

UNIVERSIDAD AUTÓNOMA DE MADRID  
FACULTAD DE CIENCIAS  
DEPARTAMENTO DE BIOLOGÍA



*Splicing* alternativo y quimerismo en genes del MHC de clase III. Relación de esta región con la Artritis Reumatoide.

*Alternative splicing and chimerism in MHC class III genes. Relation of this region to Rheumatoid Arthritis.*

Memoria presentada por:

**Raquel López Díez**

para optar al grado de Doctor en Ciencias por la  
Universidad Autónoma de Madrid

Trabajo dirigido por la **Dra Begoña Aguado Orea** y realizado en el  
Centro de Biología Molecular “Severo Ochoa” (UAM-CSIC).

**Madrid, 2014**

*DEPARTAMENTO DE BIOLOGÍA  
FACULTAD DE CIENCIAS  
UNIVERSIDAD AUTÓNOMA DE MADRID*

*Memoria presentada por Dña Raquel López Díez  
para optar al grado de Doctor por la  
Universidad Autónoma de Madrid*

*Directora: Dra. Begoña Aguado Orea*

*Tutor: Dr. José Miguel Hermoso Núñez*

*Departamento de Biología,  
Centro de Biología Molecular Severo Ochoa,  
U.A.M.-C.S.I.C.  
Madrid, 2014*

Este trabajo ha sido realizado en el Departamento de Biología de la Facultad de Ciencias y en el Centro de Biología Molecular Severo Ochoa (C.B.M.S.O.), U.A.M.-C.S.I.C., gracias a la ayuda de una beca de Formación para Personal Universitario de la Universidad Autónoma de Madrid.

# ÍNDICE

<b>Abreviaturas empleadas</b>	<b>vi</b>
<b>SUMMARY</b>	<b>ix</b>
<b>Publications</b>	<b>xv</b>
<b>Communications to meetings</b>	<b>xv</b>
<b>Submissions to the NCBI database</b>	<b>xv</b>
<b>1. INTRODUCCIÓN</b>	<b>1</b>
<b>1.1. EL FENÓMENO DEL <i>SPLICING</i> Y EL <i>SPLICING</i> ALTERNATIVO</b>	<b>3</b>
1.1.1. Mecanismo de <i>splicing</i>	4
1.1.1.1. Excepciones en el mecanismo de <i>splicing</i>	8
1.1.2. Regulación del <i>splicing</i>	9
1.1.2.1. Regulación del mecanismo de <i>splicing</i> alternativo	9
1.1.2.2. Regulación de los productos generados por <i>splicing</i> alternativo mediante Nonsense-Mediated Decay (NMD)	13
1.1.2.3. Noisy <i>splicing</i> o ruido transcripcional	14
1.1.3. Relación del AS con otros mecanismos	15
1.1.3.1. Transcripción y <i>splicing</i> alternativo	15
1.1.3.2. <i>Splicing</i> alternativo y traducción	16
1.1.4. Implicaciones funcionales del <i>splicing</i> alternativo	16
1.1.5. <i>Splicing</i> alternativo y evolución	19
1.1.6. <i>Splicing</i> alternativo y enfermedades	22
1.1.7. <i>Splicing</i> alternativo y nuevas tecnologías	24
<b>1.2. OTROS MECANISMOS DE REGULACIÓN DE LA EXPRESIÓN GÉNICA</b>	<b>26</b>
1.2.1. Quimerismo en <i>cis</i> y su mecanismo	26
1.2.2. <i>Trans-splicing</i> y su mecanismo	30
1.2.3. Implicaciones funcionales del quimerismo	32
1.2.4. Evolución del quimerismo	34
1.2.5. Asociación con enfermedades	34
1.2.6. Empleo de nuevas tecnologías	35
<b>1.3. COMPLEJO MAYOR DE HISTOCOMPATIBILIDAD (MHC) DE CLASE III</b>	<b>36</b>
1.3.1. El gen <i>RAGE</i> .	39
1.3.1.1. <i>Splicing</i> alternativo en <i>RAGE</i> e implicaciones funcionales	41
1.3.2. El gen <i>G6F</i> .	44
1.3.2.1. <i>Splicing</i> alternativo en <i>G6F</i> e implicaciones funcionales	45
1.3.3. El gen <i>LY6G6D</i> .	45
1.3.3.1. <i>Splicing</i> alternativo en <i>LY6G6D</i> e implicaciones funcionales	47
2.3.4. Quimeras entre los genes <i>G6F</i> y <i>LY6G6D</i> .	48
<b>1.4. ARTRITIS REUMATOIDE</b>	<b>49</b>
1.4.1. Relación de AR con el MHC de clase III	51
1.4.1.1. Relación de TNF $\alpha$ y NF $\kappa$ BIL1 con AR	51
1.4.1.2. Relación cluster de genes Ly6 con AR	51
1.4.1.3. Relación de <i>RAGE</i> con AR	52



<b>2. OBJETIVOS</b>	<b>53</b>
<b>3. MATERIALES Y MÉTODOS</b>	<b>57</b>
<b>3.1. ANÁLISIS BIOINFORMÁTICO</b>	<b>59</b>
3.1.1. Bases de datos	59
3.1.2. Diseño y chequeo de cebadores	59
3.1.3. Alineamiento de secuencias	60
3.1.4. Estudio a nivel proteico de las variantes de <i>splicing</i>	61
3.1.5. RNA-seq	62
3.1.6. Análisis de los resultados de qPCR	63
<b>3.2. MUESTRAS DE PARTIDA Y PROCESAMIENTO INICIAL</b>	<b>63</b>
3.2.1. Muestras de sangre de pacientes y controles	63
3.2.2. Muestras de diferentes especies de mamíferos	65
3.2.3. Procesamiento muestras de sangre de las diferentes especies y separación de las diferentes fracciones de interés	68
<b>3.3. TÉCNICAS DE LABORATORIO DE BIOLOGÍA MOLECULAR</b>	<b>68</b>
3.3.1. Técnicas de RNA	68
3.3.1.1. Extracción, precipitación y determinación de la calidad del RNA	68
3.3.1.2. Síntesis de DNA complementario (cDNA)	69
3.3.1.3. RLM-RACE	70
3.3.2. Técnicas de DNA	71
3.3.2.1. Hibridación de cebadores	71
3.3.2.2. PCR	72
3.3.2.3. Electroforesis en geles de agarosa	81
3.3.2.4. Purificación de banda de gel o de PCR	81
3.3.2.5. Digestión	81
3.3.2.6. Ligación	82
3.3.2.7. Cultivos bacterianos	83
3.3.2.8. Transformación	84
3.3.2.9. Purificación de plásmidos y báculos. Secuenciación	85
3.3.2.10. Análisis de las variantes de <i>splicing</i>	86
3.3.2.11. Clonajes	86
3.3.3. Cultivos celulares	91
3.3.3.1. Mantenimiento de los cultivos celulares	91
3.3.3.2. Transfección transitoria	92
3.3.3.3. Infección por baculovirus	93
3.3.3.4. Chequeo de las transfecciones e infecciones	94
3.3.4. Técnicas de proteínas	94
3.3.4.1. Preparación de las muestras	94
3.3.4.2. Electroforesis en geles de acrilamida	95
3.3.4.3. Western blot	95
3.3.4.4. Tinción de coomasie	97
3.3.4.5. Expresión de proteínas recombinantes mediante baculovirus	97
3.3.4.6. Estudio de modificaciones post-traduccionales	99
3.3.4.7. Ensayos de interacción proteica	101
3.3.4.8. Inmunofluorescencias	102

<b>4. RESULTADOS</b>	<b>105</b>
<b>4.1. ESTUDIO DEL <i>SPLICING</i> EN EL GEN <i>RAGE</i></b>	<b>107</b>
4.1.1. Análisis de las variantes de <i>splicing</i> del gen <i>RAGE</i> en diferentes especies de mamífero mediante RT-PCR	107
4.1.1.1. Identificación y caracterización de las variantes de <i>splicing</i> de <i>RAGE</i> humanas	107
4.1.1.2. Identificación y caracterización de las variantes de <i>splicing</i> de <i>RAGE</i> de el mono Rhesus	110
4.1.1.3. Identificación y caracterización de las variantes de <i>splicing</i> de <i>RAGE</i> de la especie bovina	113
4.1.1.4. Identificación y caracterización de las variantes de <i>splicing</i> de <i>RAGE</i> de cerdo	115
4.1.1.5. Identificación y caracterización de las variantes de <i>splicing</i> de <i>RAGE</i> de ratón	116
4.1.1.6. Identificación y caracterización de las variantes de <i>splicing</i> de <i>RAGE</i> de rata	118
4.1.2. Análisis de las variantes de <i>splicing</i> del gen <i>RAGE</i> en primates	120
4.1.3. Identificación de las variantes de <i>splicing</i> de <i>RAGE</i> mediante técnicas de secuenciación masiva	124
4.1.4. Caracterización de las principales variantes de <i>splicing</i> de <i>RAGE</i> humanas a nivel proteico	131
<b>4.2. ESTUDIO DEL FENÓMENO DE QUIMERISMO ENTRE LOS GENES <i>G6F</i> Y <i>LY6G6D</i> COMO MODELO</b>	<b>134</b>
4.2.1. Análisis de la presencia de quimeras en los genes <i>G6F</i> y <i>LY6G6D</i> en diferentes especies de mamífero.	134
4.2.1.1. Identificación y caracterización de <i>G6F</i> y sus variantes de <i>splicing</i> .	135
4.2.1.2. Identificación y caracterización de <i>LY6G6E</i> y sus variantes de <i>splicing</i> .	136
4.2.1.3. Identificación y caracterización de <i>LY6G6D</i> y sus variantes de <i>splicing</i> .	138
4.2.1.4. Identificación y caracterización de los transcritos quiméricos <i>G6F-LY6G6D</i> .	138
4.2.2. Caracterización de las principales variantes quiméricas a nivel proteico	140
<b>4.3. ESTUDIO DE LA ENFERMEDAD AR EN EL MHC DE CLASE III</b>	<b>145</b>
4.3.1. Identificación de una pareja de normalizadores adecuados para el estudio.	147
4.3.2. Estudio de la expresión génica de los genes del MHC de clase III y otros genes involucrados en inflamación en pacientes de AR.	149
<b>5. DISCUSIÓN</b>	<b>157</b>
5.1. Metodologías empleadas para el estudio de las variantes de <i>splicing</i> de un gen, el caso del gen <i>RAGE</i> .	159
5.2. Implicaciones funcionales del AS en la expresión del gen <i>RAGE</i> .	161
5.3. El mecanismo de quimerismo entre dos genes. La frontera entre la definición de gen, fenómenos de <i>splicing</i> y la generación de quimeras.	166
5.4. Implicaciones funcionales del AS y quimerismo en los genes <i>G6F</i> y <i>LY6G6D</i> .	169
5.5. La enfermedad AR en la región del MHC de clase III	171
<b>6. CONCLUSIONES</b>	<b>175</b>
<b>7. BIBLIOGRAFÍA</b>	<b>179</b>
<b>8. DATOS SUPLEMENTARIOS</b>	<b>195</b>
<b>9. ANEXOS</b>	<b>215</b>

## Abreviaturas empleadas

---

<b>AANE:</b> Aminoácidos no esenciales	<b>FR:</b> Factor Reumatoide
<b>AINE:</b> antiinflamatorio no esteroideo	<b>GAPDH:</b> gliceraldehído 3-fosfato deshidrogenasa
<b>Anti-CCP:</b> Anticuerpos anti-péptidos cíclicos citrulinados	<b>gDNA:</b> DNA genómico
<b>AoSMCs:</b> <i>Aortic smooth muscle cells</i> o células del músculo liso de la aorta.	<b>GO:</b> <i>Gene Ontology</i>
<b>APS:</b> <i>Ammonium Persulfate</i> o persulfato amónico	<b>GTC:</b> <i>Guanidinium thiocyanate</i> o tiocianato de guanidina
<b>AR:</b> Artritis Reumatoide	<b>GPI:</b> <i>glycosylphosphatidylinositol</i> o glicosilfosfatidilinositol
<b>AS:</b> <i>Alternative Splicing</i> o <i>Splicing</i> alternativo	<b>Hepes:</b> ácido N-2-hidroxiethylpiperacina-N'-2-etanosulfónico
<b>AU:</b> <i>Aury Units</i>	<b>HLA:</b> <i>Human Leukocyte Antigen</i>
<b>BF:</b> Factor B	<b>IP:</b> Inmunoprecipitación
<b>BLAST:</b> <i>Basic Local Alignment Search Tool</i>	<b>IPTG:</b> isopropiltio-β-D-galactósido
<b>BME:</b> β-mercaptoetanol	<b>ISE:</b> <i>Intronic Splicing Enhancer</i>
<b>BSA:</b> <i>Bovine Serum Albumin</i> o Albúmina de suero bovina	<b>ISS:</b> <i>Intronic Splicing Silencer</i>
<b>cDNA:</b> ácido desoxirribonucleico complementario	<b>kpb:</b> kilopares de bases
<b>CoIP:</b> Coinmunoprecipitación	<b>LB:</b> <i>Lysogeny broth</i>
<b>Ct:</b> <i>threshold cycle</i>	<b>M24:</b> placa de 24 pocillos
<b>DAS:</b> <i>Disease Activity Score</i>	<b>MCS:</b> <i>Multiple Cloning Site</i> o Sitio de clonaje múltiple
<b>DMEM:</b> Dulbecco's Modified Eagles Medium	<b>MHC:</b> <i>Major Histocompatibility Complex</i> o Complejo Mayor de Histocompatibilidad
<b>DMSO:</b> dimetilsulfóxido	<b>mRNA:</b> RNA mensajero
<b>DNA:</b> ácido desoxirribonucleico	<b>MW:</b> <i>Molecular weight</i> o Masa molecular
<b>dNTP:</b> desoxirribonucleósido trifosfato	<b>ncRNA:</b> <i>non coding RNA</i> , o RNA no codificante
<b>EDTA:</b> ácido etilendiaminotetraacético	<b>NMD:</b> <i>Nonsense-mediated decay</i>
<b>ENCODE:</b> <i>ENCYclopedia Of DNA Elements</i>	<b>NP40:</b> Nonidet P40
<b>ESE:</b> <i>ExonicSplicing Enhancer</i>	<b>O/N:</b> <i>over night</i>
<b>ESS:</b> <i>Exonic Splicing Silencer</i>	<b>ORF:</b> <i>Open Reading Frame</i> o Fase de lectura abierta
<b>EST:</b> <i>Expressed Sequence Tags</i>	<b>pb:</b> pares de bases
<b>FAME:</b> Fármaco Antirreumático Modificador de la Enfermedad	<b>PBMC:</b> <i>Peripheral Blood Mononuclear Cell</i> o células mononucleares de sangre periférica
<b>FBS:</b> <i>Fetal Bovine Serum</i> o suero fetal bovino	<b>PBS:</b> <i>Phosphate-Buffered Saline</i>
<b>FPKM:</b> <i>Fragments Per Kilobase of exon per Million fragments mapped</i>	<b>PCR:</b> <i>Polymerase Chain Reaction</i>

**PFA:** Paraformaldehido  
**PI-PLC:** *Phosphatidylinositol Phospholipase C*  
**PNGaseF:** *peptide N-glycosidase F* o péptido N-glicosidasa F.  
**PTC:** *Premature Termination Codon* o codón de terminación de la traducción prematuro  
**qPCR:** PCR cuantitativa  
**qRT-PCR:** RT-PCR cuantitativa  
**RE:** retículo endoplasmático  
**RFU:** *Relative Fluorescent Units* o Unidades de Fluorescencia Relativa  
**RIN:** *RNA Integrity Number* o Número de integridad del RNA  
**RNA:** ácido ribonucleico  
**r.p.m.:** revoluciones por minuto  
**RPMI:** *Roswell Park Memorial Institute medium*  
**RPR:** Reagina Plasmática Rápida  
**rRNA:** RNA ribosómico  
**RT:** retrotranscripción  
**RT-PCR:** ver abreviaturas RT y PCR  
**SDS:** *sodium dodecyl sulfate* o dodecilsulfato sódico  
**SIDA:** Síndrome de Inmunodeficiencia Adquirida  
**SRA:** *Sequence Read Archive*  
**STA:** *Specific Target Amplification*  
**TEMED:** N, N, N', N'-tetrametilendiamina  
**TIC:** *Transcription Induced Chimerism*  
**Tris:** tris-(hidroximetil)aminometano  
**tRNA:** RNA de transferencia  
**U:** Unidades (de enzima)  
**UTR:** *Untranslated region*  
**X-gal:** 5-bromo-4-cloro-3-indolil-β-D-galactopiranosido

## **SUMMARY**

Post-transcriptional regulations such as Alternative Splicing (AS) and Transcription Induced Chimerism (TIC) are mechanisms that expand transcriptome and proteome diversity in higher eukaryotes. These mechanisms allow the generation of different RNAs and multiple protein isoforms. Several Major Histocompatibility Complex (MHC) class III genes, such as the *Receptor for Advanced Glycosylation End Product (RAGE)* gene and the *G6F* and Leukocyte antigen *LY6G6D* genes, have been previously studied in relation to these mechanisms at RNA level, as well as some isoforms encoded by them. Moreover, this region has been related to several diseases, including Rheumatoid Arthritis (RA), although these associations have not been fully studied.

In this study we examined in detail the different RAGE, G6F, LY6G6D and chimera G6F-LY6G6D mRNAs generated by both AS and/or chimerism on different mammals from different tissues and development stages, as well as in tumor samples. By nested RT-PCR, we identified a high number of different splice variants including non-coding and predicted coding ones. In addition, the two different chimera transcripts, found between the *G6F* and *LY6G6D* genes, were conserved among mammals presenting also AS in the analysed tissues. However, analysis of RNA-seq data enabled the detection of the most abundant RAGE splice variants only when default Cufflinks parameters were altered and none of the chimera transcripts were detected, probably due to the low expression of these three genes. Over-expression of some of these protein isoforms have indicated that differences in the protein domains affect their post-translational modifications, their subcellular localization and possibly the potential ligand binding and intracellular signaling of the studied receptors. In addition, we have been able to detect the formation of G6F and Igch heterodimers when both proteins were over-expressed. In the other hand, none of these studied splice events presented an altered expression levels when RNA samples from RA patients were analyzed by nano-quantitative RT-PCR. On the contrary, the expression levels of some other MHC III genes were found altered using this method in these same RA samples compared to healthy samples, or even between samples from RA patients with the disease in remission or active.

In conclusion, alternative splicing and chimerism are involved in the generation of a considerable RNA and protein diversity of *RAGE*, *G6F* and *LY6G6D* genes in different tissues and mammalian species.

Las regulaciones post-transcripcionales tales como *splicing* alternativo (AS) y quimerismo (TIC) son mecanismos que expanden la diversidad transcriptómica y proteómica en eucariotas superiores. Estos mecanismos permiten la generación de diferentes RNAs y múltiples isoformas proteicas. Se han estudiado previamente diversos genes localizados en el complejo mayor de histocompatibilidad (MHC) de clase III, tales como el gen Receptor de Agentes de Glicosilación Avanzada (*RAGE*) y los genes *G6F* y *LY6G6D*, en relación a estos mecanismos al nivel de RNA, así como algunas isoformas codificadas por ellos. Además, esta región ha sido relacionada con varias enfermedades, entre las que se incluye la artritis reumatoide (RA), aunque estas asociaciones no han sido detalladamente estudiadas.

En este estudio hemos examinado en detalle los diferentes mRNAs de *RAGE*, *G6F*, *LY6G6D* y de la quimera *G6F-LY6G6D* que han sido generados por AS y/o por quimerismo en diferentes tejidos y estadios de desarrollo, así como en muestras de tumores, en diferentes mamíferos. Por RT-PCR anidada, hemos identificado un elevado número de variantes de *splicing* diferentes, entre las que se incluyen no codificantes y potencialmente codificantes según predicciones bioinformáticas. Además, se encontraron dos tipos diferentes de transcritos quiméricos entre los genes *G6F* y *LY6G6D*, los cuales estaban conservados en mamíferos y presentaban también eventos de AS en los tejidos analizados. Sin embargo, los análisis de los datos de RNA-seq permitieron solo la detección de las variantes de *splicing* más abundantes de *RAGE* cuando los parámetros por defecto de Cufflinks fueron alterados, en cambio no fue detectado ninguno de los transcritos quiméricos, posiblemente debido a los bajos niveles de expresión de estos genes. La sobreexpresión de alguna de estas isoformas ha indicado que las diferencias en los dominios proteicos afecta tanto las modificaciones post-traduccionales de las isoformas resultantes, como a su localización y posiblemente la unión por su potencial ligando potencial y la señalización intracelular de los receptores estudiados. Además, hemos sido capaces de detectar la formación de heterodímeros entre la proteína canónica *G6F* y la proteína quimérica Igch cuando ambas proteínas eran sobre-expresadas. Por otro lado, ninguno de los eventos de *splicing* estudiados presenta unos niveles de expresión alterados cuando analizamos muestras de RNA de pacientes con RA por nanotecnología de RT-PCR cuantitativa. Por el contrario, usando este método encontramos alterados en estas muestras de RA los niveles de expresión de otros genes localizados en el MHC de clase III, o incluso entre muestras de pacientes con AR con la enfermedad en remisión o activa.

Para concluir, el *splicing* alternativo como los fenómenos de quimerismo se han visto involucrados en la generación de una gran diversidad a nivel de RNA y de proteína tanto en *RAGE* como en *G6F*, *LY6G6D* y *G6F-LY6G6D* en diferentes tejidos y especies.

## Publications

---

**López-Díez R.**, Rastrojo A., Villate O and Aguado B. (2013). Complex tissue-specific patterns and distribution of multiple RAGE splice variants in different mammals. *Genome Biol. Evol.*, 5(12):2420-35.

Villate O., Rastrojo A., **López-Díez R.**, Hernández-Torres F. and Aguado B. (2008). Differential splicing, disease and drug targets. *Infect. Disord. Drug Targets*, 8(4):241-51. Review.

**López-Díez R.**, Rastrojo A., Hernández-Torres F., and Aguado B. Alternative splicing and transcription induced chimerism in G6F and Ly6G6D among mammals. *To be submitted*.

## Communications to meetings

---

López-Muñoz A.D., **López-Díez R.**, Rastrojo A., and Aguado B. (2013). “Caracterización funcional de la variante de splicing Tv2 del gen *RAGE* y sus posibles implicaciones en procesos inflamatorios”. Poster. *XXXVI Congreso SEBBM-50 Aniversario*, Madrid, Spain.

**López-Díez R.**, and Aguado B. (2010). “Identification of several non-coding splice variants of *RAGE* in different mammals”. Poster in the *symposium “The non-coding genome”*, EMBL-EMBO, Heidelberg, Germany.

Villate O., Rastrojo R., **López-Díez R.**, Hernández-Torres F, and Aguado B. (2008). “*Splicing* alternativo en la región de clase III del MHC”. Functional genomic group, *Red Valenciana de Genómica y Proteómica*.

**López-Díez R.**, Villate O. and Aguado B. (2008). “Estudio y caracterización de las formas de *splicing* alternativo de *RAGE*, diversidad genómica en mamíferos”. Poster and oral presentation in “Genome and Proteomic” group. *XXXI Congreso SEBBM*, Bilbao, Spain.

## Submissions to the NCBI database

---

All Sanger-sequences described in this thesis have been deposited in GenBank (NCBI database) under the following accession numbers:

✓ RAGE cDNA sequences: *Homo sapiens* (GU139370-84), *Macaca mulatta* (GU139408-22), *Bos taurus* (GU139385-407), *Sus scrofa* (GU164706-14), *Mus musculus* (GU164727-39) and *Rattus norvegicus* (GU164715-26).



## **1. INTRODUCCIÓN**

La variabilidad transcriptómica y proteómica presente en cada organismo nos confiere nuestra identidad de especie y nos asemeja o distingue del resto de las especies. Esta afirmación ampliamente aceptada en la actualidad mantuvo en incertidumbre a la comunidad científica durante décadas, ya que a pesar de presentar un número muy similar de genes entre los diferentes organismos son notorias las diferencias que les caracterizan. A este hecho se podría sumar la incongruencia dada por la presencia de una cantidad particularmente superior de DNA genómico en comparación con el RNA transcrito por cada célula (incluso si se descartaban las secuencias repetidas presentes en el DNA genómico), así como la presencia de moléculas de RNA en el núcleo de un tamaño inusualmente superior al RNA presente en el citoplasma, presentando ambas tanto la estructura del *cap* en el 5' terminal como la cola de poliA en el 3' terminal que caracterizan a las moléculas de RNA mensajero. Estos tres hechos en conjunto proporcionaban indicios de la presencia de un mecanismo post-transcripcional totalmente desconocido hasta ese momento, cuya resolución se produjo con el descubrimiento de la presencia de genes fraccionados y el RNA *splicing*.

### **1.1. EL FENÓMENO DEL *SPLICING* Y EL *SPLICING* ALTERNATIVO**

La síntesis de proteínas se realiza a partir de precursores de RNA que tienen que ser modificados mediante el proceso denominado *splicing*, mecanismo post-transcripcional de maduración del RNA en el que se eliminan fragmentos del mensajero. Este mecanismo fue descubierto hace ya tres décadas por dos grupos de investigación en paralelo a través del estudio de híbridos de DNA-RNA en adenovirus (Berget et al. 1977; Chow et al. 1977). Mediante el uso de microscopía electrónica se visualizaron estos híbridos, observándose como una molécula de RNA maduro mapeaba en diferentes regiones no contiguas y más o menos alejadas del DNA genómico. Con ello, se descubrió la presencia de secuencias en el DNA genómico que no formarían parte del RNA mensajero maduro y que, poco más tarde fueron denominadas con el término intrón (Gilbert 1978), así como a las secuencias interrumpidas que formarían parte del mensajero maduro se denominaron exones.

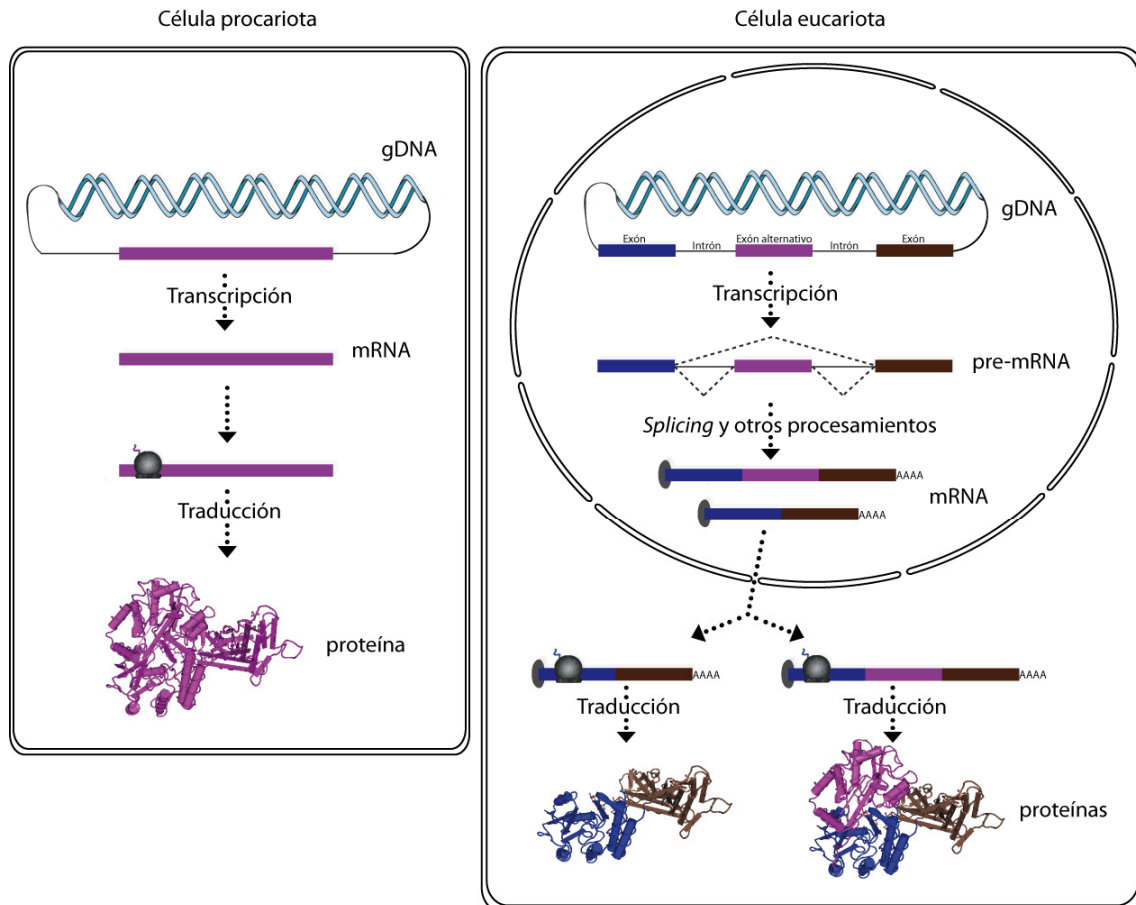
Este trascendental descubrimiento estudiado en virus fue rápidamente considerado como un proceso presente en la expresión de multitud de genes eucariotas, al mismo tiempo que se descubrió la presencia de otro mecanismo subyacente a éste denominado *splicing* alternativo (*Alternative splicing*, AS) (Calame et al. 1980; Early et al. 1980; Kole and Weissman 1982; Leff et al. 1986). Este mecanismo consistente en la eliminación diferencial

de fragmentos del pre-mensajero posibilitando la generación de diferentes mensajeros maduros que, en caso de poseer carácter codificante, resultarán en la obtención de diferentes isoformas proteicas a partir de una secuencia génica única (Caceres and Kornblihtt 2002; Kalsotra and Cooper 2011; Kelemen et al. 2013). En algunos casos, el número total de alternativas posibles que pueden generarse a partir de un solo gen es abrumador, desde cientos de mensajeros como es el caso del gen *Slo* en ratas hasta incluso miles como se producen en el gen *Dscam* en *Drosophila*. Además de generar múltiples variantes procedentes de un único gen, el AS es capaz de influir en la expresión de un gen, aumentando o disminuyendo los niveles de expresión de determinados transcritos (Blencowe 2006; T. A. Hughes 2006; Stamm et al. 2005).

Desde entonces, lo que parecía ser un fenómeno muy interesante y restringido a apenas algunos genes y organismos se ha extendido hasta considerarse un mecanismo de regulación fundamental a nivel post-transcripcional, afectando a la mayoría de los genes eucarióticos (y algunos procarióticos). Recientemente, gracias al empleo de nuevas tecnologías aplicadas al estudio del RNA se ha estimado que alrededor del 95% de los genes de humanos sufren este tipo de procesamiento (Pan et al. 2008; E. T. Wang et al. 2008). En la actualidad se conoce que el genoma humano contiene más de 200.000 intrones de un tamaño variable desde menos de 100 nucleótidos hasta intrones de un tamaño superior a 700.000 nucleótidos, con un tamaño medio de 1.800 nucleótidos. Sin embargo, los exones suelen ser con frecuencia de menor tamaño, con una media aproximada de longitud de 123 nucleótidos (Scherer 2010).

### 1.1.1. Mecanismo de *splicing*

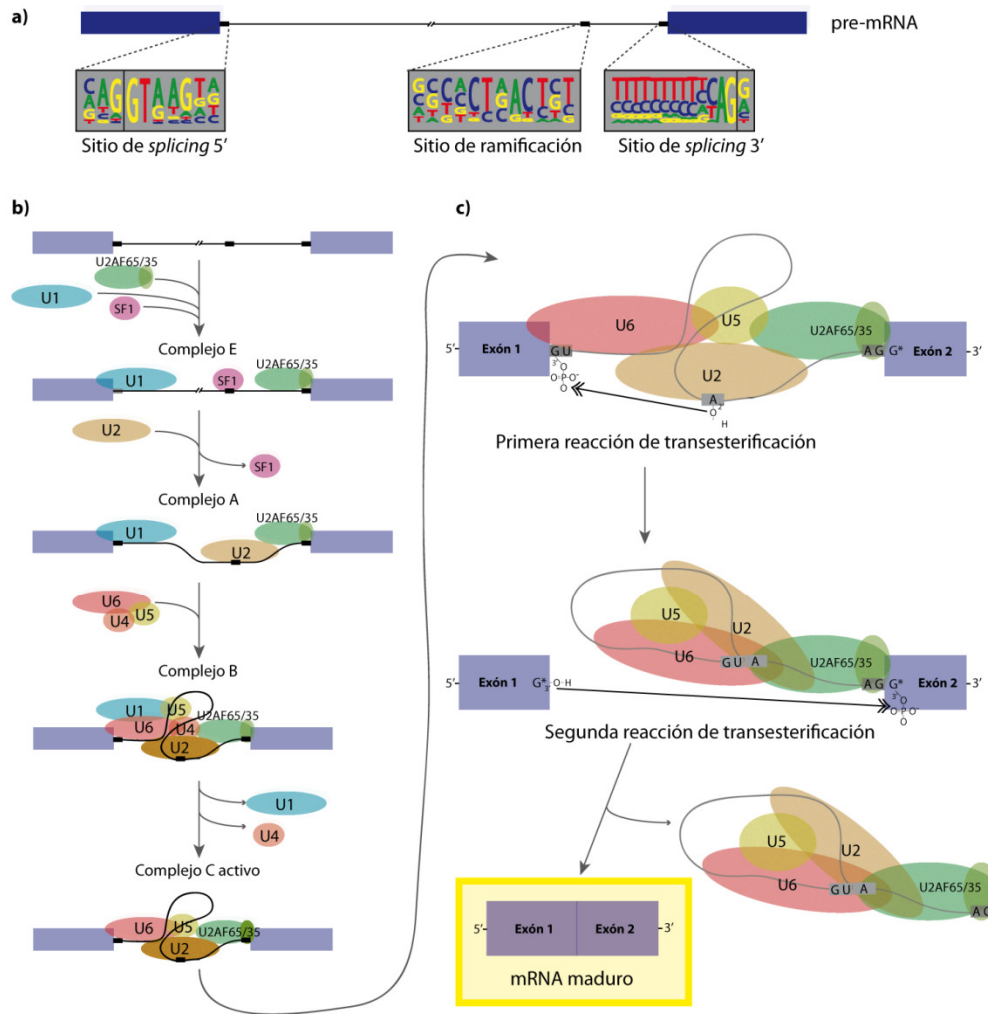
Como ocurre en los genes no interrumpidos de los organismos procariotas, los genes de los organismos eucariotas se transcriben en una sola copia de RNA con la secuencia del gen completo. Este RNA es denominado transcrito primario o pre-mRNA y contiene tanto las secuencias exónicas como las intrónicas de un gen. El pre-mRNA es posteriormente procesado en el núcleo mediante *splicing* eliminando todas aquellas secuencias intrónicas que interrumpían hasta obtener el RNA mensajero maduro (Figura 1). Las reacciones de transesterificación necesarias para catalizar la escisión de los intrones se realizan gracias a una serie de elementos en *cis* ó secuencias consenso presentes en el propio pre-mRNA, así como a un conjunto de elementos en *trans* ó complejos ribonucleoproteicos (RNP) que reconocen los elementos *cis* y al unirse provocan la formación de un complejo



**Figura 1. Representación esquemática del proceso de expresión génica en células procariotas y eucariotas.** Los eventos de AS se muestran como líneas discontinuas en la célula eucariota.

multirribonucleoproteico denominado espliceosoma (Figura 2) (Hoskins and Moore 2012; Wahl et al. 2009).

Las secuencias consenso correspondientes a los elementos *cis* se pueden clasificar en dos tipos: aquellas que definen las fronteras entre exones e intrones y aquellas implicadas en la regulación del *splicing*. En el primer grupo estarían clasificados los sitios de *splicing* 5' dador y 3' aceptor, así como el sitio de ramificación presentes en los intrones (Figura 2a). En eucariotas superiores, la amplia mayoría de los intrones siguen la regla GT-AG, de tal forma que presentan las siguientes secuencias consenso: **GURUGU** en el sitio de *splicing* 5' donador, **YNYRAY** en el sitio de ramificación y **(Y)<sub>n</sub>CAG** en el sitio de *splicing* 3' aceptor (Tarn and Steitz 1997), presentando un cierto grado de degeneración (Figura 2a). Estas secuencias serán reconocidas por el espliceosoma mayor. Además, existe otro tipo de secuencias consenso que siguen la regla AT-AC tanto en genomas de vertebrados como en no vertebrados, aunque representan a un número mucho menor de intrones, alrededor del 1%. Estas secuencias serán reconocidas por el espliceosoma menor. En este caso las secuencias consenso son de mayor tamaño, siendo **AUAUCCUU** para el



**Figura 2. Maquinaria de *splicing*.** Elementos en cis (a) y trans (b) implicados en el mecanismo de *splicing*. Se muestra en detalle las dos reacciones de transesterificación que ocurren durante el proceso en (c).

sitio 5' donador, UCCUUAAC en el sitio de ramificación y CCAC para el sitio 3' aceptor (Hall and Padgett 1994; Jackson 1991; Tarn and Steitz 1996). Por otro lado están clasificadas las secuencias implicadas en la regulación del mecanismo de *splicing*, tanto las secuencias que potenciarían como las que inhibirían el *splicing*, como se detallará más adelante. Estas secuencias se encuentran presentes tanto en exones como en intrones.

Entre los elementos *trans* más relevantes se encuentran los complejos ribonucleoproteicos nucleares pequeños (snRNPs, “*small nuclear ribonucleoproteins*”) ricos en U (como U1, U2, U4/U6 y U5 en el espliceosoma mayor; y U5, U11 y U12 en el espliceosoma menor (Tarn and Steitz 1996)), formados por un snRNA (“*small nuclear RNA*”) o dos snRNA (en el caso de la snRNP U4/U6) y un complejo variable de proteínas específicas del complejo (Hoskins and Moore) (Figura 2b). Otros elementos *trans*

implicados en la regulación del *splicing* son las proteínas SR (“*serine-arginine rich*”) y las ribonucleoproteínas heterogéneas nucleares o hnRNPs (*heterogeneous nuclear RNPs*).

