después de la transfección. Sin embargo, las células también pueden derivar de una fuente distinta del paciente o animal que se somete a la transferencia génica. En las construcciones descritas anteriormente, la expresión del péptido Colonia 1-6 puede dirigirse a partir de cualquier promotor adecuado (por ejemplo, los promotores citomegalovirus humano (CMV), virus de simio 40 (SV40) o metalotioneína) y puede regularse mediante cualquier elemento regulador de mamífero apropiado. Por ejemplo, si se desea, pueden usarse preferentemente potenciadores que se sabe que dirigen preferencialmente la expresión génica en células de músculo esquelético, para dirigir la expresión de péptido Colonia 1-6 para la vacunación *in situ*. Los potenciadores utilizados pueden incluir los que se caracterizan como específicos de tejido o células en su expresión.

Puede emplearse la práctica farmacéutica convencional para proporcionar formulaciones o composiciones adecuadas para administrar anticuerpos contra las proteínas que se desvelan en el presente documento o anticuerpos Colonia 1-6, vacunas de péptidos o ácidos nucleicos para el tratamiento o la profilaxis contra el aumento de frecuencias de linfocitos T_{reg}. Los anticuerpos contra las proteínas que se desvelan en el presente documento o anticuerpos Colonia 1-6, los péptidos de las proteínas que se desvelan en el presente documento o péptidos Colonia 1-6, las proteínas que se desvelan en el presente documento o polipéptidos Colonia 1-6 y moléculas de ácido nucleico que codifican los mismos pueden administrarse dentro de un diluyente, vehículo o excipiente farmacéuticamente aceptable, en una forma de dosificación unitaria. La administración puede comenzar antes de que un paciente sea sintomático. Puede emplearse cualquier vía de administración adecuada, por ejemplo, la administración puede ser parenteral, intravenosa, intraarterial, subcutánea, intramuscular, intracraneal, intraorbital, oftálmica, intraventricular, intracapsular, intraespinal, intracisternal, intraperitoneal, intranasal, mediante aerosol, mediante supositorios o administración oral. Las formulaciones terapéuticas pueden estar en forma de soluciones o suspensiones líquidas; para la administración oral, las formulaciones pueden estar en forma de comprimidos o cápsulas; y para las formulaciones intranasales, en forma de polvos, gotas nasales o aerosoles. Puede incluirse con la formulación un adyuvante, por ejemplo, como se ha enumerado anteriormente.

Se encuentran métodos bien conocidos en la técnica para fabricar formulaciones, por ejemplo, en *Remington's Pharmaceutical Sciences*, (18ª edición), ed. A. Gennaro, 1990, Mack Publishing Company, Easton, PA. Las formulaciones para la administración parenteral pueden, por ejemplo, contener excipientes, agua estéril o solución salina, polialquilenglicoles tales como polietilenglicol, aceites de origen vegetal o naftalenos hidrogenados. Pueden usarse polímero de lactida biocompatible biodegradable, copolímero de lactidalglicólido o copolímeros de polioxietileno-polioxipropileno para controlar la liberación de los compuestos. Otros sistemas de entrega parenteral potencialmente útiles para los péptidos, proteínas/polipéptidos y moléculas de ácido nucleico mencionadas anteriormente incluyen partículas de copolímero de etileno-acetato de vinilo, bombas osmóticas, sistemas de infusión implantables y liposomas. Las formulaciones para inhalación pueden contener excipientes, por ejemplo, lactosa o pueden ser soluciones acuosas que contengan, por ejemplo, polioxietileno-9-lauril éter, glicocolato y desoxicolato, o pueden ser soluciones oleosas para la administración en forma de gotas nasales o como un gel.

Como se ha mencionado anteriormente, además de los procedimientos de vacunación descritos anteriormente, que dan como resultado la activación de LTC *in vivo*, específicos de antígeno, restringidos por el CMH, dichas células (es decir, los LTC específicos de antígeno, restringidos por el CMH) pueden generarse *in vitro* y después administrarse a los pacientes. Puede usarse cualquier célula que exprese un gen que codifique un antígeno mayor de histocompatibilidad introducido endógenamente o exógenamente, para presentar un péptido de las proteínas de la invención o un péptido Colonia 1-6 para generar LTC específicos para las proteínas que se desvelan en el presente documento o LTC específicos de Colonia 1-6 *in vitro*. En una variación de este enfoque, una célula presentadora de péptido expresa un gen introducido endógenamente o exógenamente que codifica una proteína que se desvela en el presente documento o gen que codifica el polipéptido Colonia 1-6. La expresión de Colonia 1-6 endógeno en las células presentadoras de antígeno puede ser estimulada por citocinas, tales como IL-2, o por otras moléculas que son conocidos por los expertos en la materia para estimular la expresión de la proteína que se desvela en el presente documento o la expresión de Colonia 1-6. En otra variación, las células presentadoras de antígeno se pulsan con las proteínas que se desvelan en el presente documento o con péptidos Colonia 1-6 o de unión a CMH de las mismas y las células pulsadas después se usan para generar LTC para la administración a un paciente. Preferentemente, los LTC utilizados en estos procedimientos se obtienen del paciente al que se le van a administrar en última instancia (es decir, las células son autólogas). Como alternativa, pueden usarse en este procedimiento células donantes (es decir, las células alogenas). Por último, se incluyen en la invención procedimientos en los que se combinan cualquiera de los enfoques inmunoterápicos descritos anteriormente. Por ejemplo, un paciente puede ser tratado con un anticuerpo contra una proteína que se desvela en el presente documento o un anticuerpo Colonia 1-6, *y/o ex vivo*, los LTC activados con la proteína que se desvela en el presente documento o LTC activado por Colonia 1-6, *y/o un ex vivo*, CPA pulsado por proteína que se desvela en el presente documento o Colonia 1-6 (por ejemplo, una DC (célula dendrítica) o una célula B activada por CD40) y este tratamiento puede realizarse antes, durante o después de un enfoque de vacunación (véase anteriormente). Además de combinar los enfoques, cada enfoque (o una combinación de los mismos) puede emplear múltiples péptidos Colonia 1-6, péptidos de otros antígenos específicos de linfocitos T_{reg} o una combinación de los mismos.

Como se entiende en la técnica, un "polipéptido" es una cadena de aminoácidos unidos entre sí por enlaces peptídicos. Una "proteína" puede estar compuesta de uno o más polipéptidos, mientras que un "péptido" se entiende generalmente que es (o incluye) un fragmento de un polipéptido y consiste en una cadena de aminoácidos unidos por enlace peptídico que es más corta en longitud que un polipéptido de longitud completa del que puede derivar.

"Colonia 1-6" o "proteína Colonia 1-6" o "polipéptido Colonia 1-6", se refiere a un polipéptido de longitud completa, no fragmentado, de Colonia 1-6, mientras que un "péptido Colonia 1-6", es (o incluye) un fragmento de un polipéptido Colonia 1-6 de este tipo. Los péptidos Colonia 1-6 pueden tener cualquier longitud, hasta justo por debajo de la longitud completa de un polipéptido Colonia 1-6. Sin embargo, para su uso en el contexto de la presentación por el CMH de clase I o II, los péptidos Colonia 1-6 tienen una longitud relativamente corta, tal como, por ejemplo, ocho, nueve, diez, hasta veinte aminoácidos. Además, un péptido Colonia 1-6 puede incluir secuencias que no están presentes en un polipéptido Colonia 1-6 correspondiente, a condición de que el péptido Colonia 1-6 también incluya un tramo de al menos, por ejemplo, ocho, nueve, diez, hasta veinte aminoácidos consecutivos que

tengan una secuencia que sea idéntica a una secuencia de ocho, nueve, diez, hasta veinte aminoácidos consecutivos en un polipéptido Colonia 1-6.

Los péptidos que incluyen sustituciones de aminoácidos *también* pueden considerarse péptidos Colonia 1-6. Por ejemplo, un péptido Colonia 1-6 puede incluir una región de al menos nueve aminoácidos, de los cuales cualesquiera seis o más son idénticos a los aminoácidos dentro de un tramo de nueve aminoácidos en Colonia 1-6. Preferentemente, al menos siete, más preferentemente, al menos ocho, y, mucho más preferentemente, los nueve aminoácidos en una región de nueve aminoácidos del péptido Colonia 1-6 son idénticos a una región de nueve aminoácidos en Colonia 1-6.

Un polipéptido Colonia 1-6 correspondiente a Colonia 1-6 incluye la cantidad apropiada de aminoácidos que son sustancialmente idénticos (véase a continuación) a la secuencia de aminoácidos del gen Colonia 1-6 respectivo (N.º de referencia de Genbank: Colonia 1 NM_145235, Colonia 2 NM_145307, Colonia 3 NM_000860, Colonia 4 NM_015535, Colonia 5 NM_018166, Colonia 6 NM_207423) o un polipéptido de este tipo puede incluir la secuencia de aminoácidos de Colonia 1-6, así como secuencias adicionales.

Como se analiza adicionalmente a continuación, es de preferencia que los polipéptidos Colonia 1-6 que se desvelan en el presente documento incluyan regiones que puedan actuar como antígenos para la generación de anticuerpos policlonales o monoclonales, o puedan unirse a antígenos del complejo mayor de histocompatibilidad (CMH). Los péptidos correspondientes pueden identificarse usando procedimientos que se describen a continuación (véase también el documento PCT/US99/25438). Un péptido o polipéptido Colonia 1-6 puede fusionarse a secuencias de aminoácidos que no aparecen naturalmente en Colonia 1-6. Además, un péptido o polipéptido Colonia 1-6 puede fijarse a la superficie de una célula o a una molécula o a una macromolécula (por ejemplo, un antígeno de histocompatibilidad), o un péptido o polipéptido Colonia 1-6 puede conjugarse con inmunógenos o adyuvantes que son conocidos por los expertos en la materia, por ejemplo, hemocianina de lapa californiana (KLH, por sus siglas en inglés), con el fin de desencadenar una respuesta inmunitaria específica de Colonia 1-6, por ejemplo, la generación de anticuerpos policlonales o monoclonales. Por "molécula de ácido nucleico Colonia 1-6" se entiende una molécula de ARN o ADN (por ejemplo, ARNm) que codifica un polipéptido Colonia 1-6 o péptido Colonia 1-6, como se han definido anteriormente.

Por "linfocito T_{reg} que expresa Colonia 1-6" se entiende un linfocito T_{reg} como se ha definido anteriormente que expresa Colonia 1-6. Un linfocito T_{reg} que expresa Colonia 1-6 puede expresar un nivel de Colonia 1-6 que es igual a, o, preferentemente, mayor que el nivel de Colonia 1-6 expresado por estirpes celulares somáticas o germinales humanas normales. Los niveles de expresión de Colonia 1-6 en un linfocito T_{reg} que expresa Colonia 1-6 pueden aumentarse, por ejemplo, mediante la estimulación a través del receptor de linfocitos T y/o la coestimulación, por ejemplo, la estimulación de CD28, el aumento de la transcripción del gen Colonia 1-6, el aumento de la estabilidad o la traducción del ARNm de Colonia 1-6, el aumento de la estabilidad del polipéptido Colonia 1-6 o el aumento de la actividad enzimática Colonia 1-6. El aumento de este tipo de niveles de expresión de Colonia 1-6 puede ser útil en la invención para aumentar la probabilidad de que un linfocito T_{reg} sea reconocido como una diana de los procedimientos inmunoterápicos que se describen en el presente documento (véase a continuación).

Por "antígeno de histocompatibilidad" se entiende una molécula, tal como un complejo mayor de histocompatibilidad (CMH) de clase I, CMH clase II o antígeno de histocompatibilidad menor, que media interacciones de células del sistema inmunitario entre sí y con otros tipos celulares. Los ejemplos de antígenos de histocompatibilidad incluyen antígenos del CMH de clase I, tales como HLA-A (por ejemplo, A1, A2, A3, A11, A24, A31, A33 y A38), HLA-B y HLA-C, antígenos del CMH de clase II, tales como HLA-DR, HLA-DQ, HLA-DX, HLA-DO, HLA-DZ y HLA-DP y antígenos de histocompatibilidad menores, tales como HA1.

Por "generar LTC" se entiende un procedimiento *in vitro*, *in vivo* o *ex vivo* mediante el cual se activan (por ejemplo, se estimulan para crecer y dividirse) y/o se seleccionan LTC (por ejemplo, LTC específicos de Colonia 1-6).

Se dice que un péptido de Colonia 1-6 "se une específicamente" a un antígeno del CMH si el péptido se adhiere a un antígeno de histocompatibilidad en condiciones fisiológicas. Por ejemplo, dicha unión puede ser similar a la de un antígeno peptídico que se procesa y se presenta naturalmente en el contexto del CMH en una célula presentadora de antígeno.

Se dice que un anticuerpo o linfocito T citotóxico (LTC) "reconoce específicamente" un polipéptido que se desvela en el presente documento o un polipéptido Colonia 1-6 o un péptido Colonia 1-6 si se une al polipéptido o péptido, pero no se une sustancialmente a otros polipéptidos o péptidos no relacionados.

Se dice que un LTC "destruye específicamente" una célula si reconoce y lisa específicamente una célula que expresa un antígeno (por ejemplo, Colonia 1-6) para el que se ha activado, pero no reconoce ni lisa sustancialmente células que *no* expresan el antígeno. En el caso de Colonia 1-6, un LTC de este tipo se designa como "LTC específico de Colonia 1-6" en el presente documento.

Por "anticuerpo específico de Colonia 1-6" se entiende un anticuerpo que puede reconocer y unirse específicamente a un péptido o polipéptido Colonia 1-6, y que no reconoce ni se une sustancialmente a otras moléculas no relacionadas.

Un polipéptido Colonia 1-6 se "presenta" si se muestra un péptido de Colonia 1-6 en la superficie extracelular de una célula (por ejemplo, una célula presentadora de antígeno), de manera que pueda dar como resultado la generación *in vivo*, *ex vivo* o *in vitro* de LTC específicos de Colonia 1-6 o la lisis de un linfocito T_{reg} que expresa Colonia 1-6 por un LTC específico de Colonia 1-6. Preferentemente, el péptido Colonia 1-6 que se muestra se une a un antígeno de histocompatibilidad.

Por "condiciones fisiológicas" se entiende el entorno *in vivo* en el que se generan (se activan y/o se seleccionan) LTC específicos de Colonia 1-6 y realizan sus funciones biológicas (por ejemplo, el reconocimiento de un péptido Colonia 1-6 y la lisis restringida por el CMH de linfocitos T_{reg} que expresan Colonia 1-6), o un entorno *en vitro* o *ex vivo* que permite a los LTC específicos de Colonia 1-6 que se generen y realicen sus funciones biológicas.

Por "vacunación con Colonia 1-6" se entiende la administración de una preparación inmunógena que incluye uno o más péptidos Colonia 1-6, polipéptidos Colonia 1-6, moléculas de ácido nucleico Colonia 1-6, fragmentos de cualquiera de estas moléculas, células presentadoras de Colonia 1-6 (por ejemplo, células dendríticas o células B activadas por CD40) o mezclas de los mismos. La vacunación se realiza en un sujeto que tiene un aumento patológico de los niveles de linfocitos T_{reg}, tiene antecedentes de un aumento patológico de los niveles de linfocitos T_{reg}, es propenso a desarrollar un aumento patológico de los niveles de linfocitos T_{reg} o cualquier individuo sano para evitar un aumento patológico de los niveles de linfocitos T_{reg}, o en un sujeto en el que se han de generar células inmunitarias específicas de Colonia 1-6 (tales como LTC) para la transferencia a un paciente. Dicha vacunación estimula una respuesta inmunitaria específica de Colonia 1-6 dentro del sujeto. En sujetos que tienen un aumento patológico de los niveles de linfocitos T_{reg}, la vacunación puede dar como resultado el agotamiento parcial o completo de los linfocitos T_{reg}, a condición de que los linfocitos T_{reg} del paciente expresen Colonia 1-6. Además, la vacunación puede proporcionar profilaxis contra el desarrollo nuevos linfocitos T_{reg} que expresen Colonia 1-6.

A "vacuna", como se usa en el presente documento, es una composición inmunógena que puede administrarse en el procedimiento de vacunación descrito anteriormente. Por tanto, una vacuna incluye, por ejemplo, uno o más péptidos Colonia 1-6, polipéptidos Colonia 1-6, moléculas de ácido nucleico Colonia 1-6, fragmentos de cualquiera de estas moléculas, células presentadoras de Colonia 1-6 (por ejemplo, células dendríticas o células B activadas por CD40) o mezclas de los mismos. Opcionalmente, una composición de vacuna también puede incluir un adyuvante, que es una molécula que estimula una respuesta inmunitaria a un antígeno de vacuna coadministrado. Se proporcionan a continuación ejemplos de adyuvantes que pueden usarse en la invención. Una composición de vacuna también puede incluir otros antígenos asociados a linfocitos T_{reg} o péptidos de los mismos.

Por "célula inmunitaria" se entiende cualquier célula que desempeñe un papel en la inmunidad mediada por células o humoral, incluyendo los LTC y las células presentadoras de antígeno, por ejemplo, células B, linfocitos T auxiliares y células dendríticas.

Por "muestra" se entiende una biopsia de tumor o tejido, una biopsia de ganglio linfático, médula ósea, células, sangre, suero, orina, heces, esputo, saliva u otra muestra de ensayo obtenida de un paciente. Una muestra puede analizarse para determinar el nivel de linfocitos T_{reg} que expresan Colonia 1-6, LTC específicos de Colonia 1-6 o el nivel de cualquier otro indicador de respuesta inmunitaria (por ejemplo, una citocina) en el paciente del que se tomó mediante procedimientos que son conocidos en la técnica. Por ejemplo, puede utilizarse citometría de flujo para medir los niveles de linfocitos T_{reg} que expresan Colonia 1-6 y puede usarse ELISPOT para medir los niveles de citocinas. Además, pueden usarse ensayos de liberación de Cr⁵¹ (citotoxicidad de los linfocitos T) ensayos y ensayos que someten a ensayo la unión de los LTC a péptido Colonia 1-6 tetramérico/CMH, para medir los niveles de LTC específicos de Colonia 1-6.

Por "muestra de referencia" se entiende una muestra en la que se han medido el nivel de LTC específicos de Colonia 1-6 o el nivel de linfocitos T_{reg} que expresan Colonia 1-6 y para la se comparan cual el nivel de LTC específicos de Colonia 1-6 o el nivel de linfocitos T_{reg} que expresan Colonia 1-6 en la muestra de un sujeto de ensayo. Los niveles de referencia pueden ser mayores, menores o iguales a los niveles de la muestra del paciente. La comparación de una muestra de ensayo con una muestra de referencia proporciona una evaluación de los niveles de linfocitos T_{reg} que expresan Colonia 1-6 o la respuesta inmunitaria específica de Colonia 1-6 en el sujeto de ensayo.

Por "eliminación de linfocito T_{reg}" se entiende cualquier terapia (por ejemplo, quimioterapia, radioterapia, la administración de un LTC específico de Colonia 1-6, la administración de una CPA que presenta un péptido de Colonia 1-6 o la vacunación con Colonia 1-6, una molécula de ácido nucleico que codifica Colonia 1-6 o un fragmento del mismo, para potenciar una respuesta inmunitaria anti linfocito T_{reg}) administrada sola o en combinación con otras terapias, que influye en las frecuencias de linfocitos T_{reg} en al menos algunos pacientes a los que se administra el tratamiento. Por ejemplo, la eliminación de linfocitos T_{reg} puede reducir parcial o completamente o inhibir linfocitos T_{reg}. Además, la eliminación de linfocitos T_{reg} puede ser profiláctica, en que inhibe o impide el desarrollo de nuevos linfocitos T_{reg} en individuos sanos, en pacientes que están en remisión de cáncer, tienen cáncer metastásico o tienen un alto riesgo de desarrollar cáncer.

Por "inhibir el desarrollo de linfocitos T_{reg}" se entiende la administración de una terapia de protección (tal como LTC específicos de Colonia 1-6, CPA que presentan péptido Colonia 1-6 o una vacuna que incluye, por ejemplo, uno o

más péptidos Colonia 1-6, polipéptidos Colonia 1-6 o moléculas de ácido nucleico Colonia 1-6 o una combinación de los mismos) a un sujeto que se considera que tiene un riesgo mayor que el promedio de desarrollar altas frecuencias de linfocitos T_{reg}.

5 Por "vehículo farmacéuticamente aceptable" se entiende un vehículo que es fisiológicamente aceptable a un paciente, al tiempo que conserva las propiedades terapéuticas del compuesto con el que se administra. Un vehículo farmacéuticamente aceptable de ejemplo es la solución salina fisiológica. Otros vehículos fisiológicamente aceptables y sus formulaciones son conocidos por los expertos en la materia y se describen, por ejemplo, en *Remington's Pharmaceutical Sciences* (18ª edición), ed. A. Gennaro, 1990, Mack Publishing Company, Easton, PA.

10 La expresión "sustancialmente idéntica" se usa en el presente documento para describir un polipéptido o molécula de ácido nucleico que presenta al menos el 50 %, preferentemente al menos el 85 %, más preferentemente al menos el 90 % y mucho más preferentemente al menos el 95 % de identidad con un ácido amino de referencia o secuencia de ácido nucleico. Para los polipéptidos, la longitud de las secuencias de comparación es de al menos 8 aminoácidos, preferentemente al menos 16 aminoácidos, más preferentemente al menos 25 aminoácidos y mucho más preferentemente 35 aminoácidos. Para las moléculas de ácidos nucleicos, la longitud de las secuencias de comparación es de al menos 24 nucleótidos, preferentemente al menos 50 nucleótidos, más preferentemente al menos 75 nucleótidos y mucho más preferentemente al menos 110 nucleótidos. La identidad de secuencia normalmente se mide usando un software de análisis de secuencia con los parámetros por defecto especificados en el mismo (por ejemplo, el Paquete de Software de Análisis de Secuencias del Genetics Computer Group, Centro de Biotecnología de la Universidad de Wisconsin, 17 10, University Avenue, Madison, WI 5 3 705). Los polipéptidos, péptidos y moléculas de ácido nucleico Colonia 1-6 que se desvelan en el presente documento pueden ser idénticos o sustancialmente idénticos a moléculas de origen natural y, por tanto, pueden o no incluir secuencias de tipo no silvestre. Por "péptido sustancialmente puro" o "polipéptido sustancialmente puro" se entiende un péptido, polipéptido o un fragmento de los mismos, que se han separado de los componentes que los acompañan de forma natural. Normalmente, el péptido o polipéptido es sustancialmente puro cuando está al menos al 60 %, en peso, libre de las proteínas y moléculas orgánicas de origen natural con las que está asociado de forma natural. Preferentemente, el péptido o polipéptido es un péptido o polipéptido Colonia 1-6 que es al menos del 75 %, más preferentemente al menos el 90 % y mucho más preferentemente al menos el 99 %, en peso, puro. Se puede obtener un péptido o polipéptido Colonia 1-6 sustancialmente puro, por ejemplo, mediante la extracción de una fuente natural (por ejemplo, un linfocito T_{reg}), mediante la expresión de una molécula de ácido nucleico recombinante que codifica un péptido o polipéptido Colonia 1-6, o mediante la síntesis química del péptido o polipéptido. La pureza puede medirse mediante cualquier procedimiento apropiado, por ejemplo, mediante cromatografía en columna, electroforesis en gel de poliacrilamida o análisis por HPLC.

35 Una proteína está sustancialmente libre de componentes asociados de forma natural cuando está separada de esos contaminantes que la acompañan en su estado natural. Por tanto, una proteína que se sintetiza químicamente o se produce químicamente en un sistema celular diferente de la célula a partir de la que se origina de forma natural está sustancialmente libre de sus componentes asociados de forma natural. En consecuencia, los péptidos y polipéptidos sustancialmente puros no solo incluyen los derivados de organismos eucarióticos, sino también los sintetizados en *E. coli* u otros procariontes.

40 Por "ADN sustancialmente puro" o "ADN aislado" se entiende ADN que está libre de los genes que, en el genoma de origen natural del organismo del que deriva el DNA, flanquean el gen. Por tanto, el término incluye, por ejemplo, un ADN recombinante que se incorpora en un vector; un plásmido o virus de replicación autónoma; o el ADN genómico de un procarionte o eucariota; o ADN que existe como una molécula separada (por ejemplo, un ADNc o un fragmento genómico o de ADNc producido mediante PCR o digestión con endonucleasas de restricción) independiente de otras secuencias. También incluye un ADN recombinante que es parte de una secuencia polipeptídica adicional que codifica un gen híbrido.

45 Por "transformación", "transfección" o "transducción" se entiende cualquier procedimiento para introducir moléculas extrañas en una célula. Lipofección, transfección mediada por DEAE-dextrano, microinyección, fusión de protoplastos, precipitación con fosfato de calcio, transducción (por ejemplo, entrega por bacteriófago, adenovírica, retrovírica, lentivírica u otra), electroporación y transformación biolística son solo algunos de los procedimientos conocidos para los expertos en la materia que pueden usarse en la invención.

50 Por "célula transformada", "célula transfectada" o "célula transducida", se entiende una célula (o un descendiente de una célula) dentro de la cual se ha introducido una molécula de ácido nucleico (por ejemplo, una molécula de ADN o ARN) que codifica un polipéptido que se desvela en el presente documento por medio de técnicas de ADN recombinante.

55 Por "promotor" se entiende una secuencia mínima suficiente para dirigir la transcripción. Los elementos promotores que son suficientes para volver la expresión génica dependiente de promotor controlable, específica de tipo celular, específica de tejido, específica del tiempo o inducible por señales o agentes externos, también puede usarse en la invención; dichos elementos pueden ubicarse en las regiones de secuencias extremos 5' o 3' o intrón del gen nativo.