El mecanismo de *splicing* consiste en dos reacciones de transesterificación mediante las cuales se provoca la escisión del intrón, para lo cual son necesarias las uniones específicas de un complejo de proteínas. El proceso comienza con el reconocimiento de la secuencia consenso del sitio de *splicing* 5´ donador GT en el intrón por complementariedad de bases con la U1 snRNA perteneciente al U1 snRNP, y se genera una unión independiente de ATP. En paralelo, la proteína SF1 (“*Splicing Factor 1*”) se une al sitio de ramificación, mientras que el factor auxiliar heterodimérico U2AF65/35 se une al tracto rico en pirimidinas con su subunidad mayor de 65 kDa y a los dinucleótidos AG presentes en el sitio de *splicing* 3´ aceptor con su subunidad menor de 35 kDa. Al ensamblaje de estas proteínas se denomina estadio temprano del espliceosoma o complejo E. Tras la formación de este complejo, el sitio de ramificación es reconocido por complementariedad de bases con la U2 snRNA perteneciente al U2 snRNP, y se genera una unión dependiente de ATP, que desplaza a la proteína SF1. Este conjunto se denomina complejo A espliceosomal. Posteriormente, los complejos U4/U6 y U5 RNPs son recluidos y junto con el resto de los complejos forman el complejo B espliceosomal que, tras la liberación de los complejos U1 y U4 snRNPs se provoca la activación catalítica del espliceosoma (Figura 2b) (Wahl et al. 2009).

En este momento se forma el complejo C espliceosomal y se genera la primera reacción de transesterificación donde el 2´ hidroxilo de la adenosina presente en el sitio de ramificación ataca al grupo fosfato del primer nucleótido del intrón en el sitio de *splicing* 5´ donador (Figura 2c). Esto provoca la liberación del exón situado en el 5´ y la formación de un lazo en el intrón denominado complejo Intermedio C1. Previo a la segunda reacción de transesterificación se provocan reordenamientos conformacionales adicionales en los complejos RNPs del espliceosoma. En la segunda reacción de transesterificación el 3´OH libre del exón situado en el 5´ ataca el grupo fosfato del primer nucleótido del exón situado en el sitio de *splicing* 3´ aceptor, lo que provoca la ligación de ambos exones y la liberación del intrón ya procesado que será posteriormente degradado. Tras la liberación del intrón, el espliceosoma se disocia, liberándose los complejos U2, U5 y U6 snRNPs que serán reciclados para sucesivas rondas de *splicing* (Wahl et al. 2009).

Aquellos intrones que presentan las secuencias consenso minoritarias AT-AC son procesados de manera análoga pero empleándose diferentes combinaciones de RNPs. En

este caso, el proceso comienza con el reconocimiento de la secuencia consenso del sitio de ramificación por complementariedad de bases con la U12 snRNA. La U12 snRNP forma un dímero con la U11 snRNP que es la responsable de reconocer el sitio de *splicing* 5' donador AT en el intrón y llevar a cabo la primera reacción de trans-esterificación. Tras el corte se une un segundo complejo RNP formado por U4, U5 y U6 snRNPs. El complejo U6 snRNP estará encargado de reconocer AT tras la liberación del complejo U11, formando un lazo en la secuencia intrónica. Por otro lado, el complejo U5 snRNP reconoce por complementariedad de bases la secuencia consenso del sitio de *splicing* 3' aceptor. Tras la formación del espliceosoma menor se produce la segunda reacción de trans-esterificación y la sucesiva ligación de ambos exones (Kreivi and Lamond 1996; Tarn and Steitz 1996, 1997; Q. Wu and Krainer 1999).

#### **1.1.1.1. Excepciones en el mecanismo de *splicing***

Existen excepciones a este mecanismo, como es el caso del *splicing* llevado para la obtención de la subunidad RNA de la telomerasa en *Schizosaccharomyces pombe* (Bonnal and Valcarcel 2008; Box et al. 2008). En este caso, se produce un desacoplado funcional del espliceosoma entre la primera y la segunda reacción de trans-esterificación. La subunidad RNA funcional de la telomerasa se produce tras la primera reacción de trans-esterificación, mientras que si se completan las dos reacciones el producto generado sería una variante de *splicing* del gen inactiva. Para que produzca este suceso, el intrón de la telomerasa presenta dos condiciones que facilitan la formación de un espliceosoma desacoplado: la presencia de un tracto rico en polipirimidinas muy débil y una distancia relativamente larga entre el sitio de ramificación y el sitio de *splicing* 3' aceptor (Bonnal and Valcarcel 2008; Box et al. 2008).

Otra excepción a este proceso ha sido denominada *splicing* no convencional, y cuya principal diferencia radica en la localización del proceso, siendo realizado presumiblemente en el citoplasma en lugar de en el núcleo celular. Este proceso se ha observado en los mensajeros HAC1 en levadura y XBP1 en mamíferos como consecuencia a una respuesta a estrés por acumulación de proteínas mal plegadas en el retículo endoplásmico (RE). El proceso es mediado gracias a la quinasa Ire1p que cataliza la escisión del intrón correspondiente. La realización de este proceso alternativo permite la producción de un factor de transcripción activo en menos tiempo y presumiblemente con menor coste energético en comparación con el mecanismo de *splicing* tradicional (Calfon et al. 2002;

Gonzalez et al. 1999; Kawahara et al. 1998; Ruegsegger et al. 2001; Sidrauski et al. 1996; Sidrauski and Walter 1997; Uemura et al. 2009; Yoshida et al. 2001; Yoshida 2007). Pero no es el único caso en el que el mecanismo de *splicing* es realizado en una localización diferente. El RNA de transferencia (tRNA) sufre el proceso de *splicing* en el núcleo pero también en los compartimentos celulares del cloroplasto y las mitocondrias. Estos estudios fueron realizados en los géneros de plantas *Arabidopsis* y *Oryza*, donde se observó como en lugar de las dos reacciones de transesterificación tradicionales, el proceso se llevaba a cabo mediante una tRNA endonucleasa que removía el intrón y una tRNA ligasa que unía a los exones del tRNA (Abelson et al. 1998; Englert and Beier 2005). Este mecanismo se lleva a cabo por unos complejos proteicos diferentes a la maquinaria basal del espliceosoma.

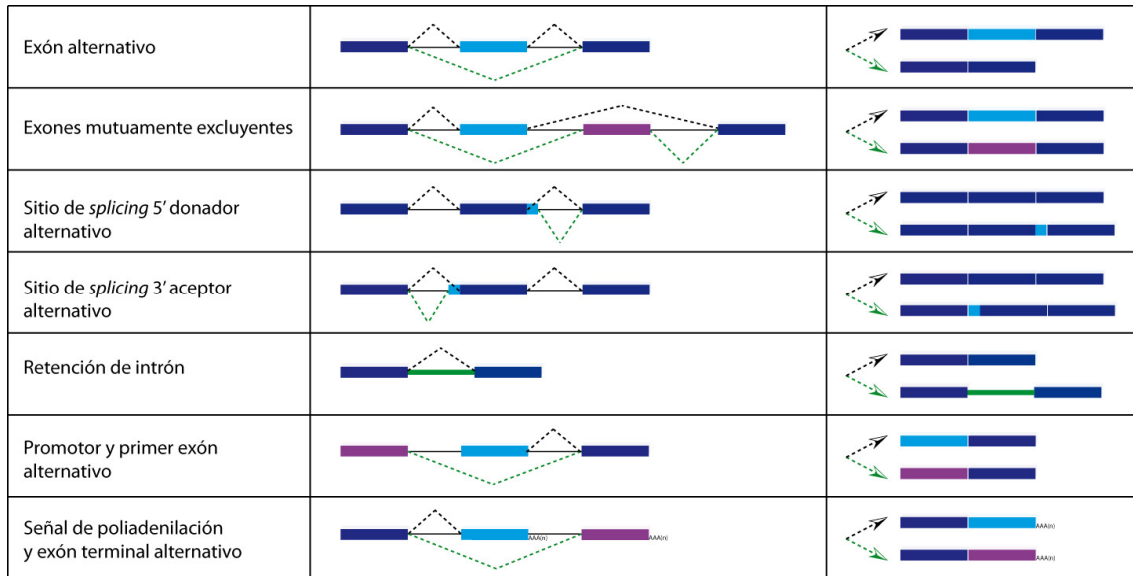
Por último, es importante destacar la presencia de tres determinados grupos de intrones que no son procesados a partir de ninguno de estos tipos de maquinarias del *splicing*, sino por un mecanismo independiente o parcialmente independiente de proteínas. Este mecanismo denominado *auto-splicing* (o *self-splicing*) ocurre en unos tipos de intrones (grupo I, grupo II o grupo III) que forman una ribozima realizando las dos reacciones de trans-esterificación de manera análoga al espliceosoma. Los grupos de intrones I y II (Cech 1990; Lehmann and Schmidt 2003) realizan el *splicing* sin requerir de ninguna proteína, mientras que los intrones del grupo III (Christopher and Hallick 1989; Montandon and Stutz 1984) requieren de la presencia de determinados factores para su procesamiento.

### **1.1.2. Regulación del *splicing***

#### **1.1.2.1. Regulación del mecanismo de *splicing* alternativo**

La expresión de los genes eucariotas es un proceso complejo que involucra numerosos pasos firmemente regulados y que finalmente generará una gran riqueza y diversidad de productos génicos. Un paso fundamental de regulación es el AS ya que, debido a la amplia diversidad de eventos de *splicing* (Figura 3) y a las diferentes combinaciones posibles entre ellos, podemos obtener múltiples productos a partir de una única secuencia génica. Los posibles eventos de *splicing* encontrados en un transcrito varían desde la presencia de un exón inicial o terminal alternativo, así como la presencia de exones alternativos (mutualmente excluyentes o no), retención de intrones y la presencia de sitios de *splicing* 5' o 3' alternativos en un exón canónico (Figura 3).





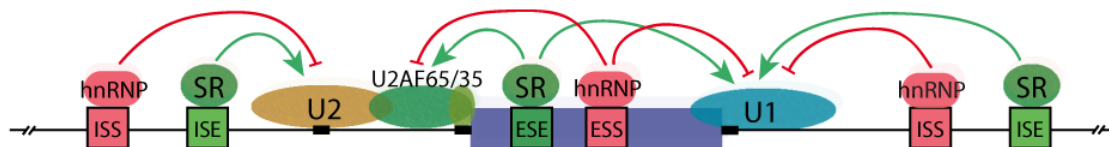
**Figura 3. Clasificación de todos los posibles eventos de *splicing* alternativo.** Las líneas finas continuas corresponden a intrones, y en discontinuas de color negro o verde se muestran los eventos alternativos de *splicing*.

En general, la regulación del mecanismo de *splicing* para la generación de esta amplia variedad de eventos de *splicing* está mediado principalmente por los elementos en *cis* o secuencias presentes en el mensajero y los elementos en *trans* o proteínas involucradas en el proceso de *splicing*.

Los elementos en *cis* pueden estar involucrados tanto en la definición de las fronteras entre exones e intrones, como en la regulación propiamente dicha del mecanismo de *splicing*. Las secuencias consenso que definen los sitios de *splicing* 5' y 3', así como el punto de ramificación son secuencias que presentan un alto grado de degeneración y, como consecuencia de ello, la definición de las fronteras entre exones e intrones es más laxa permitiendo la generación de mayor variabilidad en la combinación de los exones presente en el transcrito maduro (De Conti et al. 2013; Izquierdo and Valcarcel 2006). Únicamente aquellos exones constitutivos y, por tanto, presentes en la mayoría de los transcritos maduros, son los que presentan un grado de conservación elevado de las secuencias presentes en los sitios de *splicing*. De esta forma se favorece las interacciones más estables con los elementos *trans* gracias a la complementariedad de bases con ellos, aunque siguen precisando de otros elementos reguladores para su correcto procesamiento (De Conti et al. 2013; Izquierdo and Valcarcel 2006).

Además, en eucariotas superiores se ha desarrollado un mecanismo de regulación del AS en el que están involucrados diferentes elementos reguladores en *cis*. Dichos

elementos tienden a ser secuencias consenso relativamente degeneradas y de pequeño tamaño (normalmente entre cinco y diez nucleótidos de longitud) que son reconocidas por los dominios de unión a RNA de determinados elementos en *trans*, activadores o silenciadores, durante los estadios más tempranos de la formación del espliceosoma (Blencowe 2006; Matlin et al. 2005). Estos elementos se denominan *Exon Sequence Enhancers* (ESE) ó *Intron Sequence Enhancers* (ISE) cuando fomentan la inclusión de los exones o intrones en el transcrito maduro mediante la estabilización de la interacción de los complejos snRNPs con sus sitios de *splicing* correspondientes, ó *Exon Sequence Silencers* (ESS), *Intron Sequence Silencers* (ISS) cuando fomentan la exclusión de los exones o intrones del transcrito maduro impidiendo la unión de los diferentes complejos del espliceosoma (García-Blanco et al. 2004; Villate et al. 2008) (Figura 4).



**Figura 4. Elementos en *cis* (ISS, ISE, ESE y ESS) y *trans* (proteínas SR y hnRNPs) implicados en la regulación del proceso de *splicing* en eucariotas superiores.** Los elementos que fomentan la inclusión se muestran en verde, mientras que los que fomentan la exclusión se muestran en rojo.

Estas secuencias reguladoras serán reconocidas principalmente por dos familias de proteínas reguladoras del *splicing*: las proteínas SR y las ribonucleoproteínas heterogéneas nucleares (hnRNP) (Figura 4). La familia de proteínas SR está involucrada en el *splicing* constitutivo, pero también tienen una importante influencia en la regulación del *splicing* alternativo (Caceres and Kornblihtt 2002; X. D. Fu 1995; Graveley 2000). Estas proteínas se caracterizan por la presencia de uno o dos motivos de reconocimiento del RNA (*RNA-recognition motif*, RRM) y un dominio C-terminal rico en serinas y argininas denominado dominio SR (*serine/arginine-rich*). Los motivos RRM determinan la especificidad de unión a RNA mientras que los dominios SR median las interacciones proteína-proteína esenciales para el ensamblaje de la maquinaria del espliceosoma y el reconocimiento preciso de los sitios de *splicing* por dicha maquinaria (Wu and Maniatis 1993). La función de esta familia de proteínas se encuentra mayoritariamente relacionada con el reconocimiento de las secuencias ESEs e ISEs (Figura 4) permitiendo la estabilización del espliceosoma y, consecuentemente, la activación de los sitios de *splicing* donadores y aceptores correspondientes (Blencowe 2000; Heinrich et al. 2009). En cambio, la familia de ribonucleoproteínas hnRNPs, que presentan una composición de motivos RRM y dominios

SR similares a las proteínas SR, ejercen la función contraria, reconociendo las secuencias ESSs e ISSs (Figura 4), reprimiendo el uso del exón correspondiente gracias al bloqueo de las secuencias consenso y, por tanto, inhibiendo la unión necesaria del espliceosoma para el *splicing* (Dreyfuss et al. 1993; Heinrich et al. 2009; Pozzoli and Sironi 2005).

Sin embargo, aunque de manera general las proteínas de la familia SR fomentan la inclusión de los exones y la familia de hnRNPs fomenta su exclusión, numerosos investigadores han descrito situaciones en las que determinadas proteínas de dichas familias realizan las funciones inversas, actuando determinadas proteínas SR como represores y determinadas hnRNPs como activadores. En ocasiones, una misma proteína se ha visto involucrada en ambas funciones, dependiendo de la secuencia consenso reconocida (Blencowe 2006; Buratti et al. 2007; Goncalves et al. 2009; Martinez-Contreras et al. 2006). Además, tanto las proteínas SR como toda la familia hnRNP se expresa de manera diferencial según el tipo celular, tejido, o respuesta fisiológica, de forma que se estaría regulando la inclusión o exclusión de determinados exones a este otro nivel. Todos estos niveles de regulación fueron denominados como *Splicing Code* o código del splicing, siendo un mecanismo de gran relevancia para la diferenciación celular (Blencowe 2006).

Otro factor a tener en cuenta y que podría modular el reconocimiento de los sitios de *splicing* sería simplemente la estructura de los genes, constante en todas las células y situaciones del organismo. Es decir, la proximidad presente entre las diferentes secuencias consenso podría generar un impedimento físico entre los diversos factores en *trans* imposibilitando su unión. De esta forma, el tamaño de los exones/intrones estaría directamente modulando la inclusión/exclusión del exón en el transcrito maduro (De Conti et al. 2013). Estudios donde se ha reducido artificialmente el tamaño de un exón (<50 pb) se ha observado la exclusión del mismo en el transcrito maduro, corroborando la hipótesis anterior (Dominski and Kole 1991). Esto se genera probablemente por impedimento físico entre los complejos U1 snRNP en el sitio de *splicing* 5' y el complejo U2 snRNP en el sitio de *splicing* 3'. En el caso opuesto, estudios donde se ha aumentado artificialmente el tamaño de un exón pequeño se ha observado su inclusión en el transcrito maduro hasta un límite de 300 nucleótidos. Cuando se realizaban experimentos con aumentos de tamaño superiores a las 300 pb se inhibía la inclusión de dicho exón en el transcrito maduro o se generaban sitios de *splicing* crípticos dentro del exón provocando la inclusión de dos exones de menor tamaño (Berget 1995; Robberson et al. 1990). Por lo tanto, el tamaño de los

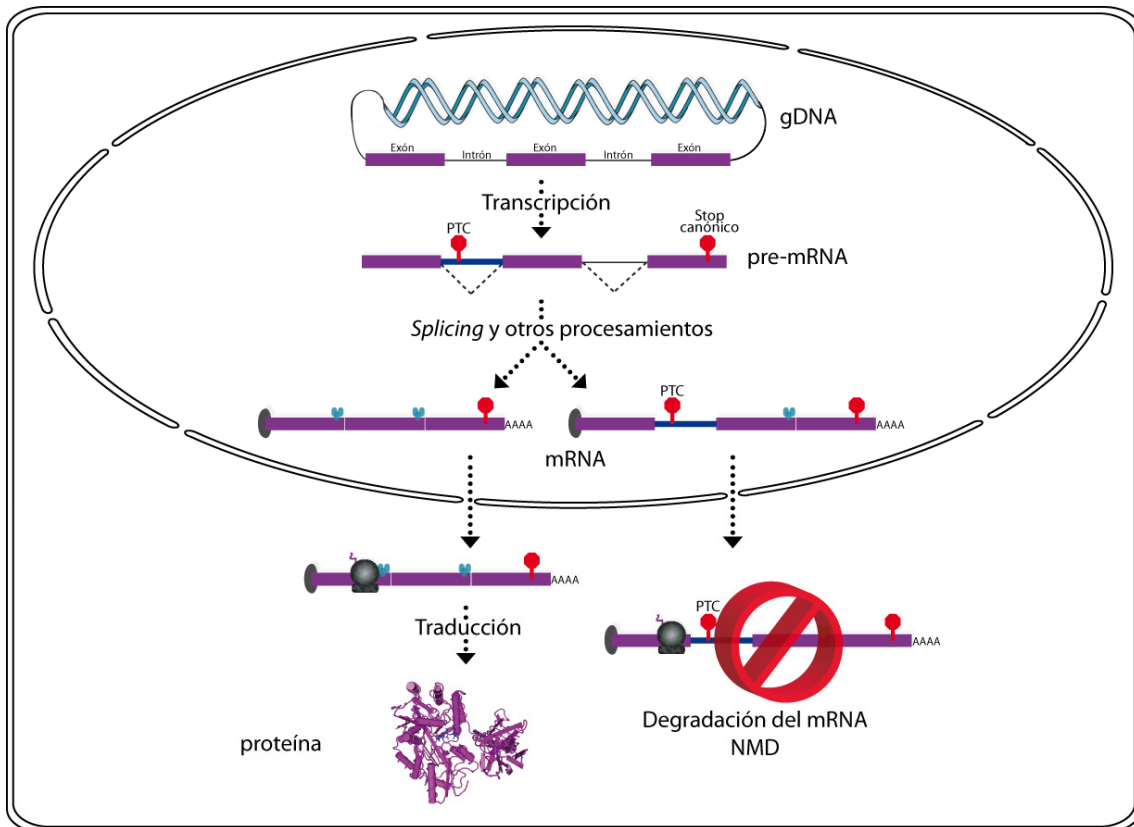
exones podría estar también determinando el mecanismo de *splicing* alternativo desde un nivel estructural génico.

En la actualidad, diversos estudios apuntan a que la propia estructura del propio RNA podría estar también involucrada en la regulación del *splicing* alternativo, sin necesidad de ser regulado por ningún factor proteico externo al propio espliceosoma (McManus and Graveley 2011; Yang et al. 2011). El pre-mRNA adopta unas complejas estructuras secundarias y terciarias *in vivo* (Russell 2008) que podrían modular el AS alterando la función de los complejos proteicos y elementos reguladores del *splicing* (Buratti and Baralle 2004; Warf and Berglund 2010). Existen numerosos ejemplos en los que estructuras en el pre-mRNA previenen el reconocimiento del espliceosoma en los sitios de *splicing* 5' donador o 3' aceptor, así como en el punto de ramificación (Buratti and Baralle 2004). Además, se han encontrado estructuras secundarias conservadas en el genoma humano que están asociadas con la selección de sitios de *splicing* alternativos (Shepard and Hertel 2008).

#### ***1.1.2.2. Regulación de los productos generados por *splicing* alternativo mediante Nonsense-Mediated Decay (NMD)***

Existe un mecanismo de regulación de mRNAs que se encarga de la degradación de mensajeros sin sentido. Este sistema es denominado *Nonsense-mediated Decay* (NMD) y su principal función es la de reducir errores en la expresión génica mediante la eliminación de aquellos mensajeros que presenten un codón prematuro de terminación de la transcripción (*Premature termination codon*, PTC) (Baker and Parker 2004; Chang et al. 2007).

El mecanismo molecular de este proceso se basa en el reconocimiento de los sitios de *splicing*, depositándose un complejo proteico (Blencowe et al. 1998; Kataoka et al. 2000; Le Hir et al. 2001; Lykke-Andersen et al. 2001; Mayeda et al. 1999; McGarvey et al. 2000; Reichert et al. 2002; Zhou et al. 2000) entre 20-24 nucleótidos por encima de las uniones de exones generadas (V. N. Kim et al. 2001; Le Hir et al. 2000a; Le Hir et al. 2000b). Estos complejos son de gran importancia para facilitar la exportación de los mensajeros al citoplasma, así como para detallar la estructura previa del mensajero (uniones de exones) (V. N. Kim et al. 2001). La posición relativa de estos complejos proteicos es revisada por el ribosoma durante la traducción (Ishigaki et al. 2001; Lejeune et al. 2002). Si el codón de terminación de la traducción se encuentra en el último exón o entre los 50 nucleótidos anteriores a la última unión de exones se realiza el proceso de traducción con normalidad.



**Figura 5. Mecanismo NMD de degradación de aquellos mensajeros que tras sufrir el proceso de AS presenten un PTC en su secuencia.** Se muestran en azul los complejos proteicos que definen las uniones de exones, y en rojo los PTC implicados en el mecanismo de NMD.

En cambio, si aparece un PTC a más de 50 nucleótidos de distancia de la última unión de exones y, por tanto del último complejo proteico, se provoca la eliminación de la estructura cap en el extremo 5' del mensajero y, consecuentemente su degradación por NMD (Czaplinski et al. 1998; Lykke-Andersen et al. 2000) (Figura 5). Este proceso suele producirse por la presencia de un evento de AS que cambia la pauta de lectura, o debido a la retención de un intrón cuya secuencia contiene un PTC.

### 1.1.2.3. Noisy splicing o ruido transcripcional

Todos estos mecanismos de regulación no son totalmente eficaces y una pequeña porción de mensajeros aberrantes son creados y conservados en la célula. Al fenómeno en el que los productos erróneos son generados por AS como consecuencia de una falta de regulación se le denomina *noisy splicing* o ruido transcripcional (Baek and Green 2005; Melamud and Moulton 2009; Pickrell et al. 2010; C. Zhang et al. 2007; Z. Zhang et al. 2009). Este fenómeno está asociado principalmente con genes de baja expresión, mientras que los genes altamente expresados parecen ser procesados de una manera más precisa. Además,

parece estar relacionado con fallos en la elección de los sitios de *splicing*, especialmente en aquellas uniones de exones no conservadas en diferentes tejidos y/o especies (Pickrell et al. 2010). Se estima que como media los intrones presentan una tasa de error en su procesamiento de alrededor del 0,7 % (Pickrell et al. 2010).

### **1.1.3. Relación del AS con otros mecanismos**

Desde su descubrimiento y durante muchos años, los mecanismos de transcripción, AS y traducción se han estudiado como rutas totalmente independientes. Sin embargo, este dogma está evolucionando tras el descubrimiento y publicación de múltiples trabajos en los que se demuestra la estrecha interrelación existente entre estos tres vitales mecanismos (Dujardin et al. 2013; Kornblihtt et al. 2009; Kornblihtt et al. 2013).

#### **1.1.3.1. Transcripción y *splicing* alternativo**

Hace más de veinte años se sugirió la primera conexión entre la arquitectura de los exones e intrones y la posición de los nucleosomas, en la base de una distribución periódica de los sitios de *splicing* (Beckmann and Trifonov 1991). Desde entonces, numerosos estudios han tratado de elucidar la relación existente entre ambos mecanismos [para revisiones; (Kornblihtt et al. 2009; Kornblihtt et al. 2013)].

La maquinaria de transcripción puede influenciar las decisiones AS afectando el tiempo en el que los elementos reguladores en *cis* son transcritos debido a la velocidad de elongación de la RNA polimerasa II (modelo cinético) (Darzacq et al. 2007), o mediante la asistencia a la hora de reclutar las proteínas involucradas como factores en *trans* del *splicing* (modelo de reclutamiento) (Munoz et al. 2010). En el modelo de reclutamiento se ha demostrado la implicación de la maquinaria transcripcional y, en especial, el dominio C-terminal de la RNA polimerasa II en el reclutamiento de diversos factores de *splicing*, así como los factores involucrados en el procesamiento del *cap* y el 3' de los mensajeros (Dujardin et al. 2013; Munoz et al. 2010). Además, el AS también está influenciado por la estructura de la cromatina afectando, a partir de diferentes propiedades, el reclutamiento de factores de *splicing* a través de proteínas adaptadoras que reconocen modificaciones específicas de histonas o, incluso, a través de la posición preferencial de los nucleosomas (Kadener et al. 2001; Kornblihtt et al. 2013).

Los últimos estudios detallan que ambos procesos, transcripción y *splicing*, se encuentran íntimamente conectados en al menos el 80% de los transcritos nuevos generados

en una célula (Dujardin et al. 2013). Otros estudios han detallado diversos casos en los que ambos mecanismos no se encuentran acoplados, permitiendo de esta forma la acumulación de mensajeros inmaduros (principalmente debidos a la retención de algún intrón). Estos mensajeros inmaduros no son procesados simultáneamente de forma que se acumulan en el núcleo a modo de reservorios (Hernandez-Torres et al. 2013; Vargas et al. 2011).

#### **1.1.3.2. *Splicing alternativo y traducción***

El caso de los mecanismos de AS y traducción es más llamativo debido a la localización subcelular de ambos procesos. Mientras que la transcripción y el AS (este en su mayoría) se producen ambos en el núcleo, el mecanismo de traducción se consideraba como un mecanismo únicamente desarrollado en el citoplasma de las células (asociados o no al Retículo Endoplasmático Rugoso o RER). Sin embargo, se han descrito fenómenos de traducción nuclear, que dependen simultáneamente de la RNA polimerasa II empleada en la transcripción (Iborra et al. 2001), acoplada asimismo al mecanismo post-transcripcional del *splicing*.

Existen diversas evidencias para el fenómeno de traducción nuclear, tales como el marcaje fluorescente de proteínas de nueva síntesis en el núcleo sin posibilidad de ser importadas desde el citoplasma tras realizarse un estudio aislando los núcleos celulares. Además, la traducción nuclear se encuentra localizada en sitios muy concretos y focalizados en el interior del núcleo, sin estar situados al azar como prevería esperar (Hentze 2001; Iborra et al. 2001). Estudios en la ruta de NMD mantienen y apoyan esta teoría debido al empleo de ribosomas para la revisión de los mensajeros. Los componentes de las maquinarias involucradas en la transcripción, traducción y el mecanismo de NMD colocalizan e interaccionan entre ellas (Iborra et al. 2004).

#### **1.1.4. Implicaciones funcionales del *splicing* alternativo**

La eliminación de los intrones y la unión de los exones colindantes mediante *splicing* es una parte esencial en la expresión génica eucariótica. Poco después de su descubrimiento se reconoció que dicho mecanismo podía ser regulado para producir múltiples variantes de RNA a partir de un único gen, aunque inicialmente se esperaba que fuera un mecanismo raro que se produjese en unos pocos genes exclusivamente. En la actualidad es ampliamente conocida la implicación que juega el AS en la expansión de la diversidad transcriptómica (Blencowe 2006; Mercer et al. 2011; Nilsen and Graveley 2010),

así como en la generación de variedad proteómica, las cuales tienen consecuencias biológicas de gran relevancia (Barash et al. 2010; Colak et al. 2013; Xing and Lee 2007).

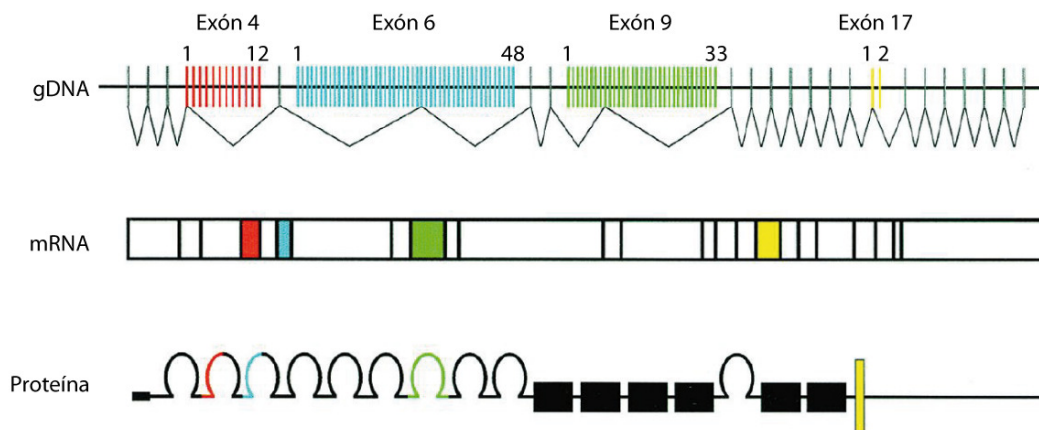
Entre los diferentes eventos de AS existentes (ver Figura 3), se podrían generar dos categorías según la parte de la secuencia que sufre la regulación: aquellos que modifican las regiones 5' y 3' no traducidas (*Untranslated regions*, UTR) del mensajero y aquellos que modifican las regiones codificantes del mensajero. En la primera categoría la modificación de los extremos UTR generalmente afecta a la expresión cuantitativa del gen (ya sea por variaciones en su eficiencia de traducción, su estabilidad o por la modificación de las secuencias de reconocimiento por microRNAs) (Elkon et al. 2013), o también podría afectar a la localización subcelular de los mensajeros (Holt and Bullock 2009; Martin and Ephrussi 2009). Un ejemplo de modificación del extremo 5' UTR es el gen *Calcium-Sensing Receptor (CaSR)* que codifica un receptor sensible a calcio. El gen humano *CASR* presenta dos promotores alternativos con un primer exón 5'UTR diferente entre ambos (exones 1A y 1B) seguido del exón 2 con una región 5' UTR común a ambas variantes de *splicing* (Hendy et al. 2013). La inclusión de uno u otro exón 1 alternativo puede modular la expresión global del gen en determinadas situaciones patológicas (Chikatsu et al. 2000; Hendy et al. 2013). Por otro lado, un ejemplo de modificación del extremo 3' UTR se da en el factor de *splicing* SC35 (Sureau and Perbal 1994). Este gen genera varios mensajeros que codifican una proteína idéntica pero que difieren en su extremo 3' UTR, lo que provoca cambios en la vida media de dichos mensajeros. Además, se observa una correlación directa entre la longitud de la región 3' UTR y la estabilidad del mensajero (Sureau and Perbal 1994).

En la segunda categoría se encontrarían todas aquellas modificaciones que afectasen la región codificante del mensajero, modulando en este caso a la expresión cualitativa del gen. Mediante AS se pueden obtener isoformas proteicas que presenten modificaciones en su secuencia de aminoácidos e, incluso que presenten variaciones en la composición de dominios proteicos (Barash et al. 2010; Colak et al. 2013; Kelemen et al. 2013; Xing and Lee 2007). Se han descrito eventos de AS que producen isoformas que actuarían como dominantes negativos de la isoforma canónica por la pérdida de alguno de los dominios funcionales (Lopez-Diez et al. 2013) o, por el contrario, isoformas constitutivamente activas por la ausencia de dominios proteicos reguladores, así como isoformas con diferente localización subcelular, cambios en la afinidad por determinados ligandos u otras proteínas



o incluso modulando la especificidad de sustrato de algunos enzimas (Kelemen et al. 2013; Stamm et al. 2005).

El ejemplo más llamativo de esta categoría sería la del gen *Dscam* en *Drosophila melanogaster* con unos niveles extraordinariamente altos de *splicing* alternativo. *Dscam* pertenece a la superfamilia de las inmunoglobulinas (Ig), presentando hasta 19.008 dominios extracelulares diferentes ligados a una de las dos regiones transmembrana que por AS se incluyen en el mensajero final, generando hasta 38.016 isoformas diferentes (Schmucker et al. 2000) (Figura 6). Todas las isoformas presentan la misma estructura de dominios pero contienen secuencias variables de aminoácidos en tres de sus diez dominios Ig y, ello provoca una alta especificidad de unión entre isoformas (Wojtowicz et al. 2004) modulando la orientación y el guiado de los axones en el sistema nervioso (Hummel et al. 2003; Wang et al. 2002a).



**Figura 6. Representación esquemática del gen *Dscam* en *Drosophila melanogaster*, y sus variantes de *splicing* y proteínas.** Los dominios proteicos Ig constantes están representados como un semicírculo en color negro, mientras que los dominios variables están codificados por exones alternativos y se indican en la figura por diferentes colores. Los transcritos maduros solo presentarán un exón alternativo de cada bloque (color). Figura adaptada de (Schmucker et al. 2000).

Otro ejemplo de modificación de la región codificante del mensajero AS es el caso del gen *Bcl-x* humano que puede ser procesado para generar dos isoformas con funciones marcadamente diferentes en la regulación de la muerte celular programada o apoptosis. El producto proteico del mensajero de mayor tamaño, la variante *Bcl-xL*, inhibe la apoptosis tras la retirada de los factores de crecimiento. Por otro lado, la isoforma codificada por el mensajero de menor tamaño, la variante *Bcl-xS*, es capaz de inducir apoptosis (Boise et al. 1993). En consonancia con estos resultados, aquellas células que están sometidas a mayores tasas de renovación (tales como los linfocitos) presentan altos niveles de expresión de la

isoforma Bcl-xS, mientras que los tejidos con tasas inferiores (tales como cerebro adulto) presentan unos niveles de expresión elevados de la isoforma Bcl-xL (Boise et al. 1993).

Las proteínas transmembrana representan una clase de proteínas que pueden ser particularmente reguladas por mecanismos de AS debido a las restricciones topológicas asociadas con su plegamiento, estructura, tráfico y función (Mittendorf et al. 2012). El número y la disposición de las regiones transmembrana (TM) en una membrana plasmática define la orientación de la proteína y, por tanto, su función y su estabilidad. Fenómenos de AS o *splicing* aberrante pueden modificar o interrumpir las regiones TM cuando los eventos de *splicing* se producen en dicha región, e incluso pueden eliminar el anclaje a la membrana generando proteínas solubles. De hecho, según estudios bioinformáticos el 84.6% de las regiones TM están codificadas por exones únicos, lo que parece indicar una presión selectiva para mantener las regiones TM sin intrones que las interrumpen en el pre-mRNA, especialmente en aquellas proteínas transmembrana de paso único (Mittendorf et al. 2012).

Dentro de la categoría de modificaciones de la secuencia codificante también se encuentran aquellas variantes que presentan alteraciones en su secuencia que generan un PTC y, por tanto, son diana para el mecanismo NMD (Popp and Maquat 2014). A través de este mecanismo, una célula podría regular los niveles de expresión de un gen generando diferentes transcritos de los cuales algunos podrían ser candidatos para NMD y otros serían traducidos con normalidad. De igual forma, la generación de múltiples variantes de *splicing* no codificantes (ncRNA, *non coding RNA*) a partir de AS podría influir en los niveles de expresión de las variantes codificantes (Lopez-Diez et al. 2013).

Además de afectar y/o alterar funciones o procesos biológicos muy concretos, el mecanismo de AS también puede ejercer una influencia global en la diferenciación celular, tisular e incluso a nivel de especie (Calarco et al. 2011; Nurtdinov et al. 2007; Pan et al. 2005), así como tener implicaciones en la generación de los distintos estados de desarrollo (Blencowe 2006; Gingeras 2007; Licatalosi and Darnell 2010; Xing and Lee 2007). Además, se está investigando el posible papel que puede tener el mecanismo del AS en el proceso de diferenciación de las células madre, lo que podría tener en un futuro cercano importantes aplicaciones terapéuticas (R. H. Fu et al. 2013).

#### **1.1.5. *Splicing* alternativo y evolución**

Se ha descrito en numerosas ocasiones que el AS podría ser uno de los principales mecanismos de evolución de los genomas ya que contribuye a incrementar la variabilidad y

la diversidad tanto transcriptómica como proteómica (Barbosa-Morais et al. 2012; Blencowe 2006; Boue et al. 2003; Merkin et al. 2012; Nilsen and Graveley 2010). Gracias a esta asociación, somos capaces de explicar las increíbles diferencias existentes entre especies que presentan un número muy similar de genes. Especialmente, habría que tener en cuenta el grupo de mamíferos, que comparten un conjunto parecido de genes y, por tanto, sus patrones de *splicing* serían los elementos principales para explicar sus diferencias (Barbosa-Morais et al. 2012; Merkin et al. 2012; Nurtdinov et al. 2007). Se ha estudiado como el mecanismo de AS se ve incrementado desde invertebrados hasta vertebrados (E. Kim et al. 2007), de tal forma que existe una correlación entre la complejidad de las especies y la proporción de genes que se regulan mediante AS, de tal forma que el porcentaje de genes que sufren AS es de alrededor del 25% en *Caenorhabditis elegans* (Gerstein et al. 2010; Hillier et al. 2009; Ramani et al. 2011), del 60% en *Drosophila melanogaster* (Graveley et al. 2011), y del 95% en *Homo sapiens* (Pan et al. 2008; E. T. Wang et al. 2008). Estudios similares se han realizado en transcriptomas de plantas tales como *Arabidopsis thaliana* (Filichkin et al. 2010) y *Zea mays* (P. Li et al. 2010), encontrándose un 42% y un 56% de genes que estaban afectados por AS, respectivamente. De esta forma nos podemos encontrar con una diversidad elevada a partir de un número muy limitado y similar de genes. Toda esta información nos indica a pensar que el AS está altamente involucrado en la diversificación de las especies (Blencowe 2006; Merkin et al. 2012; Mudge et al. 2011; Pan et al. 2008).

Además, se ha observado como individuos pertenecientes a una misma especie pueden mostrar diferencias en sus mecanismos de AS. Se compararon exones alternativos de un grupo de individuos humanos y se encontraron diferencias en la expresión de varios de estos exones (Hull et al. 2007). Posteriormente, se observaron numerosos ejemplos en los que el haplotipo podría contribuir a dichas diferencias en los patrones de AS encontrados (Douglas et al. 2009; Graveley 2008; Hull et al. 2007; Iwata et al. 2009). Estos estudios sugieren que las diferencias en los patrones de *splicing* pueden contribuir a explicar las diferencias en la expresión génica y en el fenotipo también a nivel de individuos (Kelemen et al. 2013).

En un orden inferior, estudios recientes han demostrado la existencia de un perfil de conservación de AS si comparamos el mismo tejido en diferentes especies de vertebrados (Barbosa-Morais et al. 2012; Merkin et al. 2012). En ellos se ha observado como los patrones de *splicing* están más conservados en un mismo tejido entre diferentes especies si

lo comparamos con los patrones de *splicing* presentes en diferentes tejidos dentro de una misma especie (Barbosa-Morais et al. 2012). Según esto, los patrones de *splicing* presentes en el corazón de humano y gallina serían más parecidos que aquellos presentes en dos órganos cualesquiera diferentes de humano. En un estudio en paralelo se concluyó también que existía un tipo de elementos comunes de *splicing* que son capaces de definir un tipo celular o tejido (Merkin et al. 2012).

Se han reportado también evidencias de presión selectiva funcional para las variantes de AS (Xing and Lee 2005) cuyos cambios en los patrones de *splicing* de un determinado gen podrían generar nuevas variantes de *splicing* que podrían ser seleccionadas favoreciendo la evolución de las especies sin alternar la función original del gen (Boue et al. 2003). Se ha estudiado como las variantes canónicas tienden a realizar actividades génicas críticas para la célula, mientras que las variantes minoritarias no suelen tener un impacto relevante en la fisiología (Blencowe 2006; Xing and Lee 2005). Por ello, los exones alternativos que pertenecen a esas variantes minoritarias o que son específicos de especie parecen evolucionar más rápidamente que los exones constitutivos (Xing and Lee 2005) si confieren una ventaja selectiva a la célula o individuo conservando las funciones críticas de la célula por las otras variantes. Esto sugiere que la generación de nuevos exones alternativos puede ser una fuerza impulsora en la evolución. Un ejemplo de ello sería la exonización generada a partir de los elementos móviles *Short Interspersed Nuclear Elements* (SINEs), o más específicamente de las secuencias Alu características de primates. Aproximadamente el 10% de toda la secuencia del genoma de humano está compuesta por este tipo de secuencias (Hasler et al. 2007). Estas secuencias Alu pueden exonizar, o dicho con otras palabras, puede generar exones nuevos, que contribuirán a la generación de un transcriptoma más complejo en primates (Dagan et al. 2004; Lev-Maor et al. 2003).

La comparación entre humanos y ratones ha mostrado que alrededor del 11% de los exones que son alternativos en una especie, son empleados como exones constitutivos en la otra especie, generando diferencias específicas a cada una de las especies (Claverie-Martin et al. 2005; Lei et al. 2005; Mola et al. 2007; Pan et al. 2005). La función de cada uno de estos exones individuales puede ser determinada por comparación entre organismos (Keren et al. 2010). Estas comparaciones se basan en la hipótesis de que todos aquellos exones alternativos pero funcionalmente importantes estarán conservados en la evolución. Por el contrario, los exones sin una función definida o relevante podrían terminar siendo eliminados por selección negativa. Una propiedad de estos exones alternativos es que suelen

ser simétricos, es decir, que presentan un número de nucleótidos divisible por tres y, por tanto, mantienen la pauta de lectura. Mientras que por lo general, los exones presentes en regiones UTR o aquellos exones específicos de especie muestran un mayor porcentaje de asimetría (Magen and Ast 2005; Xing and Lee 2005).

Existe una alta frecuencia de eventos de *splicing* específicos de especie que generan variantes de *splicing* de baja abundancia (niveles de expresión muy bajos) que pueden desencadenar en la generación de una nueva función o simplemente en una modificación de la función génica anterior. La acumulación de estas variantes de *splicing* de menor abundancia puede generar un alto impacto en la fisiología o quizás indirectamente influir en las características específicas de especie (Blencowe 2006). Algunos autores han relacionado estas variantes de menor expresión con el fenómeno del ruido transcripcional o *noisy splicing*, involucrándolo como una importante propiedad de la evolución génica (Pickrell et al. 2010). Las tasas de error durante el *splicing* observadas en un gen reflejan el balance entre mutaciones deletéreas y aquellas que generan una ganancia de función y, por tanto, se someten a una presión selectiva y se podrían fijar en el contexto genómico del organismo. Para todo ello habría que tener en cuenta que la tasa de error en el *splicing* de un organismo dependería de la complejidad de la maquinaria de *splicing* en esa determinada especie, así como del tamaño poblacional efectivo que permitiría o no eliminar dichas mutaciones por selección natural (Pickrell et al. 2010).

#### **1.1.6. *Splicing* alternativo y enfermedades**

La alteración del *splicing*, ya sea en una modificación de su mecanismo o en la desregulación del proceso puede desencadenar la aparición de múltiples patologías (Caceres and Kornblihtt 2002; Colak et al. 2013; Cooper et al. 2009; Garcia-Blanco et al. 2004; Irimia and Blencowe 2012; Licatalosi and Darnell 2006; Villate et al. 2008; G. S. Wang and Cooper 2007). Esto es debido a variaciones en las secuencias resultantes insertando o delecionando nucleótidos, cambiando la pauta de lectura o introduciendo PTCs en las regiones codificantes. Estas variaciones pueden afectar a la expresión de un solo gen dañando un elemento *cis* de *splicing*, ó bien puede tener un efecto en *trans* alterando un componente de la maquinaria de *splicing* que afectaría en cadena a múltiples genes. Como consecuencia, el fallo de la regulación ya sea en *cis* o en *trans* desencadenaría una patología de características muy variadas.

Las alteraciones en *cis* pueden afectar tanto a las regiones codificantes como a las no codificantes de un único gen. Aquellas mutaciones que se localizan en las regiones no codificantes, tales como los sitios de *splicing* 5', 3', punto de ramificación o las señales de poliadenilación frecuentemente causan enfermedades hereditarias (Villate et al. 2008; G. S. Wang and Cooper 2007). En cambio, mutaciones en la secuencia codificante (*Open Reading Frame*, ORF) pueden generar mutaciones sinónimas o “silenciosas” que no afectarían a la secuencia codificante, mutaciones que generan el cambio de un aminoácido (mutaciones *missense*) ó mutaciones que implican la inclusión de un PTC (mutaciones *nonsense*) que pueden provocar su degradación por el mecanismo NMD. Pero, independientemente de su resultado a nivel proteico, estas mutaciones puntuales también podrían provocar alteraciones en el *splicing* de ese gen debido a la generación de nuevos sitios de *splicing*, o el corte y/o generación de señales de elementos reguladores en los exones (ESE ó ESS) (Villate et al. 2008; G. S. Wang and Cooper 2007). De hecho, se ha propuesto que el 60% de las mutaciones que causan enfermedad son debidas a una interrupción del proceso de *splicing*, no por mutaciones que afectarían directamente a la secuencia de la proteína (Lopez-Bigas et al. 2005). Ejemplos de enfermedades con este tipo de mutaciones serían la fibrosis quística debido a una mutación en el gen *CFTR* que afecta a un elemento ESE y dos sitios de *splicing* críticos en el exón 13 (Aznarez et al. 2003), o la  $\beta$ -talasemia en humano ya que aproximadamente un 25% de todas sus mutaciones en los genes de la globina provocan defectos en el *splicing* que generan proteínas anómalas (Treisman et al. 1983).

En cambio, las alteraciones en *trans* afectan a los factores ribonucleoproteicos requeridos para el correcto procesamiento por *splicing* ya sea por mutaciones o por alteraciones no genéticas de estos factores y pueden estar implicados en el desarrollo de enfermedades tales como la Retinitis Pigmentosa (Wilkie et al. 2008), determinadas alteraciones en el Alzheimer (Hernandez et al. 2004), cáncer, ceguera y distrofias musculares, entre otras (Faustino and Cooper 2003; Garcia-Blanco et al. 2004). Hay dos categorías básicas de mutaciones de los elementos *trans*, aquellas mutaciones que afectan la función basal de la maquinaria de *splicing* y aquellas que afectan a un factor regulador del *splicing*. Las mutaciones que alteran la maquinaria basal del *splicing* estarían afectando presumiblemente a todos los pre-mRNAs, mientras que las mutaciones de los factores reguladores podrían afectar únicamente al conjunto de pre-mRNAs que fueran diana de dichos reguladores. Un ejemplo clásico de enfermedad con alteraciones en *trans* es la Distrofia Miotónica de tipo 1. Esta enfermedad está provocada por una expansión en el

tripleto CTG en la región 3'UTR del gen *DMPK*, que provoca el secuestro de dos proteínas reguladoras del *splicing*: CUGBP1 y MBNL1. Como resultado, todos los eventos de AS se alteran, lo que causa diversos defectos en el desarrollo del corazón y en el músculo esquelético (Kanadia et al. 2003; Miller et al. 2000; Ranum and Cooper 2006).

Por ello, para comprender el mecanismo de *splicing* así como las mutaciones en los motivos de DNA que, afectando al *splicing*, estén originando determinadas enfermedades es importante la búsqueda de determinadas drogas o fármacos que ayuden a corregir dichos errores y mejoren las patologías (Villate et al. 2008).

### 1.1.7. *Splicing* alternativo y nuevas tecnologías

Para el análisis de las variantes de AS se han empleado diversos métodos desde su descubrimiento. Así, unos métodos realizan un enfoque individual de un gen y sus variantes de *splicing* a través de su estudio por PCR o por híbridos DNA-RNA, y otros métodos realizan un enfoque más global tratando de estudiar el AS en todo el genoma.

Un abordaje individual es la RT-PCR cuantitativa (qRT-PCR) tradicional (Bustin 2000), empleado en estudios de análisis de las diferentes variantes de *splicing* y que permite aumentar la sensibilidad y la detección de aquellas variantes de menor expresión. Además, en la actualidad se ha desarrollado una nueva tecnología denominada *Biomark<sup>TM</sup> HD System* (*Fluidigm Corporation*) que consiste en un análisis de qRT-PCR basado en nanofluidos (Sorg et al. 2013). Esta nueva tecnología permite realizar un experimento con múltiples muestras y ensayos (parejas de cebadores) a un mismo tiempo partiendo de una cantidad ínfima de muestra, aumentando extraordinariamente la sensibilidad de detección hasta unos niveles que permite estudiar incluso el transcriptoma individual de una única célula (Flatz et al. 2011). Estas tecnologías cuantitativas son ideales para analizar la expresión global de determinados genes diseñando una pareja de cebadores en una región común a todas las variantes, así como para el estudio individual de diferentes variantes de *splicing* de un mismo gen siempre que presente pocos eventos de *splicing* y que se pudiesen diseñar parejas de cebadores específicas para cada variante individualmente. En cambio, sería difícil o prácticamente imposible emplear esta tecnología para aquellos genes que presentasen múltiples eventos de *splicing* al ser imposible diseñar una estrategia de amplificación independiente para cada una de ellas. Además, este abordaje depende de estudios previos realizados en cada uno de los genes y de las diferentes variantes descritas para ellos.

Uno de los métodos globales que se han utilizado con asiduidad para el estudio del *splicing* y sus alteraciones han sido los microarrays (Calarco et al. 2007; Cuperlovic-Culf et al. 2006; J. M. Johnson et al. 2003), aunque su abordaje presenta diversas limitaciones. Se han utilizado tanto microarrays de expresión génica para el estudio cuantitativo de la expresión de determinados genes en varias muestras, como para el análisis de las variantes de *splicing* mediante el SpliceArray (Blanchette et al. 2005; Blencowe 2006; Ule et al. 2005). Uno de los problemas emergentes de esta tecnología es la limitada sensibilidad que presentan los microarrays no permitiendo detectar variantes de *splicing* con niveles de expresión bajos. Además, las diferentes variantes son difíciles de discriminar, especialmente cuando solo se diferencian por unos pocos nucleótidos (Fehlbaum et al. 2005) y, todas las sondas están diseñadas sobre variantes previamente conocidas mediante otras tecnologías. Otra tecnología que ha sido ampliamente utilizada para el análisis global del *splicing* es el empleo de *Expressed Sequence Tags* (ESTs) (Shklar et al. 2005). Este abordaje permite el estudio “*de novo*” pudiendo describir nuevas uniones de exones y diferentes combinaciones de las mismas, pero carece de la profundidad necesaria para poder describir aquellas variantes de menor expresión.

Esta tecnología, así como la tecnología de los microarrays, han sido recientemente suplantadas por la técnica de secuenciación masiva de RNA ó también denominada RNA-seq en el estudio del transcriptoma (Barbosa-Morais et al. 2012; Merkin et al. 2012; Pan et al. 2008; E. T. Wang et al. 2008). Esta nueva tecnología nos proporciona una visión global del transcriptoma de una muestra y que, con una cobertura adecuada es capaz de detectar eventos de *splicing* poco abundantes, lo que nos ha permitido alcanzar grandes avances en la comprensión del proceso del AS. El problema que radica de estas tecnologías es la longitud de sus secuencias, siendo en los datos de ESTs secuencias de alrededor de 800-900 bases desde los extremos 5' ó 3' y que raramente abarcan el transcrito al completo, dificultando o incluso imposibilitando la detección de las diferentes formas de *splicing* generadas (Shklar et al. 2005). Igualmente ocurre con las lecturas de RNA-seq, que en la actualidad proporcionan longitudes de entre 50 y 400 pb dependiendo de las diferentes tecnologías (pirosecuenciación, PCR en puente, ligación, iónica, etc), lo que dificulta el ensamblaje “*de novo*” de estos fragmentos en los mensajeros correspondientes. A pesar de esta limitación, la tecnología de RNA-seq nos permite el análisis de todos los eventos de *splicing* presentes en una muestra sin necesidad de estar descritos previamente, lo que supone una gran ventaja en comparación con el resto de metodologías de estudio del *splicing* globales.



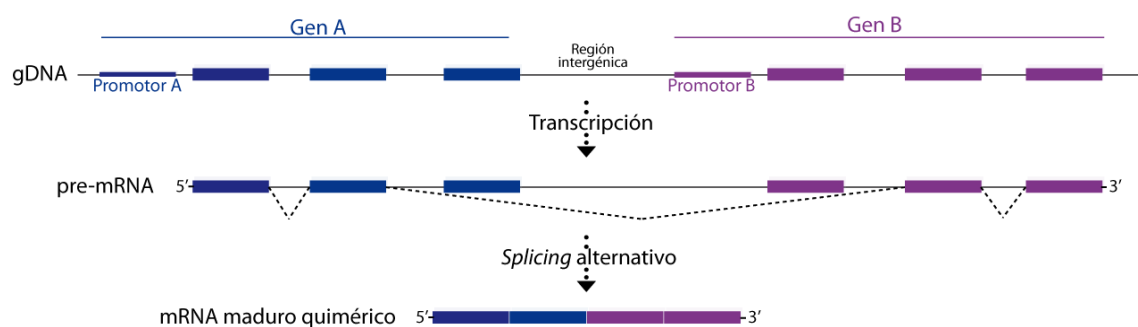
Debido a todo esto, la mayoría de los estudios a gran escala se limitan a la anotación existente en las bases de datos ignorando los eventos de *splicing* no descritos, no ensamblando las posibles variantes de *splicing* aún sin identificar. Por todo ello es imprescindible el estudio individual de la expresión de cada uno de los genes para obtener una visión completa de las posibles variantes de *splicing* que cada gen pueda generar debido a este fenómeno.

## 1.2. OTROS MECANISMOS DE REGULACIÓN DE LA EXPRESIÓN GÉNICA

En determinadas ocasiones, los precursores de RNA que son modificados mediante el mecanismo post-transcripcional de *splicing* proceden de dos genes o dos precursores de RNA diferentes, al contrario de lo que ocurre con el proceso de *splicing* tradicional (*cis-splicing*). Cuando existe un precursor de RNA único proveniente de dos genes diferentes adyacentes y situados en la misma orientación se denomina quimerismo o *Transcription Induced Chimerism* (TIC). Sin embargo, cuando el proceso se realiza a partir de dos precursores diferentes que pueden provenir del mismo o de diferente gen, el fenómeno se denomina *trans-splicing*.

### 1.2.1. Quimerismo en *cis* y su mecanismo

La formación de quimeras mediante el fenómeno de quimerismo en *cis* o TIC es un mecanismo relevante para la generación de diversidad transcriptómica, permitiendo la transcripción única de dos genes que se encuentran adyacentes en la misma orientación en el genoma y que, normalmente se transcriben también de forma independiente (Parra et al. 2006) (Figura 7).




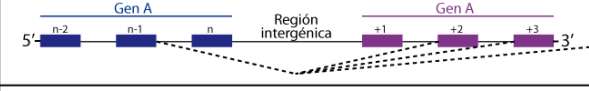
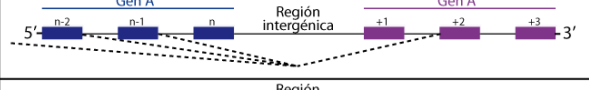


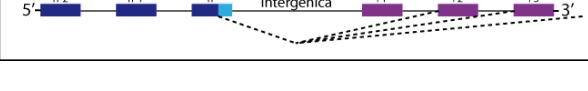
**Figura 7. Esquema de la formación de quimeras entre dos genes adyacentes.** La procedencia de los exones del transcrito quimérico se indica por la composición de colores.

Aunque estudios previos indican que este proceso parece afectar únicamente de un 2 a un 5% del total de los genes humanos (Akiva et al. 2006; Parra et al. 2006) respectivamente], análisis bioinformáticos recientes han sugerido que en realidad este mecanismo podría ser mucho más frecuente (Denoeud et al. 2007; Nacu et al. 2011). Se han descrito en humano varias quimeras encontradas tanto individualmente como en estudios globales de la expresión génica (Tabla 1).

Gen 5'	Gen 3'	Referencia
<i>MDS1</i>	<i>EVI1</i>	(Fears et al. 1996)
<i>GALT</i>	<i>IL11Ra</i>	(Magrangeas et al. 1998)
<i>CCL14</i>	<i>CCL15</i>	(Pardigol et al. 1998)
<i>HHLA1</i>	<i>OC-90</i>	(Kowalski et al. 1999)
<i>CYP2C18</i>	<i>CYP2C19</i>	(Zaphiropoulos 1999)
<i>SBLF</i>	<i>ALF</i>	(Upadhyaya et al. 1999)
<i>TSNAX (TRAX)</i>	<i>DISC1</i>	(Millar et al. 2000)
<i>Kua</i>	<i>UBE2V1 (UEV1A)</i>	(Thomson et al. 2000)
<i>PPAN (SSF1)</i>	<i>P2Y11</i>	(Communi et al. 2001)
<i>VPS72 (YL-1)</i>	<i>TMOD4</i>	(Cox et al. 2001)
<i>TNFSF12 (TWEAK)</i>	<i>TNFSF13 (APRIL)</i>	(Pradet-Balade et al. 2002)
<i>LY75 (CD205)</i>	<i>CD302 (DCL1)</i>	(Kato et al. 2003)
<i>ANKHD1 (MASK)</i>	<i>EIF4EBP3 (BP3)</i>	(Poulin et al. 2003)
<i>NME1</i>	<i>NME2</i>	(Valentijn et al. 2006)
<i>LY6G6F (G6F)</i>	<i>LY6G6D</i>	(Calvanese et al. 2008)
<i>CSNKβ</i>	<i>LY6G5B</i>	(Calvanese et al. 2008; Hernandez-Torres et al. 2013)
<i>SLC45A3</i>	<i>ELK4</i>	(Maher et al. 2009; Rickman et al. 2009)

**Tabla 1. Quimeras descritas previamente en humano.**

Estudios realizados empleando los datos disponibles de ESTs humanas y datos procedentes de RNA-seq nos muestran las proporciones en las que se encuentran en cada estudio las características más comunes en la formación de quimeras (Akiva et al. 2006; Nacu et al. 2011; Parra et al. 2006). Si consideramos aquellas quimeras que conservan la ORF, se observa que la unión entre el penúltimo exón del gen situado al 5' y el segundo exón del gen situado al 3' se encuentra presente entre el 27% y 54% de las quimeras analizadas (Tabla 2). Es decir, se emplea el último sitio de *splicing* 5' donador del gen situado en el 5' y el primer sitio de *splicing* 3' aceptor del gen situado al 3'. Este proceso puede estar relacionado con tratar de evitar la inclusión de una señal de terminación de la traducción (situada generalmente en el último exón), así como el péptido señal (situado comúnmente en el primer exón de cualquier transcrito). De hecho, de aquellas quimeras que presentan el último exón del primer gen, más de la mitad presentan un evento de *splicing* interno que evade el codón de terminación (Parra et al. 2006). Además, se observa como la mayor proporción en la formación de quimeras (79 y 80%) se produce en aquellas que presentan cualquier exón (a excepción del último) del primer gen uniéndose con el segundo exón del gen situado en el 3' (Tabla 2).

		Akiva et al 2006	Parra et al 2006	Nacu et al 2011
Exón (n-1) a (+2)		44%	27%	54%
Exón (n-1) a cualquiera		55%	---	64%
Cualquier exón a (+2)		80%	---	79%
Nuevo exón intergénico		12%	21%	---
Primer exón a cualquiera		12%	---	---
Nuevo sitio de <i>splicing</i>		29%	---	---

**Tabla 2. Esquema de frecuencias de cada uno de los eventos implicados en la formación de quimeras, según los datos disponibles por tres grupos de investigación diferentes.**

Otra forma de generación de quimeras es la contribución única de un primer exón perteneciente al gen situado al 5' y el resto de la secuencia correspondiente a exones del gen situado al 3' de la secuencia. Este primer exón contribuiría principalmente con la región 5' UTR y, ocasionalmente con el ATG iniciador y en algunos casos con el péptido señal. Por lo tanto, la quimera resultante sería la misma proteína que la codificada por el gen situado en el 3' pero regulado según los mecanismos de regulación del gen situado en el extremo 5' tanto a nivel transcripcional (por el promotor), como durante la traducción (por el 5' UTR), y regulación a nivel proteico (debido al péptido señal). La formación de este tipo de quimeras correspondería al 12% del total (Akiva et al. 2006). Además, en este mismo estudio se encontraron un mecanismo de formación de quimeras de menor abundancia (4% del total) pero con similares consecuencias de regulación, ya que determinadas quimeras estaban formadas por la parte correspondiente al gen situado en el 5' al completo pero únicamente con el último exón del segundo gen, por lo que este último principalmente contribuiría con la región 3'UTR (Akiva et al. 2006).

Además, se han realizado diversos estudios analizando los sitios de inicio y terminación de la transcripción en diferentes especies y empleando diversas técnicas (EST, librerías de cDNA completo, 5' y 3' RACE arrays) y se ha observado como en el 87% de aquellos genes que presentaban extensiones en el 5' terminal de sus transcritos, dichas extensiones alcanzaban el gen situado inmediatamente antes (Carninci et al. 2005; Denoeud

et al. 2007; Kapranov et al. 2005). Este fenómeno sorprendió a la comunidad científica al ser considerablemente superior a lo anteriormente anticipado (alrededor del 15%). Sin embargo, las proteínas individuales de los genes que formaban la quimera no presentaban relación en sus atributos funcionales (según *Gene Ontology Database*, GO) (Ashburner et al. 2000).

Un fenómeno interesante es la aparición de exones adicionales en la región intergénica entre los genes que componen la quimera, observándose éstos en el mRNA maduro (Tabla 2). Este fenómeno se observó en un 12% y un 21% en los estudios de Akiva et al. (2006) y Parra et al. (2006), respectivamente. La presencia de estos exones en la región intergénica caracterizaría a la quimera y la excluirían de un posible proceso de *trans-splicing*.

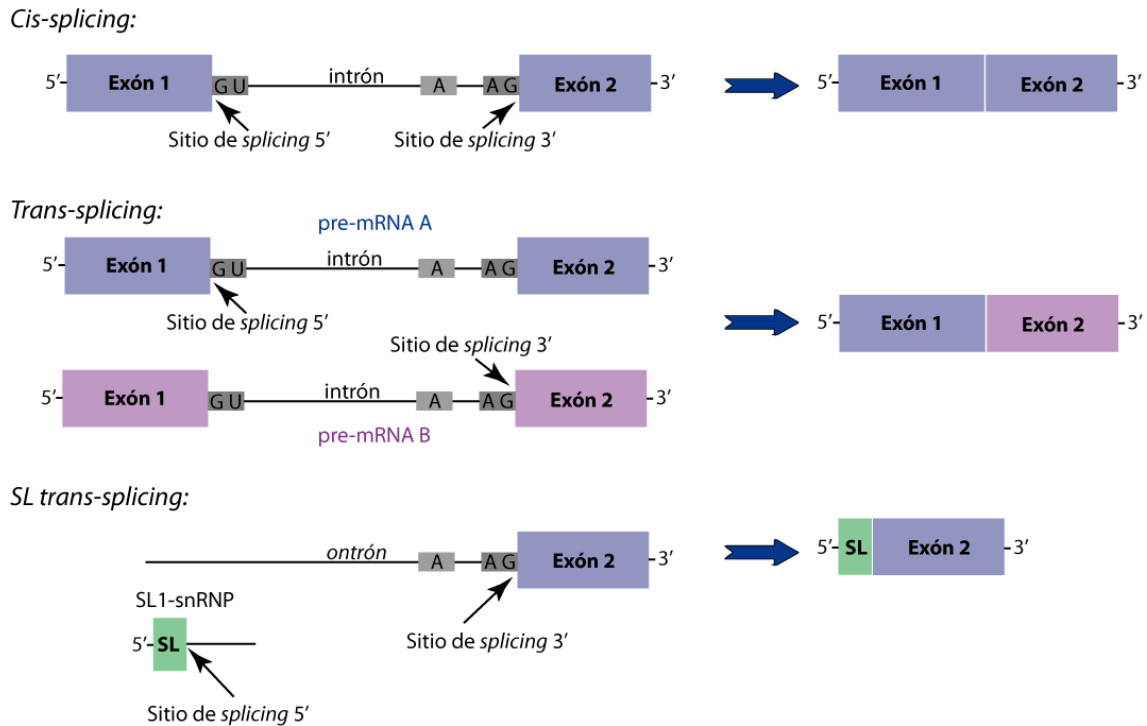
Otro dato interesante a destacar sería la escasa conservación de la pauta de lectura en los exones correspondientes al segundo gen de la quimera, ya que solo está conservada en únicamente un 35-36% de las quimeras [Nacu et al. (2011) y Parra et al. (2006), respectivamente], teniendo en cuenta que se esperaría alrededor de un 33% debido a una selección de las tres pautas de lectura al azar. De hecho, el 91% de las quimeras que presentaban un cambio en el marco de lectura también generaban un PTC y, por ello, estarían sujetas a regulación por el mecanismo de NMD (Nacu et al. 2011). Este dato se asemeja con el porcentaje del 56% de quimeras que se esperaría que fuesen reguladas por NMD según el estudio de Akiva et al. (2006).

Además, se ha estudiado la distancia presente entre los genes que generan los fenómenos de quimerismo en los datos de ESTs observándose como estos genes tienden a estar más próximos entre ellos que la media de los genes en el genoma humano. Las distancias medias obtenidas entre los genes implicados en la formación de quimeras era de aproximadamente 8,5 kbp mientras que la distancia entre los genes en la población génica completa presentaba una media de 48 kbp (Akiva et al. 2006), aunque en un 5% de los casos de quimerismo las distancias eran más largas llegando a ser superiores a 50 kbp. Estos datos se corroboraron en un estudio realizado con datos de RNA-seq en muestras de humano, donde se observaba como más de un cuarto de las quimeras presentaban una distancia media menor a 12 kpb, más de la mitad presentaban una distancia menor a 26 kpb y, más del 75% de las quimeras presentaba una distancia menor a 54 kpb. Por último, en este estudio se determinó que la distancia media entre los genes quimera era en torno a 8.866 bp (Nacu et al. 2011).

Por último, un estudio en el que se emplearon técnicas globales de secuenciación masiva de RNA-seq y espectrometría de masas determinó patrones concretos en la expresión de las quimeras (Frenkel-Morgenstern et al. 2012). Se observó como los transcritos quiméricos presentaban una especificidad de tejido superior en comparación con los transcritos normales no quiméricos. Además, se clasificó a la mayoría de las quimeras como transcritos con niveles de expresión bajos o muy bajos, observándose que aquellos que presentaban unos niveles de expresión superiores se encontraban expresados en una amplia variedad de tejidos. Otro estudio realizado en base a datos de RNA-seq también ha determinado la expresión ubicua de determinadas quimeras en diferentes tejidos, aunque mencionando la presencia de varias quimeras específicas de tejido, en concreto en próstata y cerebro (Nacu et al. 2011).

### 1.2.2. *Trans-splicing* y su mecanismo

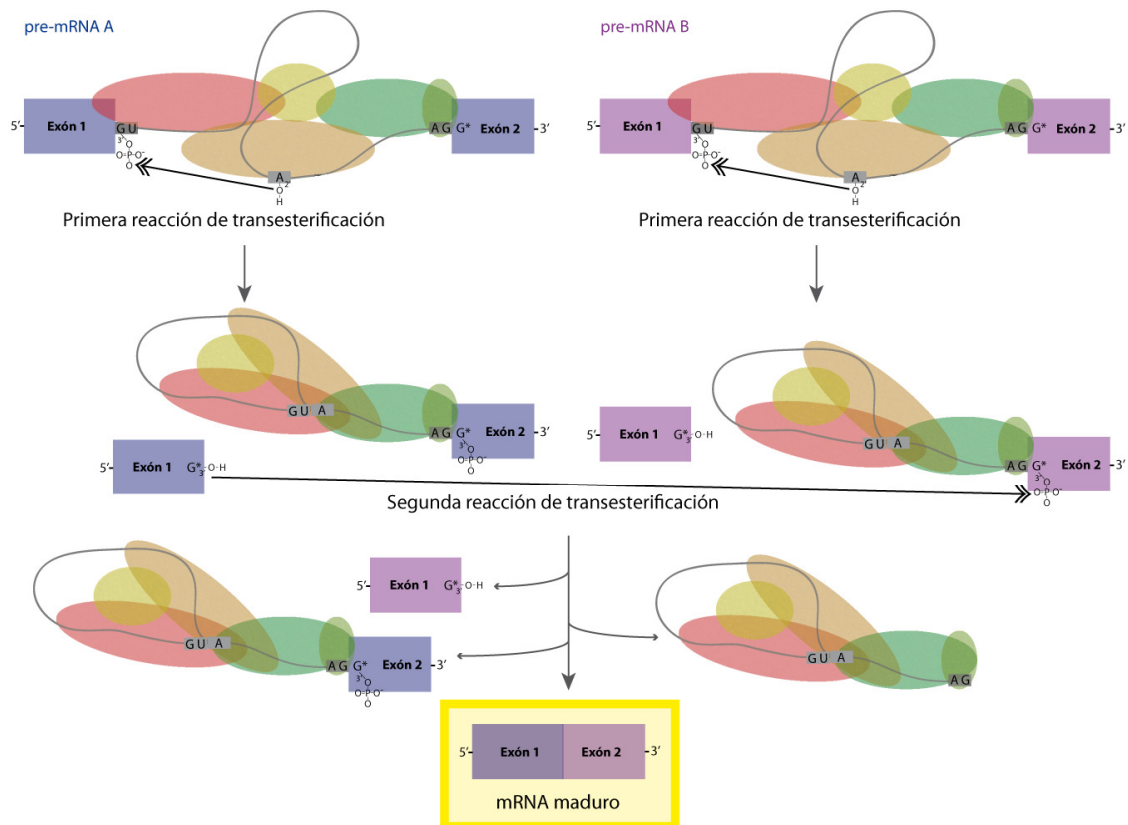
Tanto en el fenómeno de AS como en el quimerismo en *cis*, los pre-mensajeros son únicos y contienen toda la información necesaria que, tras los procesos básicos de *splicing* y su respectiva regulación generarán los mensajeros maduros correspondientes. Sin embargo, existe un mecanismo alternativo en el que exones de dos pre-mensajeros diferentes pueden llegar a unirse uno al otro en un proceso denominado *trans-splicing* (Bonen 1993; Takahara et al. 2005) (Figura 8). Se han categorizado dos tipos principales de *trans-splicing* denominados *trans-splicing* alternativo de tipo homotípico cuando los dos pre-mRNAs precursores provienen del mismo gen y de tipo intergénico cuando los precursores de RNA provienen de dos genes diferentes (Takahara et al. 2005). La primera evidencia natural demostrada en células de mamífero fue un *trans-splicing* de tipo homotípico y se observó con el gen *Carnitine octanoyltransferase* (COT) de rata (Choi et al. 1995). En concreto, se encontraron mensajeros maduros que presentaban el exón 2 duplicado y, en otros casos, se repetían los exones 2 y 3 (Caudevilla et al. 1998). Años más tarde se reportaron evidencias de *trans-splicing* de tipo intergénico en isoformas truncadas de la proteína quinasa dependiente de calmodulina (*CaM Kinase II*) y un híbrido de la colesterol acetiltransferasa-1 (ACAT-1) en humanos (Breen and Ashcroft 1997; B. L. Li et al. 1999). La eficiencia de este tipo de *trans-splicing* intergénico es extremadamente baja, por lo que difícilmente se podría considerar este segundo tipo de *trans-splicing* como un mecanismo de generación de diversidad proteica (Tasic et al. 2002).



**Figura 8. Diferentes tipos de *splicing* según los tipos y la posición de las moléculas que intervienen** (en *cis* o en *trans*). El *trans-splicing* se puede estar realizando entre dos pre-mensajeros completos, o entre una secuencia líder SL y un pre-mensajero.

Sin embargo, dentro de esta categoría se podría englobar el denominado *splice-leader* (SL) *trans-splicing* (Blumenthal et al. 2002; Hastings 2005) descubierto inicialmente en kinetoplastidos (Van der Ploeg 1986) (Figura 8). Esta forma de SL *trans-splicing* se caracteriza por la adición de exones 5' no codificantes (estructuras 5' *cap*) a la mayoría (o todos) los pre-mRNAs del organismo (R. E. Davis 1996), permitiendo el procesamiento de los transcritos policistrónicos presentes en dichas especies para convertirse en transcritos individuales ya procesados con la estructura *cap*. Por lo tanto, este mecanismo provocaría un aumento en la eficiencia de traducción de los mensajeros (Hastings 2005) y se ha encontrado presente en kinetoplastidos (Campbell et al. 2003; Mayer and Floeter-Winter 2005), nematodos (Blumenthal et al. 2002) y platelmintos (R. E. Davis and Hodgson 1997).

Para el procesamiento del mecanismo de *trans-splicing* es necesaria la generación de dos reacciones de transesterificación, de manera análoga a las realizadas mediante un fenómeno de *splicing* canónico o *splicing* en *cis*. Ambas reacciones de transesterificación generan la unión de dos exones de diferentes precursores por una unión fosfodiéster 3'-5' convencional (Figura 9). En un estudio se observó como, además de los sitios de *splicing* canónicos GT-AG, los sitios de *splicing* no canónicos también podrían funcionar en este



**Figura 9. Mecanismo de *trans-splicing*.** Se observa en detalle las reacciones de transesterificación involucradas en el proceso de *trans-splicing* entre dos pre-mensajeros diferentes (mostrados con diferente color).

mecanismo (Herai and Yamagishi 2010). Ambos sitios de *splicing* estarían generando un intrón de características especiales en el precursor de RNA situado en el 3' de la unión, considerado desde el inicio del precursor hasta el sitio 5' aceptor. Este tipo de intrón sería denominado *ontrón* (Bonen 1993) (Figura 8). En el caso del *SL trans-splicing* mencionado anteriormente, la estructura del cap que se incorporará a todos los mensajeros proviene de un precursor de RNA SL que consiste en una secuencia líder de entre 15 y 50 nucleótidos (Hastings 2005). Esta secuencia líder se convierte en el sitio de *splicing* 5' dador y se unirá a los sitios de *splicing* 3' aceptores del pre-mRNA.

### 1.2.3. Implicaciones funcionales del quimerismo

Tanto la formación de quimeras por *cis-splicing* (TIC) como por *trans-splicing* se han relacionado como mecanismos de regulación de la transcripción obteniendo una mayor diversidad de variantes de RNA en comparación con la expresión individual de cada uno de los genes en su manera tradicional. Este aumento de la diversidad de transcritos está directamente relacionado con una implicación de dichos procesos en la expansión de la

variedad proteómica. Por ejemplo, cuando la quimera resultante es codificante, obtenemos como resultado una proteína quimérica formada por exones codificantes de ambos genes y, por tanto, con una mezcla de dominios proteicos (Akiva et al. 2006; Frenkel-Morgenstern and Valencia 2012; Parra et al. 2006). Se ha observado como la formación de quimeras puede estar muy limitada a un tipo celular, encontrándose una expresión diferencial entre la quimera y los transcritos originados por los genes individuales. Un ejemplo sería la expresión diferencial que presenta la quimera LY75-CD302 predominantemente expresada en líneas celulares Hodgkin y Reed-Strenberg (Kato et al. 2003).