Por "unido operativamente" se entiende que un gen y una o más secuencias reguladoras están conectadas de una

manera que permite la expresión génica cuando las moléculas apropiadas (por ejemplo, proteínas activadoras transcripcionales) están unidas a las secuencias reguladoras.

5 Por "vector de expresión" se entiende un plásmido o virus modificado mediante ingeniería genética, derivado de, por ejemplo, un bacteriófago, adenovirus, retrovirus, lentivirus, poxvirus, herpesvirus o cromosoma artificial, que se usa para transferir una secuencia codificante de péptido o polipéptido (por ejemplo, una secuencia codificante de péptido Colonia 1-6), unido operativamente a un promotor, en una célula hospedadora, de manera que el péptido o polipéptido codificado se exprese dentro de la célula hospedadora.

Los ejemplos no cubiertos por el ámbito de las reivindicaciones son para fines ilustrativos.

Ejemplos

10 Material y procedimientos

Pacientes y parámetros clínicos: Para el aislamiento de linfocitos T_{reg} CD4⁺ CD25^{alto} y linfocitos T CD4⁺ CD25⁻ convencionales tras la aprobación por la junta de revisión institucional, se obtuvo sangre periférica de 4 individuos sanos y de 4 pacientes con leucemia linfática crónica (LLC) después del consentimiento informado. Los pacientes incluidos para el análisis fenotípico o funcional no se trataron antes de la investigación. La estadificación se realizó de acuerdo con la clasificación Binet para la LLC.

15 Para el aislamiento de linfocitos T_{reg} CD4⁺ CD127^{bajo} CD25⁺ FOXP3⁺ y linfocitos T CD4⁺ CD127⁺ CD25^{-bajo/int} FOXP3 se obtuvo sangre periférica de 4 individuos sanos después del consentimiento informado y la aprobación por la junta de revisión institucional.

20 Aislamiento de CMSP de donantes sanos y pacientes con LLC: Se obtuvieron células mononucleares de sangre periférica (CMSP) utilizando centrifugación por densidad Ficoll/Hypaque (Amersham, Uppsala, Suecia). Por tanto, se diluyeron muestras de sangre heparinizadas 1:1 con RPMI y se estratificaron sobre 15 ml de Ficoll/Hypaque. Después de la centrifugación durante 30 min a 450 g, se recogió la interfase, se lavó dos veces con RPMI y se crioconservó en la fase gaseosa de un tanque de nitrógeno líquido en DMSO al 10 % y FCS al 90 % hasta el procesamiento adicional.

25 Aislamiento de linfocitos T_{reg} CD25⁺CD4^{alto} y linfocitos T convencionales CD4⁺CD25⁻: En resumen, después de lavar dos veces con RPMI, se usaron perlas MACS® CD4 (Miltenyi Biotec) para el aislamiento de linfocitos T CD4⁺ de CMSP de acuerdo con las recomendaciones del fabricante. Después de la tinción con CD25-PE y CD4-APC (ambos de BD) de acuerdo con las recomendaciones del fabricante, los linfocitos T CD4⁺ CD25⁻ y CD4⁺ CD25^{alto} se purificaron usando un clasificador de células FACSDiVa™ (BD Biosciences). La pureza de la población de linfocitos T_{reg} CD4⁺ CD25^{alto} se comprobó de forma habitual y dio como resultado > 90 % de linfocitos T_{reg} CD4⁺ CD25^{alto} después de la purificación. Después del aislamiento de las células se lisaron directamente en el reactivo TRIzol (Invitrogen Life Technologies) y se almacenaron a -80 °C hasta su procesamiento posterior o se lisaron después de una estimulación adicional durante 20 horas.

30 Estimulación policlonal de linfocitos T_{reg} CD4⁺ CD25^{alto} y linfocitos T convencionales CD4⁺ CD25⁻ con CD3 e interleucina 2: Para evaluar el efecto de la estimulación policlonal a corto plazo sobre la expresión génica de linfocitos T_{reg} CD4⁺ CD25^{alto} y linfocitos T convencionales CD4⁺ CD25⁻, 1x10⁶ células de la respectiva población de linfocitos T se activaron en X-Vivo 15 (BioWhittaker) con anti-CD3 (0,5µg/ml, OKT-3) e IL-2 (10 UI/ml, Proleukin, Chiron) durante 20 horas.

35 Preparación de ARN, hibridación de micromatrices y tratamiento de datos de micromatrices: el aislamiento y la cuantificación del ARN se realizó de acuerdo con las recomendaciones del fabricante (Qiagen). La preparación de ARNc marcado con biotina se realizó usando el kit de amplificación de ARN Ambion Illumina (Ambion Europe, Huntington, Cambridgeshire, Reino Unido) de acuerdo con las recomendaciones del fabricante. Se hibridaron 1,5 µg de ARNc marcado con biotina a chips de perlas del genoma completo Sentrix 6 x 2 (Illumina, San Diego, CA, EE.UU.) de acuerdo con las recomendaciones del fabricante y se exploraron sobre la Illumina BeadStation 500x. Para la recolección de datos, la evaluación y el análisis estadístico se usó el software Illumina BeadStudio y dCHIP 1.3. Los siguientes criterios de filtrado se usaron para la selección de genes expresados diferencialmente: factor de cambio ≥ 2, diferencia absoluta en la intensidad de la señal entre la media del grupo ≥ 50 y valor de p < 0,05 (ensayo t pareado).

40 Aislamiento de linfocitos T_{reg} CD4⁺ CD127^{bajo} CD25⁺ y linfocitos T CD4⁺ CD127⁺ CD25^{bajo} y linfocitos T CD4⁺ CD127⁺ CD25^{int}, preparación de ARN, hibridación de micromatrices y procesamiento de datos de micromatrices: De acuerdo con el análisis de los inventores, así como dos informes publicados en 2006 CD127, la cadena α del receptor de IL-7, es la molécula de superficie más específica regulada negativamente en los linfocitos T_{reg} CD4⁺ CD25⁺. La expresión de CD127 se correlaciona inversamente con la expresión del factor de transcripción específico de linfocitos T_{reg} FOXP. Por tanto, los inventores querían confirmar la expresión de los genes expresados diferencialmente en la población de linfocitos T_{reg} CD4⁺ CD127^{bajo} CD25⁺ ya que esta población debería enriquecerse adicionalmente en linfocitos T_{reg} que la población de linfocitos T_{reg} CD4⁺ CD25⁺.

Después de la descongelación, los linfocitos T CD4⁺ se purificaron usando el sistema de separación de células BD IMag™ (BD Biosciences) con partículas de CD4 de acuerdo con las recomendaciones del fabricante. Después de la tinción con CD127-Alexa 647, CD25-PE y CD4-PerCP-Cy5.5 (ambos de BD) de acuerdo con las recomendaciones del fabricante, los linfocitos T_{reg} CD4⁺ CD127^{bajo} CD25⁺, los linfocitos T CD4⁺ CD127⁺ CD25^{bajo} convencionales y los linfocitos T CD4⁺ CD127⁺ CD25^{int} activados se purificaron usando un clasificador de células FACSDiVa™ (BD Biosciences). La pureza de la población de linfocitos T_{reg} CD4⁺ CD127^{bajo} CD25⁺ FOXP3⁺ medida por tinción FOXP3 Alexa 488 se comprobó como habitualmente y dio como resultado > 95 % de linfocitos T_{reg} CD4⁺CD25^{alto} después de la purificación. El aislamiento y la cuantificación de ARN, la preparación de ARNc marcado con biotina, la hibridación, la exploración y el análisis estadístico se realizaron como se ha descrito anteriormente. Los criterios utilizados para la selección de genes expresados diferencialmente se definieron como ya se ha mencionado.

Ejemplo 1: Identificación de genes núcleo T_{reg}.

El objetivo de este estudio era identificar los genes que muestran una expresión única en linfocitos T reguladores (T_{reg}) en comparación con otros linfocitos T CD4⁺ convencionales que se identificaron como se indica a continuación y se designaron el "transcriptoma núcleo" de los T_{reg}. Se realizó un total de 192 experimentos individuales que interrogaban los linfocitos T convencionales y T_{reg} en diferentes estados de activación. Los linfocitos T CD4⁺ se aislaron de sangre periférica y se separaron posteriormente en linfocitos CD25⁺ y CD25⁻. Los linfocitos CD25⁻ se dejaron sin estimular (en reposo) o se expusieron a diferentes estímulos. Esto incluía la activación a través de anticuerpos αCD3 y αCD28 con o sin adición de señales inhibitoras (TGFβ, CTLA4, PGE2, IL10 y PD1). Las células CD25⁺ se dejaron sin estimular, se expusieron a la activación por IL2 o se expandieron usando rapamicina. La figura 1 muestra el flujo de trabajo completo de todos los experimentos realizados. Todas las muestras están presentes en repeticiones con al menos n = 3.

Para la exploración de genes candidatos potenciales se usó un enfoque de reconocimiento de patrones. En primer lugar, se generó un gráfico de contorno virtual para cada gen sobre todo el conjunto de datos. Después, se determinó la correlación del perfil de expresión de cada gen y el perfil de expresión del control positivo FOXP3 (Figura 2). Los genes con correlación alta positiva con respecto a la correlación negativa para FOXP3 se seleccionaron como genes asociados al transcriptoma núcleo. 43 genes cumplieron con los criterios de reconocimiento de patrones y se identificaron como el transcriptoma núcleo de los linfocitos T_{reg} (Tabla 1). La Figura 3 visualiza estos 43 genes usando una representación de contorno tridimensional.

Tabla 1: 43 genes núcleo de T_{reg}.

Símbolo (SEQ ID NO) of ARNm	T _{reg}	Linfocitos T en reposo	Linfocitos T CD25 ⁻	Linfocitos T activados
ACTA2 (15)	740+/- 75	414 +/- 96	333 +/- 106	528 +/-155
BFSP2 (17)	335 +/- 125	155 +/- 15	150 +/- 31	143 +/-16
C1orf78 (11)	644 +/- 108	327 +/-66	289 +/- 68	295 +/-94
CCL5 (19)	678 +/- 433	4998 +/- 772	3404 +/- 2178	3274 +/-1774
CCR7 (21)	1491 +/- 561	4226 +/-1310	3059 +/-1012	3809 +/-1320
CD40LG (23)	197 +/- 62	568 +/- 90	577 +/- 96	616 +/- 296
CTLA4 (25, 27)	1061 +/- 313	316 +/- 165	306 +/- 84	452 +/- 440
CTSL (29, 31)	100 +/- 8	422 +/- 259	202 +/- 61	273 +/- 121
EOMES (33)	117 +/- 25	352 +/- 84	424 +/- 335	337 +/- 197
EPB41L3 (35)	89 +/- 7	212 +/-67	133 +/- 32	121 +/- 13
FAM129A (37)	1294 +/- 154	719 +/- 204	555 +/- 180	735 +/- 357
FANK1 (1)	562 +/- 118	116 +/- 21	93 +/- 2	98 +/- 15
FCGBP (39)	222 +/- 33	519 +/- 192	392 +/-73	604 +/- 350
FHIT (41)	196 +/- 40	324 +/- 46	361 +/- 91	294 +/- 96
FLOT1 (43)	543 +/- 57	1170 +/- 213	1169 +/- 156	994 +/-222
FOXP3 (45)	260 +/- 113	103 +/- 6	88 +/- 5	104 +/- 19
GNLY (47, 49)	189 +/- 45	3218 +/- 902	671 +/- 694	978 +/- 616
HPGD (7)	265 +/- 76	133 +/- 17	109 +/- 14	106 +/- 17
ICA1 (51, 53)	378 +/- 143	140 +/- 11	191 +/- 32	117 +/- 10

(continuación)

Símbolo (SEQ ID NO) of ARNm	T _{reg}	Linfocitos T en reposo	Linfocitos T CD25-	Linfocitos T activados
IL10RA (57)	1068 +/- 282	533 +/- 53	605 +/- 140	432 +/- 115
IL7R (55)	966 +/- 424	1852 +/- 762	2497 +/- 811	1766 +/- 1188
LASS6 (61)	145 +/- 54	396 +/- 39	373 +/- 91	363 +/- 94
MGST2 (65)	251 +/- 47	163 +/- 26	140 +/- 33	121 +/- 25
PTTG3 (77)	229 +/- 35	128 +/- 16	143 +/- 23	134 +/- 36
KIAA1600 (59)	782 +/- 148	544 +/- 70	452 +/- 56	445 +/- 67
C6orf190 (735)	186 +/- 57	425 +/- 90	514 +/- 66	416 +/- 162
LOC339541 (63)	205 +/- 55	422 +/- 66	508 +/- 133	398 +/- 110
TARP (103, 105)	191 +/- 54	640 +/- 94	638 +/- 393	450 +/- 225
RBMS1 (81, 83, 85)	354 +/- 34	723 +/- 225	744 +/- 77	545 +/- 65
NCF4 (67, 69)	1184 +/- 285	640 +/- 110	433 +/- 135	344 +/- 1067
NELL2 (71)	251 +/- 148	1707 +/- 560	877 +/- 152	1164 +/- 565
PLEKHK1 (3, 5)	653 +/- 109	166 +/- 31	142 +/- 19	151 +/- 18
PTPLA (73)	317 +/- 64	136 +/- 17	123 +/- 6	115 +/- 25
PTTG1 (75)	299 +/- 89	172 +/- 29	173 +/- 25	147 +/- 39
RAB6IP1 (79)	127 +/- 21	280 +/- 46	306 +/- 63	238.83 +/- 54
RHOA (87)	107 +/- 6	178 +/- 28	217 +/- 61	146 +/- 28
SATB1 (89)	787 +/- 112	1988 +/- 262	1739 +/- 439	2028 +/- 687
SELP (91)	461 +/- 169	146 +/- 9	144 +/- 24	149 +/- 21
SEMA3G (93)	194 +/- 64	105 +/- 6	109 +/- 10	105 +/- 3
SHMT2 (95)	1741 +/- 310	797 +/- 126	724 +/- 187	1275 +/- 812
STAM (97)	637 +/- 194	264 +/- 25	207 +/- 35	217 +/- 30
STOM (99, 101)	345 +/- 54	783 +/- 60	920 +/- 278	692 +/- 258
TCF7 (107, 109, 111, 113)	8270 +/- 3071	16713 +/- 1197	17162 +/- 3138	11895 +/- 5380
TNFRSF1B (115)	1950 +/- 360	861 +/- 145	832 +/- 216	826 +/- 265
TRIM16 (117)	433 +/- 125	253 +/- 18	286 +/- 32	247 +/- 24
UTS2 (119, 121)	263 +/- 84	162 +/- 28	149 +/- 21	130 +/- 4

Se muestran valores de expresión medios +/- desviación típica para los grupos de muestras distintas

Ejemplo 2: Análisis de 6 genes marcadores de Treg potenciales

5 **A. FANK1:** FANK1 humano (dominios de repetición 1 de fibronectina tipo III y anquirina; Colonia 1) se ubica en la región genómica 127575098-127688151 en el cromosoma 10 en la región 10q26.2. Consiste en 11 exones resultantes de la transcripción en un ARNm que tiene 1395 pares de bases (véase la SEQ ID NO: 1). La secuencia codificante en sí consiste en 1038 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 346 aminoácidos (SEQ ID NO: 2).

10 **B. PLEKHK1:** PLEKHK1 humana (que contiene dominio de homología a pleckstrina, familia K miembro 1, también conocida como rhotecina2; Colonia 2) se ubica en la región genómica 63698472-63622959 en el cromosoma 10 en la región 10q21.2. Consiste en 12 exones resultantes de la transcripción en un ARNm que tiene 6659 pares de bases (véase la SEQ ID NO: 3, mostrándose una variante adicional de la misma que tiene 2123 pares de bases en la SEQ ID NO: 5). La secuencia codificante en sí consiste en 1830 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 609 aminoácidos (SEQ ID NO: 4).

15 **C. HPGD:** HPGD humana (hidroxiprostaglandina deshidrogenasa 15-(NAD); Colonia 3) se ubica en la región

genómica 175680186-175647955 en el cromosoma 4 en la región 4q34-35. Consiste en 7 exones resultantes de la transcripción en un ARNm que tiene 2592 pares de bases. La secuencia codificante en sí consiste en 801 pares de bases (véase la SEQ ID NO: 7). La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 266 aminoácidos (véase la SEQ ID NO: 8).

5 D. DNATP6: DNATP6 humana (proteína 6 transactivada por ADN polimerasa; Colonia 4) se ubica en la región genómica 200879041-201051498 en el cromosoma 2 en la región 2q33.1. Consiste en 13 exones resultantes de la transcripción en un ARNm que tiene 2355 pares de bases (véase la SEQ ID NO: 9). La secuencia codificante en sí consiste en 1677 pares de bases. La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 558 aminoácidos.

10 E. C1orf78: C1orf78 humano (marco de lectura abierto 78 del cromosoma 1; Colonia 5) se ubica en la región genómica 36562342-36560219 en el cromosoma 1 en la región 1p34.3. Consiste en 3 exones resultantes de la transcripción en un ARNm que tiene 1054 pares de bases. La secuencia codificante en sí consiste en 495 pares de bases (véase la SEQ ID NO: 11). La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 164 aminoácidos (SEQ ID NO: 12).

15 F. FLJ45983: FLJ45983 humana (proteína FLJ45983; Colonia 6) se ubica en la región genómica 8135453-8132419 en el cromosoma 10 en la región 10p14. Consiste en 2 exones resultantes de la transcripción en un ARNm que tiene 2213 pares de bases. La secuencia codificante en sí consiste en 378 pares de bases (véase la SEQ ID NO: 13). La traducción de esta secuencia da como resultado una proteína supuesta con una longitud de 125 aminoácidos (SEQ ID NO: 14).

20 **Ejemplo 3: Clonación de los 6 genes candidatos en vectores de expresión bacterianos y de mamíferos**

A. Amplificación de ADNc: Se pidieron clones de ADNc de longitud completa para los 6 genes candidatos de RZPD (Deutsches Ressourcenzentrum für Genomforschung GmbH, Heubnerweg 6, D-14059 Berlin, Alemania). En primer lugar, todos los clones se sembraron en placas de agar para seleccionar un solo clon, que después se hizo crecer en cultivos durante la noche para amplificar el vector. Se aisló el ADN usando el kit de extracción de gel QIAquick de Qiagen. Usando cebadores específicos el ADNc se amplificó para la clonación adicional.

Tabla 2: Cebadores para la clonación de Col1-6

Gen	Cebador	SEQ ID NO	Secuencia	Tamaño del producto (pb)
Col1	FANK1 PCR dir	123	ATGGAGCCCCAGAGAATCAT	
	FANK1 PCR inv	124	GCAGACACAAGACTTCTTTGG	1035
	FANK1 PCR inv parada	125	TCAGCAGACACAAGACTTCTT	1038
Col2	PLEKHK1 PCR dir	126	ATGGAGGGGCCGAGCCT	
	PLEKHK1 PCR inv	127	TACTTGTGCCTGCAGCCAT	1827
	PLEKHK1 PCR inv parada	128	CTATACTTGTGCCTGCAGCC	1830
Col3	HPGD PCR dir	129	ATGCACGTGAACGGCAAAGT	
	HPGD PCR inv	130	TTGGGTTTTTGTCTTGAAATGGA	798
	HPGD PCR inv parada	131	TCATTGGGTTTTTGTCTTGAAATG	801
Col4	DNATP6 PCR dir	132	ATGGCTGAACTCAATACTCATG	
	DNATP6 PCR inv	133	GGCCACCAACGTCACAGCCG	1674
	DNATP6 PCR inv parada	134	TCAGGCCACCAACGTCACAG	1677
Col5	Col5 PCR dir	135	ATGGATGCCCCGCGAAGGGA	
	Col5 PCR inv	136	GTAATAGTGCATGCGGCCAGG	495
	Col5 PCR inv parada	137	TCAGTAATAGTGCATGCGGC	498
Col6	Col6 PCR dir	138	ATGGAGCCGGACTTTCTCC	
	Col6 PCR inv	139	GGACTTCGACCCCGGGGCTC	375
	Col6 PCR inv parada	140	TCAGGACTTCGACCCCGGG	378

Para la clonación en el vector de expresión bacteriano pTr Chis-TOPO (Invitrogen) se amplificó ADNc con codón de parada mientras que el ADNc se amplificó sin codón de parada para la clonación en el vector de expresión Lentivírico pLenti6/V5-DEST (Invitrogen).

B. Etapas de clonación: La expresión *in vitro* de estas 6 proteínas codificadas por ADNc clonado es posible usando el vector de expresión bacteriana (pTrcHis-TOPO; Fig. 11A). La expresión de estos genes en la bacteria permite la producción de grandes cantidades de proteínas para aislamiento y purificación. La proteína purificada puede usarse para la producción de anticuerpos específico (véase el Ejemplo 4). Se amplificó el ADNc de los 6 genes mediante PCR incluyendo el codón de parada, se purificó en gel (kit de extracción Jetsorb Gel, Genomed GmbH, Löhne) y después se clonó en el vector (Figuras 11 B-G). El vector se transformó en *E. coli*, se seleccionaron colonias y se analizaron para determinar transformantes positivos mediante análisis de restricción. El ADNc se verificó mediante análisis de secuenciación. Una vez que se construyeron los vectores de expresión apropiados que contenían los 6 genes, los transformantes con el inserto de ADNc correcto se cultivaron durante la noche en medio LB. La expresión de proteínas se indujo por adición de isopropil- β -D-tiogalactopiranosido (IPTG) y después se purificaron ya sea usando el sistema de purificación ProBond™ (Invitrogen) o el sistema QIAexpress®Ni-NTA Fast Start (Qiagen) para las proteínas marcadas con 6xHis. El aislamiento de las proteínas se demostró mediante transferencia de Western.

Un vector de expresión lentivírica (pLenti6/V5-Dest) se usó para transducir células de mamífero con los 6 genes. Por ejemplo, en células Jurkat, una estirpe celular de leucemia de linfocitos T aguda, la localización y función de las 6 proteínas pueden analizarse. Por tanto, el ADNc de los 6 genes se amplificó mediante PCR sin el codón de parada, se purificó en gel (kit de extracción de gel Jetsorb, Genomed GmbH, Löhne) y después se clonaron en el vector pTrcHis-TOPO como se ha descrito anteriormente. A partir de este vector el ADNc se clonó en el Vector de Entrada pENTR4™ (Figuras 13A y 13 BG) mediante clonación mediada por sitio de restricción. Este vector se modifica e incluye una molécula de eGFP de modo que se crea una proteína de fusión marcada con GFP a partir de los clones de ADNc. Las colonias se analizaron de nuevo y se confirmaron mediante análisis de restricción. Los transformantes con el inserto correcto después se clonaron en el vector pLenti6/V5-Dest usando la tecnología de puerta de Invitrogen.

El vector de expresión de mamífero (pcDNA6-HisA) se usó para expresar las 6 proteínas en células de mamífero. El ADNc de los 6 genes se amplificó por PCR incluyendo el codón de parada, se purificó en gel (kit de extracción de Gel Jetsorb, Genomed GmbH, Löhne) y después se clonó en el vector pTrcHis-TOPO como se ha descrito anteriormente (Figuras 11A y 11 BG). A partir de este vector el ADNc se clonó en el vector pcDNA6-HisA mediante clonación mediada por sitio de restricción (Figuras 15A y 15 BG). Los transformantes se analizaron mediante análisis de restricción y secuenciación.

C. Estirpes Jurkat estables: Se infectaron células Jurkat usando vectores lentivíricos (pLenti6/V5-DEST) que contenían los 6 genes candidatos marcadores acoplados a GFP. Las células se diluyeron dos días después de la infección para establecer clones estables que se seleccionaron usando blasticidina (10 μ g/ml). Cuando se recibieron estirpes celulares estables, se analizaron para determinar la expresión de GFP para verificar la expresión de la proteína de fusión integrada con el análisis por FACS e inmunofluorescencia. Posteriormente, se realizaron estudios de colocalización para definir la localización de las diferentes proteínas mediante inmunofluorescencia, así como inmunoprecipitación para caracterizar los compañeros de interacción de los 6 genes.

D. Análisis funcional de los 6 genes en linfocitos T CD4⁺ primarios humanos: Después de conseguir la primera información sobre la función y ubicación de los 6 genes en las células Jurkat, pueden usarse varios procedimientos para determinar su papel en los linfocitos T CD4⁺ primarios humanos. Con los anticuerpos específicos conseguidos (véase la parte de producción de anticuerpos) los inventores quisieron verificar si la ubicación y los compañeros de unión que encontraron en las células Jurkat infectadas eran los mismos en linfocitos T_{reg} humanos primarios. Después, planificaron determinar la función de los 6 genes. Un procedimiento es inactivar los 6 genes en linfocitos T_{reg} humanos usando ARNip. En esta etapa, los inventores usaron ARNip contra la expresión génica, la transcripción génica, la traducción génica, así como contra las proteínas de los 6 candidatos marcadores. De este modo, el potencial inhibidor de los linfocitos T_{reg} después de la delección de los genes individuales puede someterse a ensayo, el perfil de citocinas, así como el fenotipo de los linfocitos T_{reg}.

La segunda estrategia era infectar linfocitos T CD4⁺CD25⁻ convencionales con los vectores de expresión de mamífero (ya sea pcDNA6-HisA o pLenti6/V5-DEST) de los 6 genes y analizar, si la expresión de estos genes conduce al desarrollo de un fenotipo regulador y/o función.