Las quimeras también pueden presentar alteraciones en las propiedades proteicas, en comparación con las proteínas individuales correspondientes, como puede ser alteraciones en la localización de la proteína o en su función. Un ejemplo de modificación de la localización se encontraría en la proteína nuclear UBE2V1 mientras que la quimera Kua-UBE2V1 se encuentra localizada en el citoplasma, indicando que el dominio proteico Kua determina la localización citoplasmática de la quimera (Thomson et al. 2000). Por tanto, la función de la proteína podría verse alterada radicalmente. En el caso de una alteración de la función, un claro ejemplo de ello sería la quimera formada en la familia de ligandos de TNF: TNFSF12 (denominado anteriormente como TWEAK) y TNFSF13 (antes APRIL). La proteína TWEAK es un receptor transmembrana de tipo II, mientras que la proteína APRIL es una proteína soluble secretada al exterior celular. Sin embargo, la quimera TWEAK-APRIL presenta la región citoplasmática y transmembrana de la primera proteína, así como en la parte extracelular el dominio C-terminal de la segunda proteína, de forma que se convierte en un receptor-ligando activo constitutivamente expresado en células T humanas y en monocitos. Como consecuencia, este ligando activo estimula constitutivamente la proliferación en líneas celulares de linfomas T y B (Pradet-Balade et al. 2002). Además, se ha observado un enriquecimiento de regiones transmembrana en las proteínas quiméricas, encontrándose este tipo de dominio en el 51% de las quimeras a estudio (Frenkel-Morgenstern et al. 2012). Este tipo de proteínas pueden estar involucradas en funciones importantes de señalización celular y transporte, por lo que la formación de diferentes quimeras y su regulación puede conllevar repercusiones biológicas relevantes en la célula que se esté expresando.

Otra implicación funcional de vital relevancia para la célula sería la asociación de dos genes que estén involucrados en una misma ruta. Se han observado ejemplos de quimeras que están formadas de manera individual por dos proteínas que funcionan como



complejo proteico en una misma ruta. Estas dos proteínas que se encuentran separadas en una especie, están frecuentemente fusionados en una única proteína denominada “Piedra Rosetta” en otro organismo (Marcotte et al. 1999). Un ejemplo de ello sería la quimera formada por las proteínas NME1 y NME2 (Akiva et al. 2006; Valentijn et al. 2006). Estas proteínas individualmente forman un hexámero en el complejo proteico kinasa difosfato nucleósido (*nucleoside diphosphate kinase*, NDK) (Gilles et al. 1991). La proteína quimera presentaría los dímeros de NME1 y NME2 ya unidos, lo que simplificaría la formación del hexámero.

#### **1.2.4. Evolución del quimerismo**

Estudios comparativos realizados entre diferentes especies mediante el uso de las ESTs disponibles en las bases de datos han permitido indicar la presencia de al menos un nivel de conservación del 10% entre las quimeras humanas y esas mismas quimeras en otras especies (Akiva et al. 2006). Estudios similares se realizaron comparando los eventos de AS presentes entre las especies de humano y ratón, observándose como existía una conservación del 11% entre ambas (Yeo et al. 2005). Así como las variantes canónicas de cada uno de los genes implicados en la quimera tenderán a realizar actividades génicas críticas para la célula, la presencia de quimeras no parece tener un impacto relevante en la fisiología celular y, por ello, pueden evolucionar con mayor rapidez y no encontrarse conservados entre las diferentes especies. Sin embargo, se han descrito excepciones a esta teoría mostrándose un patrón de conservación elevado en la expresión de las quimeras entre seis especies diferentes de mamíferos [(Hernandez-Torres et al. 2013; Lopez-Diez et al. 2013); datos sin publicar], además de encontrarse ampliamente distribuida en los diferentes tejidos a estudio de cada una de las especies.

#### **1.2.5. Asociación con enfermedades**

En la actualidad, la comunidad científica se encuentra dividida en la posible asociación de las quimeras en el desarrollo de diferentes enfermedades, especialmente en su implicación con procesos cancerígenos. Estudios realizados mediante el análisis bioinformático de las secuencias de EST disponibles en las bases de datos muestran como la formación de las quimeras se encuentra presente en un 51% de tejidos normales sin condiciones cancerígenas (Akiva et al. 2006). Sin embargo, hay múltiples estudios realizados sobre quimeras individuales que demuestran su implicación en determinadas

patologías, así como estudios globales con la tecnología de RNA-seq [(Gingeras 2009; Herai and Yamagishi 2010; Kowalski et al. 1999; Kumar-Sinha et al. 2012; Maher et al. 2009; Pradet-Balade et al. 2002; Y. Zhang et al. 2012), entre otros]. Un ejemplo de ello sería la presencia de la quimera HHLA1-OC90 restringida exclusivamente a líneas celulares de teratocarcinomas, mientras que se encuentra ausente en células sanas (Kowalski et al. 1999). Igualmente, la expresión de la quimera TWEAK-APRIL genera un ligando constitutivamente activo que estimula la proliferación en líneas celulares de linfomas T y B (Pradet-Balade et al. 2002). En el caso de la quimera SLC45A3-ELK4, su expresión se ha relacionado con la progresión del cáncer, así como un aumento significativo de sus niveles de expresión desencadenaría en procesos metastáticos en cáncer de próstata (Kumar-Sinha et al. 2012; Y. Zhang et al. 2012).

### 1.2.6. Empleo de nuevas tecnologías

Para el estudio de estos mecanismos de regulación de la expresión génica la comunidad científica ha empleado múltiples metodologías, tanto a través de un enfoque individual como a partir de un enfoque global de su expresión. Al igual que se empleó para el estudio del *splicing*, la presencia de fenómenos de quimerismo (TIC) y de *trans-splicing* se ha llevado a cabo principalmente mediante el empleo de librerías de EST (B. D. Johnson et al. 2002; Kapranov et al. 2002; Shao et al. 2006) permitiendo el estudio “de novo” de nuevas uniones de exones entre dos genes diferentes o diferentes combinaciones de exones imposibles en una lectura normal de un gen y solo posibles debidas a un fenómeno de *trans-splicing*. El problema que radica de esta tecnología es que carece de la profundidad de lectura necesaria para describir aquellos eventos de menor expresión, como es el caso de la mayoría de estos eventos. Además, esta tecnología fue generalmente combinada con datos de RefSeq (proyecto ENCODE) (Akiva et al. 2006; Parra et al. 2006) o librerías de cDNA completo (*full-length*) (Carninci et al. 2005) con la finalidad de obtener mayor profundidad de lectura, así como estudiar las isoformas completas y no solo los eventos específicos de ellas. De hecho, un estudio global del mecanismo de *trans-splicing* se ha realizado exclusivamente empleando librerías de cDNA completo (Herai and Yamagishi 2010). En el estudio realizado en el grupo de Carninci (2005), al emplear lecturas del cDNA completo pudieron abordar también el estudio de los sitios de inicio y terminación de la transcripción. Sin embargo, otros investigadores emplearon la tecnología de 5' y 3' RACE *arrays* (Denoed et al. 2007; Kapranov et al. 2005) para determinar con mayor precisión los sitios de inicio y terminación de la transcripción.

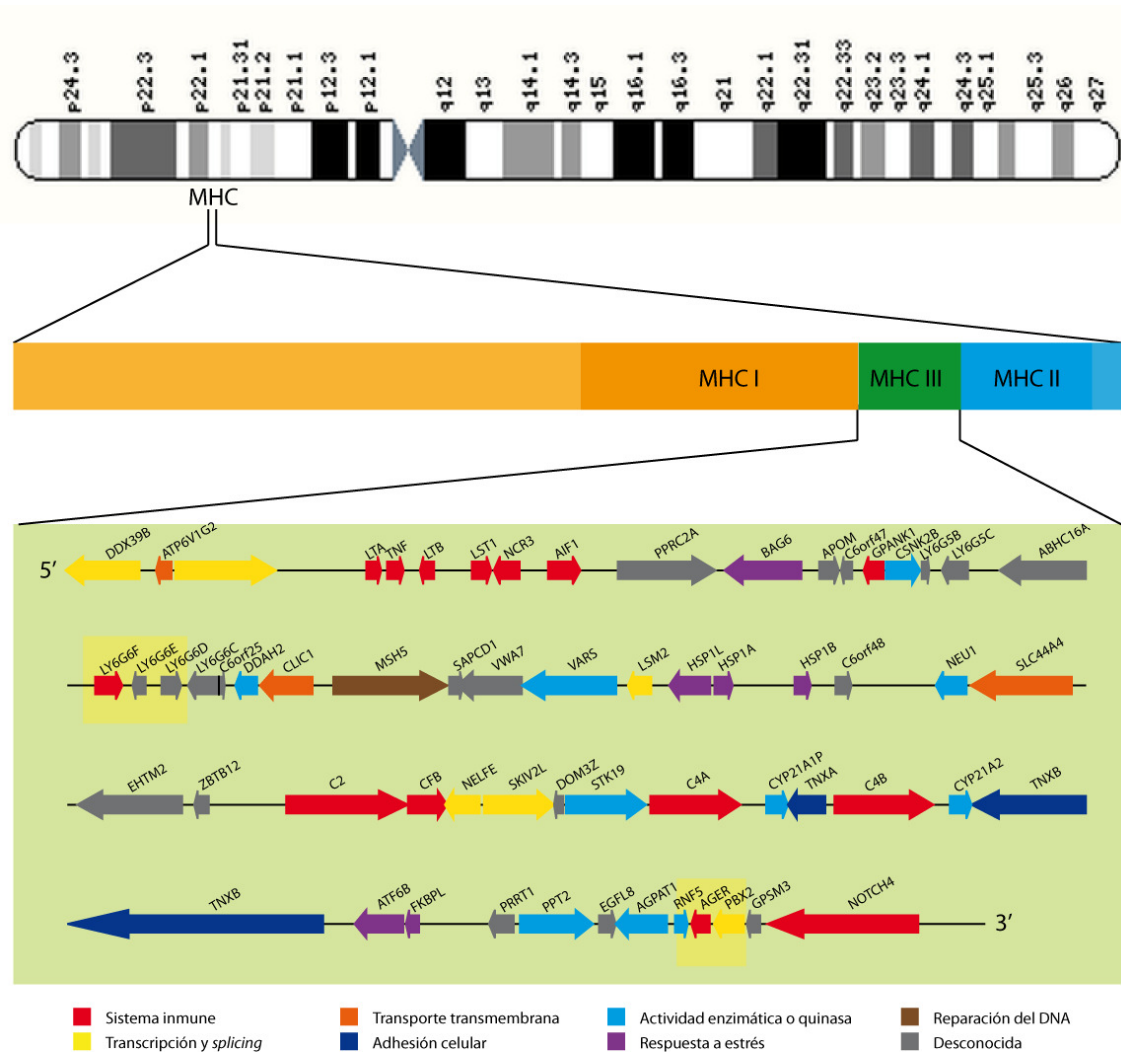
El estudio de ESTs y los *microarrays* han sido recientemente suplantadas por la técnica de secuenciación masiva de RNA (RNA-seq). Los últimos estudios globales de quimeras se han realizado empleando esta tecnología llevando a cabo una búsqueda de uniones de exones pertenecientes exclusivamente a isoformas quiméricas (Frenkel-Morgenstern et al. 2012; Maher et al. 2009; Nacu et al. 2011). Los problemas que radican de esta tecnología son principalmente la cobertura necesaria para la detección de las quimeras, pues los eventos de *splicing* relativos a ellas son por lo general poco abundantes, así como la imposibilidad de detectar o ensamblar los transcritos quiméricos al completo, pudiendo presentar alteraciones en el resto de su secuencia. Los datos de secuenciación masiva se han validado individualmente por qRT-PCR, así como globalmente por *microarrays* de hibridación (Nacu et al. 2011).

Además, la formación de quimeras también ha sido estudiada a nivel proteico tanto a nivel individual como a gran escala, aunque estos ensayos han sido realizados en menor proporción que los anteriores. Se ha realizado un estudio a nivel individual de la quimera CSNK2B-LY6G5B observándose como la proteína quimérica perdería unas posibles modificaciones post-traduccionales presentes en las proteínas individuales CSNK2B y LY6G5B, así como presenta una localización subcelular similar a la proteína CSNK2B (Hernandez-Torres et al. 2013). Se ha realizado también un estudio a nivel global mediante espectrometría de masas confirmándose la expresión de doce quimeras nuevas a nivel proteico (Frenkel-Morgenstern et al. 2012).

### 1.3. COMPLEJO MAYOR DE HISTOCOMPATIBILIDAD (MHC) DE CLASE III

La región de clase III del Complejo Mayor de Histocompatibilidad (*Major Histocompatibility Complex*, MHC) es una región del genoma humano altamente conservada. Estudios previos en esta región la consideran la región con mayor densidad génica del genoma humano (Xie et al. 2003). La alta conservación junto con la gran densidad génica alentan a una gran regulación en niveles posteriores al genómico, pudiendo ejercer una importante regulación los procesos post-transcripcionales.

La región del MHC de clase III (Figura 10) es una de las tres regiones en las que está clasificado el MHC, situándose en su región central entre el MHC de clase I (localizado hacia el telómero) y el MHC de clase II (localizado hacia la región centromérica). En humanos, el MHC es también denominado HLA (*Human Leukocyte Antigen*) al ser conocido por su papel en los trasplantes a través de los antígenos de histocompatibilidad



**Figura 10. Localización del MHC.** Representación de la localización cromosómica de la región génica del MHC humano, mostrándose en la parte inferior en detalle los genes situados en esta región. Cada gen está representado con una flecha indicando la orientación del transcrito, así como

(Vandiedonck and Knight 2009). El MHC se encuentra localizado en el cromosoma 6p21.3 y presenta una longitud de aproximadamente 7.6 Mb (Horton et al. 2004). La mayoría de los genes localizados en el MHC están involucrados en el sistema inmune del organismo, entre cuyos principales representantes se encontrarían los genes que codifican las moléculas de histocompatibilidad implicadas en la presentación de antígenos a los linfocitos T, tales como los genes HLA-A, HLA-B y HLA-C en la región de clase I, y los genes HLA-DP, HLA-DQ y HLA-DR en la región de clase II (Stephens et al. 1999; The MHC Consortium 1999).

La región del MHC de clase III en humanos presenta una longitud de tan sólo alrededor de 0.7 Mb, y en ella se encuentran 59-63 genes y 0-2 pseudogenes dependiendo del haplotipo (The MHC Consortium 1999; Xie et al. 2003), siendo una de las regiones más

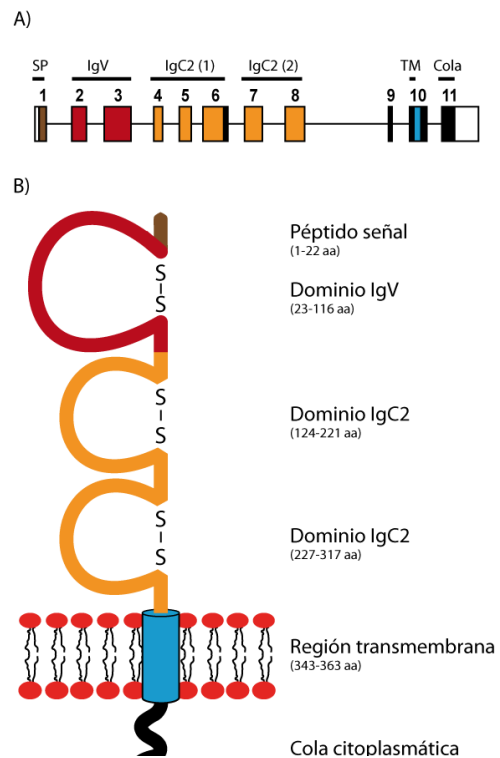
densas de todo el genoma humano. Aunque se desconoce la función de una gran parte de los genes codificantes del MHC de clase III, una mayoría de ellos parecen estar involucrados en el sistema inmune. Entre ellos nos encontramos con los componentes del complemento C4A, C4B y Factor B (BF), así como citoquinas inflamatorias tales como el Factor de Necrosis Tumoral (*Tumor Necrosis Factor*, TNF $\alpha$ ), y linfoquinas  $\alpha$  y  $\beta$  (LT $\alpha$  y LT $\beta$ ). Además, el MHC de clase III también codifica proteínas con un amplio rango de funciones celulares, tales como la sintetasa valyl tRNA (VARS2) (Hsieh and Campbell 1991), y enzimas como la ácido lisofosfatídico acil transferasa (AGPAT1) (Aguado and Campbell 1998) o la palmitoil proteína tioesterasa (PPT2) y la sialidasa Neu1 (Milner et al. 1997). También hay genes que están potencialmente involucrados en la regulación de la transcripción y el procesamiento del RNA, tales como *PBX2* (Aguado and Campbell 1995) y *BAT1*, un factor de *splicing* que además está involucrado en la exportación de los mRNAs del núcleo al citoplasma (Luo et al. 2001).

Al igual que las regiones de clase I y II del MHC, la región de clase III se encuentra asociada a diversas enfermedades o patologías en su desarrollo y/o susceptibilidad, siendo la mayoría enfermedades de carácter autoinmune (Horton et al. 2004). Debido a los desequilibrios de ligamiento en el MHC (Milot and Dausset 1978) y por tanto a la presencia de diferentes haplotipos, así como a que la mayoría de las enfermedades asociadas son multifactoriales, es extremadamente difícil la vinculación directa de un gen/isoforma de esta región con el desarrollo de la enfermedad (Horton et al. 2004). Entre las enfermedades de carácter autoinmune que se han vinculado con esta región del genoma se encuentran la Inmunodeficiencia Variable Común (CVID) (Cucca et al. 1998; Schroeder et al. 1998), la Orquitis, enfermedad autoinmune de los testículos (Snoek et al. 1998), la Artritis Reumatoide (AR) (Bali et al. 1999; Chiba et al. 2011; Hu et al. 2011; Jenkins et al. 2000; Kilding and Wilson 2005), la Espondilitis Anquilosante (Brown et al. 1998; Milicic et al. 2000), la Diabetes Tipo I (Kumar et al. 2012) y el Lupus Eritematoso Sistémico (Fernando et al. 2012). Además, también se ha observado susceptibilidad a determinadas enfermedades infecciosas como la Malaria (Delahaye et al. 2007), la Hepatitis C (Shichi et al. 2005) y el Síndrome de Inmunodeficiencia Adquirida (SIDA) (Guernon et al. 2012), entre otras. Por ello, el estudio de los diferentes genes situados en el MHC de clase III (tales como *RAGE*, *G6F* o *LY6G6D*) puede facilitarnos la comprensión y entendimiento de todas las enfermedades mencionadas anteriormente y otras aún sin asociar directamente con esta región.

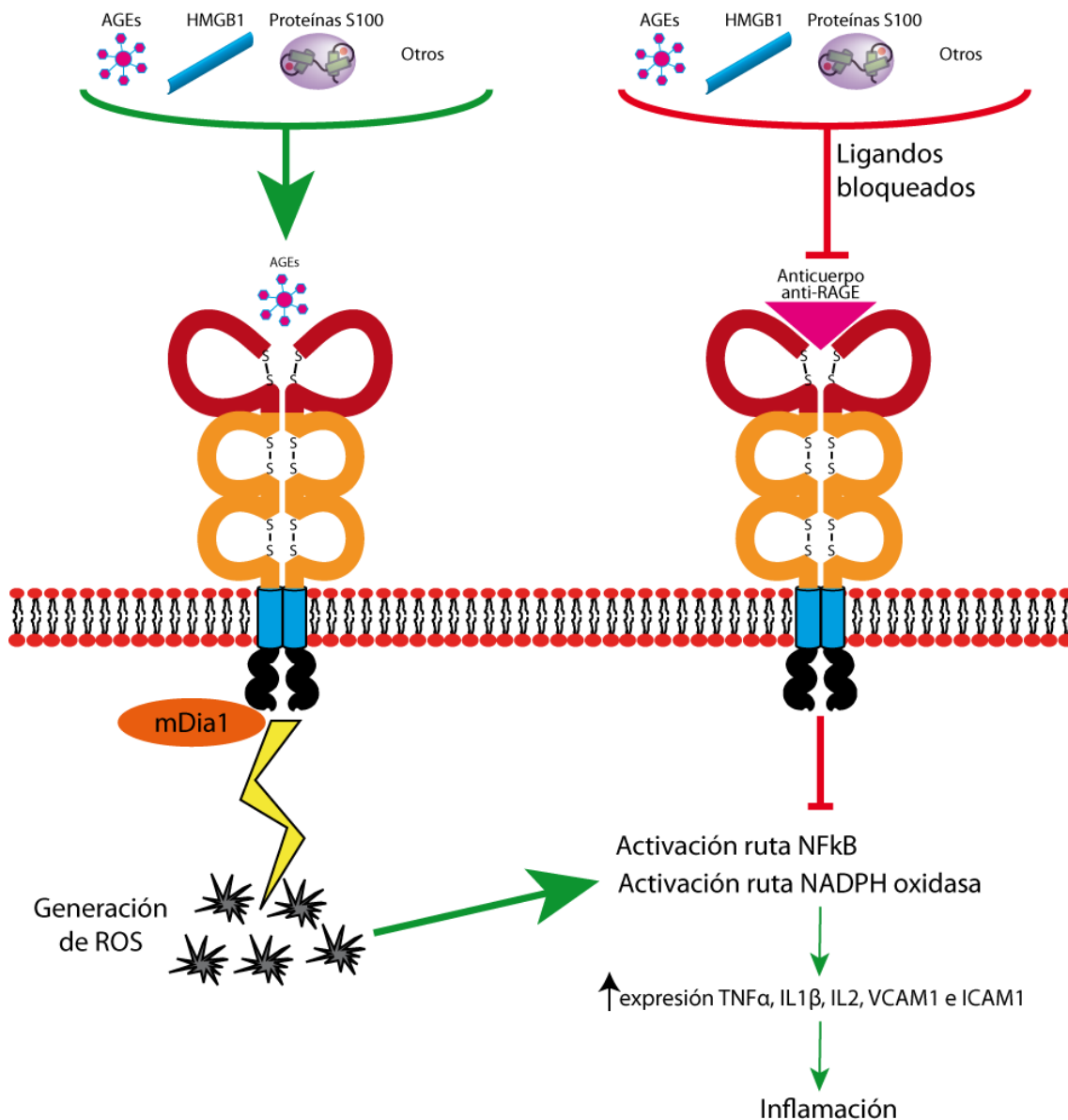
### 1.3.1. El gen *RAGE*.

Entre los genes que se encuentran en la región del MHC de clase III, se encuentra el gen *Advanced Glycosylation End-products Receptor* (*AGER*, ó más comúnmente denominado *RAGE*). El gen *RAGE* está compuesto por once exones codificantes (Figura 11a) y un mensajero resultante de alrededor de 1.4 kb que es traducido en una proteína de 404 aminoácidos y alrededor de 55 kDa (Neeper et al. 1992). La proteína *RAGE* o receptor de agentes de glicosilación avanzada pertenece a la superfamilia proteica de las Ig y está formada por un número variable de dominios proteicos dependiendo de cada isoforma. La isoforma canónica de *RAGE* presenta una región extracelular compuesta por tres dominios de Ig, incluyendo un dominio Ig variable (IgV) en el N-terminal de la proteína y, consecutivamente, dos dominios Ig constantes de tipo 2 (IgC2) responsables de la interacción con los ligandos, así como una región transmembrana única de tipo I y una pequeña cola citoplasmática (Hudson et al. 2008b) (Figura 11).

La proteína *RAGE* se clasifica como receptor de superficie celular involucrado en el reconocimiento y señalización de un amplio rango de ligandos (Figura 12), entre los que se incluyen: los productos de glicosilación avanzada (AGEs) (Schmidt et al. 1992), miembros de la superfamilia de las proteínas S100/Calgranulinas (Hofmann et al. 1999), amfoterina o también conocido como HMGB1 (Hori et al. 1995), la forma amiloidogénica del péptido beta amiloide ( $A\beta$ ) (Yan et al. 1996), LPA (Rai et al.), Mac1 (Chavakis et al. 2003) y C1q (Ma et al.). La unión de dichos ligandos al receptor *RAGE* desencadena estrés oxidativo con consecuencias específicas de tejido (Daffu et al. 2013). Esto es provocado por la generación de especies reactivas de oxígeno (*Reactive Oxygen Species*, ROS), así como un aumento en la expresión de moléculas inflamatorias tales como  $TNF\alpha$ ,  $IL1\beta$ ,  $IL2$ , VCAM1 e ICAM1 (Hofmann et al. 1999; Z. Wang et al. 2002b), mediado por las rutas de  $NF\kappa B$  (Hofmann et al. 1999; Reddy et al. 2006; Yan et al. 1994) y de NADPH oxidasa (Wautier et



**Figura 11. Estructura de *RAGE*.** Estructura del transcrito (a) y de la proteína (b) de *RAGE* con los dominios proteicos que codifica.



**Figura 12. Señalización celular de RAGE.** La interacción de los ligandos con los dímeros de RAGE provoca una cascada de señalización intracelular que involucran especies reactivas de oxígeno (ROS) y el aumento de expresión de ciertas citoquinas inflamatorias. Esta cascada es inhibida cuando se impide la unión de los ligandos al receptor (mediante anticuerpos anti-RAGE).

al. 2001) (Figura 12). Las diferentes rutas de señalización intracelular son dependientes de la naturaleza del estrés celular y del tipo celular (Daffu et al. 2013), aunque todas las rutas convergen en la interacción específica de la molécula formina mDia1 a la región citosólica de RAGE, siendo dicha interacción la responsable de la activación de la respuesta de señalización celular (Hudson et al. 2008a; Toure et al. 2012; Xu et al.) (Figura 12).

La desregulación de este proceso parece estar relacionado con diferentes enfermedades (Kalea et al. 2011) incluyendo diabetes (Goh and Cooper 2008), enfermedades inmunológicas y complicaciones inflamatorias en arterioesclerosis (Harja et



al. 2008), enfermedad del Alzheimer (Ding and Keller 2005a, 2005b; Emanuele et al. 2005), y en el aumento de invasividad en diversos cánceres (Ghavami et al. 2008; Hiwatashi et al. 2008).

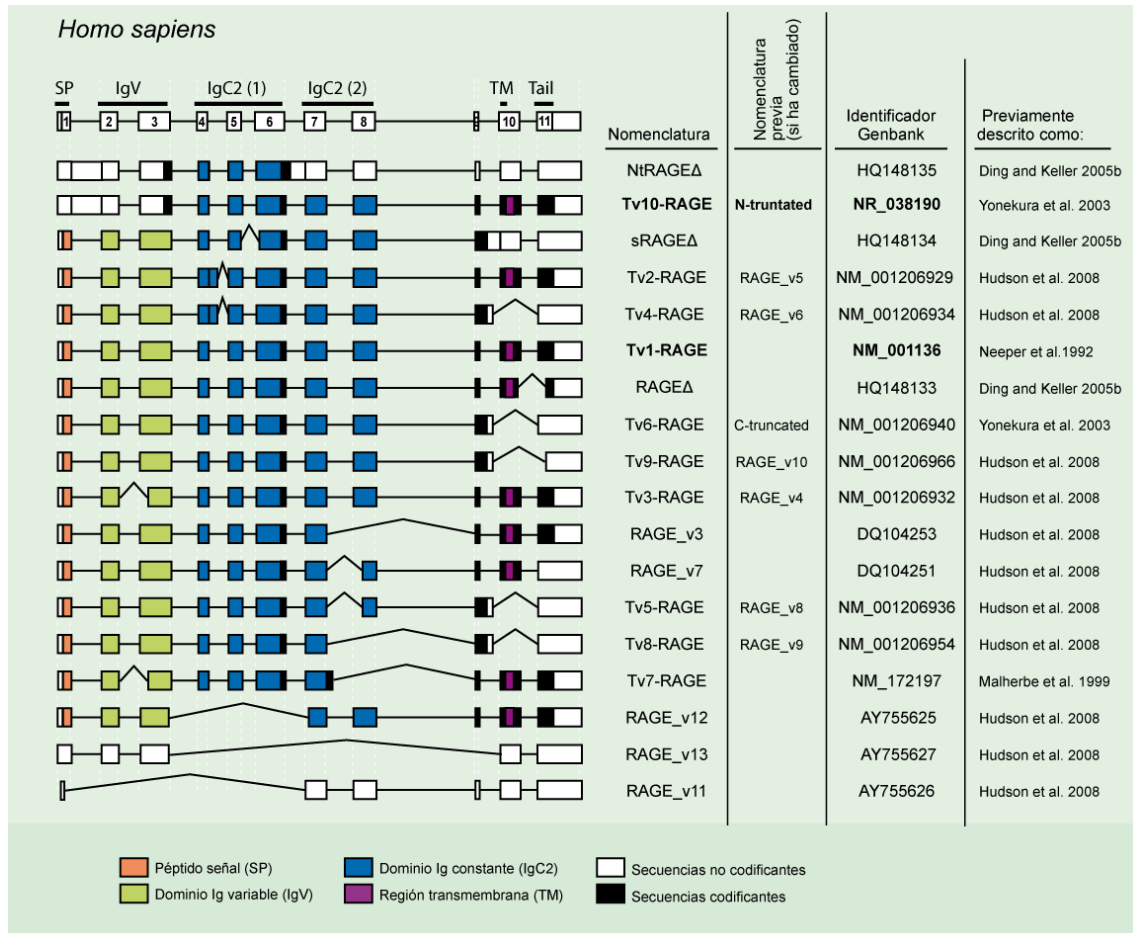
#### **1.3.1.1. *Splicing alternativo en RAGE e implicaciones funcionales***

El mecanismo del *splicing* alternativo parece ser particularmente importante en la regulación de este gen, de forma que en los últimos años se han realizado diversos estudios en los que se han descrito diferentes variantes de *splicing* de *RAGE* en humano (Ding and Keller 2005b; Hudson et al. 2008b; Malherbe et al. 1999; Yonekura et al. 2003), en ratón (Kalea et al. 2009), y en diversos tejidos y muestras de caninos (Sterenczak et al. 2009; Sterenczak et al. 2013). Las isoformas generadas a partir del AS del gen *RAGE* presentan variaciones en su secuencia codificante que alteran los dominios proteicos de *RAGE*. Cuando los fenómenos de AS se producen en los exones correspondientes a los dominios Ig extracelulares, la proteína resultante presentará modificaciones en dichos dominios pudiendo verse afectada las interacciones de los diferentes ligandos. En cambio, como se ha comentado anteriormente, cuando estos fenómenos se producen en el exón correspondiente a la región transmembrana, se pueden producir isoformas solubles que se secretarán al exterior celular (Hudson et al. 2008b; Maillard-Lefebvre et al. 2009).

En humano nos encontramos con diferentes modificaciones en los dominios de interacción a ligando. La aparición de un sitio de *splicing* 3' acceptor alternativo en el exón 3 modificaría el dominio IgV (Hudson et al. 2008b; Malherbe et al. 1999), así como la pérdida del exón 8 cambiaría la presencia y/o funcionalidad del segundo dominio IgC2 (Hudson et al. 2008b; Malherbe et al. 1999). El primer dominio IgC2 estaría afectado en las variantes que presentasen un sitio de *splicing* 3' acceptor alternativo en el exón 6 (Ding and Keller 2005b), ó que presentasen la retención parcial del intrón 4 (48 pares de bases) debido a la presencia de un sitio de *splicing* 5' donador alternativo en el exón 4 (Hudson et al. 2008b) (Figura 11a y 13). Estas modificaciones de los dominios Ig podrían estar generando variaciones en la interacción con ligando, ya fuese en la especificidad del receptor por un determinado ligando u otro, o incluso podría modular la fortaleza de su unión.

Otro evento que alteraría la funcionalidad de la isoforma resultante afecta a la cola citoplasmática. En este caso, la aparición de un sitio de *splicing* 3' acceptor alternativo en el exón 11 generaría la pérdida de la mayor parte de la cola citoplasmática, generando de esta forma un posible dominante negativo de *RAGE* (Ding and Keller 2005a). Este dominante

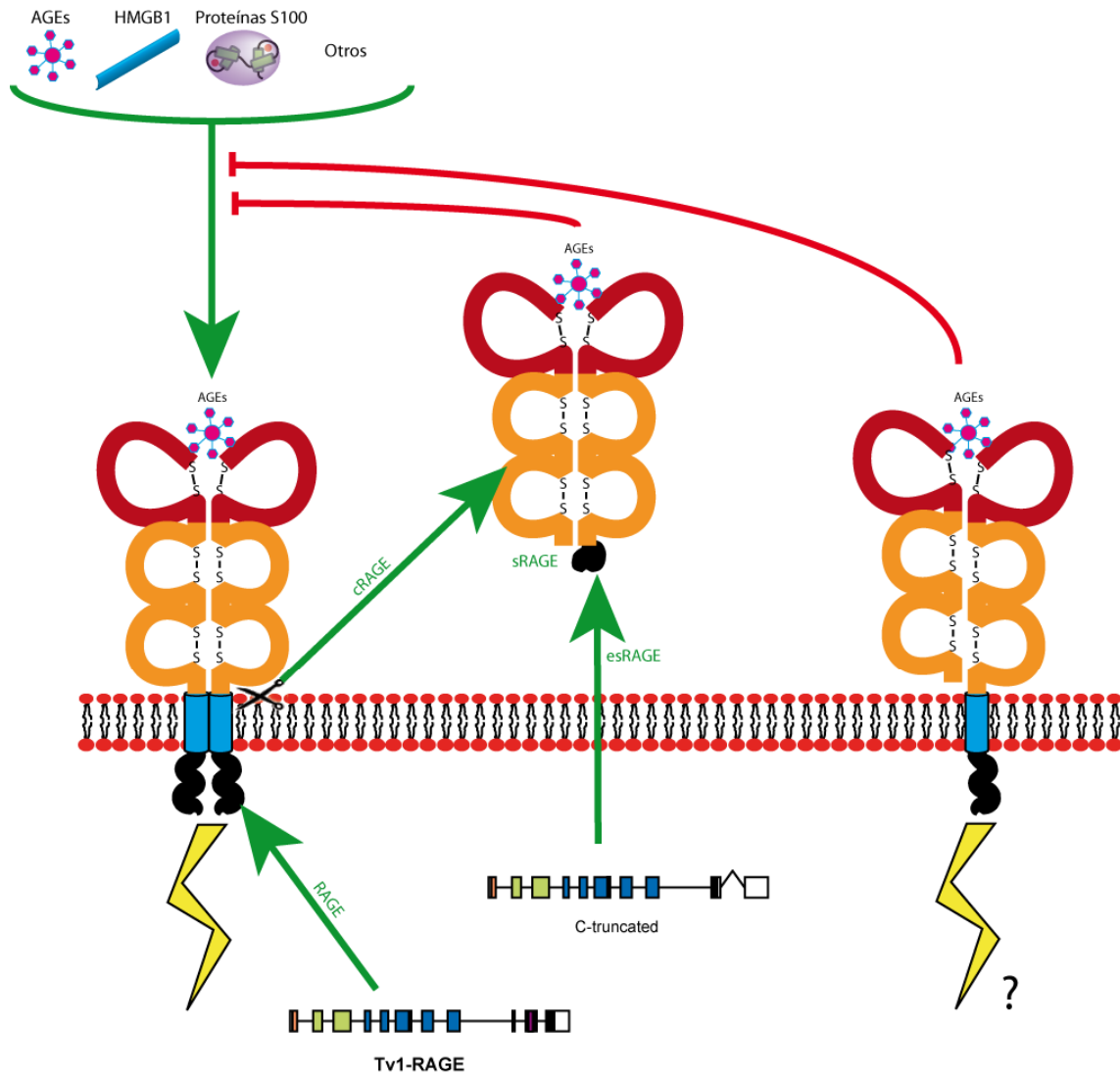




**Figura 13. Variantes de *splicing* del gen *RAGE* previamente descritas en *Homo sapiens*.**

negativo seguiría interaccionando con los ligandos pero no transmitiría una señalización celular o al menos no transmitiría una señal similar a la que transmitiría la isoforma canónica.

Además, existen otras variantes que carecen de la región transmembrana de la proteína generando isoformas solubles (sRAGE) mediante AS (esRAGE) (Maillard-Lefebvre et al. 2009) (Figura 14). Este es el caso de la retención parcial del intrón 9 debido a la presencia de un sitio de *splicing* 5' donador alternativo en el exón 9 y la pérdida del exón 10, lo que provocaría la pérdida de la región transmembrana. Este evento se ha encontrado presente en múltiples variantes (Hudson et al. 2008b; Yonekura et al. 2003). Sin embargo, se ha investigado otro método para la generación de isoformas sRAGE que consiste en el corte proteolítico de la proteína RAGE canónica transmembrana mediante la proteína shedasa ADAM10 (*cleaved RAGE*, cRAGE) (Raucci et al. 2008) (Figura 14). Estas isoformas solubles estarían inhibiendo la señalización celular al interaccionar y secuestrar



**Figura 14. Diferentes interacciones de dímeros de RAGE con sus ligandos y, correspondiente señalización intracelular.** Se observa la producción de proteína transmembrana canónica (RAGE), así como la isoforma soluble de RAGE (sRAGE) generada por *splicing* (esRAGE) o por corte proteolítico de la isoforma transmembrana (cRAGE). Dímeros de la proteína soluble o heterodímeros con una transmembrana regularían negativamente la transmisión de señales a través de RAGE.

los ligandos (Zang et al 2010) o, al generar heterodímeros afuncionales con la isoforma canónica (Sarkany et al. 2011); Zang et al 2010).