Ejemplo 4: Generación de anticuerpos

Un procedimiento para generar anticuerpos contra Colonia 1-6 implica la administración de una célula presentadora de antígeno (CPA) a animales, por ejemplo, ratón, rata, conejo, cabra. Esto da como resultado la activación de las células B para que produzcan anticuerpos que reconozcan linfocitos T_{reg} de una manera específica de Colonia 1-6. La CPA puede pulsarse con Colonia 1-6 o un péptido de Colonia 1-6 que se une a una molécula del complejo mayor de histocompatibilidad. Otro procedimiento incluye la generación de anticuerpos contra Colonia 1-6 mediante la administración de Colonia 1-6 o un péptido de Colonia 1-6 que se une a una molécula del complejo mayor de histocompatibilidad, que se procesa mediante una célula presentadora de antígeno, que, a su vez, activa las células B para que produzcan anticuerpos que reconozcan linfocitos T_{reg} de un modo específico de Colonia 1-6. El polipéptido Colonia 1-6 o péptido de Colonia 1-6 utilizado en este procedimiento pueden administrarse en asociación con un adyuvante.

Como alternativa, un procedimiento implica la administración de una molécula de ácido nucleico que codifica Colonia 1-6 o un péptido de Colonia 1-6 que se une a una molécula del complejo mayor de histocompatibilidad. La molécula de ácido nucleico se expresa de forma que pueda ser procesada por una célula presentadora de antígeno, que activa células B para producir anticuerpos que reconocen Colonia 1-6 de una manera específica de Colonia 1-6. La

molécula de ácido nucleico que codifica Colonia 1-6 o un péptido de Colonia 1-6 puede estar presentes en un vector de expresión.

Después de que un animal se hubiera expuesto varias veces con Colonia 1-6 se fusionaron células B del bazo o ganglios linfáticos con células tumorales de mieloma que pueden crecer indefinidamente en cultivo y que han perdido la capacidad de producir anticuerpos. Esta fusión se realizó haciendo las membranas celulares más permeables mediante el uso de polietilenglicol o electroporación. Las células de hibridomas fusionadas estaban suficientemente diluidas para garantizar la clonalidad y el crecimiento. Después, los anticuerpos de los diferentes clones se sometieron a ensayo para determinar su capacidad para unirse al antígeno (por ejemplo, con un ensayo tal como ELISA) o inmunotransferencia puntual y se seleccionó la más sensible. Después, se produjeron anticuerpos monoclonales en cultivo celular mediante, por ejemplo, cámaras de fermentación.

Otro procedimiento de generación de anticuerpos contra Colonia 1-6 implicó el uso de Colonia 1-6 o un péptido de Colonia 1-6 para unirse a anticuerpos expresados por una biblioteca de fagos. Se expresaron numerosos anticuerpos en la biblioteca como fusiones con la proteína de cubierta de un bacteriófago, de manera que se presentaron en la superficie de la partícula vírica. El ADN extraído de fagos interactuantes contenía las secuencias de los anticuerpos específicos que reconocían Colonia 1-6 de una manera específica de Colonia 1-6.

Lista de secuencias - texto libre

Designación de SEQ ID

- 1/2 >gi|34222185|ref|NM_145235.2| Dominios de repetición 1 de fibronectina tipo III y anquirina de Homo sapiens (FANK1; Col1), ARNm y SAC
- 3/4 >gi|26190613|ref|NM|145307.2| Familia K que contiene dominio de homología a pleckstrina, miembro 1, de Homo sapiens (PLEKHK1 variante larga; Col2), ARNm y SAC
- 5/6 >gi_19343626_gb_BC025765.1| Familia K que contiene dominio de homología a pleckstrina, miembro 1, de Homo sapiens (PLEKHK1 variante corta), ARNm y SAC
- 7/8 >gi_40226515_gb_BC018986.2_| Hidroxiprostaglandina deshidrogenasa 15-(NAD) de Homo sapiens (HPGD; Col3), ARNm y SAC
- 9/10 >gi_7661597_ref_NM_015535.1| Proteína 6 transactivada con ADN polimerasa de Homo sapiens (DNAPT6; Col4), ARNm y SAC
- 11/12 >gi_8922566|ref_NM_018166.1| Marco de lectura abierto 78 del cromosoma 1 de Homo sapiens (Clorf78; Col5), ARNm y SAC
- 13/14 >gi_46409471_ref_NM_207423.1| Proteína FLJ45983 de Homo sapiens (FLJ45983; Col6), ARNm y SAC
- 15/16 >gi|4501882|ref|NM_001613.1| Actina, alfa 2, músculo liso, aorta de Homo sapiens (ACTA2), ARNm y SAC
- 17/18 >gi|21536442|ref|NM-003571.2| Proteína estructural de filamento en cuentas 2 de Homo sapiens, facinina (BFSP2), ARNm y SAC
- 19/20 >gi|22538813|ref|NM_002985.2| Ligando 5 (CCL5) de quimiocina (motivo C-C) de Homo sapiens, ARNm y SAC
- 21/22 >gi|30795213|ref|NM_001838.2| Receptor 7 (CCR7) de quimiocina (motivo C-C) de Homo sapiens, ARNm y SAC
- 23/24 >gi|58331233|ref|NM_000074.2| Ligando CD40 de Homo sapiens (superfamilia TNF, miembro 5, síndrome de hiper-IgM) (CD40LG), ARNm y SAC
- 25/26 >gi|83700229|ref|NM_005214.3| Proteína 4 asociada a linfocitos T citotóxicos de Homo sapiens (CTLA4), variante de transcrito 1, ARNm y SAC
- 27/28 >gi|83700230|ref|NM_001037631.1| Proteína 4 asociada a linfocitos T citotóxicos de Homo sapiens (CTLA4), variante de transcrito 2, ARNm y SAC
- 29/30 >gi|125987605|ref|NM_001912.3| Catepsina L1 de Homo sapiens (CTSL1), variante de transcrito 1, ARNm y SAC
- 31/32 >gi|125987604|ref|NM_145918.2| Catepsina L1 de Homo sapiens (CTSL1), variante de transcrito 2, ARNm y SAC
- 33/34 >gi|22538469|ref|NM_005442.2| Homólogo de eomesodermina de Homo sapiens (Xenopus laevis) (EOMES), ARNm y SAC
- 35/36 >gi|32490571|ref|NM_012307.2| 3 Similar a banda 4.1 de proteína de membrana de eritrocitos de Homo sapiens (EPB41L3), ARNm y SAC
- 37/38 >gi|93277091|ref|NM_052966.2| Familia con similitud de secuencia 129 de Homo sapiens, miembro A (FAM129A), variante de transcrito 2, ARNm y SAC
- 39/40 >gi|113428621|ref|XM_001131379.1| PREDICHO: Fragmento Fc de proteína de unión a IgG de Homo sapiens (FCGBP), ARNm y SAC
- 41/42 >gi|4503718|ref|NM_002012.1| Gen de la tríada de histidina frágil de Homo sapiens (FHIT), ARNm y SAC
- 43/44 >gi|6552331|ref|NM_005803.2| Flotilina 1 de Homo sapiens (FLOT1), ARNm y SAC
- 45/46 >gi|31982942|ref|NM_014009.2| Caja forkhead P3 de Homo sapiens (FOXP3), ARNm y SAC
- 47/48 >gi|7108345|ref|NM_012483.1| Granulisina de Homo sapiens (GNLY), variante de transcrito 519, ARNm y SAC
- 49/50 >gi|7108343|ref|NM_006433.2| Granulisina de Homo sapiens (GNLY), variante de transcrito NKG5, ARNm y SAC

ES 2 666 206 T3

- 51/52 >gi|94721348|ref|NM_022307.2| Autoantígeno 1 de las células de los islotes de Homo sapiens, 69 kDa (ICA1), variante de transcrito 1, ARNm y SAC
- 53/54 >gi|94721349|ref|NM_004968.2| Autoantígeno 1 de las células de los islotes de Homo sapiens, 69 kDa (ICA1), variante de transcrito 2, ARNm y SAC
- 5 55/56 >gi|28610150|ref|NM_002185.2| Receptor de interleucina 7 de Homo sapiens (IL7R), ARNm y SAC
- 57/58 >gi|24430213|ref|NM_001558.2| Receptor de interleucina 10 de Homo sapiens, alfa (IL10RA), ARNm y SAC
- 59/60 >gi|61098051|ref|NM_020940.2| KIAA1600de Homo sapiens (KIAA1600), ARNm y SAC
- 61/62 >gi|45007001|ref|NM_203463.1| Homólogo de LAG1, ceramida sintasa 6 de Homo sapiens (*S. cerevisiae*) (LASS6), ARNm y SAC
- 10 63/64 >gi|51972195|ref|NM_001004307.1| LOC339541 hipotético de Homo sapiens (MGC33556), ARNm y SAC
- 65/66 >gi|22035639|ref|NM_002413.3| Glutación S-transferasa 2 microsómica de Homo sapiens (MGST2), ARNm y SAC
- 67/68 >gi|47519797|ref|NM_000631.3| Factor citosólico 4 de neutrófilos de Homo sapiens, 40 kDa (NCF4), variante de transcrito 1, ARNm y SAC
- 15 69/70 >gi|47519769|ref|NM_013416.2| Factor citosólico 4 de neutrófilos de Homo sapiens, 40 kDa (NCF4), variante de transcrito 2, ARNm y SAC
- 71/72 >gi|5453765|ref|NM_006159.1| 2 similar a NEL de Homo sapiens (pollo) (NELL2), ARNm y SAC
- 73/74 >gi|82659104|ref|NM_014241.3| Similar a proteína tirosina fosfatasa de Homo sapiens (prolina en lugar de arginina catalítica), miembro A (PTPLA), ARNm y SAC
- 20 75/76 >gi|11038651|ref|NM_004219.2| Transformante 1 de tumor hipofisario de Homo sapiens (PTTG1), ARNm y SAC
- 77/78 >gi|84872014|ref|NR_002734.1| Transformante 3 de tumor hipofisario de Homo sapiens (PTTG3) on chromosome 8 and CDS
- 25 79/80 >gi|44889474|ref|NM_015213.2| Proteína 1 de interacción con RAB6 de Homo sapiens (RAB6IP1), ARNm y SAC
- 81/82 >gi|46249389|ref|NM_016836.2| Proteína 1 de interacción monocatenaria, motivo de unión a ARN de Homo sapiens (RBMS1), variante de transcrito 1, ARNm y SAC
- 83/84 >gi|46249391|ref|NM_016839.2| Proteína 1 de interacción monocatenaria, motivo de unión a ARN de Homo sapiens (RBMS1), variante de transcrito 2, ARNm y SAC
- 30 85/86 >gi|46249390|ref|NM_002897.3| Proteína 1 de interacción monocatenaria, motivo de unión a ARN de Homo sapiens (RBMS1), variante de transcrito 3, ARNm y SAC
- 87/88>gi|45827773|ref|NM_021205.4| Familia de genes homólogos a ras de Homo sapiens, miembro U (RHOU), ARNm y SAC
- 35 89/90 >gi|33356175|ref|NM_002971.2| Homeobox 1 SATB de Homo sapiens (SATB1), ARNm y SAC
- 91/92 >gi|6031196|ref|NM_003005.2| Selectina P de Homo sapiens (proteína de membrana de gránulos, 140 kDa, antígeno CD62) (SELP), ARNm y SAC
- 93/94 >gi|9910361|ref|NM_020163.1| Dominio sema, dominio inmunoglobulina (Ig), dominio básico corto, secretado, (semaforina) 3G de Homo sapiens (SEMA3G), ARNm y SAC
- 40 95/96 >gi|52851445|ref|NM_005412.4| Serina hidroximetil-transferasa 2 (mitocondrial)de Homo sapiens (SHMT2), ARNm y SAC
- 97/98 >gi|21265027|ref|NM_003473.2| Molécula 1 adaptadora de transducción de señales (dominio SH3 y dominio ITAM) de Homo sapiens (STAM), ARNm y SAC
- 99/100 >gi|38016910|ref|NM_004099.4| Estomatina de Homo sapiens (STOM), variante de transcrito 1, ARNm y SAC
- 45 101/102 >gi|38016906|ref|NM_198194.1| Estomatina de Homo sapiens (STOM), variante de transcrito 2, ARNm y SAC
- 103/104 >gi|51702235|ref|NM_001003799.1| Proteína de marco de lectura alternativo de TCR gamma de Homo sapiens (TARP), proteína mitocondrial que codifica genes nucleares, variante de transcrito 1, ARNm y SAC
- 50 105/106 >gi|51702237|ref|NM_001003806.1| Proteína de marco de lectura alternativo de TCR gamma de Homo sapiens (TARP), proteína mitocondrial que codifica genes nucleares, variante de transcrito 2, ARNm y SAC
- 107/108 >gi|42518079|ref|NM_003202.2| Factor de transcripción 7 de Homo sapiens (específico de linfocitos T, caja HMG) (TCF7), variante de transcrito 1, ARNm y SAC
- 109/110 >gi|42518077|ref|NM_201632.1| Factor de transcripción 7 de Homo sapiens (específico de linfocitos T, caja HMG) (TCF7), variante de transcrito 2, ARNm y SAC
- 55 111/112 >gi|42518073|ref|NM_201633.1| Factor de transcripción 7 de Homo sapiens (específico de linfocitos T, caja HMG) (TCF7), variante de transcrito 3, ARNm y SAC
- 113/114 >gi|42714656|ref|NM_201634.1| Factor de transcripción 7 de Homo sapiens (específico de linfocitos T, caja HMG) (TCF7), variante de transcrito 4, ARNm y SAC
- 60 115/116 >gi|23312365|ref|NM_001066.2| Superfamilia de receptores del factor de necrosis tumoral de Homo sapiens, miembro 1B (TNFRSF1B), ARNm y SAC
- 117/118 >gi|48255912|ref|NM_006470.3| 16 que contiene motivo tripartito de Homo sapiens (TRIM16), ARNm y SAC
- 119/120 >gi|12056478|ref|NM_021995.1| Urotensina 2de Homo sapiens (UTS2), variante de transcrito 1, ARNm y SAC
- 65 121/122 >gi|12056478|ref|NM_021995.1| Urotensina 2de Homo sapiens (UTS2), variante de transcrito 1, ARNm

ES 2 666 206 T3

	y SAC
	123-140 cebadores para el aislamiento de Colonia 1-6
	141 vector pTrcHis FANK1
5	142 vector pTrcHis PLEKHK1 largo
	143 vector pTrcHis HPGD
	144 vector pTrcHis DNATP6
	145 vector pTrcHis Col5 (C1orf78)
	146 vector pTrcHis Col6 (FU45983)
	147 vector pTrcHis
10	Características de pTrcHis-TOPO
	Promotor <i>trc</i> y UTR 5': bases 190-382
	- región 35: bases 193-198
	- región 10: bases 216-221
	Sitio del operador <i>lac</i> : bases 228-248
15	Secuencia antiterminación <i>rrnB</i> : bases 264-333 Potenciador traduccional 10 del gen T7 bases 346-354 Sitio de unión a ribosoma: 369-373
	Sitio de cebado directo pTrcHis bases 370-390 Minicistrón: bases 383-409
	RBS de reinicio: bases 398-403
	ATG de inicio: bases 413-415
20	Marcador 6xHis bases 425-442
	Epítipo Xpress™: bases 482-505
	Sitio de cebado directo Xpress™: bases 445-463 Sitio de escisión de enterocinasa: bases 491-505 Sitio de clonación TOPO: bases 517-518
	Sitio de cebado inverso pTrcHis: bases 574-591 Secuencias de terminación de la transcripción <i>rrnB</i> T1 y T2. bases 624-781 Promotor <i>bla</i> : bases 1002-1059
25	Gen de resistencia a la ampicilina (<i>bla</i>): bases 1060-1920 Origen derivado de pBR322: bases 2065-2738
	Represor Lac (<i>lac^q</i>): bases 3392-4351
	148 vector pENTR4
	Características de pENTR4
30	Secuencia de terminación de la transcripción <i>rrnB</i> T1: bases 106-149 Secuencia de terminación de la transcripción <i>rrnB</i> T2: bases 281-308 attL1: bases 358-457 (cadena complementaria) gen <i>ccdB</i> : bases 615-920 attL2: bases 949-1048
	Gen de resistencia a la kanamicina: bases 1171-1980
	Origen pUC: bases 2044-2717
35	149 vector pENTR4 Egfp
	150 vector pLenti4/V5_DEST
	151 vector pcDNA6-V5-His
	152-255 Fragmento del CMH de Col1 (SEQ ID NO:2)
	256-435 Fragmento del CMH de Col2 (SEQ ID NO:4)
40	436-536 Fragmento del CMH de Col3 (SEQ ID NO:8)
	537-660 Fragmento del CMH de Col4 (SEQ ID NO:10)
	661-708 Fragmento del CMH de Col5 (SEQ ID NO:12)
	709-734 Fragmento del CMH de Col6 (SEQ ID NO:14)
	735-736 >gi / 85218985/ref/NM 001010923.1/ Cromosoma 6 de Homo sapiens
45	Marco de lectura abierto 190 (C6orf190), ARNm y SAC

REIVINDICACIONES

1. Un procedimiento de identificación o caracterización de linfocitos T humanos supresores y reguladores en una muestra de un paciente, en el que el procedimiento comprende detectar la expresión de una secuencia seleccionada entre las SEQ ID NO: 1 o 2.
- 5 2. El procedimiento de la reivindicación 1, en el que el procedimiento comprende detectar la expresión de una secuencia de SEQ ID NO: 1.
3. El procedimiento de la reivindicación 1, en el que el procedimiento comprende detectar la expresión de una secuencia de SEQ ID NO: 2.
4. El procedimiento de la reivindicación 1, que comprende
- 10 (a) poner en contacto una población celular obtenida a partir de sangre humana con uno o más ligandos que se unen específicamente a una o más proteínas marcadoras que comprenden una secuencia de SEQ ID NO: 2, y
(b) identificar o caracterizar los linfocitos T humanos supresores y reguladores de la población celular.
5. Un procedimiento para aislar linfocitos T humanos supresores y reguladores de una muestra del paciente, en el que el procedimiento comprende
- 15 (a) poner en contacto una población celular obtenida a partir de sangre humana con uno o más ligandos que se unen específicamente a una o más proteínas marcadoras que comprenden una secuencia de SEQ ID NO: 2, y
(b) aislar los linfocitos T humanos supresores y reguladores de la población celular.
6. El procedimiento de la reivindicación 4 o 5, en el que los ligandos son anticuerpos o fragmentos de los mismos que se unen a una o más proteínas marcadoras que comprenden una secuencia de SEQ ID NO: 2.
- 20 7. El procedimiento de la reivindicación 6, en el que los ligandos son anticuerpos monoclonales o fragmentos de los mismos que se unen a una o más proteínas marcadoras que comprenden una secuencia de SEQ ID NO: 2.
8. El procedimiento de acuerdo con una cualquiera de las reivindicaciones 4 a 7, que comprende adicionalmente poner en contacto la sangre humana con uno o más ligandos que se unen específicamente a CD4, CD25 y/o CD127 en los linfocitos T.
- 25 9. El procedimiento de acuerdo con una cualquiera de las reivindicaciones 4 a 8, que comprende adicionalmente el ensayo de la expresión de FOXP3.
10. Uso de un kit que comprende uno o más ligandos que se unen específicamente a una o más proteínas marcadoras como se define en una cualquiera de las reivindicaciones 4 a 7 para el aislamiento, la identificación o la caracterización de linfocitos T humanos supresores o reguladores de una muestra del paciente.
- 30 11. El uso de la reivindicación 10, en el que el kit comprende un anticuerpo o fragmento de anticuerpo como se define en la reivindicación 6 o 7.
12. Uso de un anticuerpo o fragmento de anticuerpo como se define en la reivindicación 6 o 7 para aislar o caracterizar linfocitos T humanos supresores y reguladores de una muestra del paciente.

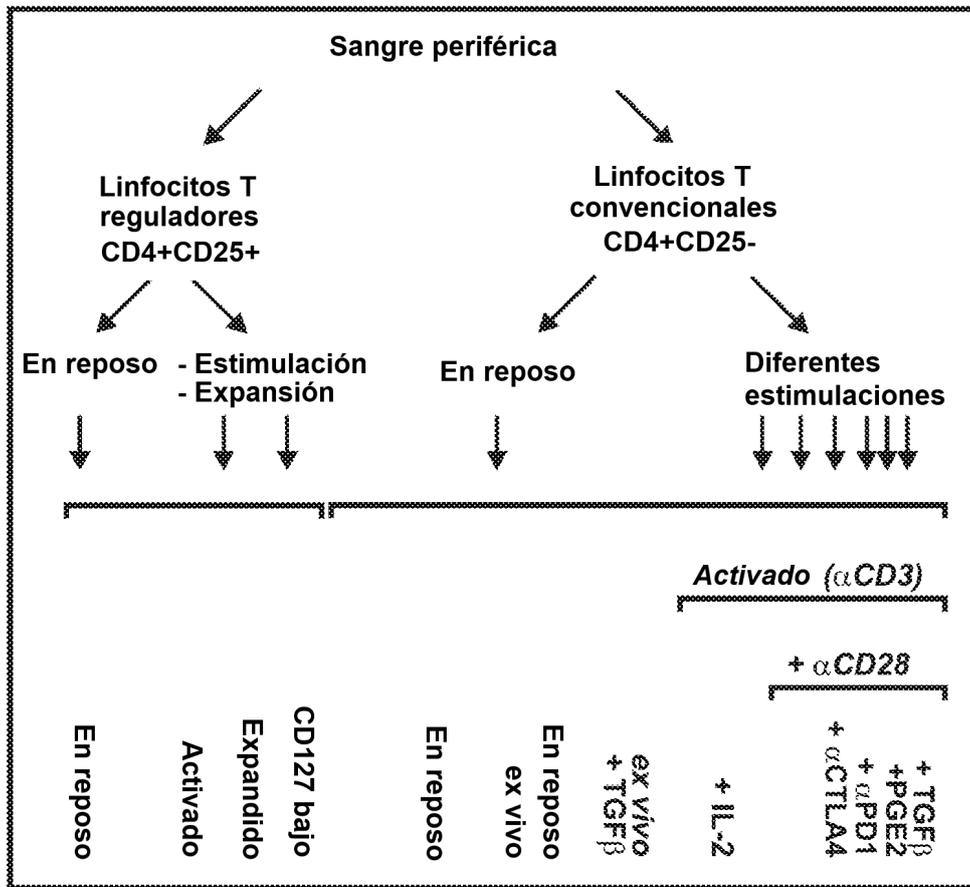


Fig.1

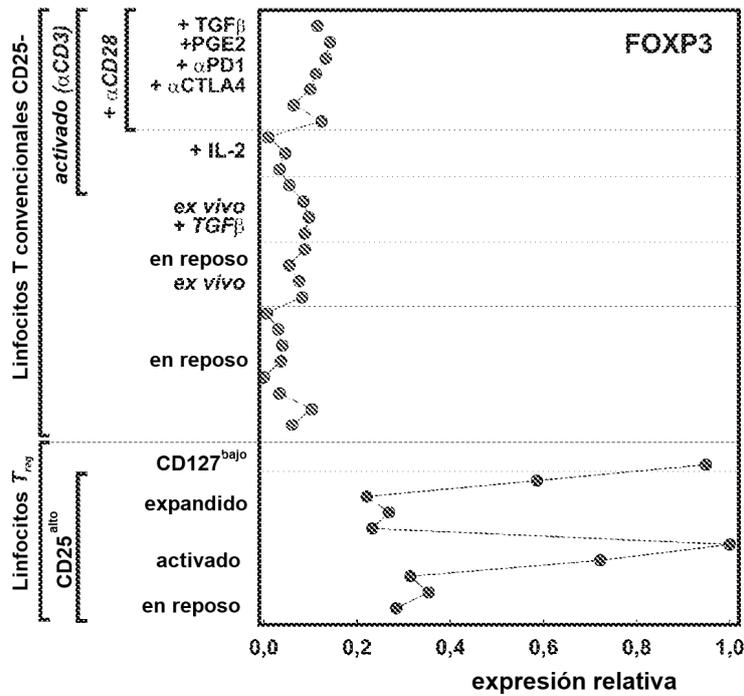


Fig.2

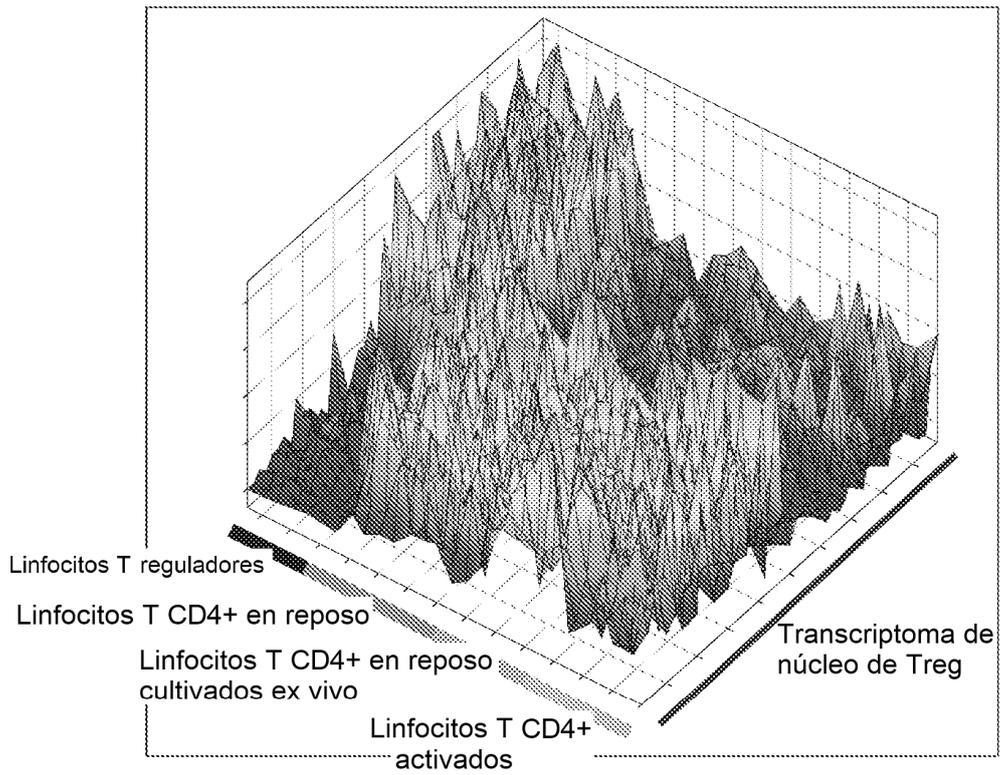


Fig.3

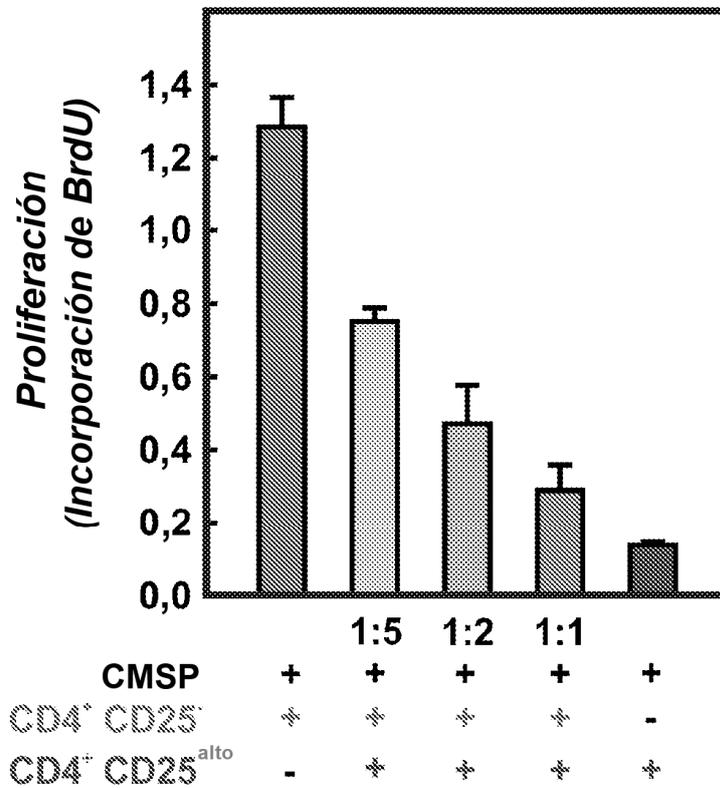


Fig.4

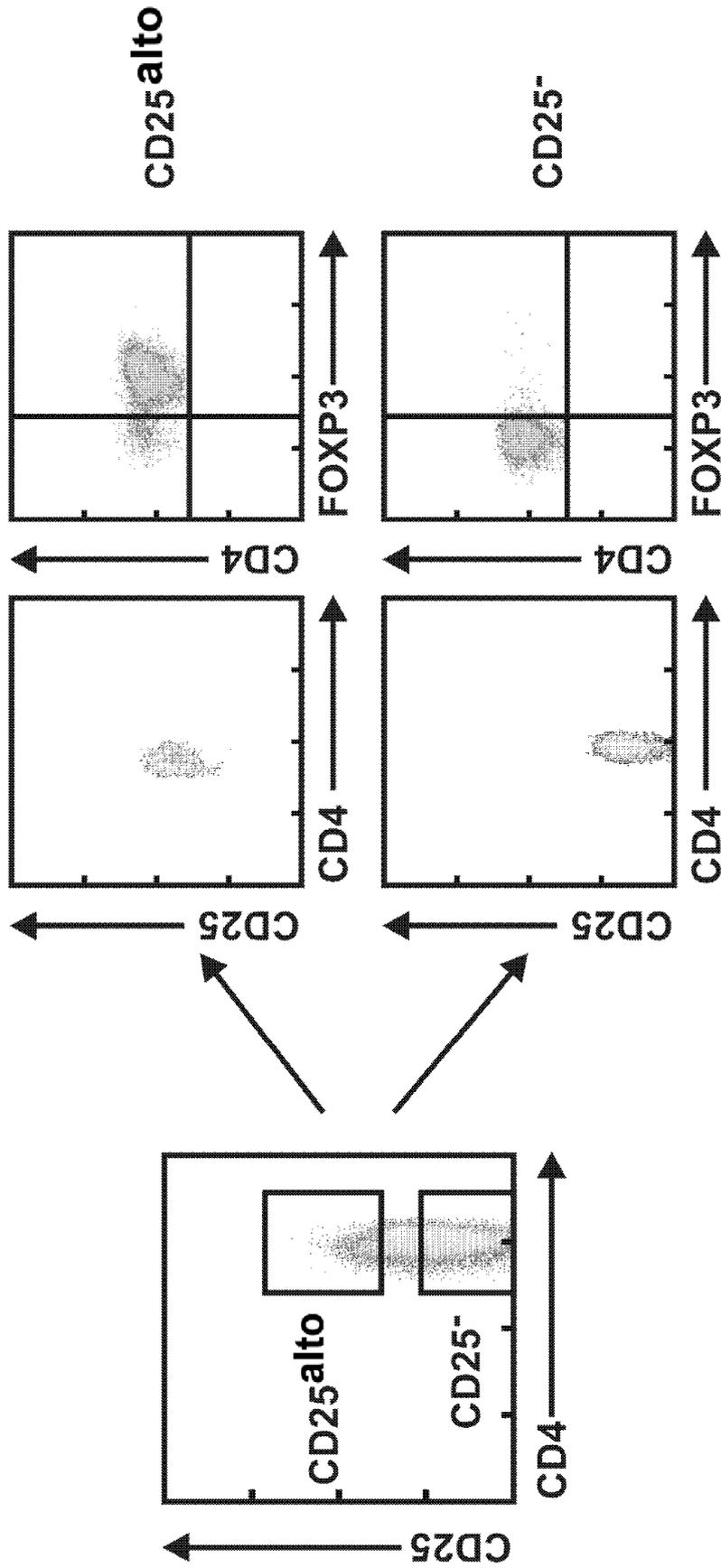


Fig.5

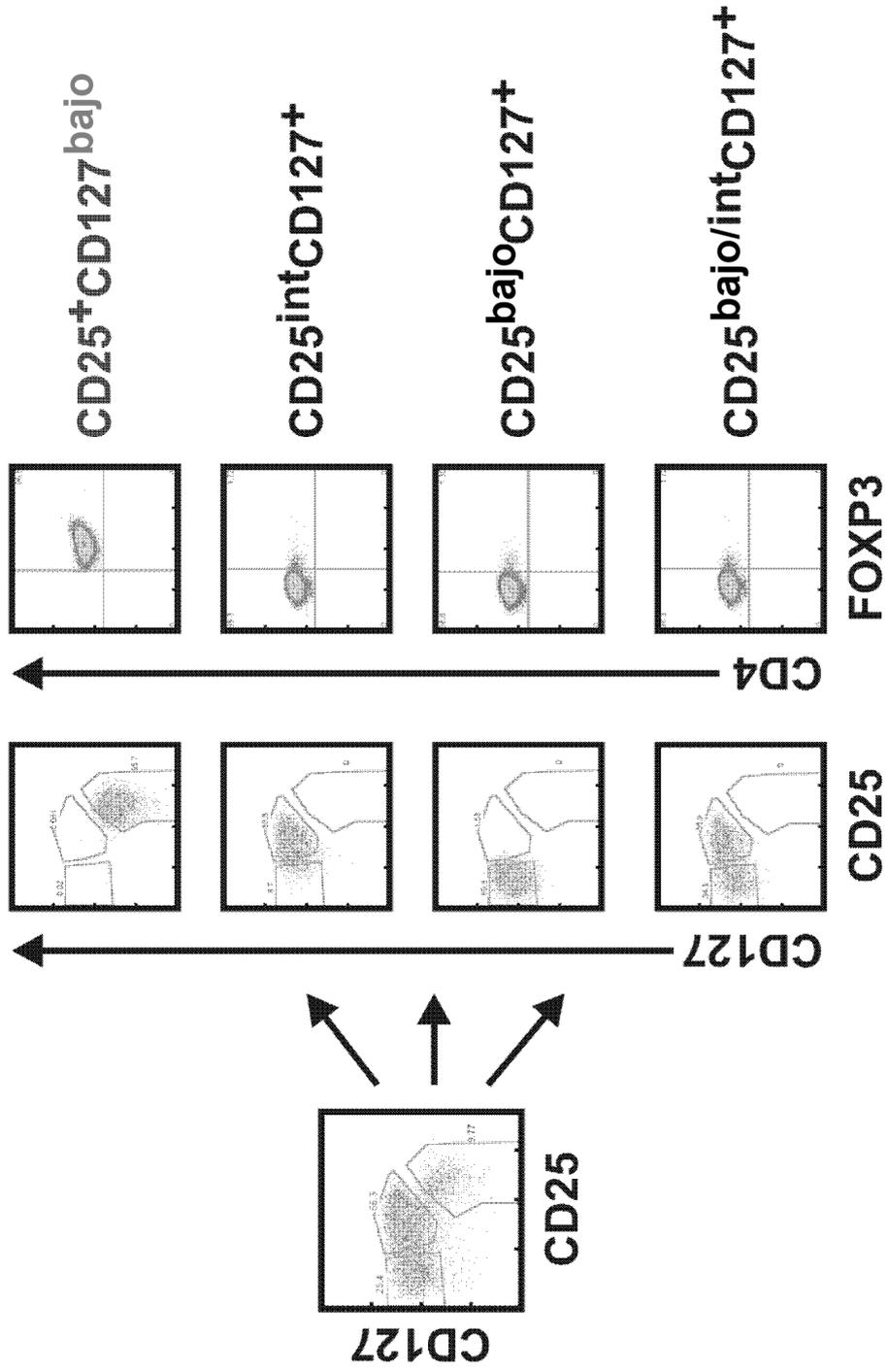


Fig.6

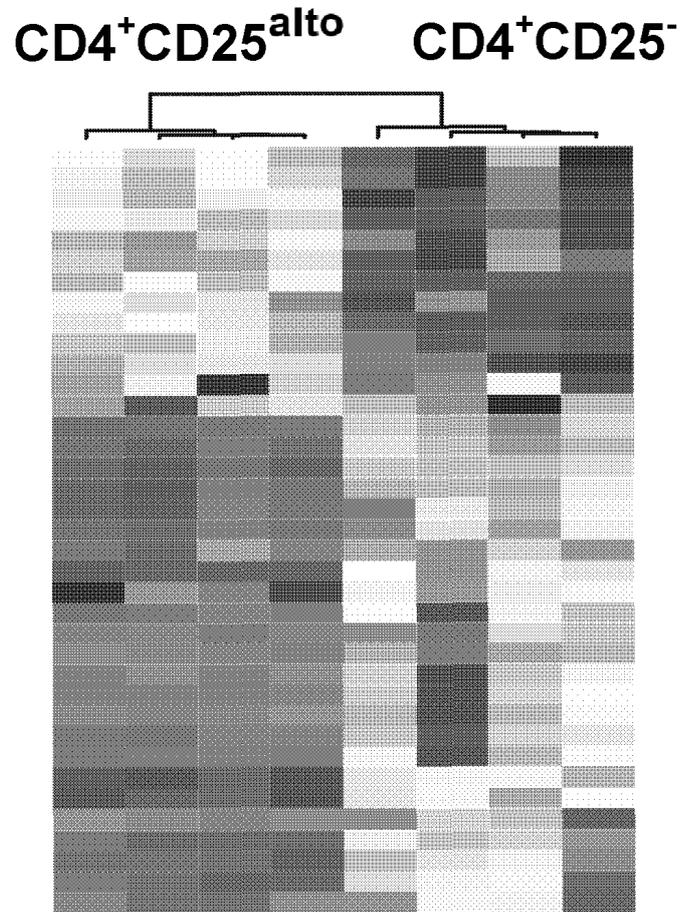


Fig.7

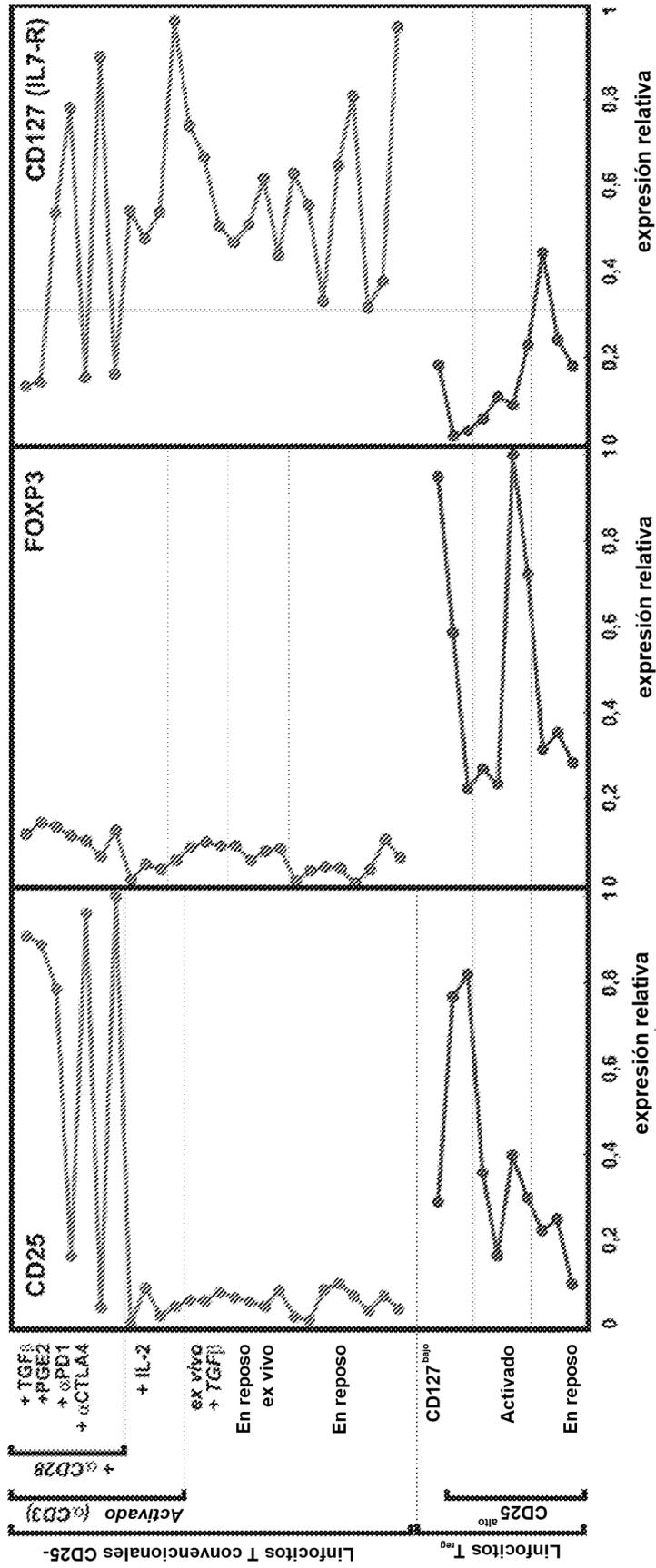


Fig.8

Colonia 1

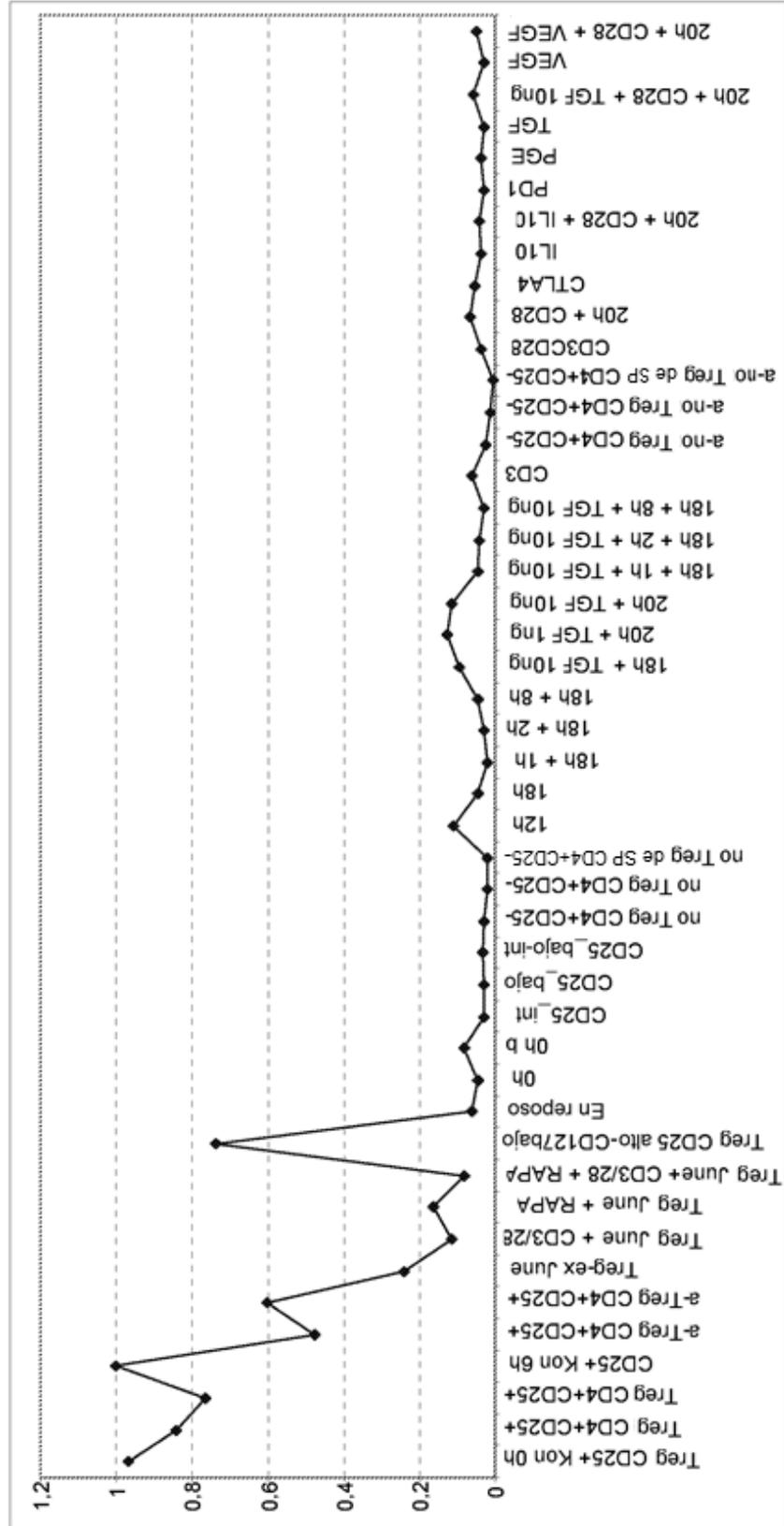


Fig.9A

Colonia 2

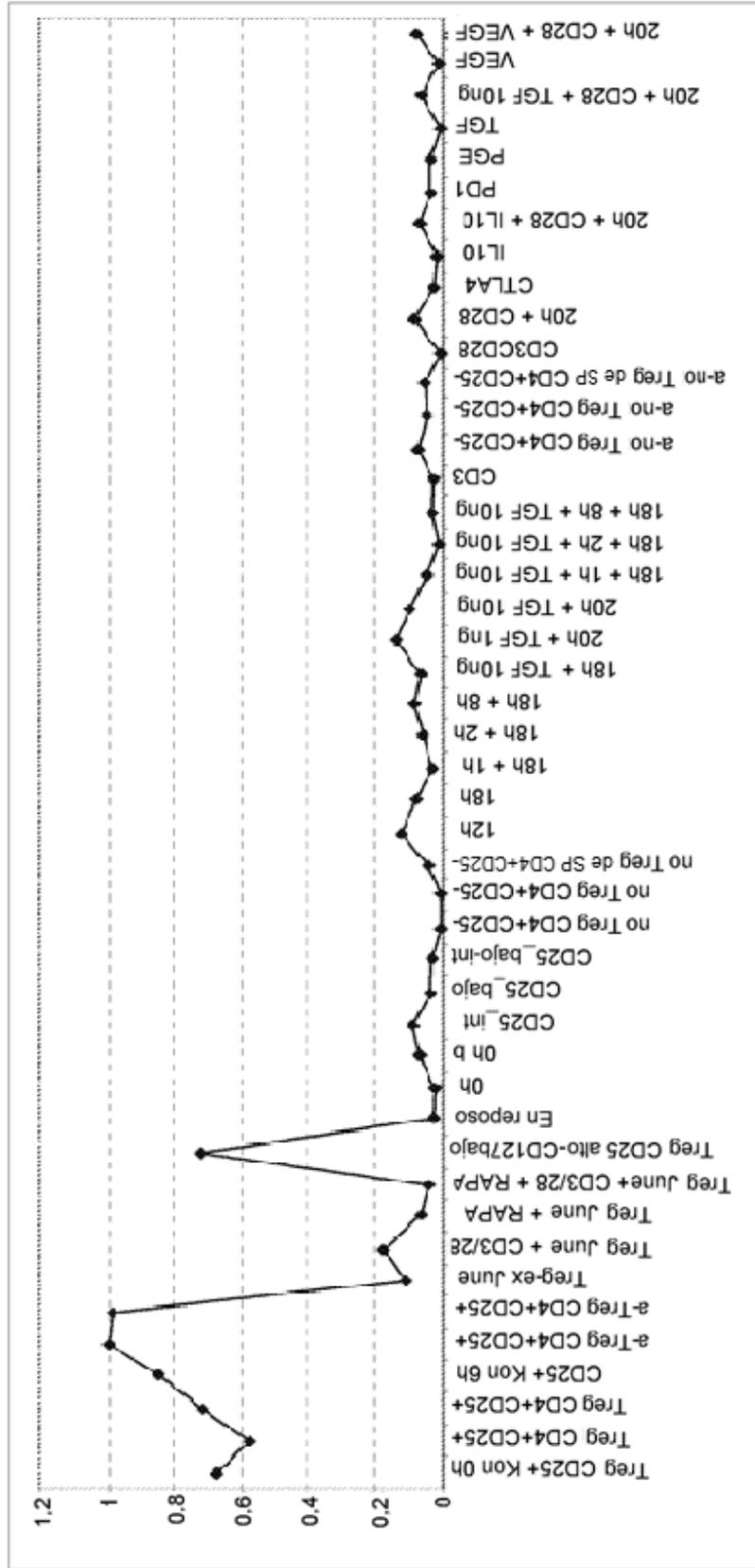


Fig.9B

Colonia 3

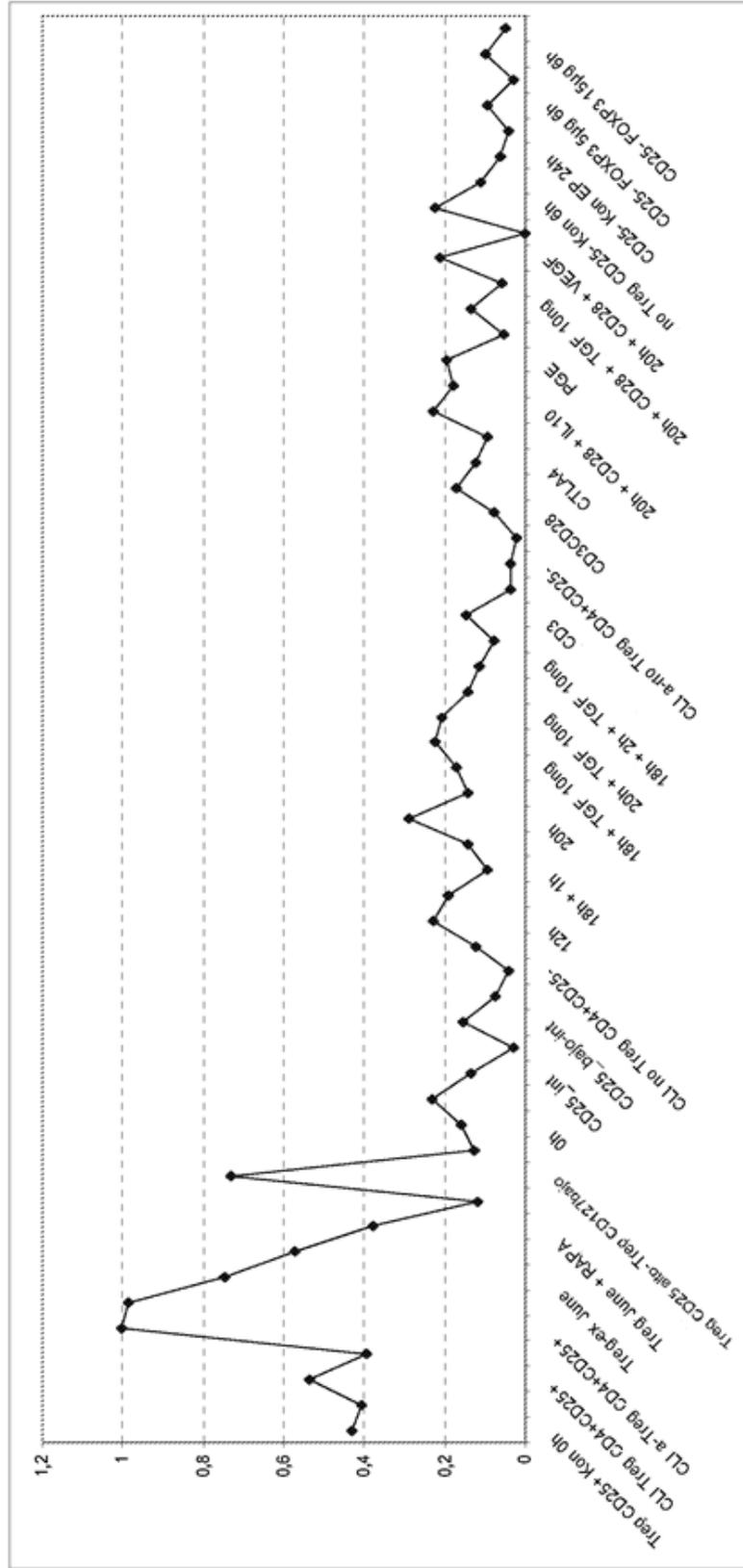


Fig.9C

Colonia 4

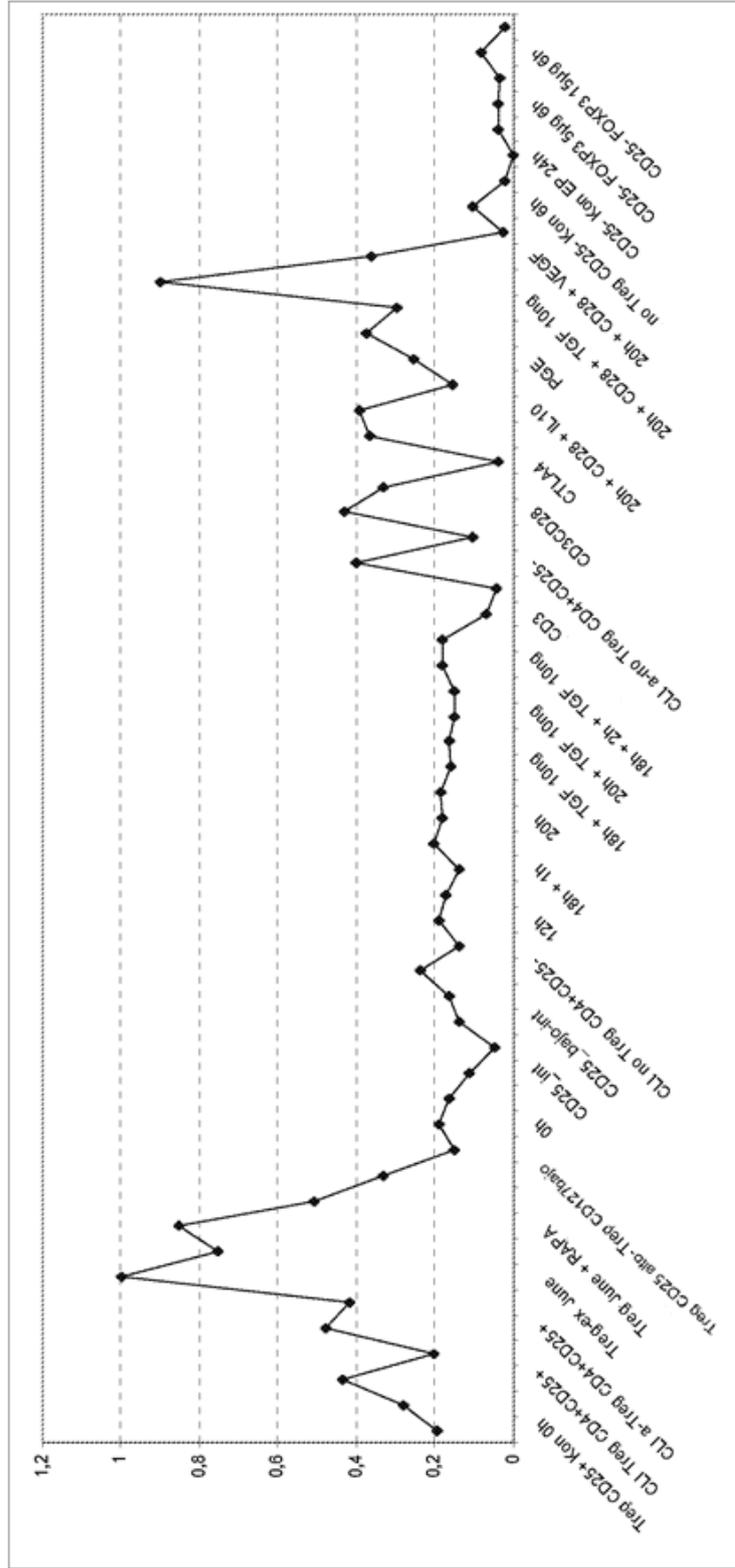