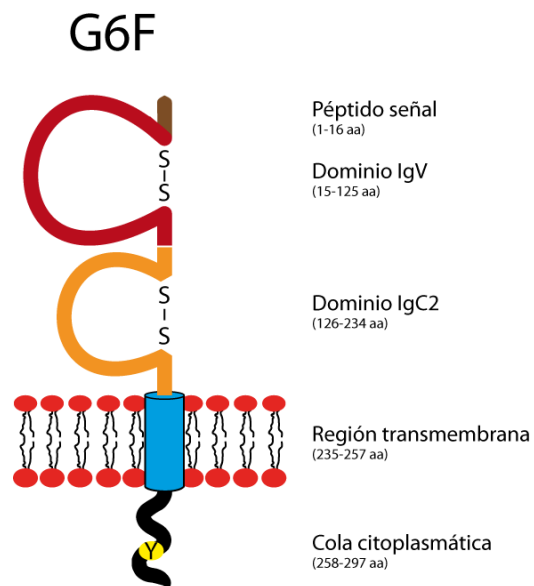
La proteína soluble sRAGE es estable en su forma monomérica en presencia de calcio pero, en cambio, en ausencia o bajas concentraciones de este catión la molécula sRAGE forma un tetrámero construido por dos homodímeros (Sarkany et al. 2011). Asimismo, la proteína transmembrana precisa de la generación de dímeros para la interacción de los ligandos y, por tanto, para la transmisión de la señal intracelular (Zong et al. 2010). En la formación del dímero participan un átomo de Zinc y enlaces por puentes de

hidrógeno situados en el primer dominio IgC2 (Koch et al. 2010) o, según otro estudio, el dominio IgV es el responsable de la interacción de dicho dímero (Sarkany et al. 2011) (Figura 14). Además, la formación de los dímeros con diferentes conformaciones puede provocar una respuesta fisiológica variada (Park et al. 2010).

### 1.3.2. El gen *G6F*.

La proteína G6F es una proteína transmembrana de tipo I perteneciente a la superfamilia de las Ig y, por lo tanto, involucrada en el sistema inmune y en el reconocimiento celular (Williams and Barclay 1988). La expresión de G6F se encuentra restringida a la línea plaquetaria (Garcia et al. 2006; Lewandrowski et al. 2006; Macaulay et al. 2007; Senis et al. 2009), careciendo de expresión en el resto de células sanguíneas, por lo que se piensa que podría estar implicada en la señalización y modulación de la función plaquetaria. Nuestro grupo ha descrito la proteína G6F humana indicando que contiene dos dominios extracelulares de Ig y una región intracelular corta (Figura 15), y que forma homodímeros (de Vet et al. 2003). Se ha observado que esta proteína se encuentra N-glicosilada (De Vet et al. 2003; Lewandrowski et al. 2006), presentando residuos susceptibles para N-glicosilación en uno de sus dominios Ig.

En relación con la región intracelular corta, se ha observado su interacción con los dominios SH2 de las proteínas *Growth-factor-receptor-bound protein* Grb2 y Grb7, siendo dicha interacción dependiente de la fosforilación del residuo de tirosina Y281 de G6F (De Vet et al. 2003). Dicho residuo se encuentra localizado en el motivo de unión consenso (YXN) en la cola intracelular de G6F. La fosforilación de este residuo se encuentra asociada con la activación plaquetaria en respuesta a *Collagen-related peptide* (CRP) o en respuesta a colágeno, mientras que no se da en respuesta a agonistas del receptor acoplado a proteína



**Figura 15. Estructura de la proteína G6F.** Se muestra en detalle la composición de dominios proteicos y los aminoácidos correspondientes a cada dominio.

G, ni a trombina, ni a ADP (Garcia et al. 2006). Las rutas de transducción de señales consecutivas incluyen la ruta de la kinasa Ras-MAP (De Vet et al. 2003).

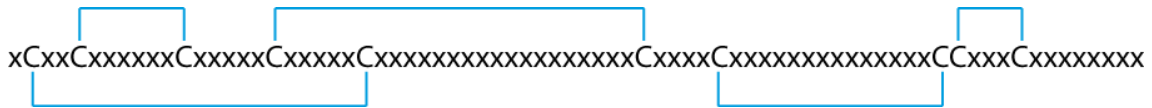
Recientemente, se han estudiado una proteína similar a G6F (*G6F-like*) en peces teleósteos (C. E. Hughes et al. 2012; Ohashi et al. 2010). Al igual que la proteína humana, esta proteína se encuentra codificada por un gen presente en el contexto genómico del MHC de clase III. Además, es una proteína transmembrana de tipo I, pero en este grupo presenta cuatro dominios extracelulares de IgV y una cola citoplasmática de mayor tamaño que contiene tanto un motivo YXN (como la humana) como un motivo de activación basado en el inmunoreceptor de tirosinas (ITAM, *Immunoreceptor tyrosine-based activation motif*) (Ohashi et al. 2010) que podría ejercer un papel importante en la activación celular. Esta proteína se encuentra expresada en los trombocitos y plaquetas de los peces, así como también en los eritrocitos de dos especies estudiadas pertenecientes a la familia de los ciprínidos (Ohashi et al. 2010). Ambas células forman la línea hematopoyética en peces teleósteos. Además, se ha observado como la proteína G6F-like es un receptor de colágeno en trombocitos de peces, y desencadena en la ruta de señalización de ITAM, similar a la presente en mamíferos (C. E. Hughes et al. 2012).

#### **1.3.2.1. Splicing alternativo en G6F e implicaciones funcionales**

La expresión de este gen se ha descrito en la línea K562, donde se caracterizó la isoforma canónica. El estudio de la presencia de eventos de AS se realizó en el mismo análisis pero no se detectaron variantes en ninguna de las líneas celulares analizadas (De Vet et al. 2003). En un estudio posterior se observa también únicamente la expresión de la isoforma canónica y, además, variaciones de AS mediante la formación de quimeras (Calvanese et al. 2008).

#### **1.3.3. El gen LY6G6D.**

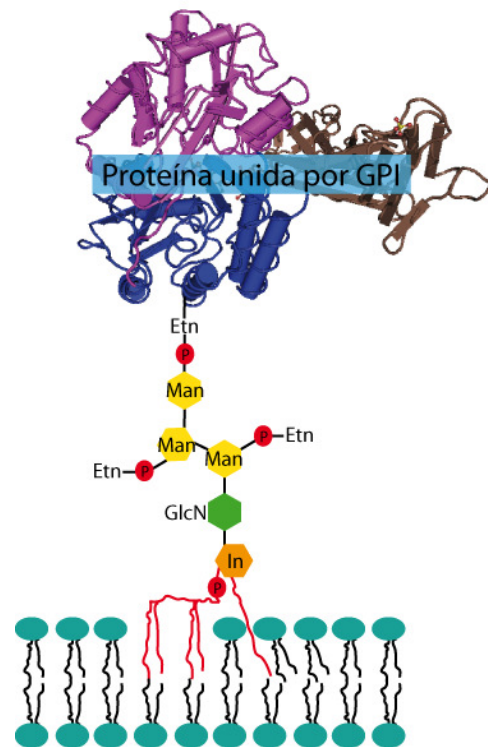
La proteína LY6G6D pertenece a la superfamilia de los antígenos 6 de linfocitos (Ly-6), encontrándose un grupo o *cluster* de genes pertenecientes a dicha familia en la región del MHC de clase III (Mallya et al. 2002; Xie et al. 2003). Estas proteínas se encuentran conservadas en ratón y se caracteriza por ser el *cluster* de mayor tamaño en el MHC de clase III (Horton et al. 2004). El dominio de proteínas de esta familia se caracteriza por presentar ocho o diez residuos de cisteína que generan un patrón de puentes disulfuro (SH) (4 ó 5 puentes disulfuro, respectivamente) muy definido (Figura 16). Se



**Figura 16. Patrón de puentes disulfuros presentes en la secuencia de las proteínas de la superfamilia Ly-6.** “C” corresponde con las cisteínas conservadas que se encuentran involucradas en los puentes disulfuro marcados por corchetes azules.

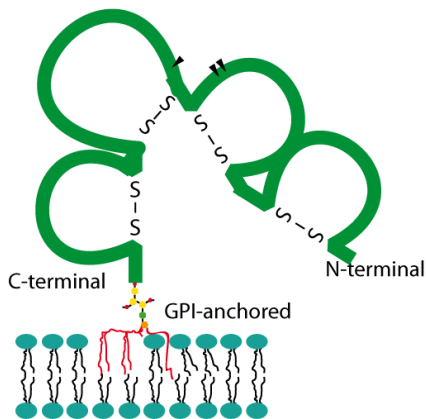
piensa que la proteína LY6G6D podría tener alguna relación potencial con la función inmune (Mallya et al. 2002), aunque en realidad la función de esta proteína es aún desconocida, al igual ocurre con la mayoría de las proteínas de su familia, a excepción de la proteína CD59 que se ha identificado como inhibidor de la cascada de complemento (Davies et al. 1989), y la proteína *urokinase Plasminogen Activator Receptor* (uPAR) (Wilhelm et al. 1999) que juega un papel importante en la proteólisis de las proteínas de la matriz extracelular.

Las proteínas de la familia Ly6 se encuentran generalmente ancladas en la superficie celular a partir de un *glycosylphosphatidylinositol* (GPI) (Figura 17). Inmediatamente después de la síntesis de la proteína, ésta queda anclada a la membrana del RE mediante una secuencia de 15-20 aminoácidos hidrofóbicos en su C-terminal mientras que el resto de la proteína se encuentra en el lumen del RE. Endoproteasas del RE (entre las que se encuentra la enzima fosfolipasa C) catalizan la adición covalente del anclaje por GPI (complejo de fosfolípidos glicosilados consistentes en dos grupos de ácidos grasos y varios azúcares, entre los que se incluyen *N-acetylglucosamine*, manosa e inositol) al C-terminal de estas proteínas destinadas a la membrana plasmática. Esta unión se forma en el lumen del RE, donde al mismo tiempo el segmento C-terminal de la proteína es cortado (Ferguson 1992; Muniz and Riezman 2000; Traub and Kornfeld 1997) de forma que la proteína queda anclada a la membrana por el GPI y no por una



**Figura 17. Anclaje por GPI a la membrana plasmática.** El anclaje se realiza mediante un complejo de fosfolípidos y varios azúcares (Etn, etanolamina; GlcN, glucosamina; In, inositol; Man, manosa).

## LY6G6D



**Figura 18. Estructura de la proteína LY6G6D.** Se observa el dominio Ly en verde, anclado a la membrana plasmática en el C-terminal de la proteína por el GPI. Las glicosilaciones se muestran como triángulos negros.

secuencia transmembrana. Se piensa que las proteínas ancladas por GPI se encuentran asociadas con los glicosfingolípidos en los *rafts* lipídicos que se forman en la región *trans* del Golgi (Jacobson and Dietrich 1999; Simons and Ikonen 1997).

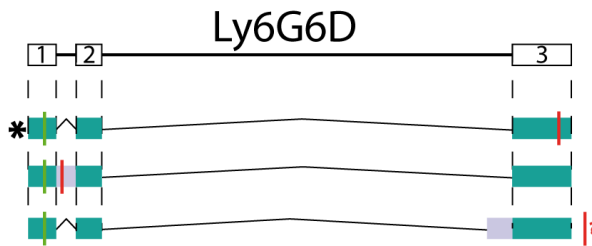
Nuestro grupo caracterizó *LY6G6D* tanto a nivel transcripcional como a nivel proteico (Mallya et al. 2002, 2006), observando como la proteína *LY6G6D* en humano presenta un potencial dominio Ly-6 con 8 cisteínas (generando únicamente cuatro puentes disulfuro), mientras que los ortólogos de ratón, cerdo y rata presentaban diez cisteínas (generando los cinco puentes disulfuro) (Mallya et al. 2006). Esta proteína formaba homodímeros uniéndose a la membrana plasmática gracias al anclaje por GPI.

Además, se describió que *Ly6G6D* se encuentra glicosilado en residuos de O-glicosilación presentando ramificaciones por ácido siálico (Mallya et al. 2006) (Figura 18). Según estudios realizados en proteínas de la misma familia, se ha observado como los patrones de glicosilación podrían ser importantes para la funcionalidad de la proteína. Ejemplos de ello sería el estado de glicosilación de la molécula de Ly-6 uPAR y su capacidad para influir en su interacción por ligando (Moller et al. 1993), o la proteína CD59 ya que parece contribuir tanto a la estructura de la proteína como a la estabilidad del sitio activo (Rudd and Dwek 1997).

### 1.3.3.1. *Splicing alternativo en LY6G6D e implicaciones funcionales*

En un estudio realizado en diferentes genes de la familia Ly6 del MHC de clase III, se analizaron los fenómenos de *splicing* alternativo en el gen *LY6G6D* humano y de ratón (Mallya et al. 2002). Además, en otro estudio realizado también en el laboratorio se observaron tanto los fenómenos de *splicing* alternativo en el gen *LY6G6D* humano, como la formación de quimeras entre este gen y el gen *G6F* situado al 5' del gen *LY6G6D* (Calvanese et al. 2008). En ambos estudios realizados en muestras de humano se observa la presencia de tres transcritos diferentes mientras que en ratón solo observamos dos variantes de *splicing* (Figura 19). Los transcritos comunes corresponderían con la variante canónica,

y con un transcrito que presenta el intrón 1 retenido. La retención de dicho intrón provoca la aparición de un PTC y, curiosamente este transcrito no es susceptible de degradación por el mecanismo de NMD (Calvanese et al. 2008) (Mallya et al. 2002). En humano se ha descrito una variante nueva que presenta un sitio de *splicing* 3' aceptador alternativo en el exón 3 (Mallya et al. 2002), que altera tanto la disposición de los puentes disulfuro como la región de anclaje por GPI en la proteína madura, afectando a su función.

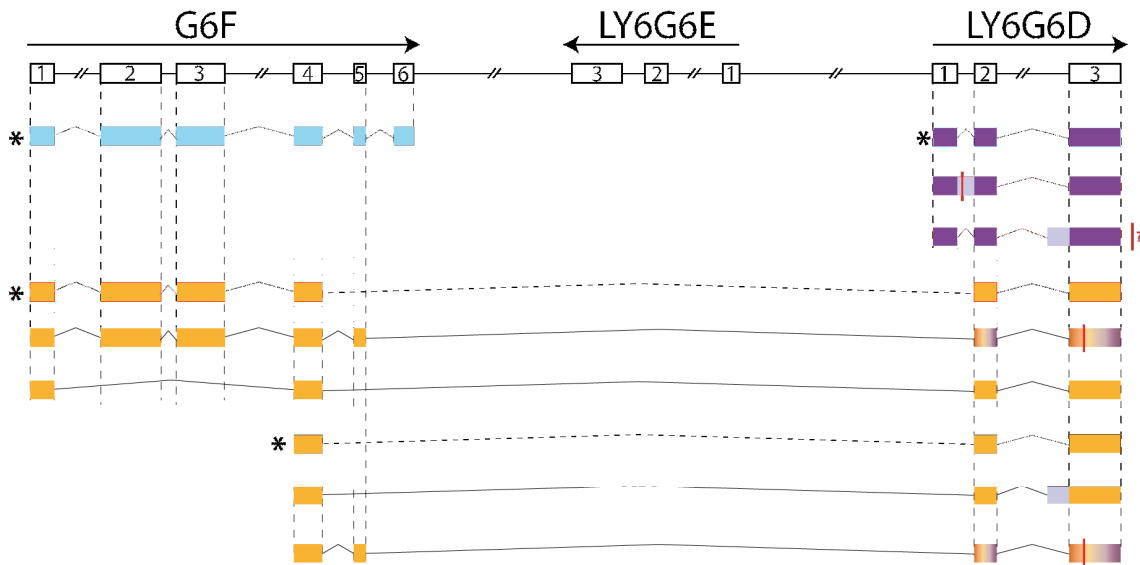


**Figura 19. Variantes de *splicing* del gen *LY6G6D* presentes en humano y ratón representadas a escala.** Las barras verdes y rojas indican el codón de inicio y de parada de la traducción, respectivamente. El asterisco corresponde con la isoforma canónica.

#### 2.3.4. Quimeras entre los genes *G6F* y *LY6G6D*.

Un estudio llevado a cabo en el laboratorio analizó la formación de quimeras entre los genes *G6F* y *LY6G6D* (Calvanese et al. 2008). En dicho estudio se observó la presencia de dos tipos de quimeras, la primera correspondería con la quimera “canónica” que iría desde el comienzo del gen *G6F* hasta el final del gen *LY6G6D*, (exceptuando los dos últimos exones de *G6F* y el primero de *LY6G6D*) y la segunda correspondería con una quimera de menor tamaño que presentaría solo el exón 4 del gen *G6F* y los dos últimos exones del gen *LY6G6D* (Figura 20). La quimera resultante “canónica” generaría una proteína que contendría dos dominios Ig y una región transmembrana (pertenecientes a la proteína *G6F*) y una secuencia correspondiente al dominio Ly-6 (perteneciente a la proteína *LY6G6D*), pero que no se plegaría como tal al encontrarse intracelularmente. La quimera de menor tamaño presentaría exclusivamente la región transmembrana de la proteína *G6F* que podría actuar como péptido señal y el dominio Ly-6 de la proteína *LY6G6D* que en este caso sí se podría plegar como tal. Además se observaron otras potenciales quimeras producidas por otros diferentes eventos de *splicing* que podrían producir alteraciones en estas proteínas (Calvanese et al. 2008) (Figura 20). En dicho estudio se observa también una expresión diferencial según el tipo de quimera, presentando una expresión limitada a la línea celular K562 de las quimeras que comienzan en el exón 1 de *G6F* y, una expresión más amplia y diversa en aquellas quimeras que comienzan exclusivamente a partir del exón 4 del gen *G6F* (Calvanese et al. 2008).





**Figura 20.** *Splicing* y quimerismo en los genes *G6F* y *LY6G6D* humanos [Modificado de: Calvanese et al (2008)]. Se muestra la orientación de los genes mediante flechas. Las líneas verticales rojas indican la presencia de PTC en la secuencia de la proteína. El asterisco corresponde con la isoforma canónica.

#### 1.4. ARTRITIS REUMATOIDE

La artritis reumatoide (AR) es una enfermedad autoinmune sistémica de etiología desconocida que se caracteriza por la inflamación sinovial crónica de las articulaciones y la subsiguiente destrucción del cartílago y del hueso, debido a una infiltración de tejido granulador denominado *pannus*. Las características principales de dicha enfermedad son sinovitis en una presentación simétrica, rigidez matutina, desarrollo de nódulos reumatoides subcutáneos y/o depósitos de fibrina en el tejido sinovial reumatoide. Existen diferentes pruebas o exámenes diagnósticos para determinar si se padece dicha enfermedad. La prueba que generalmente se realiza es la prueba del factor reumatoide (FR) positivo (pacientes AR seropositivos) o negativo (pacientes AR seronegativos), mediante pruebas de aglutinación IgM-FR, nefelometría o por ensayo inmunoenzimático (ELISA). En la actualidad, se está complementando dicha prueba con la prueba de anticuerpo anti-péptido cíclico citrulinado (anti-CCP) realizada empleando péptidos sintéticos que contienen citrulina y detectan la presencia de autoanticuerpos para péptidos citrulinados. Otras pruebas empleadas con el análisis de la proteína C reactiva, recuento sanguíneo completo, tasa de sedimentación eritrocítica, ecografías, resonancias magnéticas o radiografías de las articulaciones, así como el análisis del líquido sinovial.



La enfermedad de AR se puede controlar mediante diferentes tipos de tratamientos que se podrían clasificar en estas cuatro categorías:

1. El tratamiento mediante anti-inflamatorios no esteroideos (AINES/NSAID) que alivian el dolor y la inflamación a corto plazo. En este grupo se incluirían los AINES clásicos (Ibuprofeno, naproxeno...) que actúan inhibiendo la enzima ciclooxigenasa (COX) de manera genérica (afectan tanto a COX-1 como a COX-2), y los AINES selectivos o inhibidores de COX-2.
2. El tratamiento mediante glucocorticoides (Prednisona, metilprednisolona...) mejoran la calidad de vida a largo plazo. Estas sustancias lipofílicas se unen al receptor citosólico provocando la expresión de genes de respuesta a glucocorticoides que inhiben el reclutamiento celular en los sitios inflamados y bloquean los mediadores proinflamatorios tanto en su producción como en sus efectos.
3. El tratamiento mediante fármacos antirreumáticos modificadores de la enfermedad (FAMEs o *Disease Modifying Antirheumatic Drugs*, DMARDs) como terapias de segunda línea que disminuyen la actividad de la enfermedad a largo plazo. Los principales tratamientos de segunda línea se realizan con el metotrexato (Mtx, Rheumatrex<sup>®</sup>) que provoca un efecto antiinflamatorio y la leflunomida (Arava<sup>®</sup>) como inmunomodulador de la AR.
4. Los nuevos tratamientos biológicos o fármacos modificadores de la respuesta biológica generan los mejores resultados en relación con la evolución y prognosis de la AR y se están empleando más recientemente. Existen tres tipos de tratamientos, los inhibidores de TNF $\alpha$  (Infliximab, Remidace<sup>®</sup>; Etanercept, Enbrel<sup>®</sup>; Adalimumab, Humira<sup>®</sup>), los inhibidores de otras citoquinas implicadas en inflamación (IL6: Tocilizumab, Actemra<sup>®</sup>; IL1: Anakinra, Kineret<sup>®</sup>; CD20: Rituximab o RTX, Mabthera<sup>®</sup>) y los moduladores de leucocitos (Abatacept, Orencia<sup>®</sup>). Estos tratamientos bloquean los efectos inflamatorios de las moléculas inflamatorias diana.

La cuantificación de la actividad de la enfermedad AR se realiza según diferentes criterios. La puntuación DAS (*Disease Activity Score*) es el criterio más utilizado en la práctica clínica e incluye el número de articulaciones dolorosas, el número de articulaciones tumefactas, la velocidad de sedimentación globular, y la evolución de la actividad global de la enfermedad por el paciente en una escala analógica visual (EVA) de 10 (donde “0”

correspondería con muy bien y “10” como muy mal). Sin embargo, debido a este último factor de índice de dolor, esta medición presenta cierto grado de subjetividad inherente a cada paciente. El grado de desarrollo de la enfermedad AR también se puede cuantificar mediante modalidades de imagen estructural (por resonancia magnética, tomografía de rayos X y perfusión, ecografías de contraste...) y de imagen funcional (imagen óptica, nuclear o PET, ecografías de angiogénesis en las articulaciones dañadas).

#### **1.4.1. Relación de AR con el MHC de clase III**

En diversos estudios realizados con pacientes de AR se ha observado como la AR es un ejemplo de enfermedad cuyo mapeo está definido en el MHC (Bali et al. 1999), presentando una fuerte asociación con alelos particulares de la región de clase II, aunque también se han descrito polimorfismos en la región de clase III que podrían jugar un papel importante en la susceptibilidad a dicha enfermedad (Jenkins et al. 2000; Kilding et al. 2004; Kilding and Wilson 2005; Mulcahy et al. 1996). Existen tres asociaciones directas observadas dentro de la región del MHC de clase III en función de su localización génica. Nos encontramos con la región más telomérica del MHC III (*TNF $\alpha$*  y *NF $\kappa$ BIL1*), una región central (cluster de genes *Ly6*) y una región situada hacia el centrómero próximo al MHC de clase II (región del gen *RAGE*).

##### **1.4.1.1. Relación de *TNF $\alpha$* y *NF $\kappa$ BIL1* con AR**

Empleando los estudios de mapeo por polimorfismo anteriores, se estudiaron los genes concretos observándose la asociación de hasta doce genes (*PADI4*, *FCGR3*, *TNFRSF1B*, *ITGAV*, *BTLA*, *SLC22A4*, *IL3*, *VEGF*, *TNF*, *NFKBIL1*, *TRAF1-C5*, and *MIF*) en el desarrollo de la AR, encontrándose entre los genes candidatos los genes *TNF $\alpha$*  y *NF $\kappa$ BIL1* (Huang et al. 2009). Estudios similares destacaron la implicación de *I $\kappa$ BL* (inhibidor de la ruta de NF $\kappa$ B) (Okamoto et al. 2003) e *NF $\kappa$ BIL1* (Greetham et al. 2007) como genes candidatos a influir en la susceptibilidad de padecer AR.

##### **1.4.1.2. Relación cluster de genes *Ly6* con AR**

El cluster Ly-6 está localizado en un segmento de DNA de unos 80 Kb entre *TNF $\alpha$*  y *HSP70*. Esta región génica concreta se encuentra asociada con la susceptibilidad a padecer la enfermedad AR (Newton et al. 2004; Zanelli et al. 2001), así como se encuentra

también asociada a otras enfermedades como es el caso de la diabetes tipo I (Schroeder et al. 1998) y la deficiencia de IgA (Schroeder et al. 1998).

#### ***1.4.1.3. Relación de RAGE con AR***

Estudios publicados han examinado los marcadores de estrés oxidativo y los niveles de esRAGE en plasma en 54 pacientes con AR y 20 individuos controles sanos, observándose unos niveles significativamente superiores de estrés oxidativo y unos niveles menores de esRAGE en los pacientes de AR con respecto a los controles (Daffu et al. 2013; Ferrante et al. 2010). Interesantemente, la correlación directa entre los marcadores de estrés oxidativo se observó únicamente en los pacientes que no estaban siendo tratados con los tratamientos biológicos inhibidores de TNF $\alpha$ . Por el contrario, los pacientes tratados con los diferentes tipos de inhibidores de TNF $\alpha$  mostraban unos niveles significativamente reducidos de los marcadores de estrés oxidativo en relación con los individuos controles (Ferrante et al. 2010).

## **2. OBJETIVOS**

Para el desarrollo de esta tesis doctoral se plantearon los objetivos que se detallan a continuación:

- 1.- Caracterización experimental de las variantes de *splicing* del gen *RAGE* en diferentes especies de mamíferos y tejidos.** Identificación y comparación a nivel de tejido de los patrones de *splicing* en distintas especies mediante RT-PCR.
- 2.- Estudio detallado de las variantes de *splicing* del gen *RAGE* en datos de RNA-seq.** Análisis de las uniones de exones y el ensamblaje de mensajeros para identificar variantes de *splicing* en las especies y tejidos disponibles en las bases de datos.
- 3.- Caracterización a nivel proteico de las variantes de *splicing* seleccionadas del gen *RAGE* en humano.** Estudios funcionales de las variantes Tv1 y Tv2, y su interacción con el ligando S00A12.
- 4.- Caracterización experimental de las quimeras existentes entre los genes *G6F* y *LY6G6D* en humano y otros mamíferos.** Identificación y comparación de las quimeras G6F-Ly6G6D a nivel de tejido en diferentes especies.
- 5.- Caracterización a nivel proteico de las principales quimeras existentes entre los genes *G6F* y *LY6G6D* en humano.** Estudios comparativos de las modificaciones post-transcripcionales de las diferentes quimeras respecto a las isoformas canónicas de ambos genes.
- 6.- Artritis Reumatoide y su relación con el MHC de clase III.** Estudio de la expresión diferencial de *RAGE*, *G6F* y *LY6G6D*, así como de los otros 60 genes del MHC de clase III, en pacientes con AR.

### **3. MATERIALES Y MÉTODOS**

### 3.1. ANÁLISIS BIOINFORMÁTICO

#### 3.1.1. Bases de datos

El estudio de los genes detallados en este trabajo se basó en la información a nivel de DNA genómico, RNA mensajero y proteínas disponible en las principales bases de datos públicas, tales como el *National Center for Biotechnology Information* (NCBI; <http://www.ncbi.nlm.nih.gov/>), Ensembl (<http://www.ensembl.org/>), *The European Bioinformatics Institute* (EBI; <http://www.ebi.ac.uk/>), Universidad de California Santa Cruz (UCSC) *Genome Browser* (<http://genome.ucsc.edu/>), GeneCards (<http://www.genecards.org/>) y *HUGO Gene Nomenclature Committee* (HGNC; <http://www.genenames.org/>). Se estudiaron tanto las secuencias conservadas, como estudios de ESTs y predicciones bioinformáticas, identificando todas las posibles variantes de *splicing* presentes en cada especie.

#### 3.1.2. Diseño y chequeo de cebadores

Los cebadores utilizados en e trabajo se diseñaron mediante el programa *Primer3Plus* ([www.bioinformatics.nl/primer3plus/](http://www.bioinformatics.nl/primer3plus/)), que se encuentra públicamente disponible online. Se diseñaron los cebadores teniendo en cuenta una temperatura de *melting* de aproximadamente 60°C, un tamaño de 20 pb (18-25) y un porcentaje de GC de al menos 40%. Se diseñaron dos tipos diferentes de cebadores dependiendo de su aplicación. Cuando los cebadores eran diseñados para clonajes, se empleó la herramienta “*Cloning*” del programa *Primer3Plus* añadiéndose las dianas de restricción correspondientes y unas grapas en el extremo 5´ de ambos cebadores para permitir el corte por las enzimas de restricción. Para el análisis previo requerido de secuencias diana de enzimas de restricción se empleó la aplicación gratuita de NEBcutter V2.0 (<http://tools.neb.com/NEBcutter2/>). Cuando los cebadores eran diseñados para el análisis de las variantes de *splicing* se empleó la herramienta “*Detection*”. En el caso de los cebadores empleados en el análisis de las variantes en primates, se realizó inicialmente un estudio de conservación de las secuencias disponibles en las bases de datos. Dichas secuencias se alinearon mediante los programas MultAlin (<http://multalin.toulouse.inra.fr/multalin/>) y ClustalW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>), diseñando las parejas de cebadores en las regiones conservadas.

Por otro lado, cuando los cebadores eran diseñados para qRT-PCR también se empleó la herramienta “*Detection*”. En este caso, la mayoría de los amplicones fueron diseñados de un tamaño inferior a 150 pb y abarcaban uno o más intrones intermedios con un tamaño mínimo de 500 pb (*intron-spanning primers*), o incluso diseñados en las uniones de exones, de tal forma que se evitara la amplificación inespecífica del DNA genómico. Se escogieron aquellas parejas de cebadores que formasen menos estructuras secundarias entre ellos y consigo mismos. Esto se comprobó tanto por el programa *Primer3Plus* con la herramienta “*Primer\_Check*”, como por el programa *Oligo Calc* de Biotools ([www.basic.northwestern.edu/biotools/](http://www.basic.northwestern.edu/biotools/)). Además, mediante el programa *Primer-BLAST* del NCBI (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) se comprobaron sus especificidades, de forma que ningún cebador diseñado reconociera específicamente ninguna otra secuencia en al menos las últimas cinco bases del extremo 3'. Algunas parejas de cebadores que cumplían con las condiciones requeridas diseñados para las amplificaciones de ciertos normalizadores fueron obtenidos de otros estudios, tales como: *UBC*, *YWHAZ* y *B2M* (Vandesompele et al. 2002) y *ATP5B*, *HSP90AB1* y *FLOT2* (Andersen et al. 2004), así como parejas de cebadores de genes involucrados en AR, tales como: *HLA-DRB1* (Jawaheer and Gregersen 2002), *PTPN22* (Kallberg et al. 2007), *CIITA* (Swanberg et al. 2005), *IRF5* (Sigurdsson et al. 2007). Una vez diseñados, chequeados y seleccionados los cebadores finales se ordenó su síntesis a la compañía Sigma.

### 3.1.3. Alineamiento de secuencias

Las secuencias de las diferentes variantes de *splicing* obtenidas experimentalmente mediante secuenciación Sanger fueron procesadas por el programa BioEdit (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>) y alineadas contra todas las secuencias disponibles en las bases de datos usando el programa *Basic Local Alignment Search Tool* (BLAST del NCBI; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>), o contra secuencias conocidas de interés mediante *EBI Pairwise Alignment* (<http://www.ebi.ac.uk/Tools/psa/>). Se empleó también la herramienta Spidey del NCBI (<http://www.ncbi.nlm.nih.gov/spidey/>), para alinear las secuencias de mensajeros obtenidas contra el DNA genómico de referencia, obteniendo la composición exónica de cada isoforma y poder compararla con la canónica. Alineamientos múltiples entre todas las variantes de *splicing* con variantes conocidas se



realizaron usando los programas MultAlin (<http://multalin.toulouse.inra.fr/multalin/>) y ClustalW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>).

#### 3.1.4. Estudio a nivel proteico de las variantes de *splicing*

La traducción de cada variante de *splicing* a proteína fue realizada usando la herramienta ORF Finder del NCBI (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) y el cálculo de la masa molecular (MW) teórica de los péptidos resultantes fue realizada a través de la aplicación de Biopeptide Co. (<http://www.biopeptide.com/PepCalc/>). Para el estudio de los dominios proteicos de estas isoformas se emplearon diversas herramientas de predicción como Pfam (<http://pfam.sanger.ac.uk/>), SMART (<http://smart.embl-heidelberg.de/index2.cgi>) y Prosite (<http://prosite.expasy.org/>). Además, se realizó una búsqueda de los péptidos señales que potencialmente tendrían las isoformas a estudio mediante la herramienta SignalP 4.1 Server (<http://www.cbs.dtu.dk/services/SignalP/>), así como se realizó una predicción de las potenciales regiones transmembrana mediante la herramienta TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>) y su orientación en TMpred ([http://www.ch.embnet.org/software/TMPRED\\_form.html](http://www.ch.embnet.org/software/TMPRED_form.html)). La predicción de los sitios de anclaje por GPI fueron estudiados mediante la herramienta *GPI Modification Site Prediction* ([http://mendel.imp.ac.at/gpi/gpi\\_server.html](http://mendel.imp.ac.at/gpi/gpi_server.html)). Los sitios de glicosilación fueron estudiados con la herramienta *GPP Prediction Server* (<http://comp.chem.nottingham.ac.uk/cgi-bin/glyco/bin/getparams.cgi>) y diferentes herramientas (NetNGlyc y NetOGlyc, entre otras) presentes en *CBS Prediction Servers* (<http://www.cbs.dtu.dk/services>).

Aquellas variantes de *splicing* con menos de un dominio proteico, y uno o más PTC fueron considerados como variantes no codificantes. Las variantes descritas en este trabajo fueron nombradas como *SpRAGE\_VN*, donde Sp corresponde a la especie abreviada como Gsss (primera letra del género y las tres primeras letras de la especie, como por ejemplo: *Hsap* como *Homo sapiens*) y VN es el número de la variante de mRNA (que se encuentran ordenados por tamaño de las variantes, de mayor a menor tamaño).

### 3.1.5. RNA-seq

El estudio de las variantes de *splicing* de los genes implicados en este trabajo también se abordó desde la tecnología RNA-seq, para lo que se utilizaron datos públicamente accesibles volcados por otros investigadores. En concreto, se emplearon datos de RNA-seq de los tejidos cerebro e hígado en diversas especies obtenidos de la base de datos NCBI *Sequence Read Archive* (SRA). Se obtuvieron lecturas simples de SRP007412 (Barbosa-Morais et al. 2012), y lecturas pareadas de SRP016501 (Merkin et al. 2012). Además, se obtuvieron lecturas pareadas de una muestra de Monocitos-CD14+ (SRX179434) procedentes del proyecto *ENCyclopedia of DNA Elements* (ENCODE). Se empleó el comando `fastq-dump` del paquete SRA para la conversión del formato de estas secuencias al convencional `fastq`, y se comprobó la calidad de las lecturas mediante el programa `FastQC` ([bioinformatics.babraham.ac.uk/projects/fastqc/](http://bioinformatics.babraham.ac.uk/projects/fastqc/)). Las lecturas fueron alineadas contra el genoma de referencia de cada especie obtenido de la base de datos del Ensembl (*Homo sapiens*: GRCh37, *Macaca mulatta*: MMUL\_1, *Bos taurus*: UMD3.1, *Mus musculus*: GRCm38, *Rattus norvegicus*: Rnor\_5.0), empleando los parámetros por defecto de `TopHat` y `Bowtie` (Langmead et al. 2009; Trapnell et al. 2009). Los ficheros de salida con las uniones de exones o *junctions* (ficheros BED, *TopHat outfile*) fueron normalizados por millón de lecturas alineadas para comparar diferentes eventos de *splicing* entre el total de las muestras a estudio. Se empleó el *software Samtools* (H. Li et al. 2009) para indexar los ficheros BAM (*TopHat outfile*), convertir los formatos de los alineamientos y seleccionar una región génica de interés. Las lecturas alineadas de *RAGE* fueron usadas para ensamblar transcritos “*de novo*” usando el programa `Cufflinks` (Trapnell et al. 2010) mediante los parámetros por defecto (-F 0.1 y -j 0.15) o con modificaciones de ellos, empleando los parámetros -F 0 y -j 0.5. Adicionalmente, se suministraron al programa `Cufflinks` un fichero GTF con las coordenadas anotadas manualmente de todos los transcritos descritos experimentalmente en este estudio, y sus abundancias relativas fueron estimadas en base a los datos de RNA-seq. La expresión de cada transcrito fue medida en *Fragments Per Kilobase of exon per Million fragments mapped* (FPKM). El navegador IGV (<http://www.broadinstitute.org/igv/>) se empleó para la visualización de las lecturas alineadas (Robinson et al. 2011).

### 3.1.6. Análisis de los resultados de qPCR

Los valores resultantes de la cuantificación de la qPCR se obtuvieron de los programas SDS 2.2.1 (Applied Biosystems) y *Fluidigm Gene Expression Software* (Fluidigm) dependiendo de la procedencia de los datos, qPCR convencional o *BioMark™ HD System*, respectivamente. El parámetro Ct fue establecido de forma automática para cada serie de datos. El análisis de los resultados se realizó empleando el software GenEx® y/o una hoja de cálculo de Excel. Se analizaron los valores de expresión absoluta generando una recta patrón con los Ct de amplificación del plásmido linearizado de cantidad previamente conocidas y extrapolando el resto de los datos a dicha recta. Para el estudio de expresión relativa, se utilizaron los algoritmos geNorm y NormFinder (Andersen et al. 2004) del software GenEx® para detectar la pareja de normalizadores óptimos para nuestro estudio. Una vez seleccionados los dos normalizadores, las muestras fueron calibradas en referencia a la muestra control C1 a la que se asignó el valor arbitrario 1, y a los valores de los dos genes normalizadores (*CFL1* y *TMBIM6*). Para el estudio estadístico y la representación gráfica de los datos se empleó el programa GraphPad Prism5®, aplicándose el test no paramétrico Kruskal-Wallis test que no asume distribuciones normales de las muestras y el test Dunnett como procedimiento de comparación múltiple. Los p-valores menores a 0,01 fueron tomados como significación estadística.