Fig.9D

Colonia 5

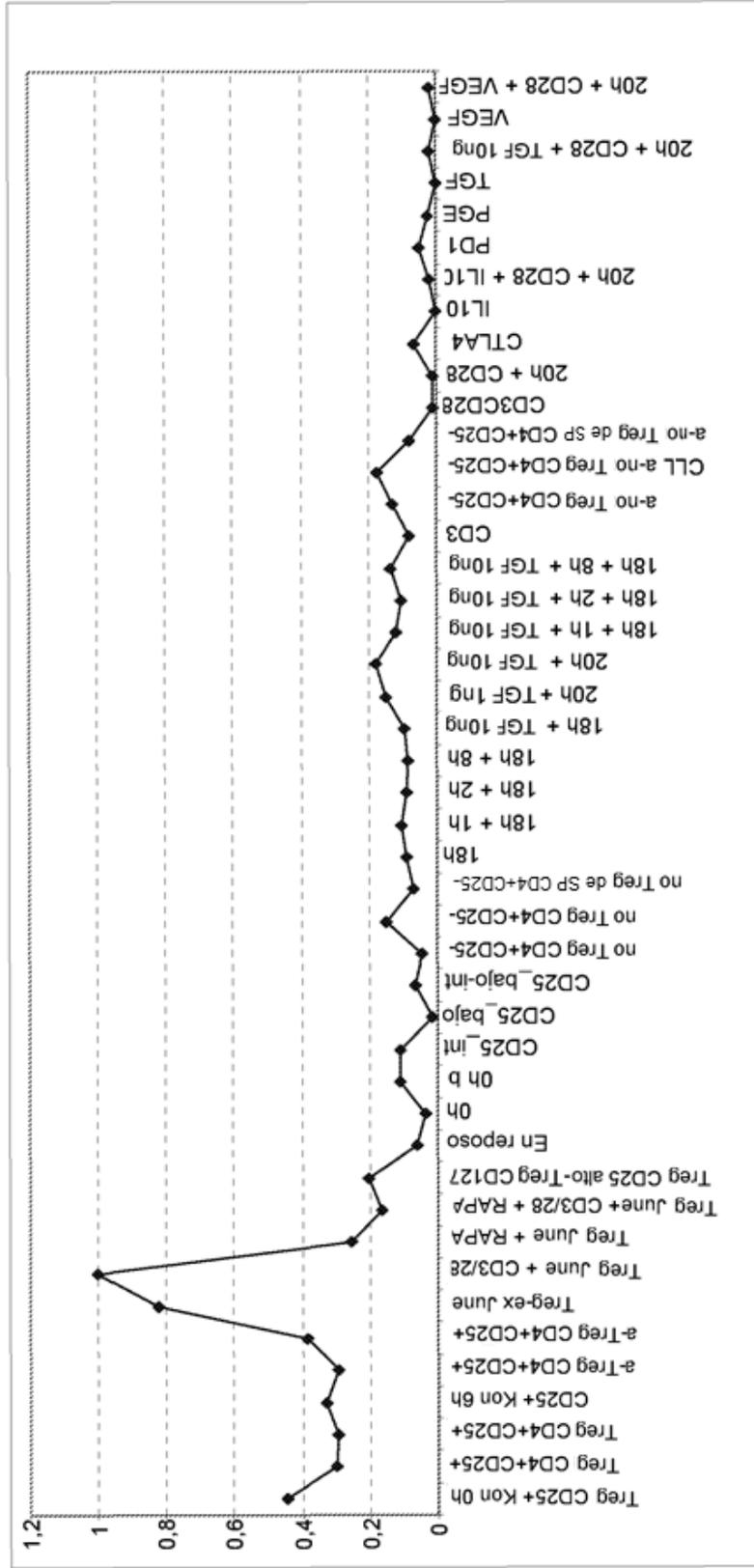


Fig.9E

Colonia 6

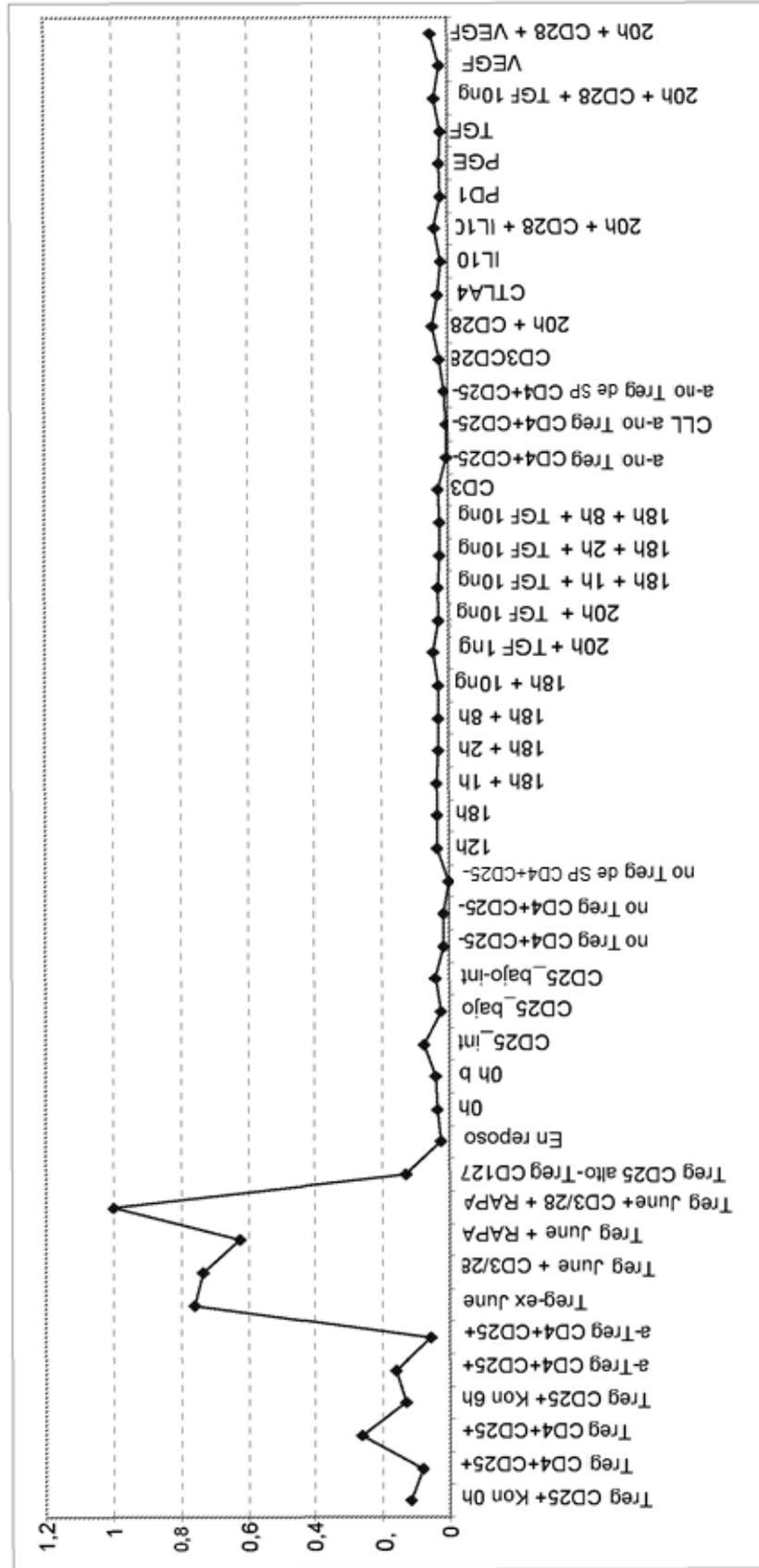


Fig.9F

Colonia 1 (cromosoma 10)

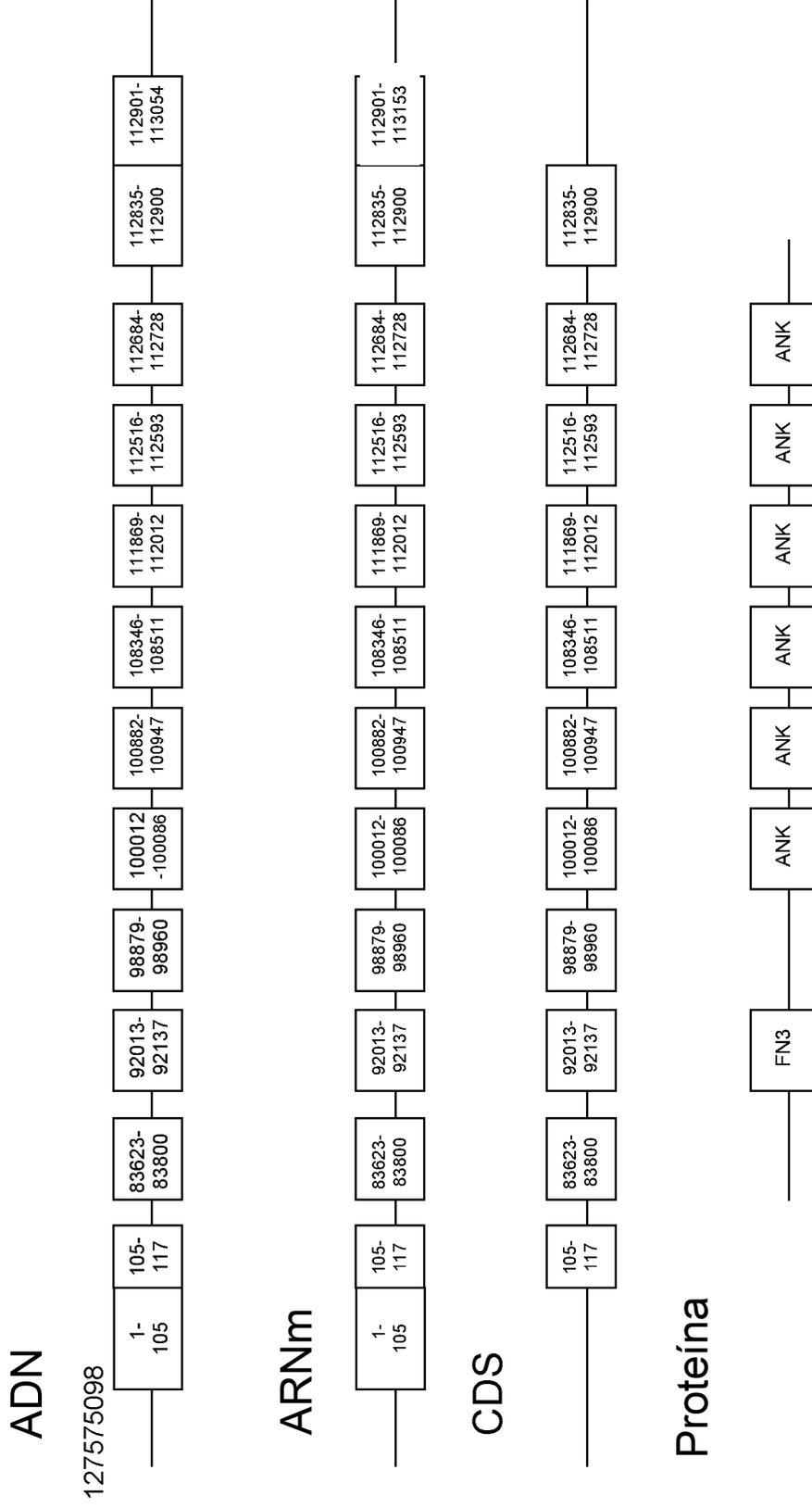
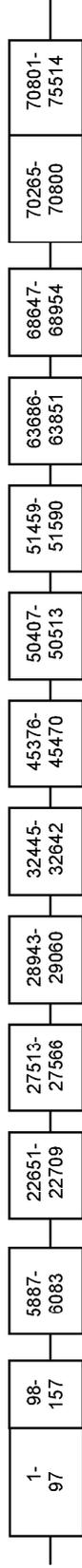


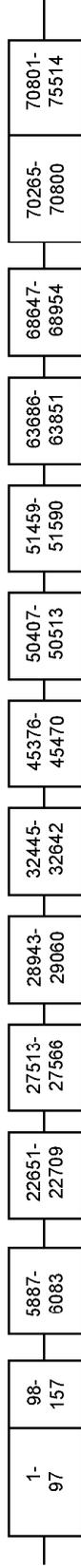
Fig.10A

Colonia 2 (cromosoma 10)

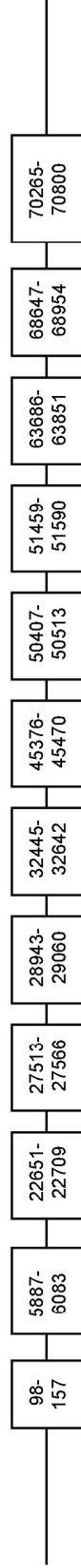
ADN



ARNm



CDS



Proteína

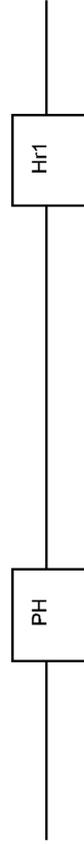


Fig.10B

Colonia 3 (cromosoma 3)

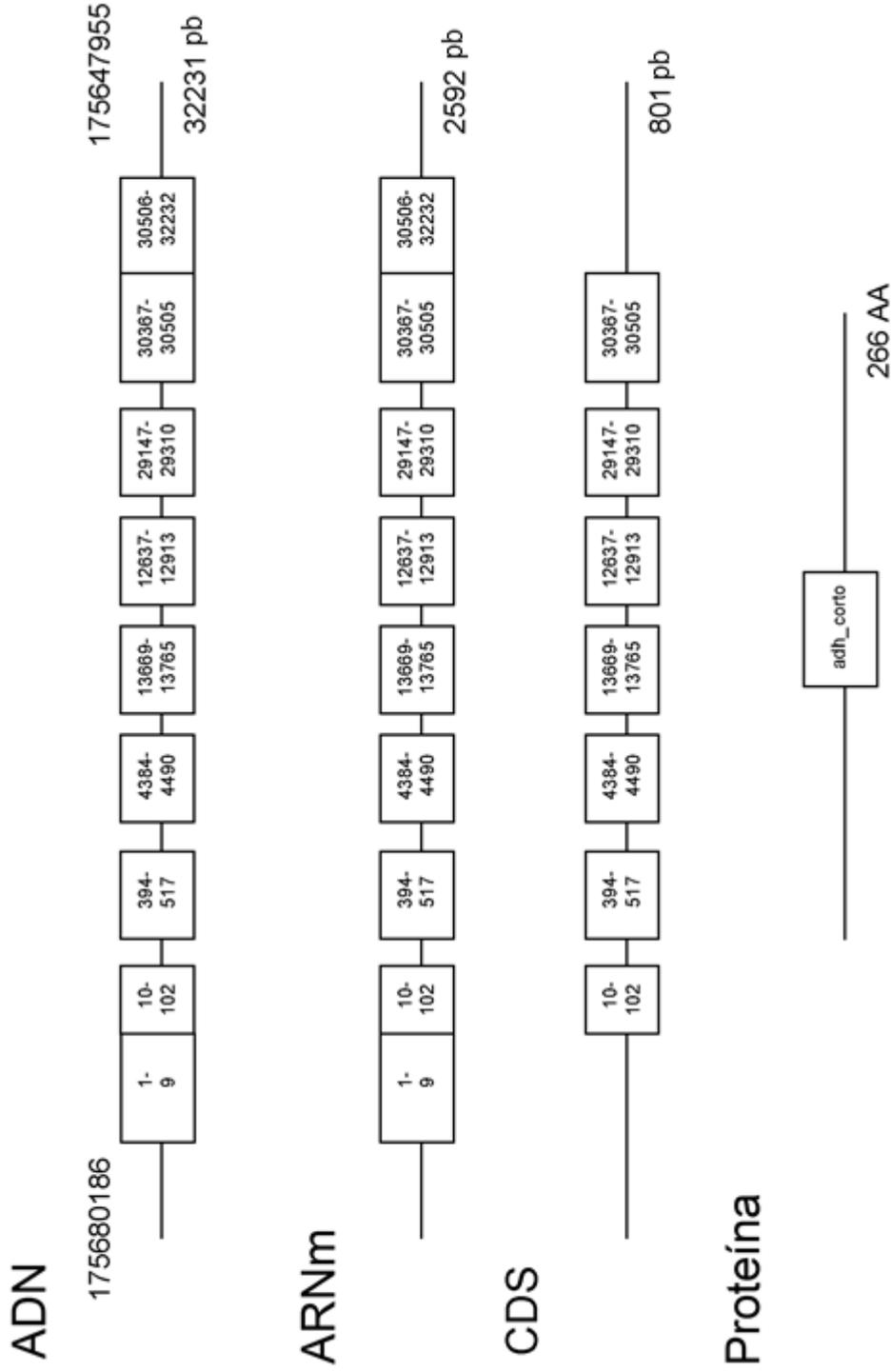


Fig.10C

Colonia 4 (cromosoma 2)

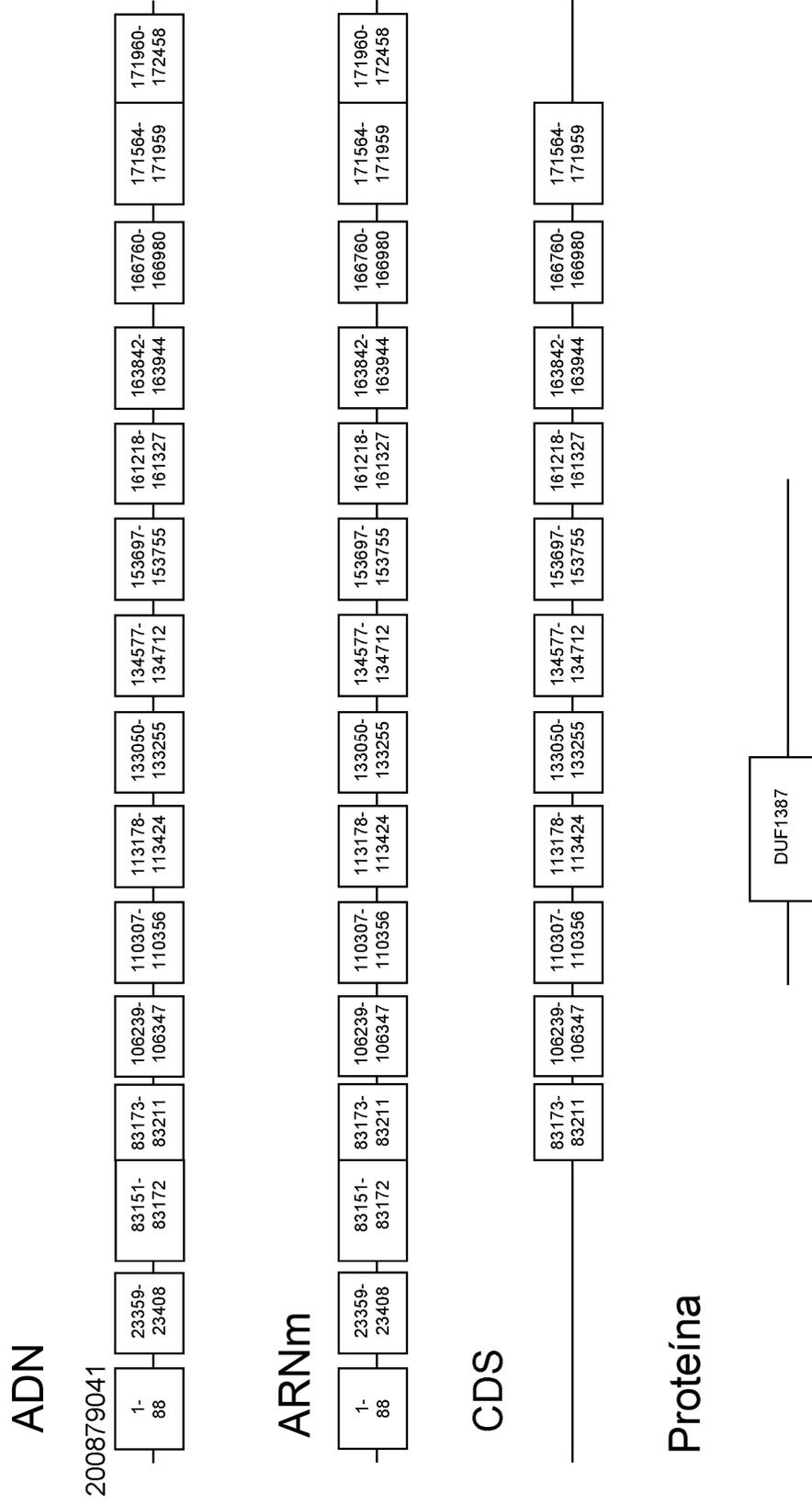


Fig.10D

Colonia 5 (cromosoma 1)