## 3.2. MUESTRAS DE PARTIDA Y PROCESAMIENTO INICIAL

### 3.2.1. Muestras de sangre de pacientes AR y controles

Las muestras de sangre de controles sanos fueron proporcionadas por el Banco de Donantes de sangre del Hospital Universitario la Paz (Madrid) gracias a la colaboración con la Dra. Aurora Viejo. Las muestras de sangre de pacientes con AR fueron suministradas gracias a una colaboración con el Hospital Universitario Ramón y Cajal (Madrid) por el Dr. Antonio Zea Mendoza, la Dra. Mónica Vázquez Díaz, la residente María Ángeles Blázquez y el enfermero Eduardo Fernández Ulloa. Se obtuvo el consentimiento informado de todos los pacientes y los proyectos fueron evaluados por los correspondientes Comités Éticos. Las muestras fueron suministradas con identificación numérica para asegurar el anonimato de los pacientes. Cada muestra proporcionada disponía de un volumen de 5 ml de sangre en tubos cerrados conteniendo ácido etilendiaminotetraacético (EDTA) como anticoagulante y se mantuvieron a 4°C durante 24 horas. Esta espera era debida a que en las muestras de

donantes sanos los resultados de los análisis de rutina no se reciben hasta 24 horas después de la extracción. Por consistencia en el procesamiento de las muestras, las muestras de pacientes con AR también se mantenían 24 horas a 4°C. Los análisis rutinarios se realizaban en el centro de transfusión de la Comunidad Autónoma de Madrid para el estudio de posibles infecciones de transmisión sexual (ITS), tales como: el virus de la Hepatitis B (VHB), Hepatitis C (VHC) y de la Inmunodeficiencia Humana (VIH). También se realizó la prueba de Reagina Plasmática Rápida (RPR) para detectar *Treponema pallidum*.

Las muestras de sangre obtenidas de controles y de dos tipos de pacientes se detallan en las Tablas 3-5, mostrándose en todas las muestras la edad y sexo de las personas donantes. Además, se indica en las muestras provenientes de los pacientes las características de su enfermedad AR (FR o factor reumatoideo, y anti-CCP o anticuerpos antipéptidos cíclicos citrulinados), así como su tratamiento con FAMEs (Metrotrexato o Mtx, Leflunomida o Lef, y Doliquine) solos o en combinación con glucocorticoides (Prednisona o Pred, y Zamene). Ninguno de estos pacientes fue tratado con tratamientos biológicos o fármacos modificadores de la respuesta biológica. Se consideraron pacientes con la enfermedad AR en remisión aquellos que presentaban una o ninguna articulación tumefaciente y, como máximo dos articulaciones dolorosas, considerándose pacientes con la enfermedad activa cuando estos indicadores se superaban. El estudio contó con quince muestras de individuos controles (Tabla 3) y 31 muestras de pacientes con AR, de los cuáles 16 pacientes presentaban la enfermedad en remisión (Tabla 4) y 15 la enfermedad activa (Tabla 5).

Nombre muestra	Número de identificación	Sexo	Fecha nacimiento	Edad a fecha de extracción
C1	E0016.10.305935V	Mujer	21/11/60	50 años
C2	E0016.11.303987M	Mujer	20/04/49	61 años
C3	E0016.11.305404V	Mujer	11/09/57	52 años
C4	E0016.11.305415P	Mujer	20/11/52	58 años
C5	E0016.11.305782G	Mujer	11/01/60	51 años
C6	E0016.11.305784C	Mujer	17/07/58	53 años
C7	E0016.11.309597S	Mujer	03/08/49	62 años
C8	E0016.11.309669S	Mujer	14/11/56	55 años
C9	E0016.11.309871O	Mujer	13/09/52	59 años
C10	E0016.11.309876E	Mujer	08/08/57	54 años
C11	E0016.11.309883G	Mujer	29/05/60	51 años
C12	E0016.11.309909S	Mujer	01/12/49	62 años
C13	E0016.11.309926Q	Mujer	20/09/59	52 años
C14	E0016.11.309934Q	Mujer	08/10/60	51 años
C15	E0016.11.309946I	Mujer	02/03/60	51 años

**Tabla 3. Características de los controles incluidos en el estudio.**

Nombre muestra	Número de identificación	Sexo	Fecha nacimiento	Edad a fecha de extracción	FR	Anti-CCP	FAMEs	Tratamiento combinado con glucocorticoides
R1	20003	Mujer	15/05/67	43 años	+	+	Mtx	No
R2	20007	Mujer	16/01/65	45 años	+	+	Mtx	No
R3	20008	Mujer	16/08/31	79 años	+	+(50)	Mtx	No
R4	20010	Mujer	15/09/51	59 años	++	+	Mtx	No
R5	20012	Mujer	17/08/39	71 años	+	+	Mtx	No
R6	20016	Mujer	30/06/45	65 años	+	+	Mtx	Pred. 1 mg/dia
R7	20030	Mujer	20/08/45	66 años	+	+	Mtx	No
R8	20036	Mujer	08/12/43	68 años	+	+	Lef	No
R9	20038	Mujer	27/02/40	72 años	+	+	Mtx	Si
R10	C-1	Mujer	09/01/46	66 años	+	+	Mtx	No
R11	20002	Mujer	12/07/46	64 años	+	+	Lef y Mtx	No
R12	20055	Mujer	09/05/32	80 años	+	+	Mtx	Pred.
R13	20057	Mujer	06/01/68	44 años	+	+	Mtx	No
R14	20058	Mujer	05/06/54	58 años	+	+	Mtx	No
R15	20059	Mujer	12/10/24	87 años	+	+	Mtx	Si
R16	20051	Mujer	24/02/74	38 años	+	+	Mtx	No

**Tabla 4. Características de los pacientes con la enfermedad en remisión (R) incluidos en el estudio.** Se indican los tratamientos con FAMEs (Metrotrexato o Mtx, Leflunomida o Lef), así como con glucocorticoides (Prednisona o Pred).

Nombre muestra	Número de identificación	Sexo	Fecha nacimiento	Edad a fecha de extracción	FR	Anti-CCP	FAMEs	Tratamiento combinado con glucocorticoides
A1	20004	Mujer	09/05/59	51 años	+	+	Mtx	Pred 2.5
A2	20013	Mujer	04/02/37	73 años	+	+	Mtx	No
A3	20018	Mujer	03/04/52	59 años	+	+	Mtx	Pred.
A4	20020	Mujer	1953	58 años	+	+	Mtx	Pred.
A5	20022	Mujer	1958	53 años	+	+	Mtx	Pred. 10
A6	20024	Mujer	1942	69 años	+	+	Mtx	Pred.
A7	20032	Mujer	01/07/73	38 años	+	+	Mtx y Lef	Si
A8	20037	Mujer	18/08/48	63 años	+	+	Mtx	No
A9	20039	Mujer	22/05/50	61 años	+	+	No	Si
A10	20042	Mujer	27/05/71	40 años	+	+	Mtx	No
A11	20045	Mujer	31/01/65	47 años	+	+	No	Si, Zamene
A12	20046	Mujer	20/09/44	67 años	+	+	Mtx	Si
A13	20047	Mujer	22/12/32	79 años	+	+	Doliquine	Pred.
A14	20048	Mujer	29/05/53	58 años	+	+	Mtx	Zamene
A15	20056	Mujer	17/07/47	64 años	+	+	Mtx	No

**Tabla 5. Características de los pacientes con la enfermedad activa (A) incluidos en el estudio.** Se indican los tratamientos con FAMEs (Metrotrexato o Mtx, Leflunomida o Lef, y Doliquine), así como con glucocorticoides (Prednisona o Pred, y Zamene).

### 3.2.2. Muestras de diferentes especies de mamíferos

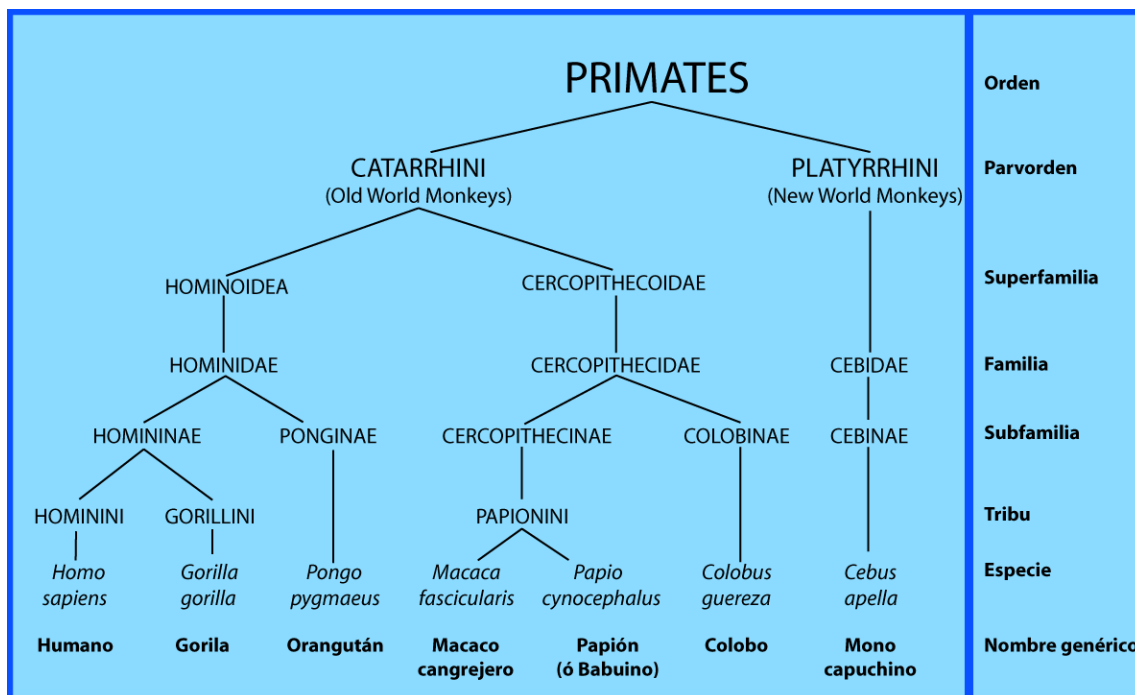
Se utilizó RNA total de muestras comerciales (Biochain) de diferentes tejidos y varias especies: *Homo sapiens*, *Macaca mulatta*, *Sus scrofa*, *Bos taurus*, *Mus musculus* y *Rattus norvegicus* (Tabla 6) para el análisis de las formas de *splicing* de *RAGE*, *G6F*, *Ly6G6D* y las quimeras entre los dos últimos. La calidad de estos ARNs comerciales fue testada en el Bionalizador (Agilent) corroborándose que su integridad era óptima para los estudios que se iban a realizar posteriormente.

## MATERIAL Y MÉTODOS

Especies	Nombre del tejido	Tipo de tejido	Información del donante			Catálogo No.
			Sexo	Edad	Conjunto	
<i>Homo sapiens</i>	Sangre	Adulto Normal	Masculino	50 años	1 donante	R1234148-50
<i>Homo sapiens</i>	Cerebro	Adulto Normal	Masculino	30 años	1 donante	R1234035-50
<i>Homo sapiens</i>	Riñón	Adulto Normal	Masculino	65 años	1 donante	R1234142-50
<i>Homo sapiens</i>	Hígado	Adulto Normal	Masculino	64 años	1 donante	R1234149-50
<i>Homo sapiens</i>	Pulmón	Adulto Normal	Masculino	24 años	1 donante	R1234152-50
<i>Homo sapiens</i>	Pancreas	Adulto Normal	Masculino	44 años	1 donante	R1234188-50
<i>Homo sapiens</i>	Bazo	Adulto Normal	Masculino	29 años	1 donante	R1234246-50
<i>Homo sapiens</i>	Timo	Adulto Normal	Masculino	21 años	1 donante	R1234264-50
<i>Homo sapiens</i>	Riñón	Adulto Tumoral	Femenino	28 años	1 donante	R1235142-50
<i>Homo sapiens</i>	Hígado	Adulto Tumoral	Masculino	65 años	1 donante	R1235149-50
<i>Homo sapiens</i>	Pulmón	Adulto Tumoral	Femenino	39 años	1 donante	R1235152-50
<i>Homo sapiens</i>	Cerebro	Fetal	Femenino	28 semanas	1 donante	R1244035-50
<i>Homo sapiens</i>	Corazón	Fetal	Femenino	33 semanas	1 donante	R1244122-50
<i>Homo sapiens</i>	Riñón	Fetal	Femenino	30 semanas	1 donante	R1244142-50
<i>Homo sapiens</i>	Hígado	Fetal	Masculino	38 semanas	1 donante	R1244149-50
<i>Homo sapiens</i>	Pulmón	Fetal	Femenino	37 semanas	1 donante	R1244152-50
<i>Homo sapiens</i>	Bazo	Fetal	Masculino	40 semanas	1 donante	R1244246-50
<i>Homo sapiens</i>	Timo	Fetal	Masculino	24 semanas	1 donante	R1244264-50
<i>Macaca mulatta</i>	Cerebro	Normal	Masculino	Desconocido	1 donante	R1534035-50
<i>Macaca mulatta</i>	Corazón	Normal	Femenino	4,5 años	1 donante	R1534122-50
<i>Macaca mulatta</i>	Riñón	Normal	Femenino	4,5 años	1 donante	R1534142-50
<i>Macaca mulatta</i>	Hígado	Normal	Masculino	4,5 años	1 donante	R1534149-50
<i>Macaca mulatta</i>	Pulmón	Normal	Masculino	4,5 años	1 donante	S1534152
<i>Macaca mulatta</i>	Pancreas	Normal	Masculino	2,5 años	1 donante	R1534188-50
<i>Bos taurus</i>	Cerebro	Normal	Femenino	2 años	1 donante	R1B34035
<i>Bos taurus</i>	Corazón	Normal	Femenino	2 años	1 donante	R1B34122
<i>Bos taurus</i>	Riñón	Normal	Femenino	2 años	1 donante	R1B34142
<i>Bos taurus</i>	Hígado	Normal	Femenino	2 años	1 donante	R1B34149
<i>Bos taurus</i>	Pulmón	Normal	Femenino	2 años	1 donante	R1B34152
<i>Bos taurus</i>	Bazo	Normal	Femenino	2 años	1 donante	R1B34246
<i>Sus scrofa</i>	Cerebro	Normal	Femenino	7,5 años	1 donante	R1934035
<i>Sus scrofa</i>	Corazón	Normal	Femenino	7,5 años	1 donante	R1934122
<i>Sus scrofa</i>	Riñón	Normal	Femenino	7,5 años	1 donante	R1934142
<i>Sus scrofa</i>	Hígado	Normal	Femenino	7,5 años	1 donante	R1934149
<i>Sus scrofa</i>	Pulmón	Normal	Femenino	7,5 años	1 donante	R11934152
<i>Sus scrofa</i>	Bazo	Normal	Femenino	7,5 años	1 donante	R1934246
<i>Mus musculus</i>	11 días	Embrionario	Desconocido	11 días	5 donantes	R1334XI-50
<i>Mus musculus</i>	15 días	Embrionario	Desconocido	15 días	2 donantes	R1334XV-50
<i>Mus musculus</i>	17 días	Embrionario	Desconocido	17 días	2 donantes	R1334XVII-50
<i>Mus musculus</i>	Cerebro	Adulto Normal	Masculino	12 semanas	1 donante	R1334035-50
<i>Mus musculus</i>	Corazón	Adulto Normal	Desconocido	Desconocido	15 donantes	R1334122-50
<i>Mus musculus</i>	Riñón	Adulto Normal	Femenino	3-4 semanas	20 donantes	R1334142-50
<i>Mus musculus</i>	Hígado	Adulto Normal	Femenino	8 semanas	1 donante	R1334149-50
<i>Mus musculus</i>	Pulmón	Adulto Normal	Femenino	3-4 semanas	20 donantes	R1334152-50
<i>Mus musculus</i>	Pancreas	Adulto Normal	Femenino	8,5 semanas	1 donante	R1334188-50
<i>Rattus norvegicus</i>	19 días	Embrionario	Desconocido	19 días	2 donantes	R1434000-50
<i>Rattus norvegicus</i>	Cerebro	Adulto Normal	Masculino	10 semanas	4 donantes	R1434035-50
<i>Rattus norvegicus</i>	Corazón	Adulto Normal	Femenino	8-11 semanas	7 donantes	R1434122-50
<i>Rattus norvegicus</i>	Riñón	Adulto Normal	Masculino/ Femenino	12 / 8-9 semanas	3 / 3 donantes	R1434142-50
<i>Rattus norvegicus</i>	Hígado	Adulto Normal	Femenino	13,2 semanas	1 donante	R1434149-50
<i>Rattus norvegicus</i>	Pulmón	Adulto Normal	Masculino	24 semanas	3 donantes	R1434152-50
<i>Rattus norvegicus</i>	Pancreas	Adulto Normal	Femenino	9 semanas	1 donante	R1434188-50
<i>Rattus norvegicus</i>	Bazo	Adulto Normal	Femenino	13 semanas	10 donantes	R1434246-50

**Tabla 6. Información de las muestras de RNA analizadas.**

Además, se obtuvieron muestras de sangre de diferentes especies de primates gracias a una colaboración con el Zoo Aquarium de Madrid mediante su director Jesús Fernández Morán y la veterinaria Eva Martínez Nevado, quienes nos suministraron muestras de sangre de diferentes especies de primates: *Gorilla gorilla*, *Pongo pygmaeus*, *Papio sp.*, *Colobus guereza* y *Cebus apella* (Figura 21).



**Figura 21. Muestras de primates provenientes del Zoo Aquarium de Madrid organizadas según su clasificación científica.**

Asimismo, se obtuvieron muestras de sangre de dos tipos de macacos diferentes (*Macaca mulatta* y *Macaca fascicularis*) gracias a una colaboración con Carmen Fernández Criado, directora del Gabinete Veterinario de la Universidad Autónoma de Madrid (UAM) y responsable del Servicio de Animalario del Instituto de Investigaciones Biomédicas (IIB) de Madrid. Además, muestras de sangre de *Mus musculus* y *Rattus norvegicus* fueron suministrados por Javier Palacín Urquijo, responsable del Servicio de Animalario del Centro de Biología Molecular Severo Ochoa (CBMSO) de Madrid, así como una muestra de sangre de *Sus scrofa* obtenida en colaboración con el Servicio de Animalario del Centro Nacional de Investigaciones Cardiovasculares (CNIC) de Madrid. Las muestras consistían en 2-5 ml de sangre en tubos conteniendo EDTA que mantuvimos a 4 °C durante 24 horas siguiendo el mismo procedimiento que con las muestras humanas de sangre de controles.

### 3.2.3. Procesamiento muestras de sangre de las diferentes especies y separación de las diferentes fracciones de interés

Con el fin de obtener una separación óptima de las células mononucleares de las muestras de sangre total se realizó un gradiente de centrifugación con Ficoll-Paque (GE Healthcare). Para ello se emplearon tubos Leucosep<sup>®</sup> que se rellenaron con 3 ml de Ficoll-Paque y se centrifugaron durante un minuto a 1000 x g. Posteriormente, se añadieron entre 2-5 ml de sangre y se centrifugaron durante 15 minutos a 1000 x g, quitando el freno a la centrífuga. Durante la centrifugación se van formando diferentes capas, de abajo a arriba nos encontraríamos el sedimento compuesto principalmente por granulocitos y eritrocitos, por encima estaría la capa de Ficoll-paque, e inmediatamente encima se encontraría la fracción celular enriquecida en células mononucleares de sangre periférica (*Peripheral Blood Mononuclear Cell*, PBMC) consistente en monocitos, linfocitos y células NK; finalmente sobre esta capa se encontraría el plasma y las plaquetas. Se extrajeron las capas de células mononucleares y el plasma de cada una de las muestras de sangre, así como 500 µl de la muestra de sangre total inicial para su futura extracción de DNA genómico (gDNA). La fracción de células PBMCs se extrajo con cuidado y se centrifugo durante 5 minutos a 5000 x g. Se realizaron dos lavados con *Phosphate-Buffered Saline* (PBS) 1X y se centrifugó de nuevo con el fin de eliminar toda contaminación de Ficoll-Paque en la muestra. A partir del pellet se procedió seguidamente a la extracción de RNA.

## 3.3. TÉCNICAS DE LABORATORIO DE BIOLOGÍA MOLECULAR

### 3.3.1. Técnicas de RNA

#### 3.3.1.1. Extracción, precipitación y determinación de la calidad del RNA

La extracción de RNA de las diferentes líneas celulares anteriormente citadas, así como de las células mononucleares procedentes de las muestras de sangre suministradas, se realizó empleando SV Total RNA Isolation System (Promega), según las indicaciones del fabricante. Este sistema de extracción de ARN emplea una combinación de tiocianato de guanidina (GTC, agente caotrópico) y β-mercaptoetanol (BME, agente reductor de puentes disulfuro) como agentes desnaturalizantes y degradantes de proteínas tales como RNasas y DNasas. Asimismo, se combinan las propiedades del GTC junto con dodecilsulfato sódico (SDS), de forma que se rompen los complejos nucleoprotéicos y se permite la liberación del RNA en solución pudiendo ser aislado sin proteínas. Posteriormente, el protocolo incluye



un tratamiento con la enzima DNasa I para digerir el gDNA contaminante y se eluye en 100  $\mu$ l, según las indicaciones del fabricante. La concentración del RNA se midió por espectrofotometría en un NanoDrop ND-1000 (Thermo Scientific) del Servicio de Genómica y Secuenciación Masiva del CBMSO.

Aquellas muestras que presentasen una concentración inferior a 100 ng/ $\mu$ l se precipitaron empleando 3  $\mu$ l de glicógeno (Ambion<sup>®</sup>) a una concentración de 5 mg/ml, acetato sódico 3M pH5 (1/10 del volumen total) y etanol absoluto. Las muestras se precipitaron a -70°C durante toda la noche. A continuación, las muestras se centrifugaron a 15.000 x g durante 15 minutos a 4°C, se eliminó cuidadosamente el sobrenadante y se resuspendió el pellet seco en el volumen deseado de agua libre de nucleasas. Posteriormente se volvió a medir su concentración por espectrofotometría y se evaluó la calidad del RNA mediante electroforesis usando el kit RNA 6000 Nano Chip en el Bioanalizador Agilent 2100 (Agilent Technologies) en el Servicio de Genómica y Secuenciación Masiva del CBMSO. Analizando los electroferogramas correspondientes se puede visualizar la integridad del RNA mediante el *RNA Integrity Number* (RIN) que corresponde con la relación 28S/18S de cada muestra, así como también tiene en cuenta la presencia de DNA genómico contaminante o RNA degradado. Se consideraron viables para el estudio únicamente aquellas muestras que presentaban un valor de RIN superior a 7, en una escala de 1 a 10 en la cual el valor 1 corresponde a una muestra de RNA totalmente degradada y un valor de 10 a una muestra intacta.

### **3.3.1.2. Síntesis de DNA complementario (cDNA)**

Las muestras de RNA se retrotranscribieron a cDNA mediante el kit *Improm-II<sup>TM</sup> Reverse Transcription System* (Promega), empleando Oligo (dT)<sub>15</sub> y siguiendo las indicaciones del fabricante. El programa consistía en un paso de desnaturalización del RNA y del oligo (dT)<sub>15</sub> de 5 minutos a 70°C y después inmediatamente en hielo. Posteriormente, se añade la mezcla de la retrotranscriptasa compuesta por tampón de reacción 1X, dNTPs a una concentración final de 0.5 mM, MgCl<sub>2</sub> 3 mM y 1U de retrotranscriptasa. Ambas mezclas se incuban 5 minutos a 25°C para permitir el anillamiento del cebador al RNA, a continuación se incuban 1 hora y media a 42°C para que se produzca la síntesis y, por último, se desnaturaliza la muestra a 70°C durante 5 minutos para inactivar el enzima. Dependiendo del experimento las retrotranscripciones (RT) se realizaron en un total de 20

µl normalmente con 1 µg de RNA total para las RT-PCR, o únicamente con 250 ng de RNA total en las muestras preparadas para la PCR cuantitativa con Biomark™. Para las muestras que se emplearían en PCR cuantitativa se realizó un control negativo (nombrada como RT-) que consistía en la misma mezcla de reacción que la retrotranscripción normal pero sin retrotranscriptasa reversa, lo que nos permitió verificar la ausencia o presencia de contaminación de gDNA en las muestras.

Se comprobó la síntesis de todos los cDNAs, así como también se chequearon posibles contaminaciones de gDNA en cada muestra, mediante PCR amplificando el gen normalizador actina (*ACTB*). Para ello se emplearon los cebadores ACTB\_F (5'-CTTCGCGGGCGACGATGC-3') y ACTB\_R (5'-TGGTGGTGAAGCTGTAGCC-3') comunes a todas las especies. Al ser cebadores diseñados con la presencia de un intrón entre ambos (*intron spanning primers*) el producto amplificado del mRNA y del gDNA difiere en más de 500 pb y permite discriminar entre ambos.

### 3.3.1.3. RLM-RACE

Para analizar el extremo 5' del gen *G6F* se realizó un *RNA Ligase Mediated Rapid Amplification of cDNA Ends* (5'RLM-RACE) mediante el kit First Choice® RLM-RACE (Ambion), siguiendo las indicaciones del fabricante. Básicamente, se trataron 10 µg de RNA total procedente de la línea celular K562 con fosfatasa alcalina de intestino de ternero (CIP, *Calf Intestine alkaline Phosphatase*) para eliminar los 5' fosfatos de los RNA mensajeros (mRNA) parcialmente degradados, RNA ribosómicos (rRNA), RNA de transferencia (tRNA) y moléculas de DNA contaminantes. El mRNA posteriormente se trató con pirofosfatasa ácida de tabaco (TAP, *Tobacco Acid Pyrophosphatase*) para eliminar la estructura cap de los mRNAs intactos, dejando un 5' monofosfato libre de tal forma que la ligación de los adaptadores se realizara solo en los mensajeros completos. Para ello, se procedió a la ligación de un adaptador 5' RACE a partir de un cebador RNA (5'-GCUGAUGGCGAUGAUAUGAACACUGCGUUUGCUGGCUUUGAUGAAA -3') a los extremos 5' del mensajeros sin el cap.

Posteriormente se realizó una retrotranscripción con un primer específico para *G6F* (HsG6Fe6R: 5'-GGTCACACTTCCCAGCAGAT-3'), seguida de una reacción de PCR anidada a partir de la cual se consiguió amplificar el extremo 5' del transcrito *G6F* de humano. La PCR anidada (*nested PCR*) se realizó siguiendo el programa y condiciones de

la PCR convencional empleando los cebadores 5' *RACE Outer primer* (5'-GCTGATGGCGATGAATGAACACTG-3') y HsG6Fe4fR (5'-CTGCCTCCAGAGCACGAT-3') en la primera reacción de PCR, y los cebadores 5' *RACE Inner primer* (5'-CGCGGATCCGAACACTGCGTTTGCTGGCTTTGATG-3') y HsG6Fe4R (5'-GAGGGCCAGGATGACAACACT-3') en la segunda reacción de PCR. Los cebadores correspondientes a la secuencia del adaptador 5' RACE se suministraban con el kit First Choice® RLM-RACE (Ambion), mientras que los cebadores reversos se diseñaron para la amplificación específica de *G6F*. Los productos se cargaron en geles de agarosa del 1% y las bandas individuales fueron purificadas, clonadas en el vector pGEM-T y posteriormente secuenciadas.

### 3.3.2. Técnicas de DNA

#### 3.3.2.1. Hibridación de cebadores

Para el clonaje de fragmentos de pequeño tamaño (hasta 100 pb) se procedió a realizar un *Touch-down* de cada pareja de oligonucleótidos con la secuencia completa del fragmento a clonar (ver Tabla 7). Para ello se empleó de cada oligonucleótido una cantidad final de entre 25-50 picogramos en 10 µl, empleándose la concentración mayor cuando formasen estructuras secundarias fuertes y, por ello, se requiriera una concentración superior para asegurarnos que hubiese suficientes oligonucleótidos sin estructuras secundarias después del choque térmico. Se utilizó un programa de termociclador cíclico donde la temperatura de anillamiento es gradualmente reducida, partiendo desde 100°C durante 15 minutos como choque térmico para desnaturalizar los cebadores y descendiendo hasta 4°C a razón de 0.5°C por minuto.

<b>Epítopo V5</b>	V5_F	5'- <b>GGG</b> <u>GCGGCCGC</u> <b>ATG</b> GGAAAGCCGATCCCAAACCCCTCTATTAGG TCTGGACTCCACCACCGG <b>TGA</b> GATATCGGC -3'
	V5_R	5'- <b>GCC</b> GATAT <b>CTCA</b> ACCGGTGGTGGAGTCCAGACCTAATAGAGGGTT TGGGATCGGCTTTCC <b>CAT</b> GCGGCCGCCCC -3'
<b>Epítopo cMyc</b>	cMyc_F	5'- <b>GGG</b> <u>GCGGCCGC</u> <b>ATG</b> GAACAAAACTTATTTCTGAAGAAGATCTG ACCGG <b>TGA</b> GATATCGGC -3'
	cMyc_R	5'- <b>GCC</b> GATAT <b>CTCA</b> ACCGGTCAGATCTTCTCAGAAATAAGTTTTTG TT <b>CAT</b> GCGGCCGCCCC -3'
<b>Péptido señal LY6G6D</b>	SignalP_LY6G6D_F	5'- <b>GGG</b> <u>CTCGAG</u> <b>ATG</b> AAACCCAGTTTGTGGGATCTTGCTCAGCTCC CTGCTAGGGGCTGCCTTGGGAAACC <u>GCGGCCGCGGG</u> -3'
	SignalP_LY6G6D_R	5'- <b>CCC</b> <u>GCGGCCGCGG</u> TTTCCCAAGGCAGCCCCTAGCAGGGAGCTG AGCAAGATCCCAACAACTGGGGTTT <b>CAT</b> CTCGAGCCC -3'

**Tabla 7. Lista de oligonucleótidos empleados en la hibridación de cebadores.** Los oligonucleótidos directo y reverso se identifican como F y R, respectivamente. Nucleótidos en verde corresponden con ATG iniciador y en rojo con la señal de Stop de la traducción. Nucleótidos en cursiva y subrayados corresponden a dianas de restricción. Nucleótidos en negrita corresponden a la grapa añadida en los extremos libres de las dianas de restricción.

3.3.2.2. PCR

A) PCR CONVENCIONAL

La reacción en cadena de la polimerasa o *Polymerase Chain Reaction* (PCR) se empleó para múltiples amplificaciones de DNA partiendo de diferente material de partida (cDNA, plásmido ó genómico). Las amplificaciones convencionales se realizaron con *GoTaq® Green Master Mix* (Promega). Esta mezcla previamente preparada contiene la polimerasa *Taq*, así como los diferentes compuestos necesarios para la PCR: dNTPs, MgCl<sub>2</sub> y buffer. Las reacciones de PCR contenían el molde, cuya cantidad de DNA dependía de su

Cebadores			
<b>CD33</b>	Péptido señal	HsCD33F_ <i>Xho</i> I	5'- CCGCTCGAGATGCCGCTGCTG -3'
		HsCD33R_ <i>Not</i> I	5'- CAGGCGGCCGCACATAGCCAG -3'
<b>RAGE</b>	Péptido señal	HsSpAGERe1F_ <i>Xho</i> I	5'- GGCCTCGAGATGGCTGCCGGAAC -3'
		HsSpAGERe2R_ <i>Not</i> I	5'- GGC GCGGCCGCTGTTTTGAGCAC -3'
		HsAGERe1F_ <i>Xba</i> I	5'- GGCCTCTAGAATGGCTGCCGGAAC -3'
	Isoformas	HsAGERe2F_ <i>Age</i> I	5'- GGCACCGGTATCACAGCCCGGA -3'
		HsAGERe2F_ <i>Xho</i> I	5'- GGCCTCGAGATCACAGCCCGGA -3'
		HsRAGEe2F_ <i>Nde</i> I	5'- GCCCATATGATCACAGCCCGGA -3'
		HsAGERe2F_ <i>Not</i> I	5'- GCTGCGGCCGCATCACAGCCCGGA -3'
		HsRAGEi9pR_ <i>Bst</i> BI	5'- GCCTTCGAACATGTGTTGGG -3'
		HsRAGEe10pR_ <i>Bst</i> BI	5'- GCCTTCGAAGGCTAGAGTTC -3'
		HsAGERe11pR_ <i>Not</i> I	5'- GGC GCGGCCGCAAGGCCCTCCAG -3'
		HsAGERe11R_ <i>Eco</i> RI	5'- GCGGAATTC <del>CA</del> AGGCCCTCCAG -3'
		<b>Quimeras (G6F y LY6G6D)</b>	Péptido señal (y epítipo de 6xHis)
HsSpG6Fe1R_ <i>Not</i> I	5'- GGGGCGGCCGCTGTCTGCAGC -3'		
Hs_His_SpG6Fe1R_ <i>Age</i> I	5'- GGGACCGGTATGGTGTATGGTGTATGGTCTGCAGC -3'		
Isoformas	HsSpG6Fe4F_ <i>Xho</i> I		5'- GGGCTCGAGATGCCTTGGATTCC -3'
	HsSpG6Fe4R_ <i>Not</i> I		5'- CCCGCGGCCGCCTGGAGCCCA -3'
	HsG6Fe1F_ <i>Age</i> I		5'- GGGACCGGTAACATGCAGGCCATC -3'
	HsG6Fe4F_ <i>Age</i> I		5'- GGGACCGGTAGAGGAAACCGAATG -3'
	HsG6Fe6R_ <i>Eco</i> RI		5'- GCGGAATTC <del>CA</del> CCTGGGCTTGTG -3'
	HsLy6G6De1F_ <i>Age</i> I		5'- GGGACCGGTATGCGGTGCTACAAC -3'
	HsLy6G6De3R_ <i>Eco</i> RI		5'- GCGGAATTC <del>CA</del> TCCGCTCCACAG -3'
<b>Proteínas S100</b>	S100A8	HsS100A8_F_ <i>Bam</i> HI	5'- GCCGGATCCATGTTGACCGAGCTG -3'
		HsS100A8_pR_ <i>Age</i> I	5'- GCCACCGGTCTCTTTGTGGCTTTC -3'
	S100A12	HsS100A12_F_ <i>Bam</i> HI	5'- GCCGGATCCATGACAAAAC -3'
		HsS100A12_pR_ <i>Age</i> I	5'- GCCACCGGTCTCTTTGTGG -3'
	S100B	HsS100B_F_ <i>Bam</i> HI	5'- GCCGGATCCATGCTGAGCT -3'
		HsS100B_pR_ <i>Age</i> I	5'- GCCACCGGTCTCATGTTCAA -3'
	S100P	HsS100P_F_ <i>Bam</i> HI	5'- GCCGGATCCATGACGAACTAGA -3'
		HsS100P_pR_ <i>Age</i> I	5'- GCCACCGGTTTTGAGTCTCTGC -3'

**Tabla 8. Lista de cebadores empleados en el clonaje de las diferentes proteínas recombinantes.** Los cebadores directo y reverso están identificados en el nombre del cebador como F y R, respectivamente. Nucleótidos en verde corresponden con ATG iniciador y en rojo con la señal de Stop de traducción. Nucleótidos en cursiva y subrayados corresponden a dianas de restricción. Nucleótidos en negrita corresponden a la grapa añadida en los extremos libres de las dianas de restricción

naturaleza: 25 ng plásmido, 50-100 ng genómico, o 12,5-50 ng (1µl) de cDNA, 0.5 µM de cada cebador y 10 µl de *GoTaq<sup>®</sup> Green Master Mix 2x*, en un volumen de reacción de 20 µl. Para las reacciones en los que no se podían permitir errores en la secuencia amplificada, como en el caso de los clonajes para expresión de proteínas, se emplearon las DNA polimerasas de alta fidelidad *Pwo SuperYield* (Roche) o *Phusion<sup>®</sup>* (New England Biolabs, NEB) en un volumen final de 20 µl y suplementando la mezcla de reacción con *GC-Rich Resolution Solution* (Roche) ya que la región del genoma del MHC de clase III posee un alto contenido en GCs. Los cebadores empleados en dichos clonajes se muestran en la Tabla 8. Las PCR se realizaron principalmente en un termociclador MJ Mini (Bio-Rad) bajo las siguientes condiciones: 95°C durante 5 minutos seguido de 30 ciclos de 95°C durante 1 minuto, 58-65°C (dependiendo de la T<sub>m</sub> de los cebadores empleados) durante 1 minuto, 72°C durante 1-5 minutos (dependiendo del tamaño a amplificar), y finalmente a 72°C durante 5 minutos.

#### B) PCR de colonia

Para el chequeo de los clonajes se realizaron PCRs de colonia (*colony PCR*) de las colonias de bacterias transformadas. Para ello se empleó la polimerasa *GoTaq<sup>®</sup> Green Master Mix* (Promega) y la mezcla de reacción seguida para una PCR convencional (ver Apartado 3.3.2.2.A), excepto por un volumen final de 10 µl y por el molde que no es una solución líquida sino simplemente el contacto de una punta estéril con la colonia de interés. El programa de PCR empleado fue el convencional (ver Apartado 3.3.2.2.A), con una T<sub>m</sub> de entre 58-65°C (dependiendo de la T<sub>m</sub> de los cebadores empleados), y un tiempo de elongación de entre 1 a 5 minutos (dependiendo del tamaño a amplificar, usando 5 minutos en el chequeo de los báculos). Para el chequeo de los clonajes se emplearon diferentes cebadores dependiendo de los plásmidos empleados. En el caso de los plásmidos recombinantes de pGEM-T, los cebadores que se emplearon para la PCR de colonia fueron los mismos que se emplearon para la amplificación del inserto correspondiente. En cambio, para los plásmidos pcDNA3.1Zeo(-) y pAL7 se diseñaron unos cebadores en ambos extremos del sitio de clonaje múltiple (MCS) del plásmido, como se muestran en la Tabla 9. En la tabla se muestran también los cebadores diseñados para el chequeo del báculo recombinante usado para la expresión en baculovirus.