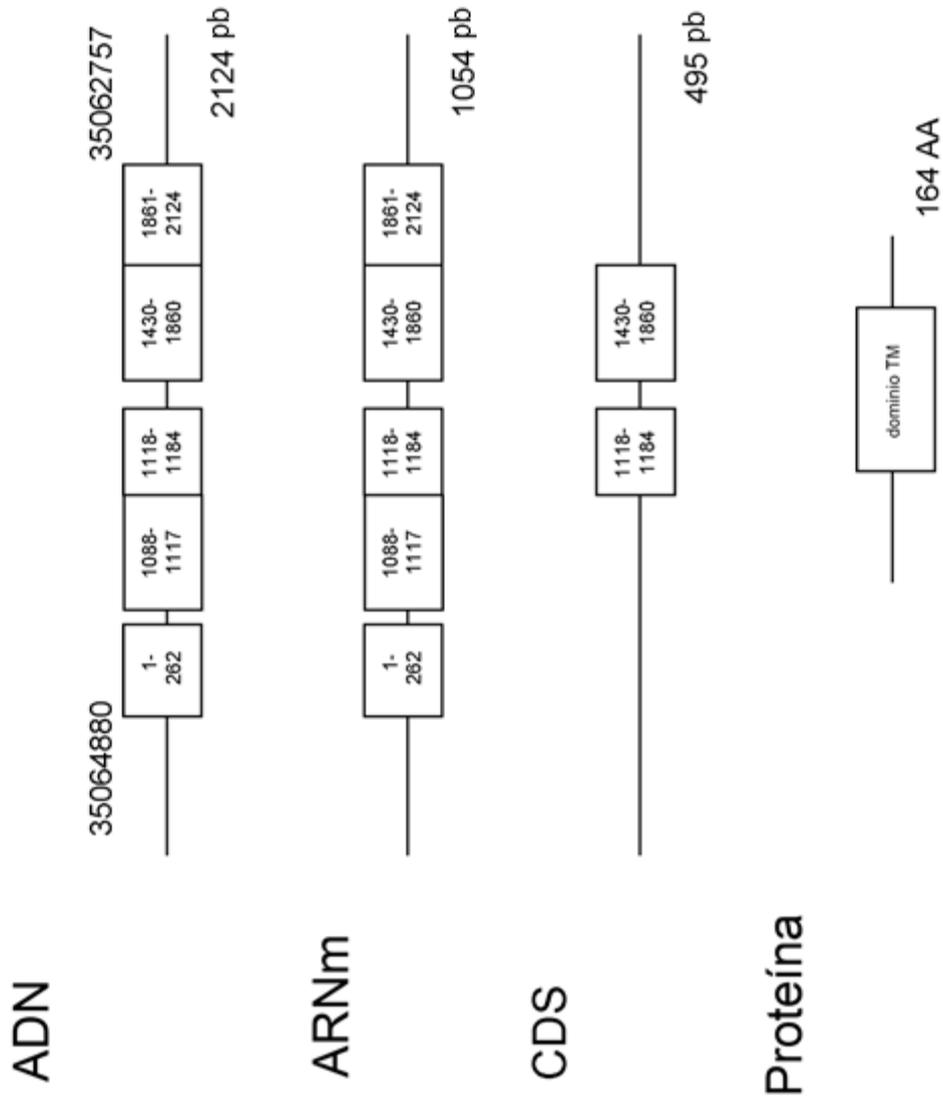


Fig.10E

Colonia 6 (cromosoma 10)

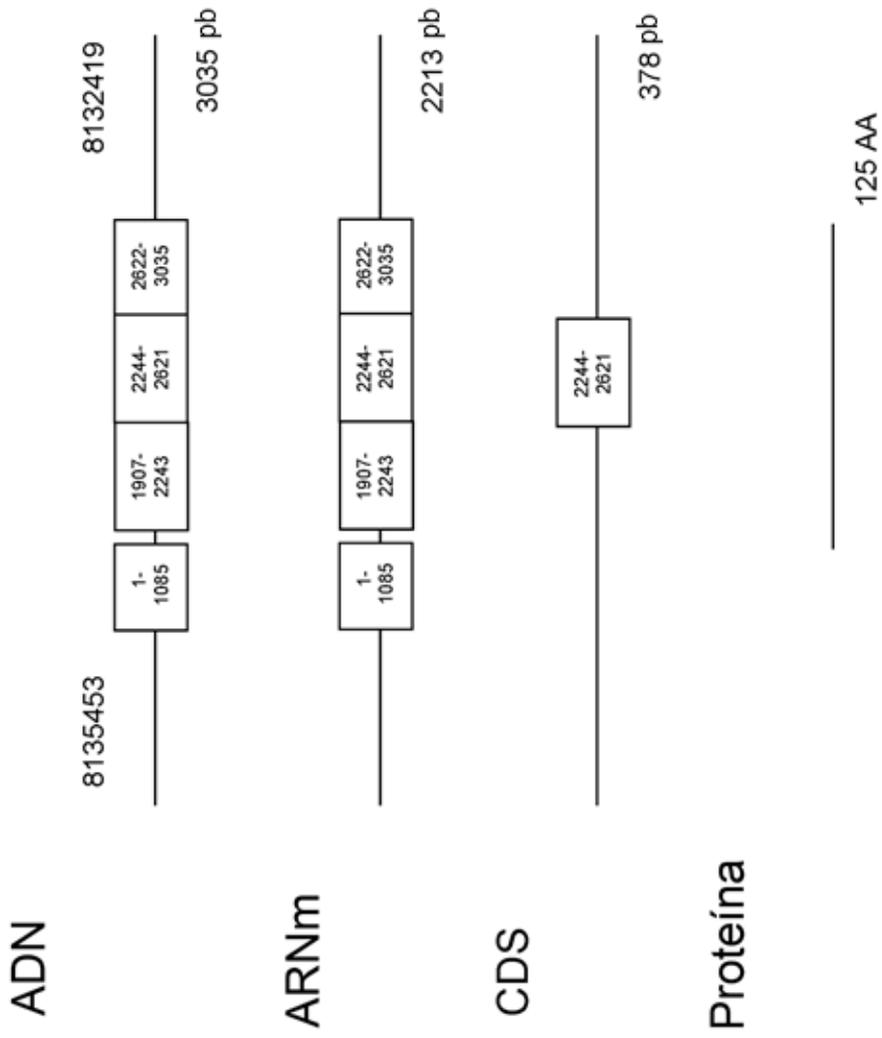


Fig.10F

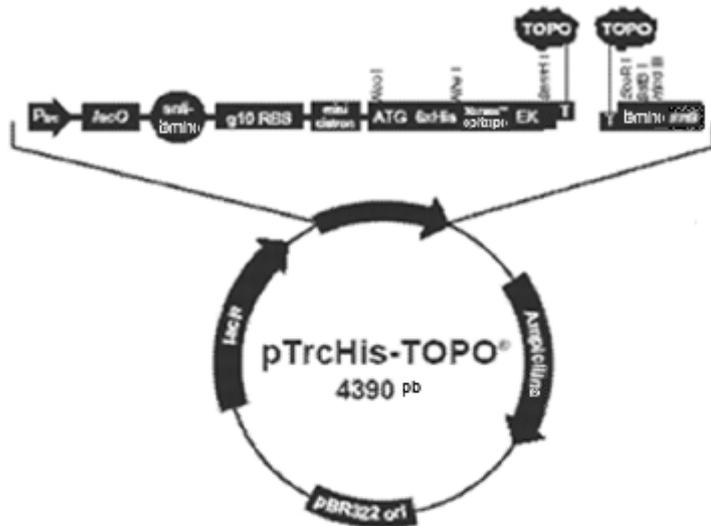


Fig.11A

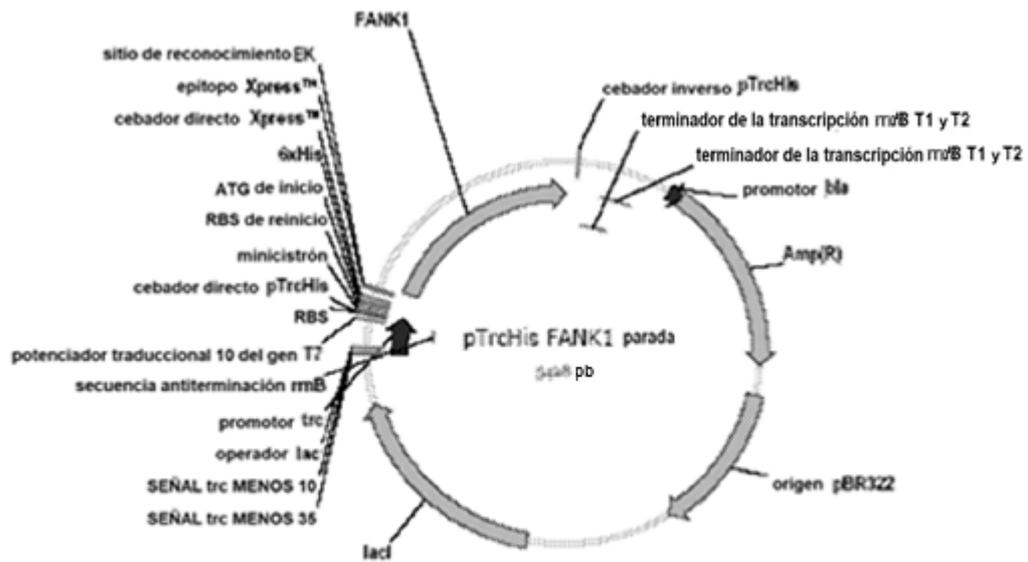


Fig.11B

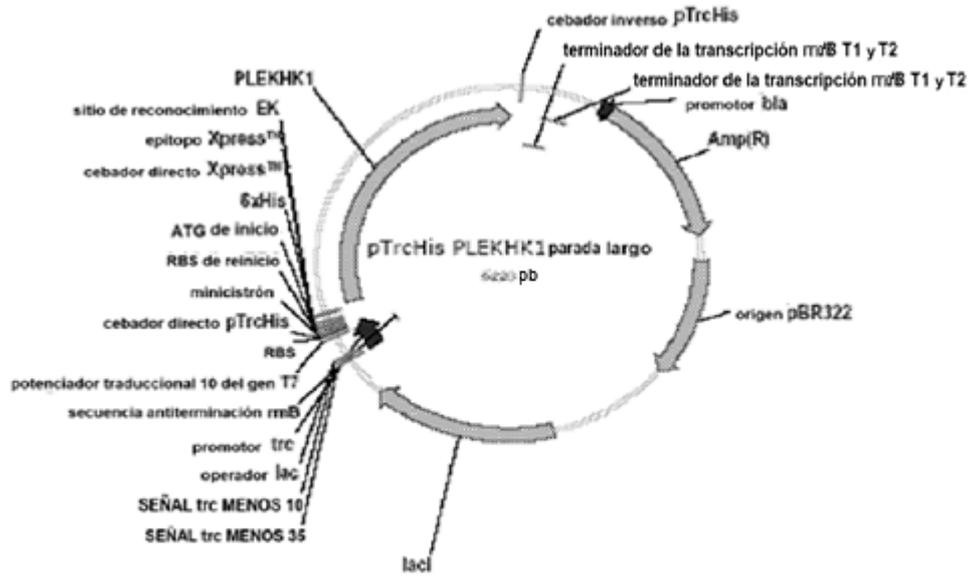


Fig.11C

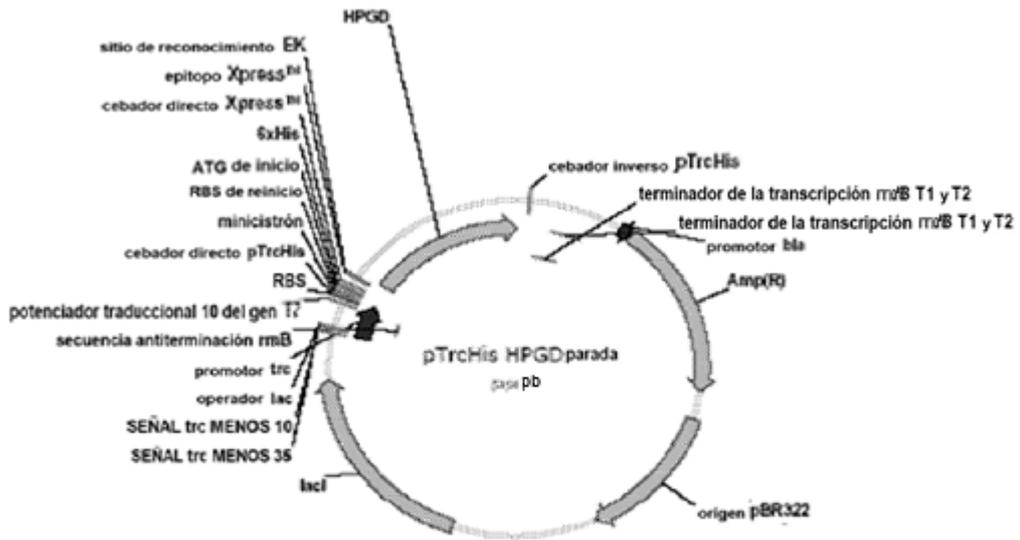


Fig.11D

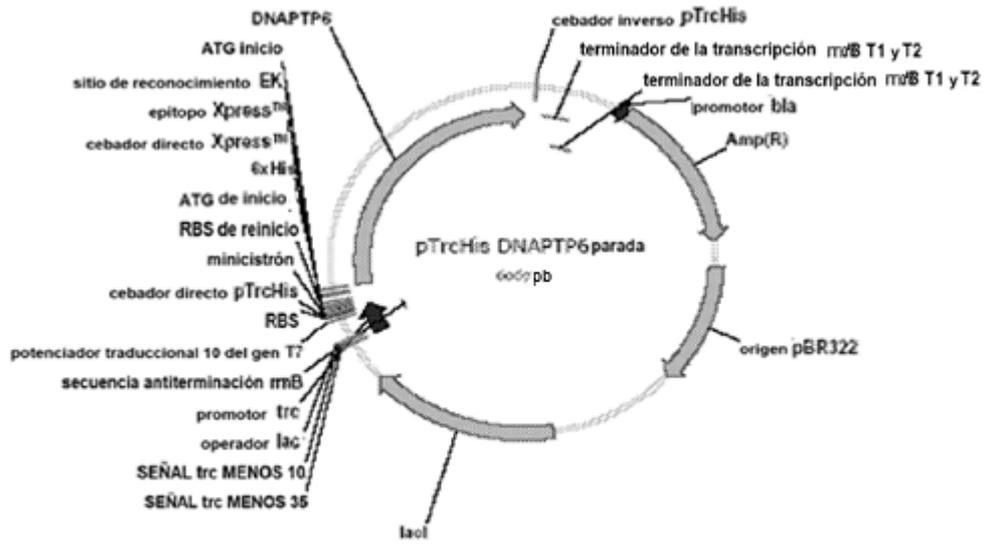


Fig.11E

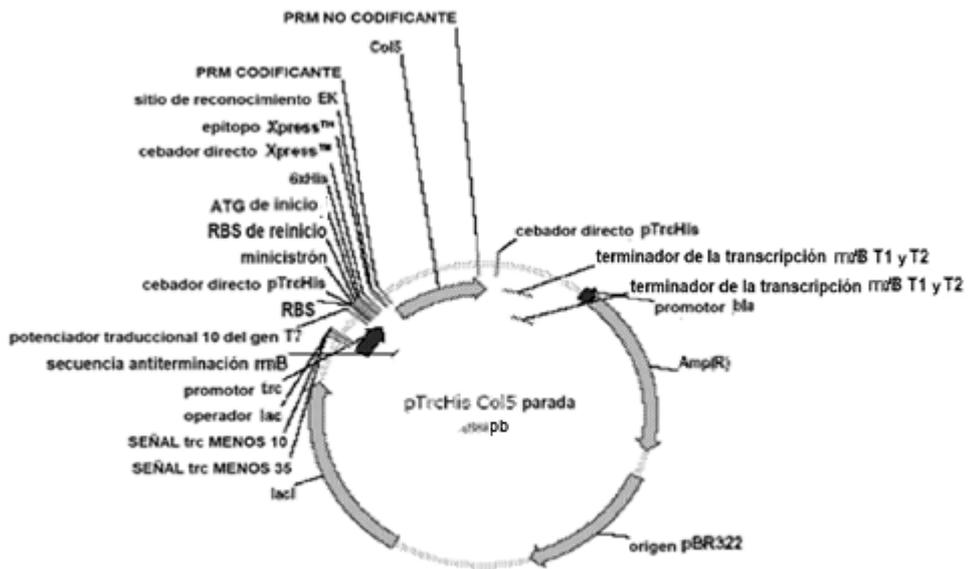


Fig.11F

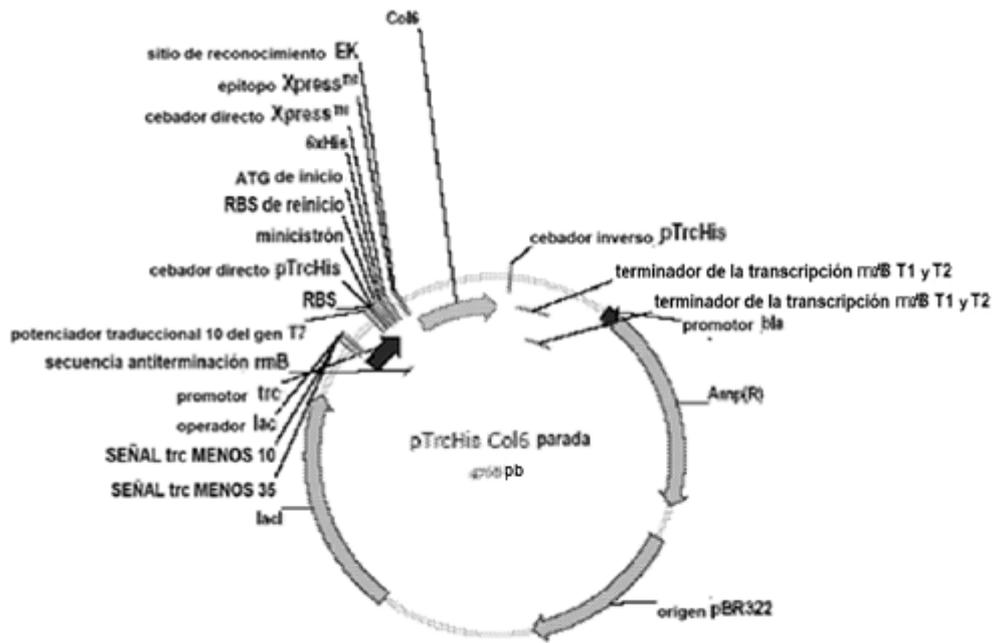


Fig.11G

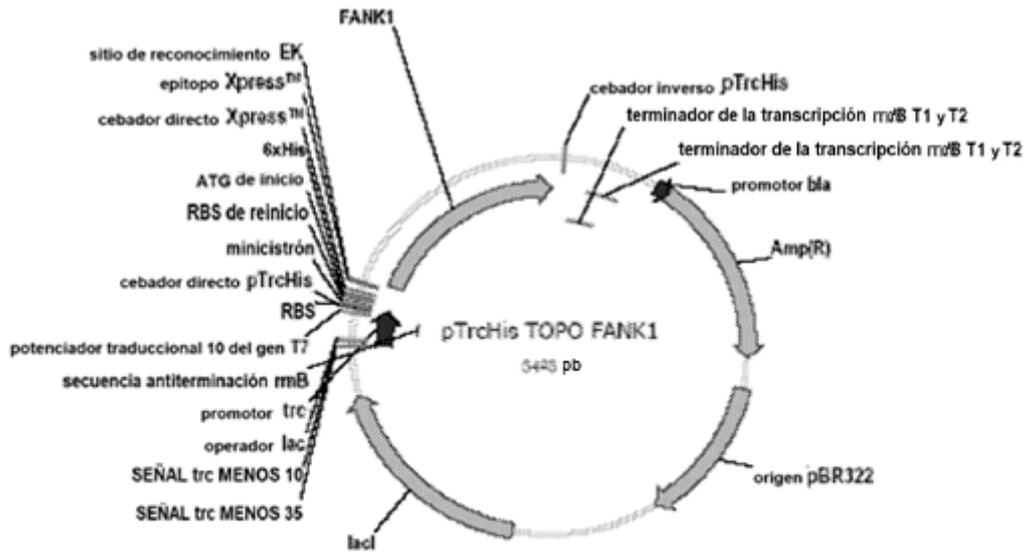


Fig.12A

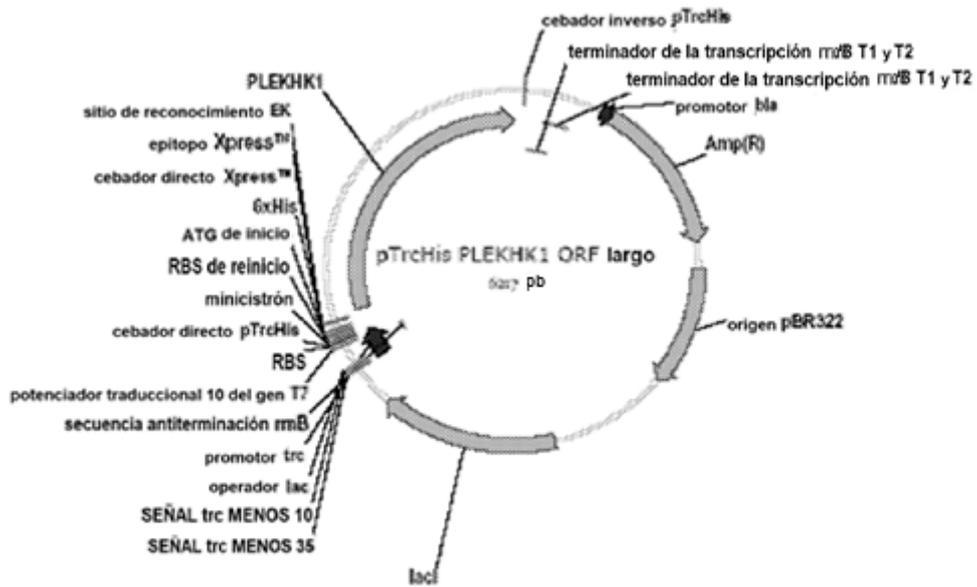


Fig.12B

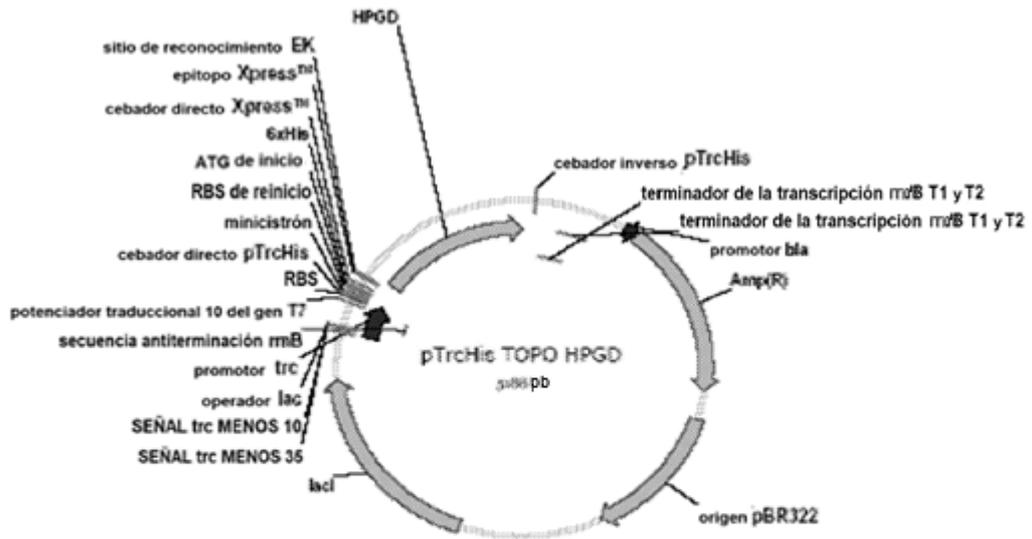


Fig.12C

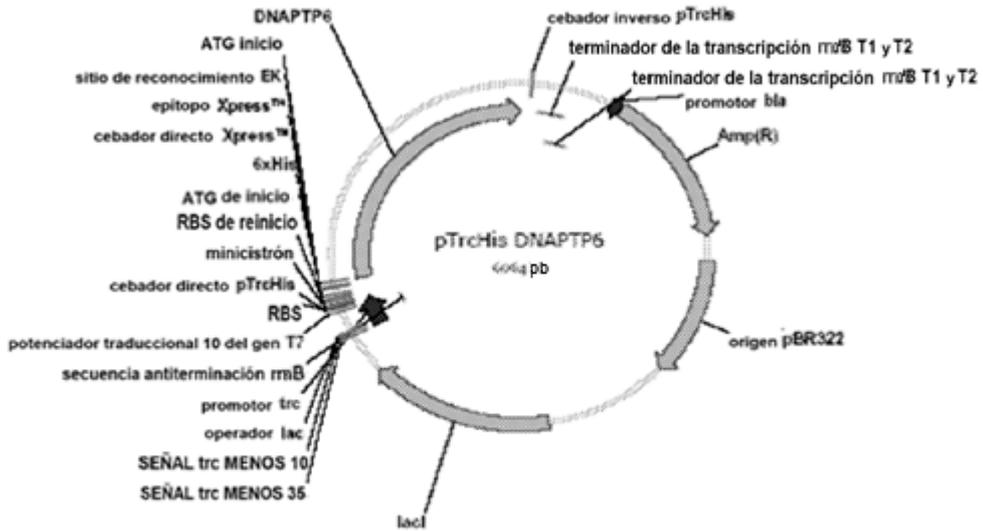


Fig.12D

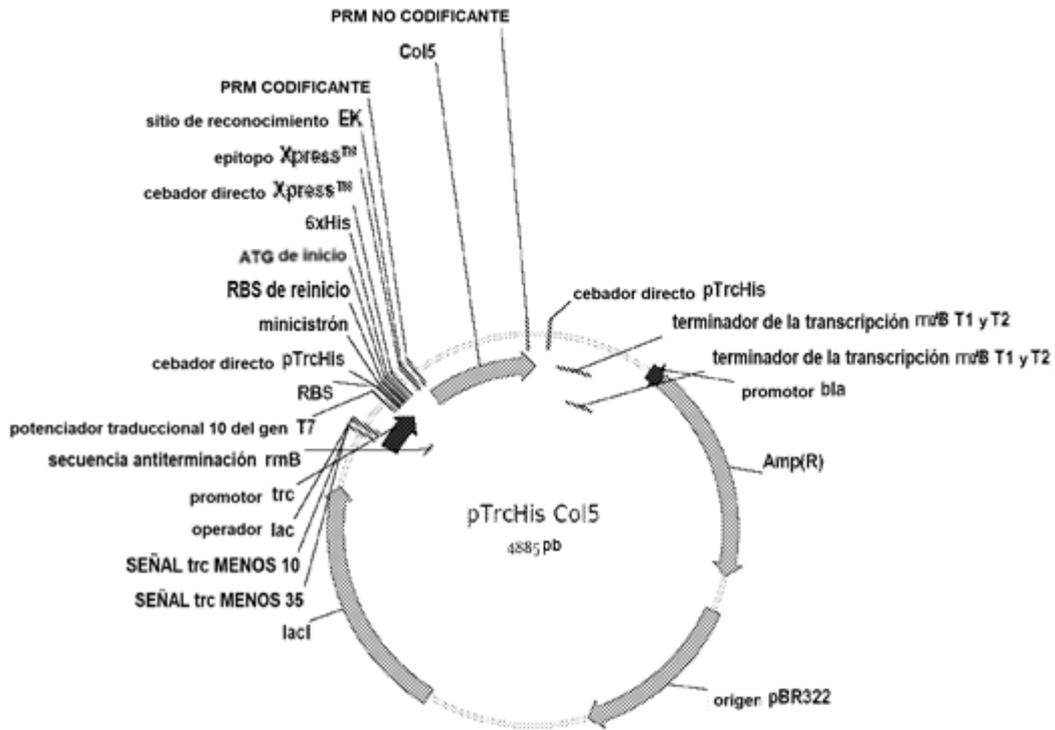


Fig.12E

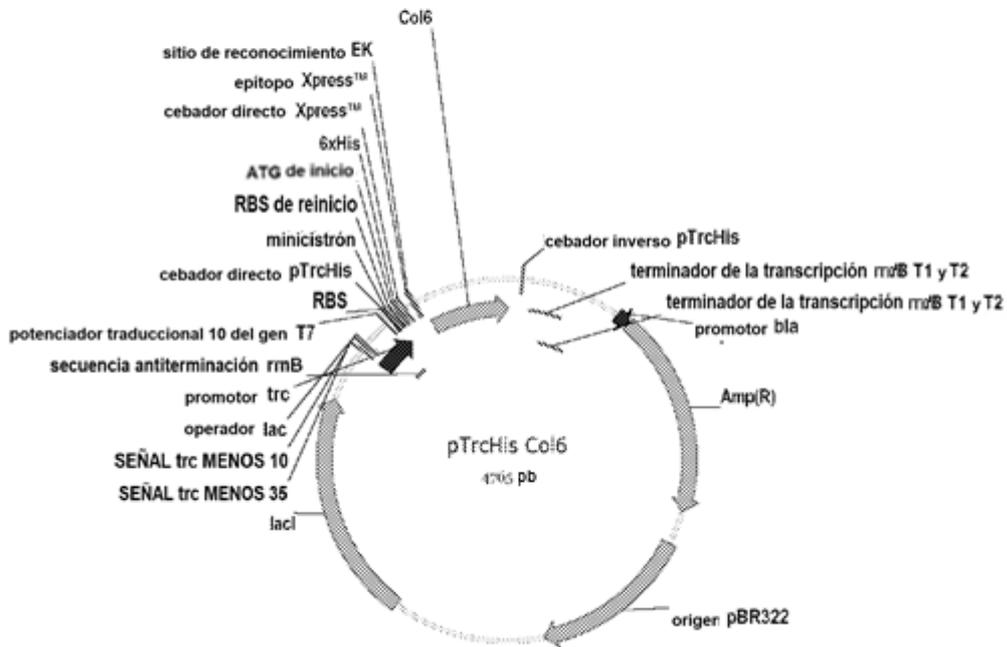


Fig.12F

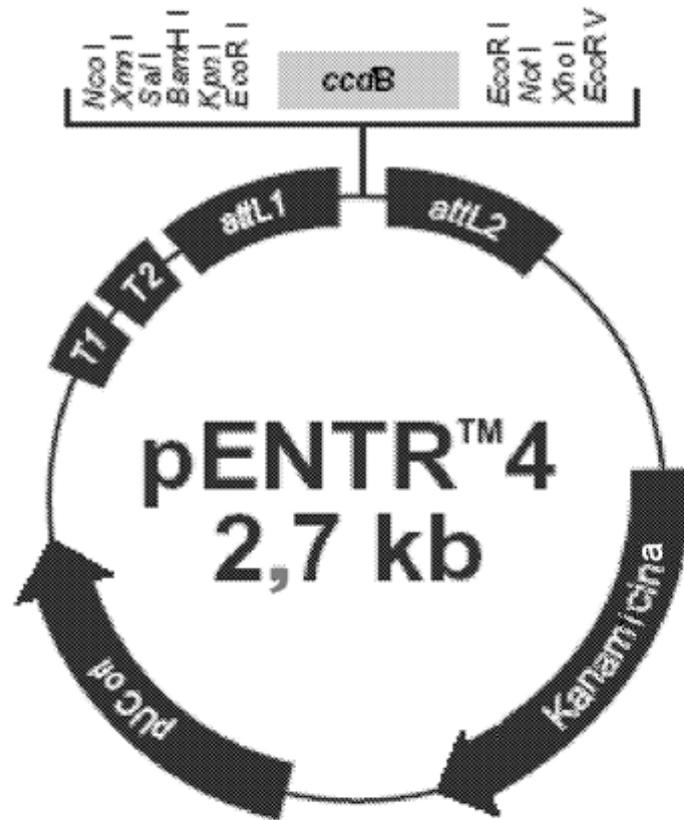


Fig.13A

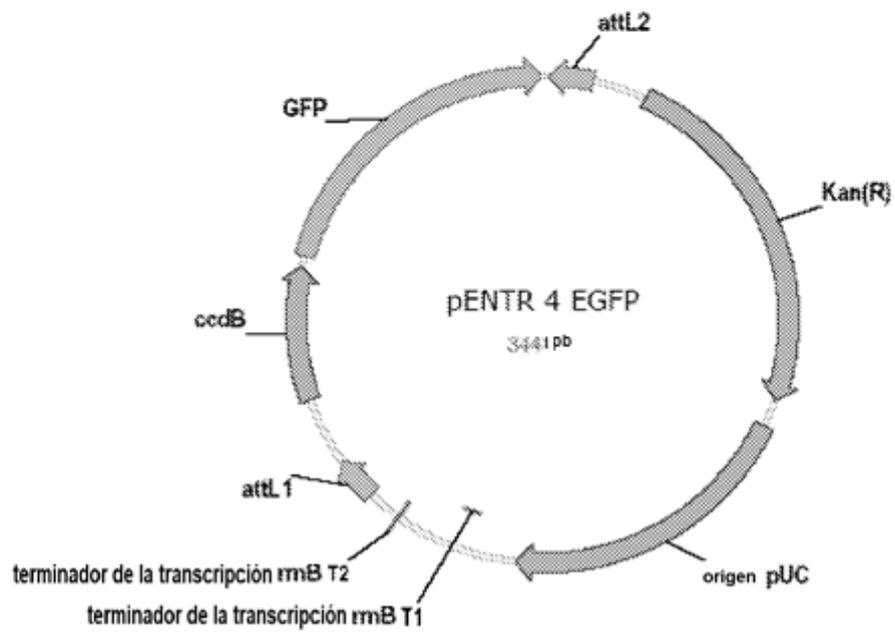


Fig.13A2

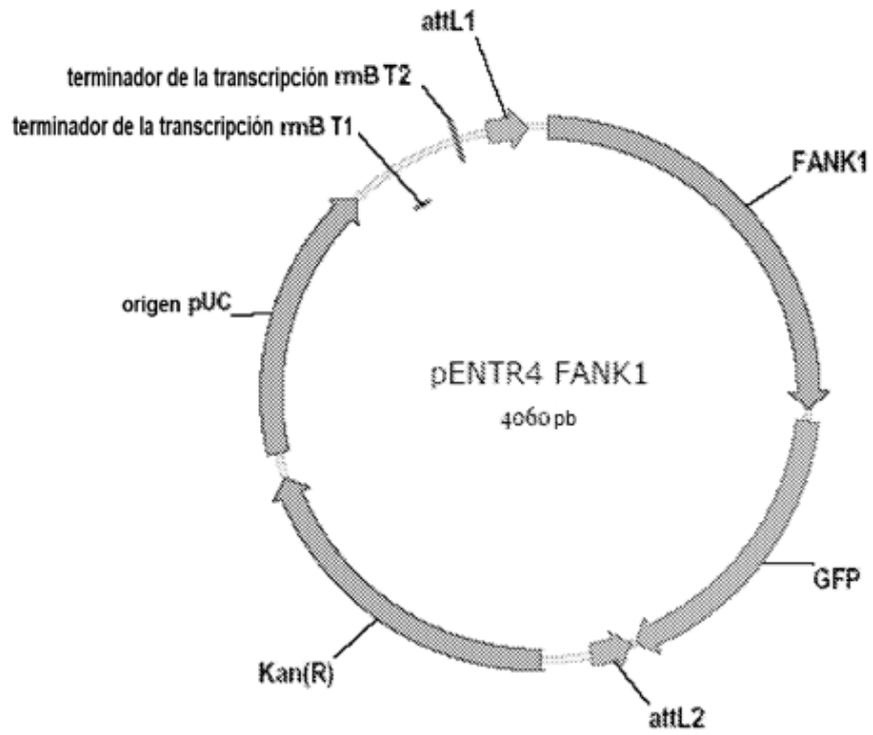


Fig.13B

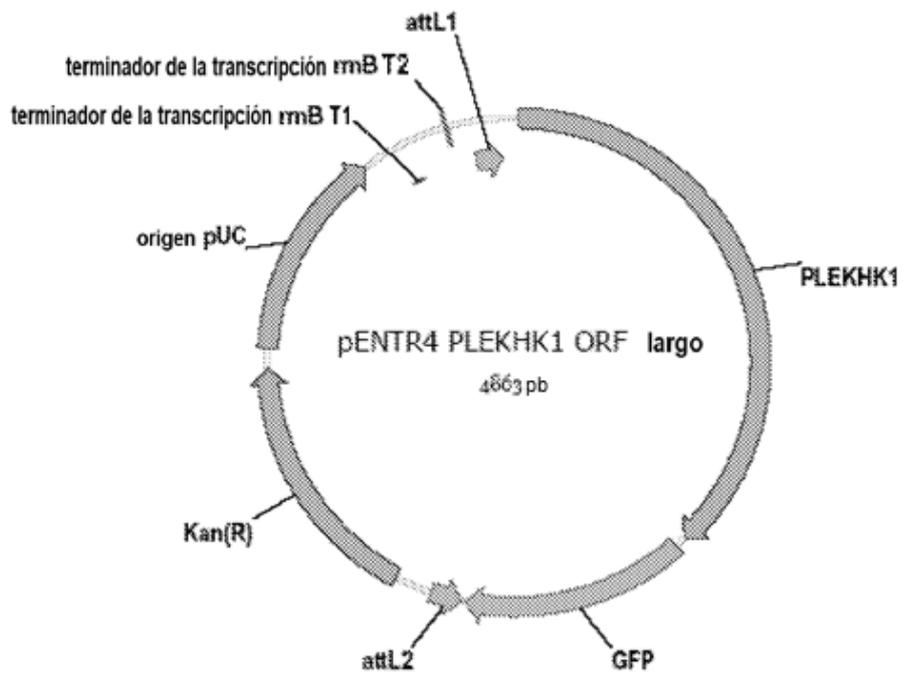


Fig.13C