Cebadores		Tamaño producto negativo	Tamaño producto positivo
pGEM-T	Empleados en la amplificación previa del inserto.	-----	I
pcDNA3.1Zeo(-)	pZ_5MCS_F 5'-CACCAAAATCAACGGGACTT-3'	860 pb	860 + I pb
	pZ_Ori_R 5'-CCACTACGTGAACCATCACCC-3'		
	pAL7_MCS_F 5'-ATGAAGTGGTTCGCATCCTC-3'	660 pb	660 + I pb
	pAL7_MCS_R 5'-CAAACCACAACACTAGAATGCAGTG-3'		
Bácmido (bMON14272)	pUC/M13_F 5'-CCCAGTCACGACGTTGTAAAACG-3'	300 pb	~ 3 kpb
	pUC/M13_R 5'-AGCGGATAACAATTCACACAGG-3'		

**Tabla 9. Lista de cebadores empleados en las diferentes PCRs de colonia.** Los cebadores directo y reverso están identificados en el nombre del cebador como F y R, respectivamente. I corresponde al tamaño del inserto.

### C) PCR anidada

La PCR anidada (*nested PCR*) consiste en una PCR con doble ronda de amplificación empleando como molde de la segunda PCR el producto de la primera, de forma que genes de baja expresión pueden llegar a detectarse. Las reacciones de PCR se realizaron empleando *GoTaq® Green Master Mix* (Promega) como se detalla en las PCR convencionales (ver Apartado 3.3.2.2.A). En la primera ronda de amplificación se empleó 1 µl del correspondiente cDNA molde proveniente de las retrotranscripciones del RNA, en un volumen de 20 µl finales y empleando unos cebadores externos en su secuencia (Tabla 10). En la segunda ronda de amplificación se empleó 1 µl de una dilución 1:10 del producto de la primera amplificación en el caso de la amplificación de las diferentes variantes de *splicing* de un gen, mientras que se empleó directamente 1 µl del producto de la primera amplificación en el ensayo de 5'RLM-RACE. En esta segunda amplificación se emplearon unos cebadores internos (Tabla 10) diseñados en una región de la secuencia de los transcritos más interna en comparación con los primers externos y, además, se redujo el número de ciclos a 20. Para verificar la reproducibilidad de una técnica tan sensible se realizaron dos experimentos de PCR completamente independientes para cada experimento y, solamente se consideraron aquellos productos que aparecían en ambas amplificaciones.

**Tabla 10 (siguiente página). Lista de cebadores empleados en el análisis de las diferentes variantes de *splicing*.** Los cebadores directo y reverso están identificados en el nombre del cebador como F y R, respectivamente. El exón en el que están diseñados las parejas de cebadores está identificado como “e”. Todos los cebadores están indicados en orientación 5'→3'.

			Cebadores externos	Cebadores internos	
<b>Homo sapiens</b>	RAGE	RAGEe1F	CAGGACCCTGGAAGGAAGCAGG	AGCCGGAACAGCAGTTGGAGCC	
		RAGEe11R	CTGGTTGTAGAAGAAAGCTTGGC	GCCCTCCAGTACTACTCTCGCC	
		G6Fe1F	GAGTGTGGGCCTTTATCTGG	CCAGGTCTTGAGCAAGAGA	
	G6F	G6Fe4F	CTCCATCGATGCTTCTCCTG	CTGATGCTGCTGCTCACAAT	
		G6Fe6R	CCCTTGCTAAGGTTGTCGAG	GGTCACACTTCCCAGCAGAT	
		LY6G6Ee1F	GCAGATCACCTCTCCCAGAT	ATGGGCACCTCCAGCATC	
	LY6G6E	LY6G6Ee3R	GTCTGTGCAGGTGTTTCCAG	CACGGAAGCTAGTGGAGGAG	
		LY6G6D	LY6G6De1F	GGATATAAGGCAGGCAGAGC	CGATGAAACCCCAGTTTGT
			LY6G6De3R	TGTTCAAGATGTCCCTTGTCC	GCTCCCTTGTTCCTTCTCT
<b>Macaca mulatta</b>	RAGE	RAGEe1F	CAGGACCCTGGAAGGAAGCAGG	AGCCGGAACAGCAGTTGGAGCC	
		RAGEe11R	CTGGTTGTAGAAGAAAGCTTGGC	GCCCTCCAGTACTACTCTCGCC	
		G6Fe1F	GAGTGTGGGCCTTTATCTGG	CCAGGTCTTGAGCAAGAGA	
	G6F	G6Fe4F	GATGCTTCTCCTGCTTCTG	CTGATGCTGCTGCTCACAAT	
		G6Fe6R	GCCAAGGTTGTCGAGGAATA	GGTCACACTTCCCAGCAGAT	
		LY6G6Ee1F	GTTTCCTGGTTCCCTCCCTGT	ATCACCTCTCCCCAATTCTCT	
	LY6G6E	LY6G6Ee3R	GTCTGTGCAGGTGTTTCCAG	GTGAGCTGCCATGAGGAAGT	
		LY6G6D	LY6G6De1F	TCAAGGCAAGGGGATATAA	CGATGAAACCCCAGTTTGT
			LY6G6De3R	AAGGATGCCTGTCTTCTCACA	AGATGTTCCGATGTCCCTTG
<b>Bos taurus</b>	RAGE	BiRAGEe1F	AGCCTGGGAAGGAAGCAGG	GGGGCAGTGGTCCGAGCC	
		BiRAGEe11R	GGTTGTGGAAGAAAGTTTGGG	CTCCTGTGCTGCTCTCTGCC	
		G6Fe1F	AGCAGGTCCCAGACAAGAGA	CTCTCCCATGGCTGTCTTA	
	G6F	G6Fe4F	TTCCGCAGACTGGGATGT	GGTTCACCATTGTGGTCTCT	
		G6Fe6R	TCCCCACTGTCAAGCTTTT	TCCTCGAGATCACCTGGTTT	
		LY6G6Ee1F	CTCAGCGTGCAGATCACATT	CATGGGTACCTCCAGCATCT	
	LY6G6E	LY6G6Ee3R	AATGCCTGGAGGAATCAGGT	AGTCAGGTCTTGGGCAGAT	
		LY6G6D	LY6G6De1F	GATAAAAAGCCGGCAGAGTG	AGAGGCGCTACAGACACAGG
			LY6G6De3R	CAGCTCTGGAGTGGTCCGAG	TTCTTGTGGCTTCTCTCTC
<b>Sus scrofa</b>	RAGE	SsRAGEe1F	AGCCTGGGAAGGAAGCAGG	GCAGCAGGGACAACCGCC	
		SsRAGEe11R	GGGTGAACTGGTCTGGGGCC	GCTCCTGCACTGCTCTCAGCC	
		G6Fe1F	GAGTGTGGGCCTTTATCTGG	CAGGTCCCAGCAAGAGAAC	
	G6F	G6Fe4F	TTCTGCAGGCTGGGATGT	ATCCTGGTCTCAGCATCAT	
		G6Fe6R	AGGGAGACAGCAGGTCACAC	CTTCCCAGCAGGTATCATCG	
		LY6G6Ee1F	TCTCTGATCTCAGGGTGCAG	ATGGGCACCTCCAGCATC	
	LY6G6E	LY6G6Ee3R	CGGAAGACTCTTGGGTTGAA	TGGTCCACATGAGGCTAGTG	
		LY6G6D	LY6G6De1F	AGAGGTGCTACCACACAGG	CGGACACAATGAACCCAGTT
			LY6G6De3R	ATCTCCTTCGTTCCCTTGCT	CTCCCACTACCCTCTCCACA
<b>Rattus norvegicus</b>	RAGE	RnRAGEe1F	AGCCTGGGAAGGAAGCACC	ACAGCAGCTAGAGCCTGGG	
		RnRAGEe11R	ATCATGTGGGCTCTGGTTGG	CCCGGCACCATTCTCTGGC	
		G6Fe1F	GGGACTGTGGGCCTTTATCT	CCCATGGCGATGGTAGTATT	
	G6F	G6Fe4F	ACTGTTTGTGCCCTCTCC	CCAAGGGGTCACCATCATAG	
		G6Fe6R	CAGGTTCCAGCAACAATCT	GGGATTTAAGGCTCCCTCCT	
		LY6G6Ee1F	GGGCAGGTGTGTCTGACTTT	AGACTGCAGAGCCGTTCTCT	
	LY6G6E	LY6G6Ee3R	CTGTGCAGGTGTTTTCATGC	ACGGGCTAGAGGACGATATG	
		LY6G6D	LY6G6De1F	GGATATAAGCCCGGTGGAGT	GGCATAAGGGAACGTGATGA
			LY6G6De3R	GATGTCCAGTCCCTCACTC	TCCCTCAGTCCCTTCCCTAC
<b>Mus musculus</b>	RAGE	MmRAGEe1F	AGCCTGGGAAGGAAGCACC	ACAGCAGCTAGAGCCTGGG	
		MmRAGEe11R	ATCATGTGGGCTCTGGTTGG	CCCGGCACCATTCTCTGGC	
		G6Fe1F	CCTCCTCTCTGAAACACCA	GCTCTTCCATGGCAGTTGTAG	
	G6F	G6Fe4F	CCCTGGATACTGGTGTCTGTT	CCAAGGGGTCACCATCATAG	
		G6Fe6R	GAATCCTTCAGGTGCCAGTG	AGGATCACCTGGTCTTGTGG	
		LY6G6Ee1F	ACTCGGGGACTGGTTTGTG	CCAGACTGCTGAGCTGTACCT	
	LY6G6E	LY6G6Ee3R	ACACTGGAATCCTGGGTTGA	GATGGGCTAGAGGAAGTCA	
		LY6G6D	LY6G6De1F	GCTGGTGGAGTAGAGCCTGA	GGCATAAACGAACGTGATGA
			LY6G6De3R	GATGTCCAGTCCCTCACTC	CCACCCTCAGCTCTACAGT



### D) PCR cuantitativa (qPCR)

La PCR cuantitativa (qPCR) o PCR en tiempo real es un método empleado para la amplificación y cuantificación de DNA. Esta tecnología se basa en el uso de fluoróforos intercalantes de DNA para la cuantificación de DNA inmediatamente después de cada ciclo. En los experimentos realizados en esta Tesis se empleó el fluoróforo SYBR<sup>®</sup> Green (Applied Biosystems), que al intercalarse en el DNA y excitarse a 497 nm es capaz de emitir fluorescencia a 520 nm y, por tanto, es posible cuantificar la amplificación del DNA en el termociclador ABI Prism 7900HT (Applied Biosystems). La fluorescencia captada se mide en unidades relativas de fluorescencia (RFU, *Relative Fluorescent Units*) y se determina el Ct (*threshold cycle* o ciclo umbral) a partir del ciclo en el que la fluorescencia de la muestra comienza a ser superior a la fluorescencia basal. El valor Ct es inversamente proporcional a la cantidad inicial de moléculas diana en la muestra, de manera que cuanto mayor número de moléculas iniciales presente menor será nuestro valor Ct.

Se incorporó un ciclo final de disociación en el programa de PCR que consistía en un aumento progresivo de temperatura desde 60°C a 95°C en el que se cuantificaban los valores de fluorescencia obteniendo la temperatura de *melting* de cada uno de los productos amplificados. La temperatura de *melting* se define como la temperatura a la cual se encuentra desnaturalizada la mitad de una molécula de DNA de banda doble. La presencia de este pico de *melting* es específico para cada producto de PCR, de manera que se pueden distinguir contaminaciones por dímeros de cebadores o bandas inespecíficas. Además, mediante curvas de dilución del DNA molde se descartaron tanto contaminaciones con etanol u otros compuestos (de la muestra de partida) que inhibirían la amplificación por la polimerasa, así como se calculó la eficiencia de las parejas de cebadores empleadas (Pfaffl 2001). En algunos casos, se empleó el método  $2^{-\Delta\Delta C_t}$  desarrollado por Livak y Schmittgen (Livak and Schmittgen 2001) en el que se asumía que la eficiencia de amplificación de las diferentes parejas de primers era igual a 2.

Existen dos abordajes diferentes de qPCR que se pueden llevar a cabo, la cuantificación absoluta y la cuantificación relativa. La qPCR absoluta cuantifica los niveles de expresión totales de un mensajero y, nos permite comparar los niveles de expresión de varios mensajeros en una misma muestra. Para ello es necesario el empleo de curvas de dilución de estándares con una concentración conocida para poder extrapolar la concentración de nuestro amplicón problema. Se emplearon como estándares plásmidos en



## MATERIAL Y MÉTODOS

Nombre	Cebadores directo	Cebadores reversos	Intron-spanning	Longitud amplicón	Chip
<b>Genes del MHC de clase III</b>					
<i>BAT1</i>	ACAGCTCTGGCTTTTCGTGA	GGATGCACTCATGCTGGAC	Si	102	2
<i>ATP6V1G2</i>	GACTGAAGCAGGCAAAGGAG	GACACGCTCTCGGTTTCTCT	Si	191	2
<i>NFkBIL1</i>	CCAGATGCCTACACCGATT	TGGAGGCATCATCTTCTTCC	Si	154	2
<i>LTA</i>	ACAGCACCTCAAACCTGCT	TGGCAAGGAGAAACCATCC	Si	114	2
<i>TNF</i>	CAGCCTCTTCTCCTCTGAT	GCCAGAGGGCTGATTAGAGA	Si	123	2
<i>LTB</i>	GAGGAGGAGCCAGAAACAGA	TCAGAAACGCCTGTTCCTTC	Si	115	2
<i>LST1</i>	GACTATGCCTGCATTGCTGA	GCTCGAGATGGACTGAGAGG	No	152	2
<i>NCR3</i>	TACGTGTGCAGAGTGGAGGT	GGCCACAGAGAGAAAGCTGA	Si	150	2
<i>UQCRHP †</i>	GCTGACTGGGCCTAGAGATG	TCTCGCTTTTGTAGGAATTGA	No*	78	2
<i>AIF1</i>	ATGTCCTGAAACGAATGCT	TTTTAGGATGGCAGATCTTGG	No	157	2
<i>BAT2</i>	AGTTGGCCTGGGTAGGAGAT	GGGGAGGTACTGGAGGAGTC	Si	129	2
<i>BAT3</i>	GGATTCCCACCAGAGTGTG	GGAGCACTCGGAACACCAC	Si	91	96
<i>APOM</i>	CTCCACCTTCGTGCTACCAT	TAACCTGGCCTGTCTCATT	Si	179	96
<i>C6orf47</i>	GGCATCCCAAGACTAAGGACT	GCTCCTCCTTGCTGGTTC	No*	72	96
<i>BAT4</i>	CGGAAGCTACACAGCAACAC	CTGTAGCAACCACAGCCTCA	Si	168	96
<i>CSNK2B</i>	TGTCCTGGATTTCTGGTTC	CAGGCTCCAGGTCCAAGAT	Si	156	96
<i>LY6G5B</i>	AGGATGCATTTCAAGCTCAG	AGGTGTTGCGTTTTCTTGG	Si	149	96
<i>LY6G5C</i>	GCTGCATCACTCTCCACAAA	AGGGTCATTGCAGAAATCCA	Si	164	96
<i>BAT5</i>	CTGTGATGGCAATGAGATTGA	CCACCTCATAAAACCAGCA	Si	113	96
<i>G6F</i>	GAGCCGAAGACCAAGAATCA	AGGGCCAGGATGACAACCTC	Si	177	96
<i>LY6G6E †</i>	ATCTTCTCTGCGTGCTGTT	GGTTTGGTGAAGCCACAGAT	Si	130	96
<i>LY6G6D</i>	AATGGCGTGCTACAACCTGTG	AGGCTGGATGTTGGTGAATC	Si	149	96
<i>LY6G6C</i>	GTGCTGGGCTGTGTGGAC	ACGGTTGGTTTGGTTGAAG	Si	156	96
<i>C6orf25</i>	AACCATGGCTGTGTTTCTGC	AGCGGATGGGATGAGAGACT	Si	134	96
<i>DDAH2</i>	TCTTTCTTCGCTCCTGGGTTG	CCAGTCTGAGCAGGACAC	Si	135	96
<i>CLIC1</i>	AGTTTTTGGATGGCAACGAG	CAGGTGGAAGCGAATTCTTC	Si	172	96
<i>MSH5</i>	ACTATCGTAGTGCCGAACC	GAGGCAAGGTCCAATACTCG	Si	151	96
<i>C6orf26</i>	GCAGGCTCTGGAGAGAGAAC	CGACCAGGCTCTGAGTGTA	Si	165	96
<i>C6orf27</i>	CTTACCAAAGACCAGCACA	GTGGATCCGGACTGTGATCT	Si	160	96
<i>VAR5</i>	CGGGCTATCCTGTCAAGGT	GGGTGAGGGGTGAAGCTG	Si	149	96
<i>LSM2</i>	AAGGATGTGGTCTGTTGAACT	CAGGGTCTGTGACACTGATGT	Si	109	96
<i>HSPA1L</i>	CTGGGCTGCGTAATCTG	GATGCCTATGGCGATTCC	Si	102	96
<i>HSPA1A</i>	CCCACCATGAGGAGGTAGA	AAAACAAAACAGCAATCTTGA	No*	54	96
<i>HSPA1B</i>	GGACTGTTGGGACTCAAGGA	TGAAGCAGCAAAGAGCTGAA	No*	71	96
<i>C6orf48</i>	TTCCTGCAACTTGACCTTCA	CCAGGCGGAAGATATTGTGT	Si	96	96
<i>NEU1</i>	ACCGGGCTCTGGTATTGAG	CTGACCCCACTCCGTAGC	Si	144	96
<i>SLC44A4</i>	GGCTGTGAGAAAGTGCCAAT	TTGGATTAGCCTTTGGATG	Si	117	96
<i>EHMT2</i>	CGGCTACAGCTCTACCGAAC	GAGATCAGTCCCCGACATA	Si	104	96
<i>ZBTB12</i>	GGTGGGAAAATAAGGGATT	GTTTCATGTTCCGTAGCGTTG	Si	140	96
<i>C2</i>	TCTCCAAAGACAGCTGTTGA	CCTTCTGGACCTGACTCA	Si	142	96
<i>CFB</i>	ATGCTGACCCCAATACTTGC	TCCACTACTCCCAGCTGAT	Si	103	96
<i>RDBP</i>	CTCAGGCTTCAAGCGTTCTC	CTGAAAGTGGGGACTGGT	Si	76	96
<i>SKIV2L</i>	CCATCCTCAAGGAGATCGTG	TTACTCCCATGGCAAAGGTC	Si	87	96
<i>DOM3Z</i>	TGTTTGAATATGTCAGGAATGACC	CAGGCTCCCAAGAGAAGAGA	Si	141	96
<i>STK19</i>	GTACAGCCTTGTGCCTGACA	GATAATCCATGGGCATCCA	Si	115	96
<i>C4A</i>	GGACCCTGTCCAGTGTAG	GCTGTGAGTGCCACAGTCTC	Si	75	96
<i>C4B</i>	GCTCGTTCCAGGACCTCTCT	GCTGTGAGTGCCACAGTCTC	Si	85	96
<i>CYP21A2</i>	ACGTGATTCCCTTTCTCAGG	TAGTCCATCATGTCCCTCCAC	Si	148	96
<i>TNXB</i>	CCCCTACAGGTTCTCTCTCT	GGTGGTCACGTCAGTCACAG	Si	151	96
<i>ATF6B</i>	CAATTGACCGACGGGAAG	CACGGCCTGACAGGGTCT	Si	150	96
<i>FKBPL</i>	TCACCATGCTAGGCACTTTG	GGTCAGAAAGGGATGAACCA	Si	64	96
<i>PRRT1</i>	CTACGTAGCGCCCGGATAC	AGTGGAGGTCACTCCTGTTCC	Si	118	96
<i>PPT2</i>	CGGAAGAATTTCTGCGTGT	CATCTCCAGGACGGTCTCAT	Si	123	96
<i>EGFL8</i>	GACCGGATCGAATCTCTCAG	GGCTTCTTATCGATGATTGACG	Si	111	96
<i>AGPAT</i>	TCTTGCCTAATGCTGCTC	CCTGGCAGTACCTCCATCAT	Si	154	96
<i>RNF5</i>	ATGCCATGAGCCTTTCC	CGAGAAACAGGAAGAGGGAAT	Si	93	2
<i>AGER</i>	CCCTGGTGCCTAATGAGAAG	CTGGCTGAAGCTACAGGAG	Si	150	2
<i>PBX2</i>	ACTTCAGCAAACAGGCCACT	CCTCTTGTGGCCAAACCAGT	Si	149	2
<i>GPSM3</i>	CCACATCCCTGGATCTGAAT	CCCTGCTCACCATCTCTT	Si	120	2
<i>NOTCH4</i>	CCTCTCTGCAACCTTCCACT	GCCTCCATTGTGGCAAAG	Si	87	2

<b>Variantes de <i>splicing</i> del gen <i>NFkBIL1</i></b>					
<i>NFkBIL1_A</i>	TGTGAATGCGGTGAAAAGC	TCAAACCTCCCCATGACTTC	Si	514	48
<i>NFkBIL1_B</i>	AAGCCTCCACATCTGTCTG	TCAAACCTCCCCATGACTTC	Si	519	48
<i>NFkBIL1_C</i>	TGTGAATGCGGTGAAAAGC	GGAGGCATCACCCCTGG	Si	483	48
<i>NFkBIL1_D</i>	AAGCCTCCACATCTGTCTG	GGAGGCATCACCCCTGG	Si	488	48
<b>Quimeras de los genes <i>G6F</i> y <i>Ly6G6D</i></b>					
<i>Shch</i>	CCTCTGGTCTGGCCCTTACT	ACCGCATTCCGGTTTCCTC	Si	195	96
<i>Lgch</i>	CACAACAAAGGGGTCAGCTT	ACCGCATTCCGGTTTCCTC	Si	202	96
<b>Variantes de <i>splicing</i> del gen <i>RAGE</i></b>					
<i>RAGE total</i>	TGCTGATCCTCCCTGAGATAG	CTCCTCGCCTGGTTCGAT	Si	125	2
<i>Intron 1 ret.</i>	CCCTTCTACCCTCTACCA	CTGTCCGGCCTGTGTTC	Si	159	96
<i>Intron 4 ret.</i>	AGGCAATGAACAGGAATGGA	GGGCCGTTTTCTACTTCTCC	Si	158	96
<i>Intron 6 ret.</i>	CACCTCAAACCCCTCCAAC	ATCTCAGGGAGGATCAGCAC	Si	207	96
<i>Intron 9 ret.</i>	TGCTGATCCTCCCTGAGATAG	CTTTATCAAACCCCTCACCTG	Si	156	2
<b>Genes relacionados con la proteína <i>RAGE</i></b>					
<i>ADAM 10</i>	TCCCTTGCAACGATTTTAG	AATACTGCCACCAATGAGC	Si	158	96
<i>ADAM 17</i>	CCTTTCTGCGAGAGGGAAC	CACCTTGCAAGGATGTCAGT	Si	69	96
<i>SP100</i>	GAACCACCAAGCACAAGAAAA	TGACTGAAGCATCTGGGTCT	Si	103	96
<i>HMGB1</i>	TCACAGCCATGTCAGTACATT	AGGATCTCCTTTGCCATGT	Si	125	96
<i>S100A12</i>	CACATTCCTGTGCATTGAGG	GGGTCAAAATGCCCTTC	Si	122	96
<i>S100B</i>	AGACCAGGAAGGGGTGAGA	CCTCCTTCCAGAATATTGGT	Si	97	96
<i>S100P</i>	GGGAGCTCAAGGTGCTGAT	AGCAATTATCCACGGCATC	Si	82	96
<i>S100A1</i>	CCATCTGTCCAGAACCTGCT	GCTCCTTCTGTCTCAGTTG	Si	156	96
<i>S100A8</i>	CAAGTCCGTGGGCATCAT	GACGTGATGATAGAGTTCAAGG	Si	61	96
<i>S100A9</i>	GTGCGAAAAGATCTGCAAAA	TCAGCTGCTTGTCTGCATTT	Si	103	96
<b>Genes relacionados con <i>AR</i></b>					
<i>HLA-DRB1</i>	CGACCACGTTTCTGTGG	CACGGACTCCTCTGGTTAT	Si	105	48
<i>PTPN22</i>	TTTATCTACAACCCTCTGGACTT	TCCATGCATGCCATAACAAT	Si	78	48
<i>CIITA</i>	CAGGGAGGCTTATGCCAATA	CTGGGAGTCTTGAAGACAT	Si	58	48
<i>IRF5</i>	TCTACGAGGTCTGTCCAATG	TGCACAAAAGAGTAATCCTCA	Si	71	48
<b>Genes relacionados con la ruta de <i>NFkB</i></b>					
<i>IL1β</i>	TACCTGTCTCCTGCGTGTGAA	TCTTTGGGTAATTTTGGGATCT	Si	76	96
<i>IL6</i>	GGCTGAAAAGATGGATGCT	GCTCTGGCTTGTTCCTCACT	Si	141	96
<i>IL8</i>	CTCTCTTGGCAGCCTTCT	TCTTTAGCACTCCTTGGCAAAA	Si	70	96
<i>IL10</i>	CATCGATTCTTCCCTGTGA	TGTCAAACCTCACTCATGGCTTT	Si	109	96
<i>COX2</i>	CTTCACGCATCAGTTTTCAAG	ACCCTGAAATATGATTTAAGTCCAC	Si	96	96
<b>Genes normalizadores</b>					
<i>CSNK2B</i>	TGCTCCTGGATTCTCTGGTTC	CAGGCTCCAGGTCCAAGAT	Si	156	48
<i>UBC</i>	ATTTGGGTGCGAGTTCTTG	TGCCTTGACATTCTCGATGGT	Si	133	48
<i>YWHAZ</i>	GCTTCAAAGGGCCAGTGTA	CCGCCAGGACAAACCAGTAT	No*	77	2
<i>B2M</i>	TGCTGTCTCCATGTTTGTATCT	TCTCTGTCTCCCCACCTCTAAGT	No	86	48
<i>ATP5B</i>	TCACCCAGGCTGGTTCAGA	AGTGGCCAGGGTAGGCTGAT	Si	80	48
<i>HSP90AB1</i>	AAGAGAGCAAGGCAAAGTTGAG	TGGTCACAATGCAGCAAGGT	Si	120	48
<i>FLOT2</i>	TGCCGTGGTGCAGAGAGA	GTGTCTGCCATGAACCTCACA	Si	113	48
<i>TMBIM6</i>	TGCTGGATTGCATTCCTTACA	GAGTGCATGAGGGTGAAGC	Si	136	2
<i>CFL1</i>	GCAAGAAGGAGGATCTGGTG	CCTTGGAGCTGGCATAAATC	Si	84	2
<i>TPT1</i>	CTACAAGATCCGGGAGATCG	GAGGCATTTCCACCAATGAG	Si	108	48
<i>TUBA1C</i>	GTTAGTGGGAGATCCTTGTTC	GCATTGCCAATCTGGACAC	Si	125	48
<i>ACTB</i>	ACAGAGCCTCGCCTTTGC	GCGGCGATATCATCATCC	Si	75	2
<i>S100A6</i>	AAGGAGCTGATCCAGAAGGAG	TTGTTCCGGTCCAAGTCTTC	Si	89	48
<i>MRPL19</i>	GGTCATCGTGGACAAGCAC	TTCTTCCCCTTCGAGGAATG	Si	80	48
<i>TBP</i>	AGGGTTTCTGGTTTGCCAAG	GCCATAAGGCATCATTGGAC	Si	140	48
<i>TFRC</i>	TCTGGAATCCCAGCAGTTTC	CGTGCCACTTTGTTCAACTC	Si	122	48
<i>ESR1</i>	CAGTGAAGCTTCGATGATGG	GAGGGTCAAATCCACAAAGC	Si	106	48
<i>ERBB2</i>	GTGTGGACCTGGATGACAAG	AACCACCGCAGAGATGATG	Si	77	48
<i>GAPDH</i>	GAGTCAACGGATTTGGTCGT	TTGATTTTGGAGGGATCTCG	Si	238	2
<i>POLR2A</i>	GCAAATTCACCAAGAGAGACG	CACGTGACAGGAACATCAG	Si	73	2

**Tabla 11 (página anterior). Lista de cebadores empleados en los estudios de qPCR y Fluidigm.** \* indica que dicho gen solo tiene un exón. † corresponde a un pseudogen. Chip de 48, 96 ó 2 indica que las parejas de cebadores han sido empleadas en la placa de 48\*48, 96\*96 o en ambas placas, respectivamente.

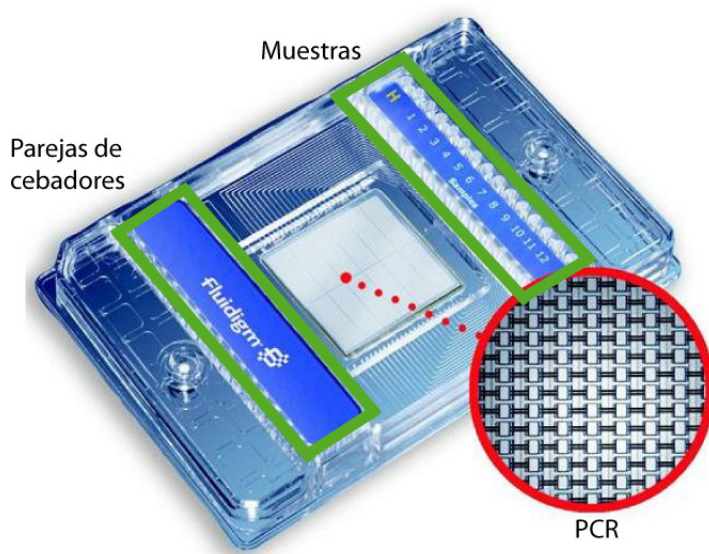
los que habían sido previamente clonados el producto que se estaba amplificando por qPCR. Por otro lado, la cuantificación relativa nos permite comparar los niveles de expresión de un mensajero en diferentes muestras mediante el empleo de determinados genes normalizadores o *housekeeping genes*. Estos normalizadores permiten corregir posibles diferencias existentes en la cantidad y degradación del RNA de partida entre las diferentes muestras.

Se diseñaron cebadores para amplificar a todas las variantes de los genes del MHC de clase III conjuntamente, así como también se diseñaron cebadores para amplificar y discriminar determinadas variantes de *splicing* de interés (Tabla 11). Para ello se realizaron ensayos cruzados de qPCR con curvas de dilución de estándares respectivas a cada una de las variantes a estudio, empleando diluciones seriadas de 10 a 10<sup>6</sup> moléculas. Con ello se comprobó que cada pareja de cebadores amplificaba únicamente la variante correspondiente o había una diferencia de al menos diez ciclos, por lo que no se verían afectados los Ct ni el resultado final. Cada mezcla de reacción contenía *IX Power SYBR<sup>®</sup> Green Master Mix* (Applied Biosystems) y 250 nM de cada cebador. Las condiciones del termociclador fueron: 95°C durante 10 minutos, 40 ciclos de 15 segundos a 95°C, 30 segundos a 60°C y 30 segundos a 72°C y, posteriormente, se realizó el ciclo final de disociación comentado anteriormente. La detección de la fluorescencia se realizó a 72°C. Todos los puntos se realizaron por triplicado. El análisis de los resultados de qPCR se realizó empleando el programa SDS 2.2.1 (Applied Biosystems). Se emplearon los criterios establecidos por Bustin et al. (2013) para aportar la información técnica necesaria y favorecer de este modo la transparencia y reproducibilidad de los experimentos de qPCR (Bustin et al. 2013).

#### E) PCR cuantitativa de alto rendimiento (plataforma Biomark™)

Además de los ensayos de qPCR convencionales, en esta Tesis también se realizaron ensayos de qPCR con la tecnología *BioMark™ HD System* (Fluidigm Corporation) que nos permitía trabajar con muy poca cantidad de muestra y a su vez analizar un elevado número de sondas diferentes. Esta tecnología consiste en un sistema de

análisis por RT-PCR basado en nanofluidos que se realiza en *chips* de 48x48 ó 96x96, de tal forma que permite enfrentar 48 ó 96 muestras simultáneamente con 48 ó 96 sondas diferentes (genes o transcritos diferentes) (Figura 22). En este trabajo se realizaron dos ensayos de Fluidigm, uno de 48x48 y otro de 96x96. El chip de 48x48 fue realizado en el laboratorio de investigación molecular de la Fundación de Investigación del Hospital General Universitario de Valencia (FIHGUV), a través del Dr. Manuel Mata Roig, y el chip de 96x96 en la Unitat de Genòmica Funcional del Institut D'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS) de Barcelona a través de la Dra. Magdalena Pinyol. En el chip de 48x48 se incluyeron múltiples genes normalizadores para sangre periférica para determinar la mejor pareja de normalizadores para el segundo experimento en el que se incluyeron todos los genes situados en el MHC de clase III, así como genes relacionados con inflamación (Tabla 11). Estos experimentos fueron realizados con los consejos técnicos del Dr. Juan José Fuentes del Servicio Técnico de *Fluidigm Corporation*.



**Figura 22. Chip de BioMark™ System, Fluidigm Corporation.** En la parte izquierda se cargan las parejas de cebadores (48 ó 96) y en la parte derecha se cargan las muestras (48 ó 96). Posteriormente, en la parte central se cruzan todos los cebadores con todas las muestras, de manera que en cada cruce se produce una qPCR en tan sólo 5 nL de volumen de reacción.

El protocolo de qPCR en la plataforma Biomark™ requiere de un paso previo de amplificación conocido como *Specific Target Amplification* (STA), que permite el enriquecimiento de las muestras en los genes de interés. Para cada STA se emplearon 1X *TaqMan® PreAmp Master Mix* (Applied Biosystems) junto con 0.2X del conjunto de parejas de cebadores (preparado en 10 mM Tris pH 8, 0.1 mM EDTA) (Tabla 11) y 12.5 ng del cDNA correspondiente en un volumen final de 5 µL. La polimerasa *TaqMan® PreAmp* fue amablemente cedida por el Dr. Ricardo Ramos de la Fundación Parque Científico de Madrid (FPCM). El programa empleado consta de una fase desnaturalización a 95°C

durante 10 minutos y 15 ciclos de 95°C durante 15 segundos y 60°C durante 4 minutos. Posteriormente, se diluyen las reacciones 1:10 y se cargan en el chip correspondiente. Los experimentos realizados con esta tecnología se enmarcan dentro de la categoría de qPCR relativa por lo que se requiere el uso de normalizadores, así como también se emplearon calibradores interplaca para poder comparar los resultados en ambos chips de Fluidigm.

### **3.3.2.3. Electroforesis en geles de agarosa**

Los productos obtenidos en las diferentes amplificaciones fueron analizados en geles de agarosa al 1%, salvo en el caso del análisis de las variantes de *splicing* que se realizó a un porcentaje de agarosa del 2%. Las reacciones de PCR con la mezcla *GoTaq*<sup>®</sup> *Green Master Mix* (Promega) poseen la suficiente densidad para cargarlo directamente en los geles de agarosa. Además, contiene dos colorantes (azul y amarillo) que permiten monitorear el progreso durante la electroforesis. Los productos amplificados procedentes de las enzimas de alta fidelidad fueron cargados en un gel junto con un buffer de carga que les aporta la densidad necesaria y el colorante azul de bromofenol para monitorear el frente de corrida de la electroforesis.

### **3.3.2.4. Purificación de banda de gel o de PCR**

Las bandas cortadas del gel, así como directamente los productos de PCR, fueron purificados empleando el kit *Wizard*<sup>®</sup> *SV Gel and PCR Clean Up System* (Promega), según las indicaciones del fabricante. Se eluyeron en 30 µl de agua libre de nucleasas. Las bandas de gel cortadas fueron siempre bandas individuales a excepción de las bandas procedentes de las amplificaciones de *RAGE* en humano y macaco Rhesus. En estos casos, las bandas entre 1 y 1.5 kpb fueron extraídas y purificadas al mismo tiempo. La concentración del DNA se midió por espectrofotometría en un NanoDrop ND-1000 (Thermo Scientific) del Servicio de Genómica y Secuenciación Masiva del CBMSO.

### **3.3.2.5. Digestión**

El DNA o los plásmidos purificados se digirieron con diversas enzimas de restricción, clasificadas por los extremos que generan tras la digestión. Hemos empleado enzimas de restricción que generaban extremos 5´ protuberantes tales como *AgeI*, *BamHI*,

*EcoRI*, *NheI*, *NotI*, *XbaI* y *XhoI* (Promega), así como *BstBI* y *NdeI* (NEB), y enzimas de restricción que generaban extremos romos tales como *EcoRV* y *SmaI* (Promega), con sus correspondientes buffers y las condiciones óptimas indicadas por los fabricantes. Las digestiones se realizaron preferentemente durante una hora a 37°C, a excepción de aquellos clonajes más difíciles que su periodo de incubación se prolongó de dos a cuatro horas. Para facilitar el clonaje, se procedió posteriormente a tratar los plásmidos digeridos con la enzima fosfatasa alcalina (*Alkaline Phosphatase, shrimp*, Roche) defosforilando los 5' fosfatos. Con ello se evita la religación del plásmido original. Para ello se trataron las muestras 10 minutos a 37°C con 0.5 U de la enzima y el buffer correspondiente (o suplementando el buffer de la enzima de restricción empleada anteriormente hasta igualar ambos). En los casos de posteriores ligaciones con extremos romos, la defosforilación se amplió hasta a una hora de tratamiento.

Las muestras digeridas se precipitaron y purificaron posteriormente mediante el protocolo de extracción con fenol-cloroformo. Se completaron hasta un volumen de 50 µl con agua libre de nucleasas y se añadieron 1/10 volumen de acetato potásico 3M (pH 4.8-5.2) y un volumen de fenol-cloroformo (Amresco®) para precipitar las proteínas. Se centrifugaron las muestras dos minutos a 13.000 r.p.m. y se recogieron las fases acuosas superiores. Posteriormente, un volumen de Isopropanol (Merck) fue añadido a cada muestra para precipitar los ácidos nucleicos y éstas se mantuvieron diez minutos a -20°C. Se centrifugaron y lavaron las muestras con 150 µl etanol al 70% frío y, por último, se resuspendieron en el volumen deseado.