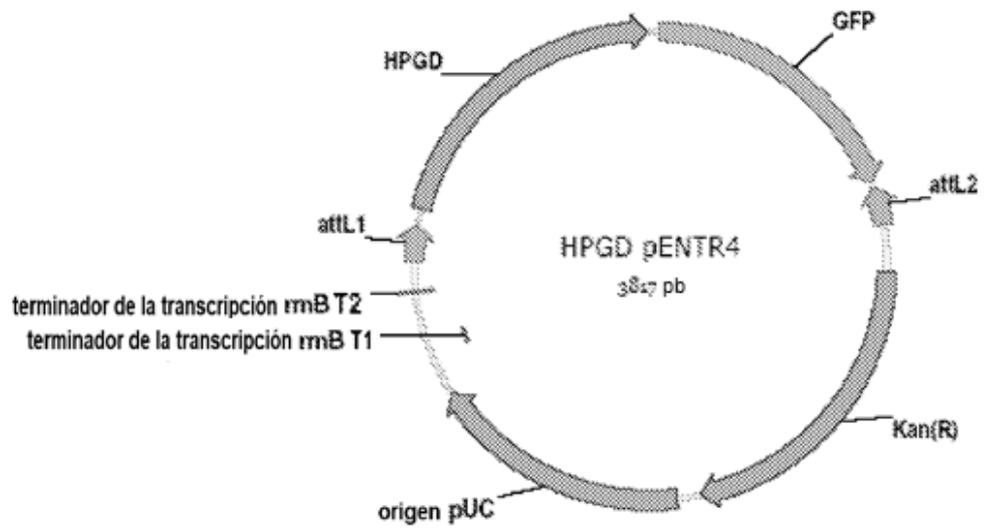


Fig.13D

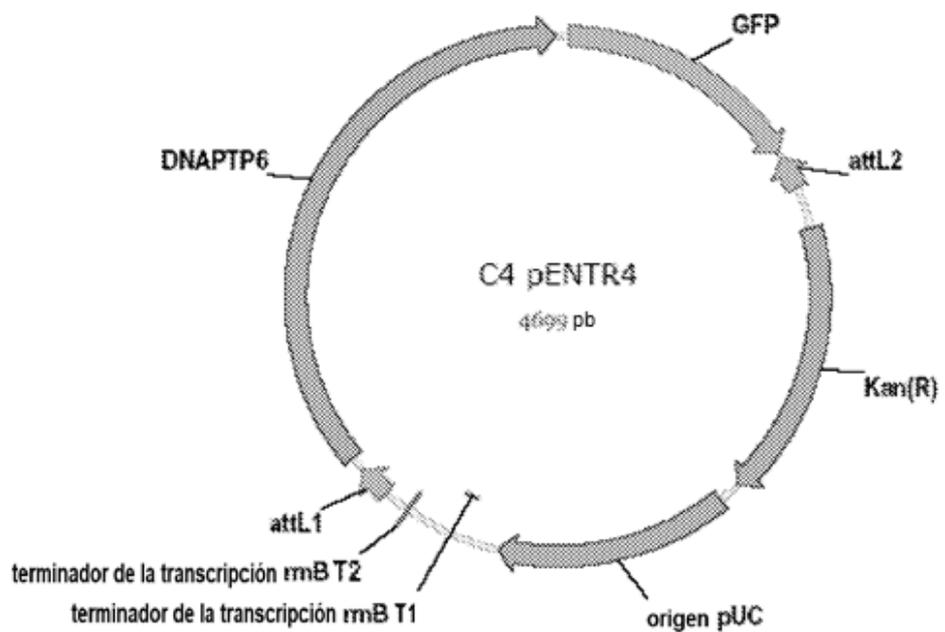


Fig.13E

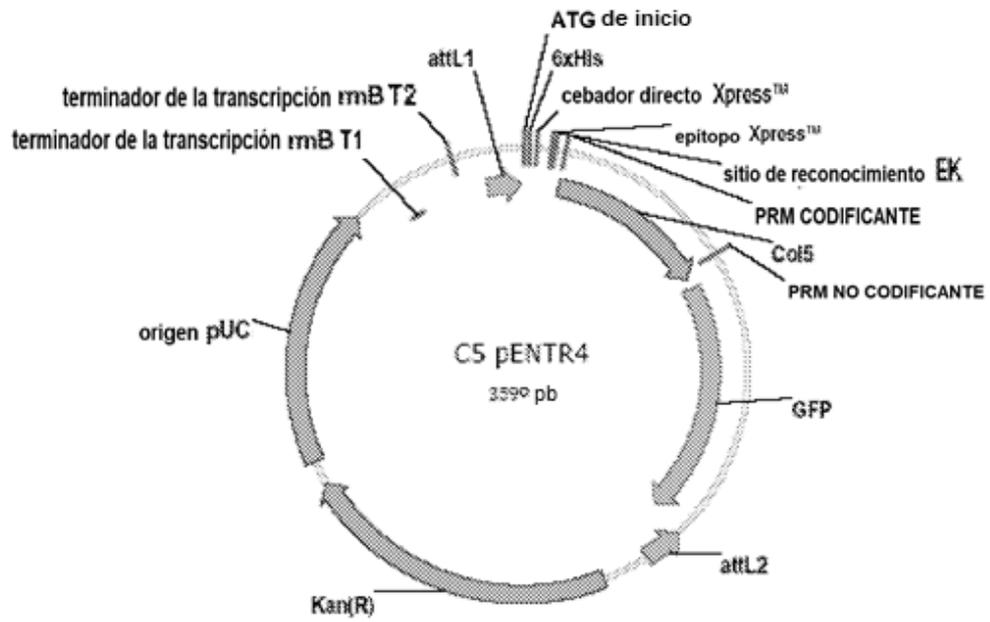


Fig.13F

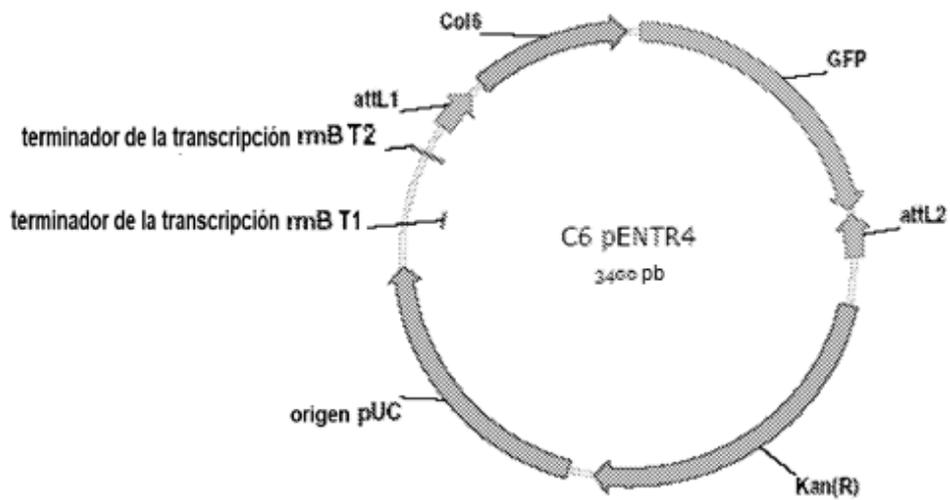


Fig.13G

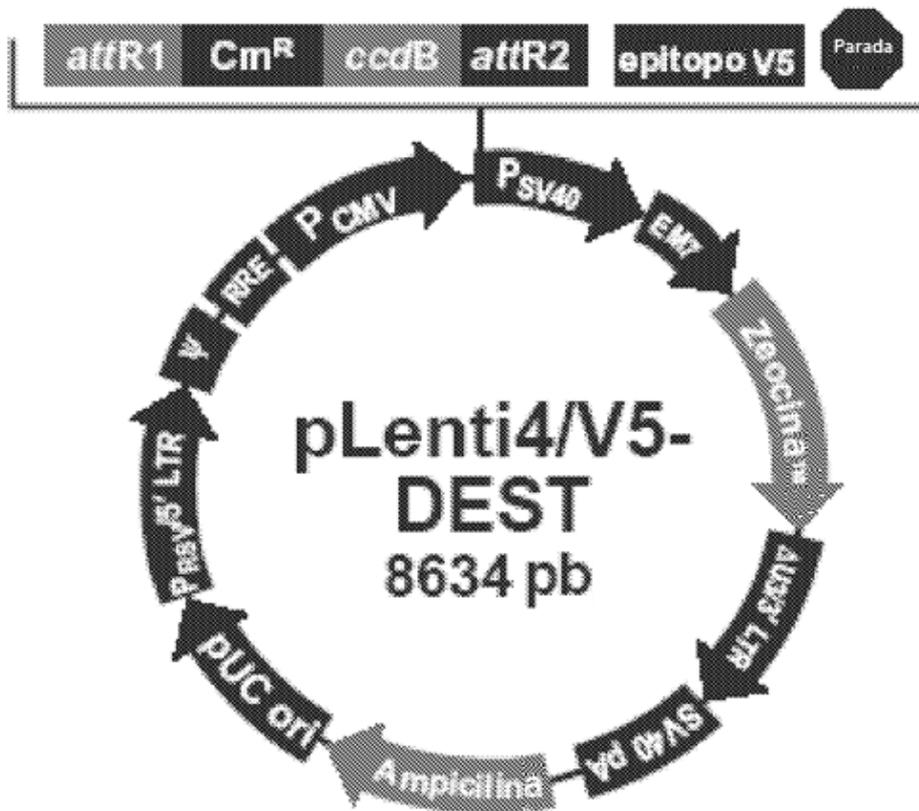


Fig.14A

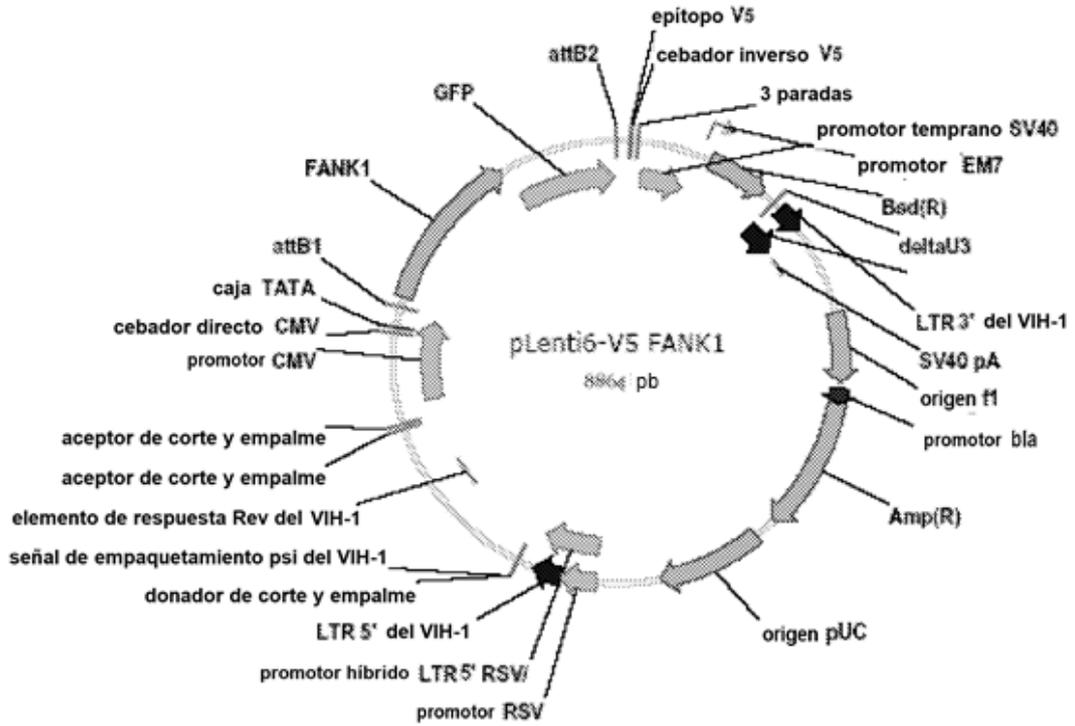


Fig.14B

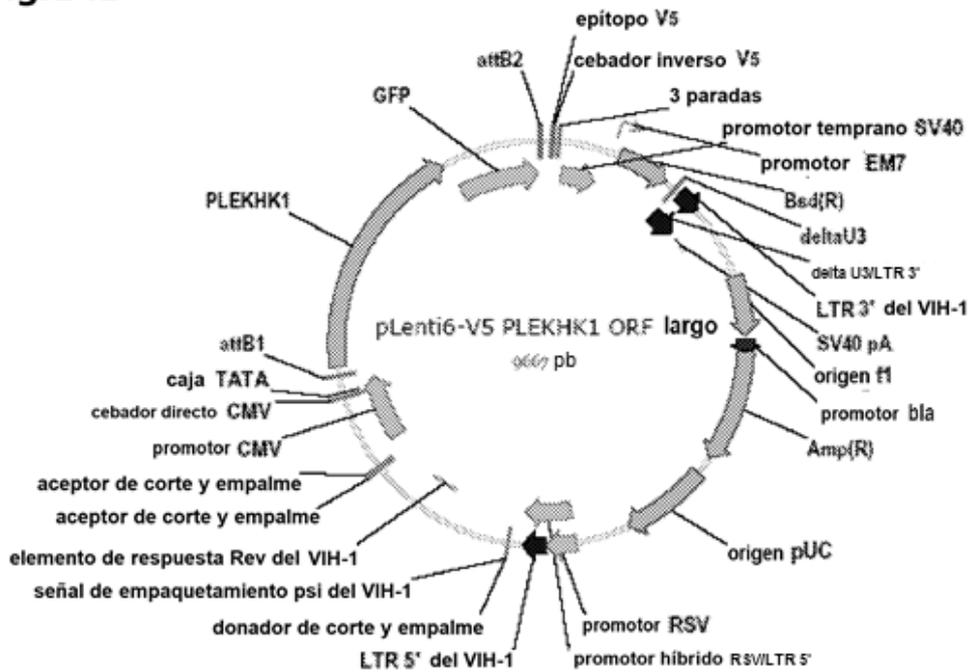


Fig.14C

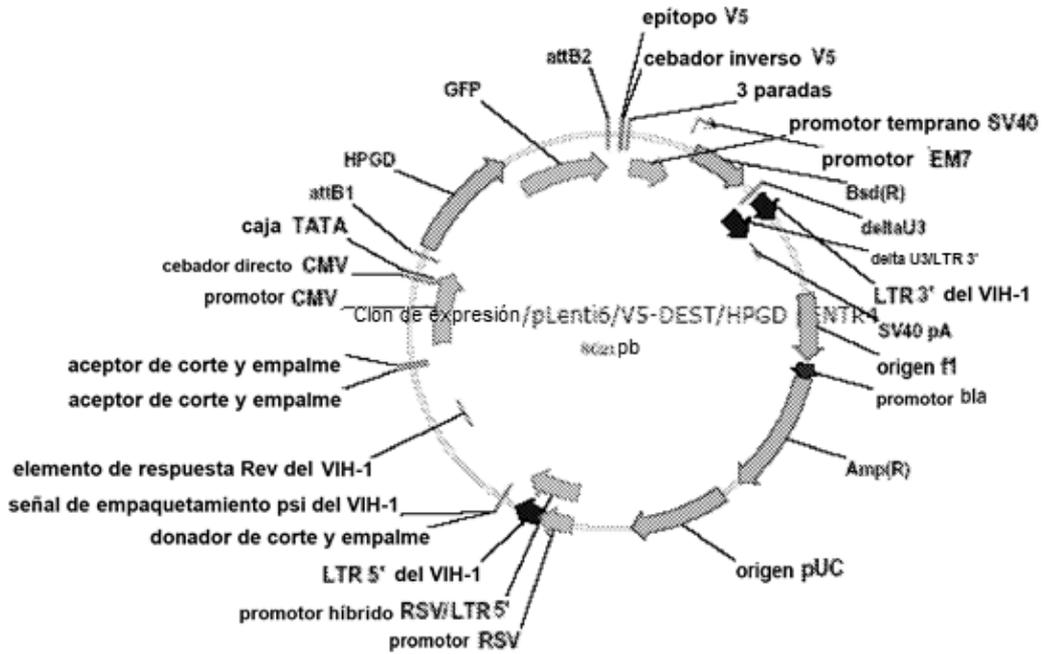


Fig.14D

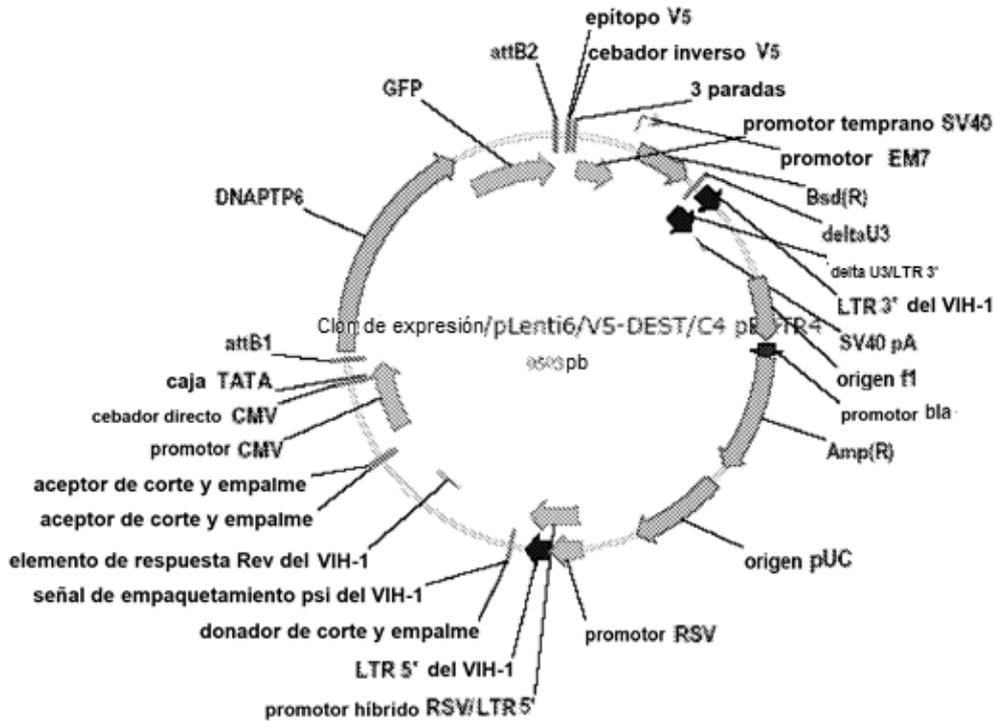


Fig.14E

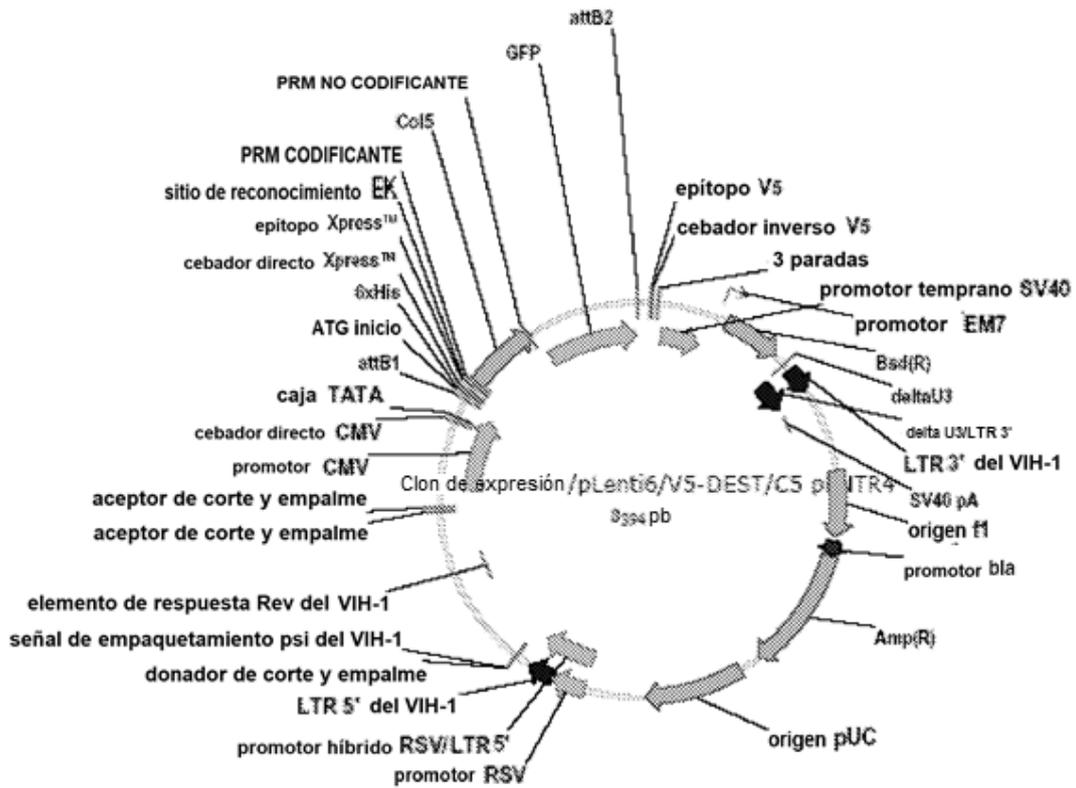


Fig.14F

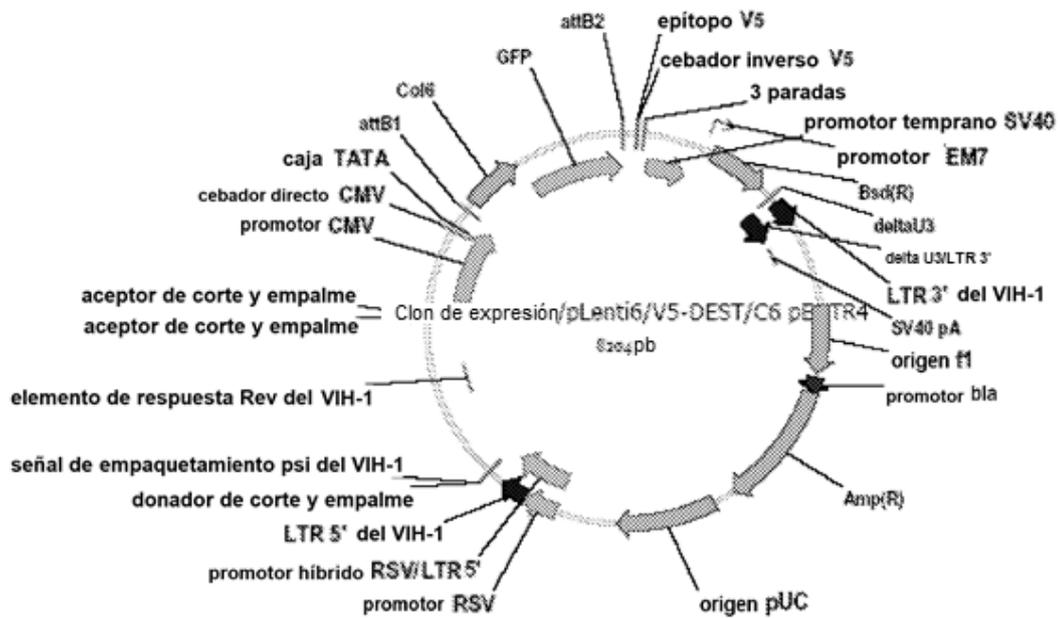


Fig.14G

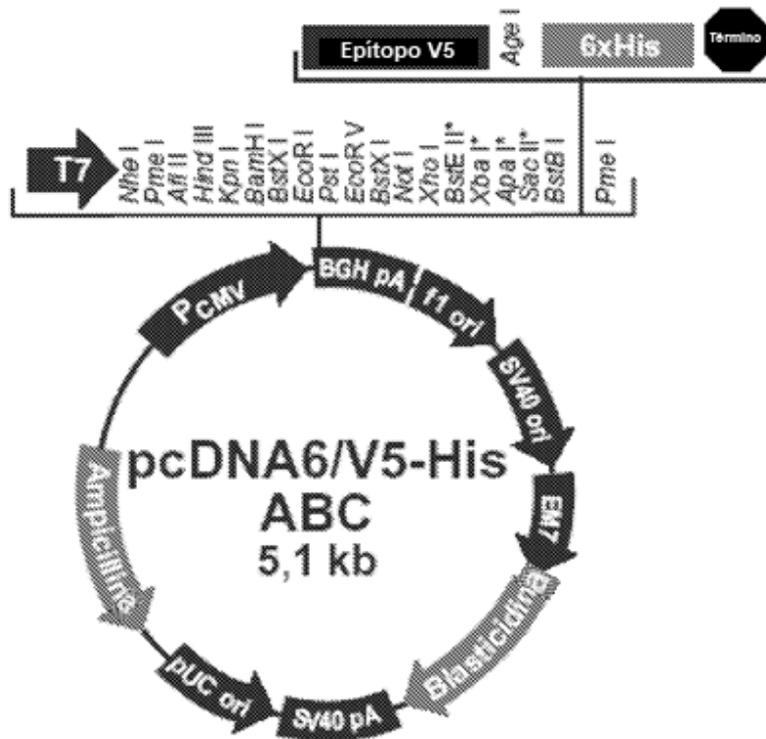


Fig.15A

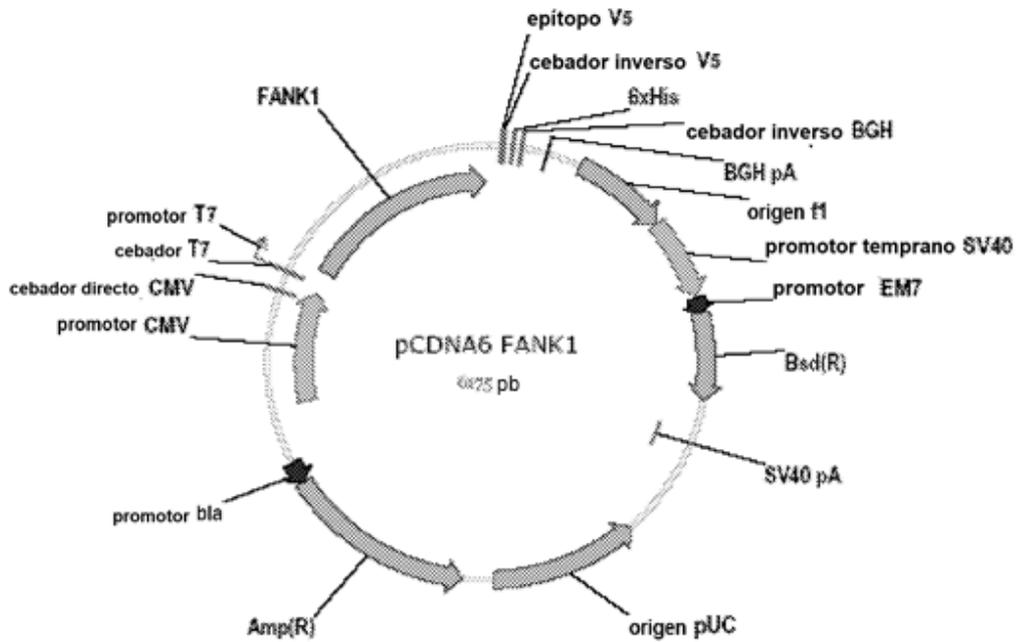


Fig.15B

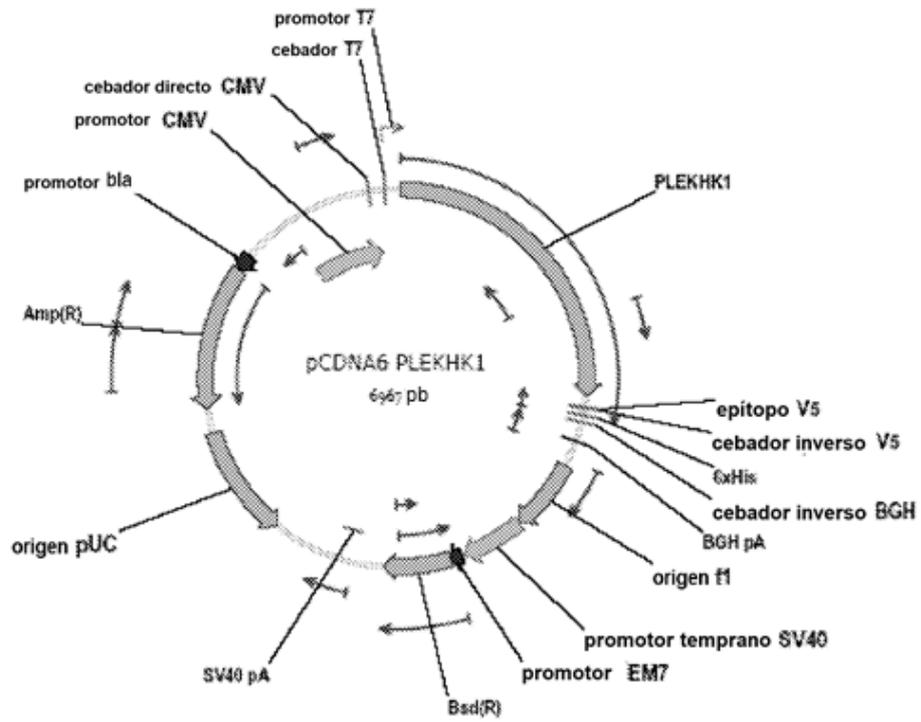


Fig.15C

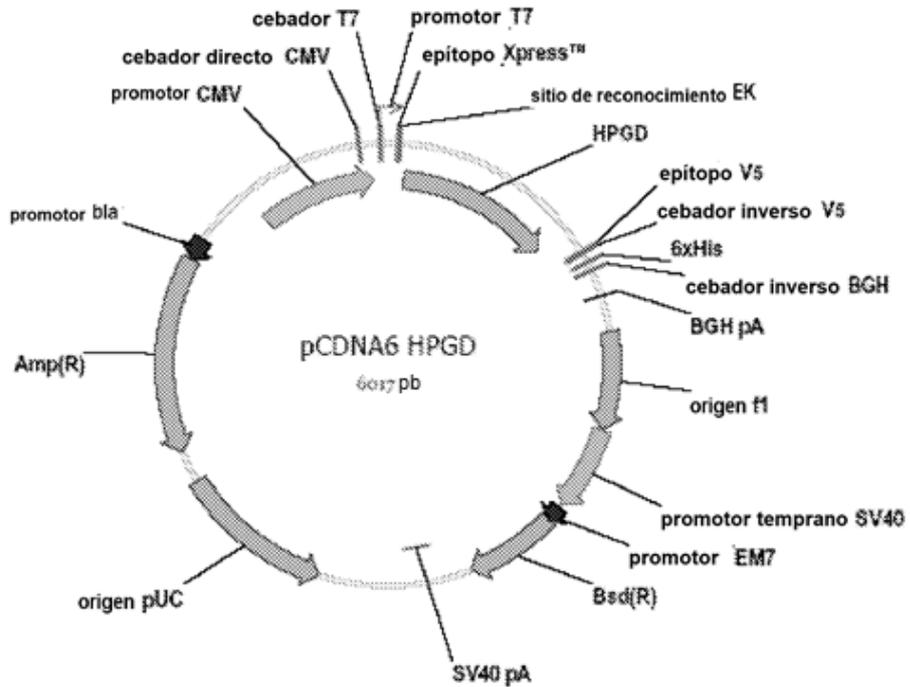


Fig.15D

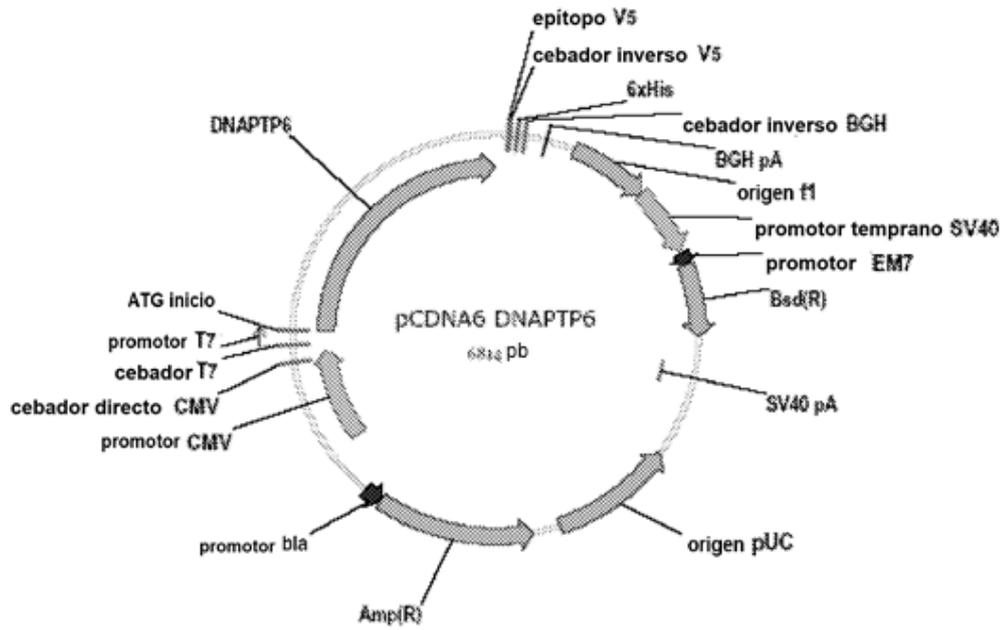


Fig.15E

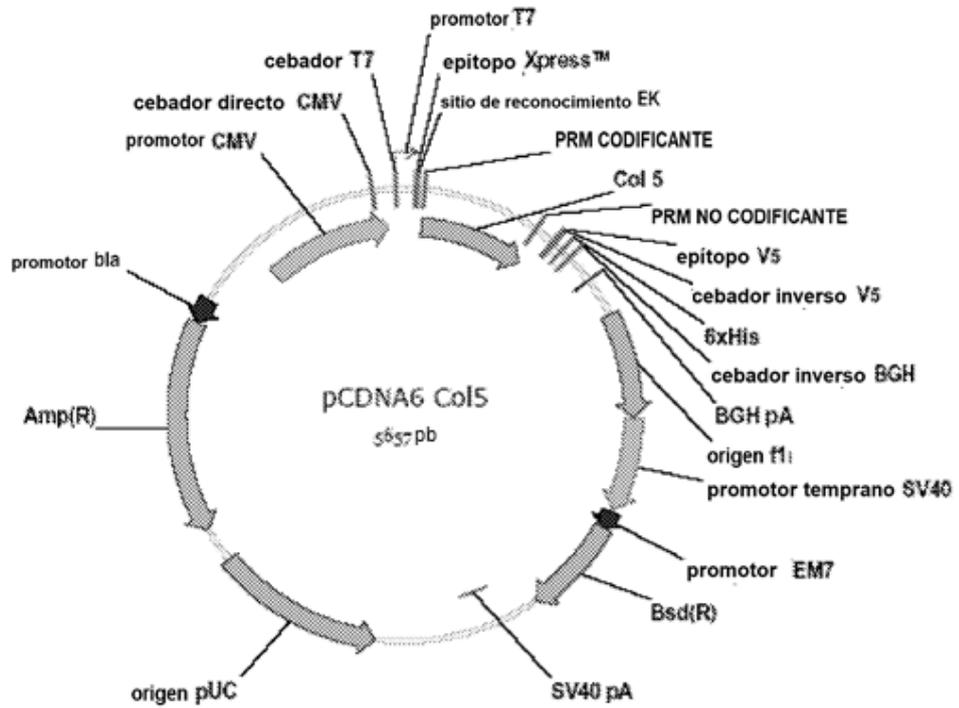


Fig.15F

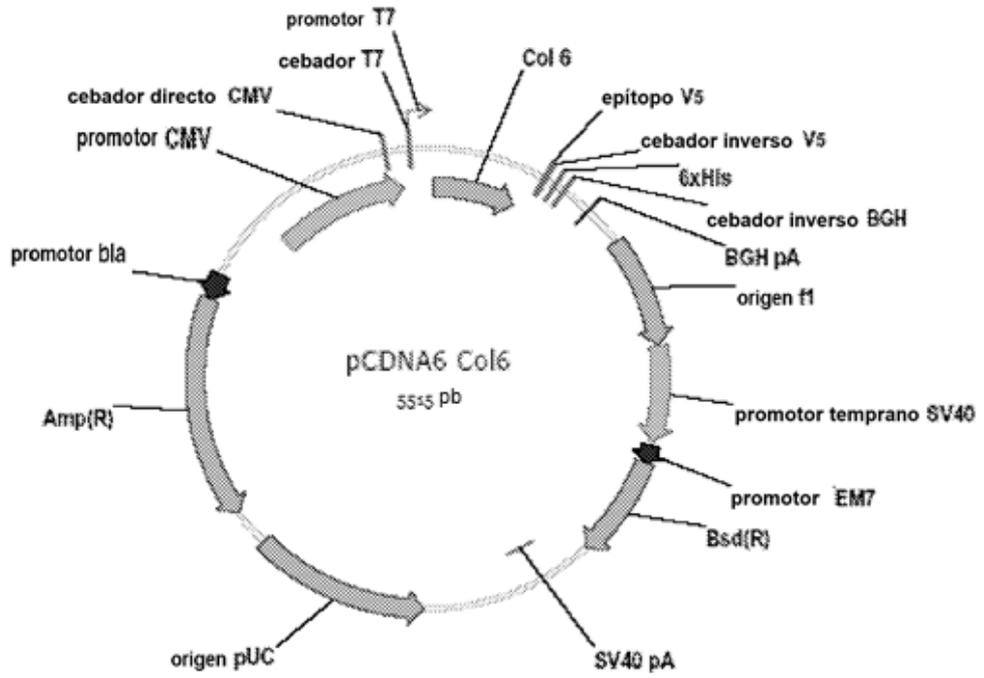


Fig.15G