### **3.3.2.6. Ligación**

Las ligaciones se realizaron en un volumen final de 20 µl con la enzima DNA ligasa del fago T4 (Promega) y su tampón correspondiente a una concentración final de 30 mM Tris-HCl pH 7.8, 10 mM MgCl<sub>2</sub>, 10 mM DTT y 1 mM ATP. Para las ligaciones se emplearon 50 ng de plásmido provenientes de un plásmido de interés digerido o pGEM-T Easy Vector (Promega) y una cantidad variable de inserto. Dicha cantidad se determinó calculando un ratio molar vector: inserto de 1:3 o 1:9, que depende del tamaño tanto del plásmido como del fragmento a insertar. Las ligaciones se realizaron incubando la mezcla de ligación toda la noche (*over night O/N*) a 4°C para mejorar el rendimiento de la ligasa (según recomendaciones del fabricante) o incubando durante una hora a 37°C.



### 3.3.2.7. Cultivos bacterianos

#### A) Mantenimiento de los cultivos bacterianos

Para el crecimiento y sobre-expresión de los plásmidos pGEM-T y pcDNA3.1 se emplearon diferentes cepas de bacterias *Escherichia coli* (*E.coli*) competentes, tales como DH5 $\alpha$  y XL1 Blue MRF'. Estas cepas bacterianas se caracterizan por ser versátiles y adecuadas para el clonaje de plásmidos, permitiendo la selección por color (blanco/azul) en placas de medio con X-Gal e IPTG. Las bacterias competentes fueron suministradas por el Servicio de Fermentación del CBMSO. Para su crecimiento se empleó el medio *Lysogeny broth* (LB), compuesto por 10 g/L de triptona, 5 g/L de extracto de levadura y 10 g/L de cloruro sódico. Se utilizó tanto en su forma líquida como el medio sólido añadiendo 15 g/L de agar a la mezcla. Los medios fueron esterilizados por calor en autoclave en ciclo de líquidos a 121°C durante 20 minutos. El medio LB-agar fue posteriormente añadido y enfriado en placas Petri. La selección por color se realizó con X-Gal (0.6 mg/placa) e IPTG (0.96 mg/placa). Ambos medios fueron suministrados por el Servicio de Cultivos del CBMSO. Para la sobreexpresión de ambos plásmidos, tanto el medio LB como las placas LB-agar fueron preparadas con ampicilina (50  $\mu$ g/mL) acorde a la resistencia que presentan estos plásmidos a dicho antibiótico.

Para el crecimiento y sobre-expresión de los bÁcmidos recombinantes, necesarios para el sistema de sobreexpresión de proteínas por baculovirus, se emplearon las bacterias competentes *MAX Efficiency DH10bac<sup>TM</sup>* (Invitrogen). Esta cepa de bacterias *E.coli* presenta un bÁcmido (bMON14272) y un plÁsmido ayudante (pMON7142) que pueden recombinar con el plÁsmido donante pFastBac<sup>TM</sup> que se genera con la construcción deseada, de forma que se obtiene un bÁcmido recombinante con el gen de interés. Las bacterias competentes *DH10bac<sup>TM</sup>* presentan también selección por color (blanco/azul) en placas de medio con X-Gal e IPTG. Para la sobre-expresión de dicho plÁsmido se empleó medio LB, tanto en medio líquido como en medio sólido, preparado con los antibióticos ampicilina (50  $\mu$ g/mL), kanamicina (50  $\mu$ g/mL), tetraciclina (10  $\mu$ g/mL) y gentamicina (7  $\mu$ g/mL).

Los clones bacterianos de interés fueron conservados a -70°C en medio LB líquido con los antibióticos correspondientes a cada cepa bacteriana y un 15% de glicerol estéril.

### B) Chequeo de los clones de bacterias transformadas

La selección de los clones de bacterias transformadas positivas fue realizada mediante PCR de colonia (ver Apartado 3.3.2.2.B). Para ello, se picaron colonias individuales de las placas LB-agar que habíamos dejado creciendo O/N tras sembrar la transformación. Las colonias individuales fueron sembradas y extendidas en una zona anotada de una nueva placa LB-agar con los antibióticos correspondientes, así como se añadieron como molde en la mezcla de PCR. La siembra y extensión individual de cada una de las colonias se realizó para poder cultivarla de nuevo en medio LB líquido en caso de que fuera positiva.

### 3.3.2.8. Transformación

Los plásmidos de interés fueron transformados mediante el método de choque térmico a 42°C. Para transformar los plásmidos provenientes de una ligación se incubaron 100 µl de bacterias competentes y 10 µl de la ligación durante 50 segundos a 42°C. Posteriormente se atemperó la mezcla de unos segundos a 20 minutos (dependiendo de la dificultad de la transformación) en hielo. Para la expresión de los genes de resistencia a antibióticos que las bacterias transformadas presentan se añadieron 700 µl de medio LB e incubó a 37°C con agitación intensa durante 25 minutos. Posteriormente la mezcla se sembró en placas de LB con el antibiótico de selección adecuado, así como X-Gal e IPTG en caso de un plásmido transformado con selección por color (ver Apartado 3.3.2.7.A.) y se dejó crecer a 37°C durante toda la noche. Para retransformar plásmidos previamente purificados se transformaron las bacterias competentes con únicamente 25 ng del plásmido de interés, se sometieron a un choque térmico a 42°C y directamente se incubaron a 37°C con medio LB y agitación intensa durante 30 minutos.

Las transformaciones de las bacterias DH10bac<sup>TM</sup> (Invitrogen) con los plásmidos pAL7 para la generación de los bÁcmidos recombinantes se realizaron con ligeras variaciones al protocolo anterior. Dichos plásmidos pAL7 provienen del plásmido pFastBac (Life Technologies<sup>TM</sup>) modificados para contener el péptido señal de la proteína melitina de abeja en la región 5' y los epítotos V5 e 6xHis en el C-terminal de la futura proteína recombinante. En este caso se incubaron 100 µl de bacterias competentes DH10bac<sup>TM</sup> y 10 ng del plásmido pAL7 durante 30 minutos en hielo previo al choque térmico a 42°C. Posteriormente se mantuvieron las mezclas durante dos minutos en hielo, se añadieron 900



μl de medio LB y se incubaron a 37°C con agitación intensa durante 4 horas para la expresión de los diferentes genes de resistencia presentes en el bÁcrido recombinante. Posteriormente la mezcla se sembró en placas de LB con los antibióticos de selección kanamicina, tetraciclina y gentamicina (ver Apartado 3.3.2.7.A.) y se dejó crecer a 37°C durante uno o dos días. Los plÁsmidos y bÁcridos recombinantes se chequearon por PCR de colonia (ver Apartado 3.3.2.2.B).

### 3.3.2.9. Purificación de plÁsmidos y bÁcridos. Secuenciación

Las colonias transformadas se chequearon por PCR de colonia (ver Apartados 3.3.2.7.B. y 3.3.2.2.B) y aquellas que fueron positivas se crecieron en cultivos de 2-3 mL de LB líquido con el/los antibiótico/s (correspondientes a su resistencia) O/N a 37°C con agitación. Los plÁsmidos se purificaron con el kit *Wizard® Plus SV Minipreps* (Promega), siguiendo las indicaciones del fabricante y se eluyeron en 50 μl de agua libre de nucleasas. Los bÁcridos recombinantes se purificaron siguiendo el protocolo indicado por los fabricantes del sistema *Bac-to-Bac® Baculovirus Expression System* (Invitrogen). Los plÁsmidos purificados se secuenciaron en el servicio de Secuenciación del Instituto de Investigaciones Biomédicas (IIB) utilizando cebadores localizados en los plÁsmidos correspondientes (Tabla 12)

	Cebadores	
pGEM-T	T7	5'-TAATACGACTCACTATAGGG-3'
	Sp6	5'-GATTTAGGTGACACTATAG-3')
pcDNA3.1Zeo(-)	pZ_5MCS_F	5'-CACCAAAATCAACGGGACTT-3'
	pZ_Ori_R	5'-CCACTACGTGAACCATCACC-3'
	pAL7_MCS_F	5'-ATGAAGTGGTTCGCATCCTC-3'
	pAL7_MCS_R	5'-CAAACCACAACACTAGAATGCAGTG-3'

**Tabla 12. Lista de cebadores empleados en secuenciación según el plÁsmido de origen.** Los cebadores directo y reverso están identificados en el nombre del cebador como F y R, respectivamente.

Los plÁsmidos positivos y secuenciados fueron crecidos nuevamente para obtener mayor cantidad, concentración y pureza. Para ello, se picaron las colonias de la placa de LB-agar correspondiente (Apartado 3.3.2.7.B.) y se inocularon en 5 ml de medio LB líquido con el antibiótico específico (ver Apartado 3.3.2.7.A.) durante aproximadamente 8 horas a 37°C con agitación (alrededor de 300 r.p.m.). Posteriormente, los 5 ml de medio se inocularon en 200 ml de medio LB y se crecieron a 37°C con agitación O/N. Una vez centrifugado el cultivo, se procedió a purificar el plÁsmido recombinante con QIAGEN Plasmid Maxi Kit (Qiagen) eluyendo en 500 μl de agua libre de nucleasas. Se siguieron las

indicaciones del fabricante. En todos los casos, la concentración del DNA se midió por espectrofotometría en un NanoDrop ND-1000 (Thermo Scientific) del Servicio de Genómica y Secuenciación Masiva del CBMSO.

### **3.3.2.10. Análisis de las variantes de *splicing***

La abundancia relativa de cada variante de *splicing* se midió como resultado de la intensidad de cada banda única en el gel usando el programa ImageJ. En el caso de las bandas de 1-1.5 kbp, la abundancia relativa de cada una de las variantes de *splicing* se midió comparando el número de clones positivos secuenciados correspondientes a esta variante en relación con el número de clones totales analizados (48 colonias/clones totales por cada tejido y cada especie), de tal forma que inferior al 20% del total de las clones fue considerado como baja expresión, entre un 20% y un 40% fue considerado como expresión moderada, entre un 40 y un 50% fue considerado como moderadamente alta expresión, y finalmente, una presencia mayor al 50% del total de clones fue considerada como un nivel de expresión alta.

### **3.3.2.11. Clonajes**

Se emplearon los vectores comerciales *pGEM®-T Easy Vector* (Promega) y *pcDNA<sup>TM</sup>3.1/Zeo(-)* (Invitrogen) para los diferentes clonajes y construcción de plásmidos recombinantes que se han utilizado en este trabajo. El vector *pGEM®-T Easy Vector* fue empleado para el clonaje directo a partir de PCR de las diferentes variantes de *splicing* gracias a los extremos en timinas presentes en el plásmido que pueden ligar con los extremos en adeninas que generan las polimerasas (tales como la polimerasa de la *GoTaq®* Green Master Mix empleada en este tipo de clonajes, ver Apartado 3.3.2.2.C). En cambio, el vector *pcDNA<sup>TM</sup>3.1/Zeo(-)* fue empleado para el clonaje de diferentes isoformas para su sobre-expresión en cultivos celulares. En este tipo de clonajes se emplearon polimerasas de alta fidelidad (ver Apartado 3.3.2.2.A.) que no dejan libre adeninas en sus extremos y, por lo tanto, se clonó mediante digestión con distintas enzimas y posterior ligación de los insertos y vectores. Se confirmaron las secuencias de todos los clonajes por secuenciación (ver Apartado 3.3.2.9.). Se realizaron diferentes construcciones como se detallan a continuación:

El **plásmido pZCM** corresponde al plásmido pcDNA<sup>TM</sup>3.1/Zeo(-) con la secuencia del epítipo cMyc (EQKLISEED), obtenido a partir de la hibridación de los cebadores cMyc\_F y cMyc\_R (Tabla 7), clonado a través de las dianas de restricción *NotI* y *EcoRV*, generando una diana de restricción interna *AgeI* inmediatamente antes de una señal de terminación de la traducción. Esta diana de restricción interna se empleó para el clonaje de isoformas que necesitaban tener el epítipo en el extremo N-terminal de la proteína recombinante y, por otro lado, se mantendrá en la secuencia cuando sea necesario mantener el epítipo en el extremo C-terminal. Además, se añadió la secuencia codificante del péptido señal de la proteína CD33 humana en las dianas de restricción *XhoI* y *NotI*, mediante amplificación con la pareja de cebadores HsCD33F\_*XhoI* y HsCD33R\_*NotI* (Tabla 8) del plásmido pZC-V.

A partir de este plásmido se clonaron las secuencias codificantes de las isoformas de RAGE Tv1 y Tv2, generando proteínas recombinantes con el epítipo cMyc tanto en el extremo N-terminal como en el extremo C-terminal. Para los clonajes con cMyc en el extremo N-terminal se clonaron las secuencias codificantes en el 3' del epítipo en las dianas de restricción *AgeI* y *EcoRI*, empleando la pareja de cebadores HsAGERe2F\_*AgeI* y HsAGERe11R\_*EcoRI* (Tabla 8). Los plásmidos finales generados fueron: **pZCM-Tv1** y **pZCM-Tv2** (Figura 23), amablemente cedidos por Alberto D. López de nuestro grupo de investigación. Para los clonajes con el epítipo cMyc en el extremo C-terminal las secuencias codificantes fueron clonadas en el extremo 5' del epítipo con la diana de restricción *NotI*. Para ello se empleó la pareja de cebadores HsAGERe2F\_*NotI* y HsAGERe11pR\_*NotI* (Tabla 8), que mantienen la fase de lectura hasta la señal de terminación de la traducción presente al final del epítipo. Los productos fueron chequeados por secuenciación para comprobar su adecuada orientación. Los plásmidos finales generados son: **pZC-Tv1-M** y **pZC-Tv2-M** (Figura 23a).

El **plásmido pZCV** fue amablemente cedido por el Dr. Alberto Rastrojo de nuestro grupo. Fue diseñado a partir del plásmido pcDNA<sup>TM</sup>3.1/Zeo(-), al que fueron añadidas la secuencia codificante del péptido señal de la proteína CD33 humana en las dianas de restricción *NheI* y *XbaI* y la secuencia codificante del epítipo V5 (GKPIPPLLGLDST) en las dianas de restricción *XbaI* y *XhoI*. Partiendo de este plásmido se clonaron las secuencias codificantes de dos isoformas de RAGE (Tv1-RAGE y Tv2-RAGE) en las dianas de restricción *XhoI* y *EcoRI*, de tal forma que la proteína recombinante presentara el epítipo

V5 en su extremo N-terminal. Para ello se emplearon la pareja de cebadores HsAGERe2F\_*XhoI* y HsAGERe11R\_*EcoRI* (Tabla 8). Los plásmidos finales generados se denominaron: **pZCV-Tv1** y **pZCV-Tv2** (Figura 23a).

El **plásmido pZV** corresponde al plásmido pcDNA<sup>TM</sup>3.1/Zeo(-) con la secuencia del epítipo V5, clonada a partir de los cebadores V5\_F y V5\_R hibridados (Tabla 7), y clonados mediante las dianas de restricción *NotI* y *EcoRV*. A partir de este plásmido se clono la variante de *splicing* N-truncada de *RAGE* (Tv10-RAGE). Para ello se emplearon la pareja de cebadores HsAGERe1F\_*XbaI* y HsAGERe11pR\_*NotI*, que contenían la secuencia de las dianas de restricción *XbaI* y *NotI*, respectivamente. El cebador reverso mantiene la fase de lectura abierta hasta la señal de Stop presente al final del epítipo V5. El plásmido final generado se denominó: **pZ-Tv10-V** (Figura 23a).

A partir del plásmido pZV también se clonaron las diferentes secuencias codificantes de las quimeras, incluyendo el péptido señal endógeno de cada una de ellas para su expresión y el estudio de su localización, y presentando el epítipo en el N-terminal de las proteínas recombinantes resultantes. Para ello había que clonar el péptido señal de cada quimera delante del péptido V5. Las secuencias codificantes del péptido señal de G6F y de la quimera larga (exón 1 de G6F, a partir de la pareja de cebadores HsSpG6Fe1F\_*XhoI* y HsSpG6Fe1R\_*NotI*, Tabla 8), así como del péptido señal de la quimera corta (localizado en el exón 4 de G6F, a partir de la pareja de cebadores HsSpG6Fe4F\_*XhoI* y HsSpG6Fe4R\_*NotI*, Tabla 8) y del péptido señal de LY6G6D (exón 1 de LY6G6D, a partir de la pareja de cebadores SignalP\_LY6G6D\_F y SignalP\_LY6G6D\_R, Tabla 7) fueron clonados en las dianas de restricción *XhoI* y *NotI*. Posteriormente, el resto de la secuencia codificante fue clonado en el 3' del epítipo en las dianas de restricción *AgeI* y *EcoRI*, de tal forma que la proteína recombinante presentará el epítipo V5 en su N-terminal después de ser procesado el péptido señal. Para la amplificación de la secuencia codificante de G6F se emplearon los cebadores HsG6Fe1F\_*AgeI* y HsG6Fe6R\_*EcoRI*, para la amplificación de la quimera larga se emplearon los cebadores HsG6Fe1F\_*AgeI* y HsLy6G6De3R\_*EcoRI*, para la quimera corta se emplearon los cebadores HsG6Fe4F\_*AgeI* y HsLy6G6De3R\_*EcoRI*, y para la amplificación de la secuencia codificante LY6G6D se emplearon los cebadores HsLy6G6De1F\_*AgeI* y HsLy6G6De3R\_*EcoRI* (Tabla 8). Los plásmidos finales generados fueron: **pZSpV5-G6F**, **pZSpV5-lgch**, **pZSpV5-shch** y **pZSpV5-LY6G6D** (Figura 24)

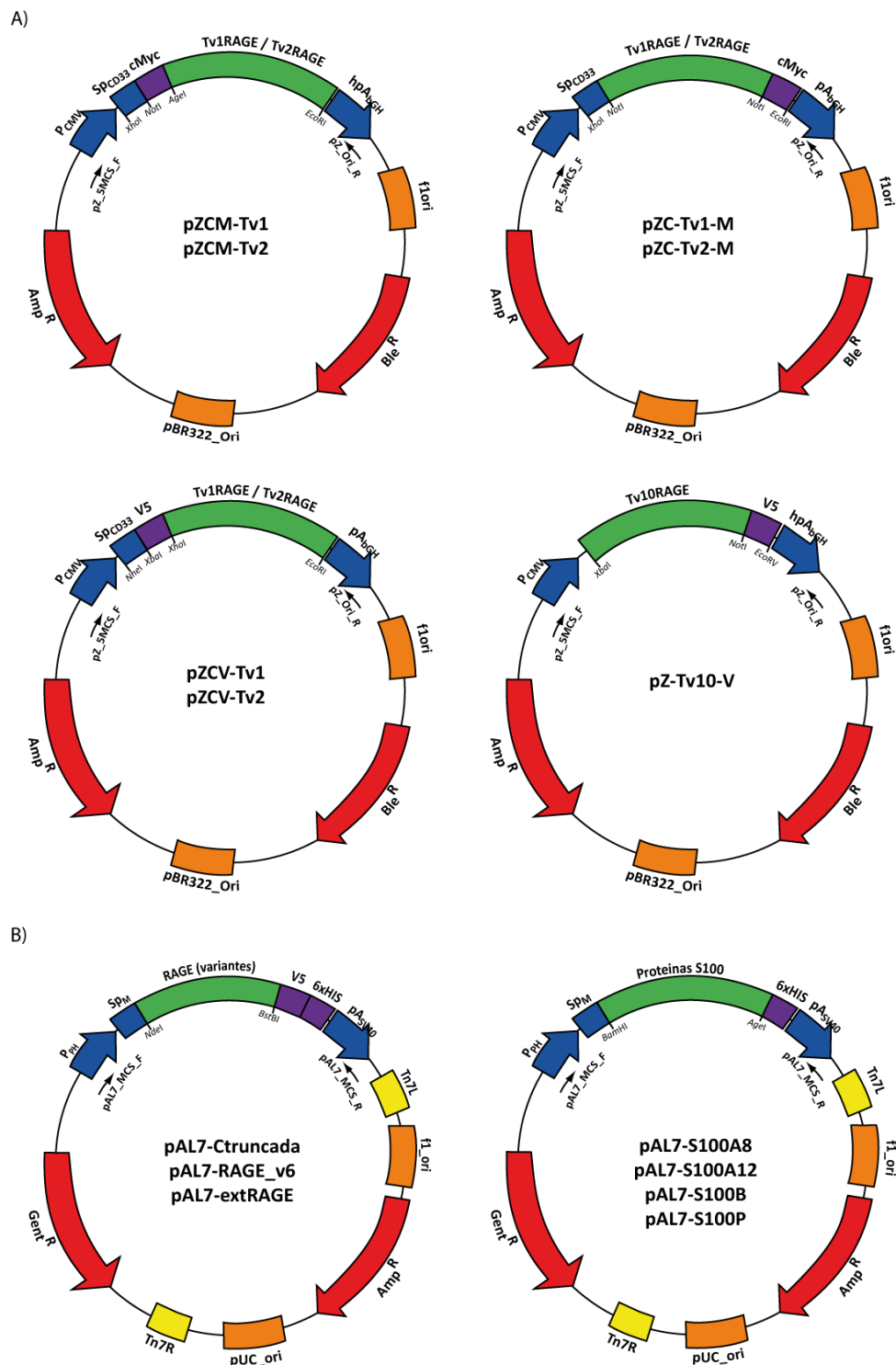


Figura 23. Construcciones de los plásmidos recombinantes relacionados con la proteína RAGE o sus ligandos que se han generado y utilizado en este trabajo. Construcciones empleadas para la sobre-expresión de RAGE en células humanas en cultivo (a) y para la sobre-expresión mediante baculovirus en células de insecto (b).

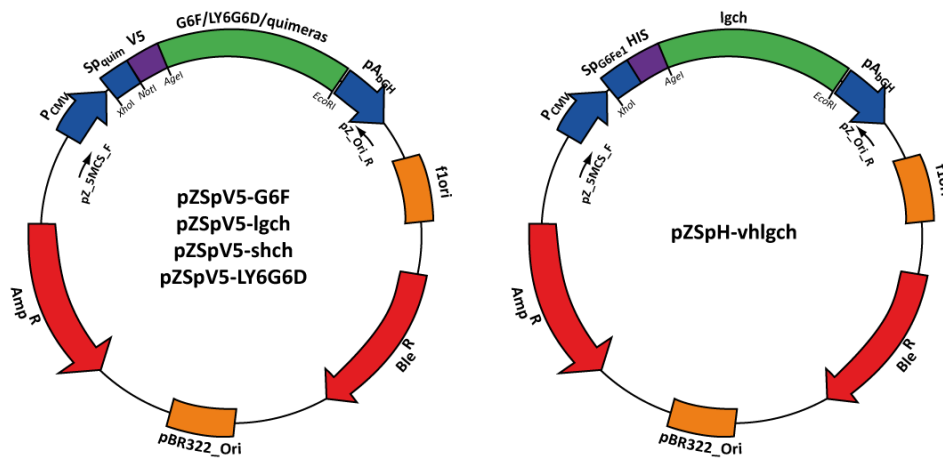


Figura 24. Construcciones de los plásmidos recombinantes relacionados con las proteínas quiméricas que se han generado y utilizado en este trabajo.

El **plásmido pZSpH** corresponde al plásmido pcDNA<sup>TM</sup>3.1/Zeo(-) con la secuencia codificante del péptido señal de la quimera larga (localizado en el exón 1 de G6F) y la secuencia del epítipo 6xHis (HHHHHH) clonado inmediatamente después del péptido señal a partir de los cebadores HsSpG6Fe1F\_ *Xho*I y Hs\_His\_SpG6Fe1R\_ *Age*I (Tabla 8), en las dianas de restricción *Xho*I y *Age*I. A partir de este plásmido se generó el **plásmido pZSpH-lgch** (Figura 24), clonando la secuencia de la quimera larga con las dianas de restricción *Age*I y *Eco*RI, de tal forma que la proteína recombinante presentara el epítipo 6xHis en su N-terminal una vez procesado el péptido señal. Para ello se empleó la pareja de cebadores HsG6Fe1F\_ *Age*I y HsLy6G6De3R\_ *Eco*RI (Tabla 8).

El **plásmido pZ-EGFP** fue amablemente cedido por el Dr. Alberto Rastrojo de nuestro laboratorio. Fue también diseñado a partir del plásmido pcDNA<sup>TM</sup>3.1/Zeo(-), al que se le incorporó la secuencia codificante de la proteína EGFP.

El **plásmido pAL7** fue amablemente cedido por el laboratorio del Dr. Antonio Alcamí (CBMSO). Este plásmido proviene del plásmido pFastBac (Life Technologies<sup>TM</sup>) modificado para contener el péptido señal de la melitina (*Apis mellifera*) en la región 5' y las secuencias codificantes de los epítipos V5 y 6xHis en el C-terminal de la futura proteína recombinante. A partir de este plásmido pAL7 se clonaron las diferentes proteínas que posteriormente fueron sobre-expresadas y purificadas por el método de baculovirus. La secuencia codificante de las diferentes isoformas de RAGE fue clonada en las dianas de restricción *Nde*I y *Bst*BI. Para ello se emplearon los cebadores HsRAGEe2F\_ *Nde*I y

HsRAGEi9pR\_ *Bst*BI para la amplificación y clonaje de las isoformas C-truncada y RAGE\_v6, y los cebadores HsRAGEe2F\_ *Nde*I y HsRAGEe10pR\_ *Bst*BI para la isoforma extRAGE (Tabla 8). Los plásmidos finales generados fueron: **pAL7-Ctruncada**, **pAL7-RAGE\_v6** y **pAL7-extRAGE** (Figura 23b). La secuencia codificante de las diferentes proteínas S100 fue clonada en las dianas de restricción *Bam*HI y *Age*I, para lo que se emplearon las parejas de cebadores listadas en la Tabla 8. Los plásmidos finales generados fueron: **pAL7-S100A8**, **pAL7-S100A12**, **pAL7-S100B** y **pAL7-S100P** (Figura 23b).

### 3.3.3. Cultivos celulares

#### 3.3.3.1. Mantenimiento de los cultivos celulares

Las líneas celulares de tipo adherente como HeLa (células epiteliales obtenidas de un adenocarcinoma de cuello de útero) y HEK293T (células embrionarias de riñón) fueron crecidas en *Dulbecco's Modified Eagles Medium* (DMEM) y la línea celular en suspensión K562 (granulocitos indiferenciados derivados de una leucemia mieloide crónica) en *Roswell Park Memorial Institute medium* (RPMI). Ambos medios contenían suero fetal bovino (*Foetal Bovine Serum*, FBS) (10% v/v) descomplementado a 56°C durante 30 minutos, 2 mM de L-glutamina y una mezcla de antibióticos de Estreptomicina Sulfato y Penicilina (1% v/v). Todas las líneas celulares se mantuvieron en un incubador a 37°C, en una atmósfera con un 5% de CO<sub>2</sub>, y humedad relativa del 95%. La línea celular High-Five<sup>TM</sup> (células de ovario del insecto *Trichoplusia ni*) se creció de forma adherente para la expansión del virus empleando el medio TC100 suplementado con 10% de FBS (v/v), y en suspensión (agitación 180 r.p.m.) para la producción de proteínas recombinantes mediante el medio ExpressFive<sup>®</sup> STM (Invitrogen). Ambos medios se suplementaron con antibióticos (75 U/ml de estreptomicina, 75 µg/ml de penicilina G, 10 µg/ml de tetraciclina (Sigma) y 7 µg/ml de gentamicina), 2 mM de L-glutamina, 45 µM de aminoácidos no esenciales, así como antimicóticos y fungizina. Las células High-Five<sup>TM</sup> se crecieron a 28°C en una atmósfera con un 5% de CO<sub>2</sub>, y humedad relativa del 95%.

Para la criopreservación de las líneas celulares, las células se resuspendieron en 1 mL de FBS conteniendo 10% de dimetilsulfóxido (DMSO), y se pasaron a criotubos estériles. Para asegurar una congelación lenta necesaria para la supervivencia de las células, los criotubos se guardaron en un contenedor de congelamiento celular (Nalgene<sup>TM</sup>) con isopropanol a -80 °C durante 2 días, y posteriormente se trasladaron y conservaron

definitivamente en un contenedor de nitrógeno líquido (-196 °C). La descongelación se realizó a 37°C en un baño con una ligera agitación. Se depositaron las células en 10 mL del medio de cultivo y se centrifugaron durante 5 minutos a 1200 r.p.m. con el fin de eliminar el DMSO existente en el medio de congelación. Las células se resuspendieron en medio de cultivo fresco y, a continuación, se sembraron o se mantuvieron en suspensión a la densidad requerida sobre la superficie de cultivo deseada.

Se realizaron los pases necesarios en los medios correspondientes al tipo celular hasta alcanzar la confluencia y el crecimiento normal de cada línea celular para realizar los experimentos correspondientes. Las células adherentes HeLa y HEK293T se lavaron con EDTA/Tripsina 0.25% y se levantaron al incubarse con 2 ml de Tripsina a 37°C durante cinco minutos. La tripsinización se paró al añadir medio DMEM completo y las células se centrifugaron 5 min a 1000-1500 r.p.m. Tras eliminar el sobrenadante se resuspendieron en medio DMEM completo fresco y se sembraron a la densidad necesaria. Las células en suspensión se mantuvieron entre  $0.5-5 \times 10^5$  células/ml, diluyéndose en medio completo fresco cada 2 días. Para el conteo de las células se empleó un hemocitómetro o cámara de Neubauer utilizando el método de tinción por exclusión. Dicho método consiste en realizar una tinción con azul tripán con el fin de poder determinar la viabilidad de las células y únicamente contar las células vivas que no hayan incorporado dicho colorante al tener la membrana celular intacta.

### **3.3.3.2. Transfección transitoria**

Para la transfección transitoria de las células adherentes HEK293T o HeLa se sembraron entre  $1-2 \times 10^5$  células en placas de 24 pocillos (P24), dependiendo del tiempo de incubación. En cada pocillo se sembraron. Se transfectaron con 0.5 µg de DNA utilizando *TransIT<sup>®</sup>-LT1 Transfection* (Mirus) o *Lipofectamine<sup>®</sup> 2000 reagent* (Invitrogen - Life Technologies), siguiendo las indicaciones del fabricante. En algunas ocasiones también se empleó el método de fosfato cálcico para transfectar células adherentes. Para ello se añadieron 0.5 µg de DNA sobre una solución con cloruro cálcico (CaCl<sub>2</sub> 250 mM) y posteriormente se añadió esta mezcla sobre una solución salina HEPES (HEPES 250 mM, Na<sub>2</sub>HPO<sub>4</sub> 1,5 mM, NaCl 140 mM, pH = 7,05) burbujeando. Tras una incubación de 45 segundos se añadió la mezcla sobre el medio de las células adherentes previamente cambiado a medio sin suero, y las placas se incubaron a 37°C durante 24 horas.



Las células en suspensión K562 fueron contadas, separadas entre 1 y  $15 \times 10^6$  células y centrifugadas. Las células se resuspendieron en 400  $\mu$ l de medio RPMI suplementada con un 10% de FBS y 2 mM de L-glutamina, sin antibióticos y con 50  $\mu$ g del plásmido correspondiente. Se incubaron durante 15 minutos a temperatura ambiente (RT) y se electroporaron con un choque a 250 V y 950  $\mu$ F. Se incubaron de nuevo 15 minutos a RT y, posteriormente se resuspendieron en 20 ml de medio RPMI complementado pero sin antibióticos. Se mantuvieron las células en cultivo durante 48-72 horas, según el experimento.

Para la transfección de células High-Five<sup>TM</sup> con bácmidos recombinantes (ver Apartado 3.3.4.5.), se sembraron dichas células a una densidad de  $4 \times 10^5$  células/pocillo en placas de seis pocillos (P6). Se añadieron 2  $\mu$ g del correspondiente bácmido y 6  $\mu$ l de Cellfectin<sup>®</sup> (Invitrogen) en 200  $\mu$ l de medio TC100 completo sin FBS. Esta mezcla se incubó durante 45 minutos a RT y, posteriormente, se añadió sobre las células con medio completo sin FBS. Las células se incubaron con la mezcla alrededor de cinco horas y, posteriormente, se lavaron las células y se mantuvieron con medio completo suplementado con suero. Se incubaron durante 72 horas y se recogió el sobrenadante de las células conteniendo el baculovirus recombinante de interés, denominado como stock de virus 0 (P0).

### ***3.3.3.3. Infección por baculovirus***

Para amplificar el stock de baculovirus se realizaron infecciones en cantidades crecientes de células High-Five<sup>TM</sup> adherentes. Primeramente se infectaron células sembradas al 50% de confluencia en placas p100. Para ello se emplearon 30  $\mu$ L de cada virus (P0) en 3 mL de medio TC100 sin suero durante 2 horas a 27°C. Posteriormente se eliminó el medio, se reemplazó con medio TC100 suplementado con suero y se incubaron las células durante 72 horas (o hasta ver signos de infección viral). El medio de las células infectadas se recolectó y se centrifugó durante 5 minutos a 900 x g, conservándose el sobrenadante que contenía los baculovirus recombinantes o stock de virus 1 (P1). La siguiente infección se realizó sobre una placa p150 de células High-Five<sup>TM</sup> sembradas al 50% de confluencia. Se infectaron las placas con 100  $\mu$ L de cada virus (P1) en 10 mL de medio TC100 sin suero durante 2 horas a 27°C. Nuevamente, se procedió como se ha descrito anteriormente para obtener los virus P1 y se conservó el sobrenadante

denominándose stock de virus 2 (P2). Los stock de virus deben estar en todo momento protegidos de la luz y para largo periodo de almacenaje es recomendable conservarlos a -70°C.

Una vez que se ha chequeado que se está amplificando el virus correspondiente (ver Apartado 3.3.3.4.) se procede a infectar células High-Five™ en suspensión para la purificación de la proteína recombinante. Para ello, infectamos 200 millones de células High-Five™ en suspensión a una densidad de 1.5 millones/mL con 10 mL del stock de virus 2. Se mantiene la infección durante 72 horas y finalmente se recoge el medio para realizar la purificación de proteínas (ver Apartado 3.3.4.5.).

### ***3.3.3.4. Chequeo de las transfecciones e infecciones***

Las transfecciones de las diferentes construcciones de plásmidos e infecciones fueron chequeadas por diferentes técnicas dependiendo del tipo de proteína recombinante sobreexpresada. Cuando se requería un chequeo rápido de la transfección se incluyó un pocillo transfectado con el plásmido pcDNA3.1-EGFP (ver Apartado 3.3.2.11.), de tal forma que su sobreexpresión se visualizó con el microscopio invertido Axiovert 135 (Zeiss) para observación de placas de cultivo que se encuentra disponible en el servicio de Microscopía Óptica y Confocal del CBMSO. Cuando las variantes que se estaban sobreexpresando eran proteínas de membrana y presentaban el epítipo (cMyc, V5, 6xHIS...) en la región extracelular de la proteína, la sobreexpresión fue chequeada mediante citometría de flujo. En cambio, cuando esta técnica no era posible, la sobreexpresión y los stock de virus fueron controlados mediante un Western blot específico para cada proteína.

## **3.3.4. Técnicas de proteínas**

### ***3.3.4.1. Preparación de las muestras***

Para la preparación de las muestras de proteínas provenientes de extractos celulares o proteínas purificadas se lavaron primero las células dos veces con una solución de PBS tras lo cual se añadió 50-100 µl de tampón reductor Laemmli-SDS (60 mM TrisHCl pH 6.8, 2% SDS, 5% BME, 10% glicerol, 0.01% azul de bromofenol). Para los estudios en condiciones no reductoras se empleó el mismo tampón sin BME pero con 0.2 mM de

Iodoacetamida (Sigma). Las muestras se hirvieron durante cinco minutos a 100°C e, inmediatamente después se cargaron en los geles de acrilamida.

### 3.3.4.2. Electroforesis en geles de acrilamida

Los extractos celulares o proteínas purificadas y el marcador de peso molecular *Precision Plus Protein™ Dual Color Standards* (Bio-Rad) se cargaron en geles de poliacrilamida (SDS-PAGE) preparados como se detalla en la Tabla 13. Se empleó un porcentaje del gel separador del 10 y del 12% dependiendo del tamaño de las proteínas que quisiéramos separar (mayor y menor de 100 kDa, respectivamente). Las electroforesis se realizaron a 100 voltios (o 30mA de amperaje constante) durante 1-2 horas con tampón Tris-Glicina (25 mM Tris-Base, 250 mM Glicina, 0.1% SDS).

	Gel concentrador		Gel separador			
	5%		10%		12%	
	[ ] <sub>f</sub>	Volumen	[ ] <sub>f</sub>	Volumen	[ ] <sub>f</sub>	Volumen
Agua destilada	-	2.92 mL	-	3.55 mL	-	3.05 mL
40% Solución Acrilamida/Bis, 37.5:1 (Bio-Rad)	5%	0.5 mL	10 %	2.5 mL	12 %	3 mL
Tris 1M pH 6.8	125 mM	0.5 mL				
Tris 1M pH 8.8			375 mM	3.75 mL	375 mM	3.75 mL
SDS 10%	0.1%	40 µL	0.1%	100 µL	0.1%	100 µL
APS 10%	0.1%	40 µL	0.1%	100 µL	0.1%	100 µL
TEMED	-	4 µL	-	4 µL	-	4 µL
Azul de bromofenol	-	4 µL				
Volumen final =		4 mL		10 mL		10 mL

**Tabla 13. Componentes y concentraciones utilizadas en la preparación de los geles de poliacrilamida.**

### 3.3.4.3. Western blot

Tras correr la electroforesis, los geles se transfirieron a membranas de nitrocelulosa Protran® (Whatman) mediante un sistema de transferencia húmeda en el *Mini Trans-Blot® Electrophoretic Transfer Cell* (Bio-Rad). Las transferencias se realizaron durante una hora a 100 voltios constantes en el tampón de transferencia Tris-Glicina (25 mM Tris-Base, 250 mM Glicina, 0.1% SDS) con un 20% de metanol. Cuando las muestras cargadas procedían de un extracto celular, las membranas se tiñeron con una solución de Rojo Ponceau (0.1% v/v) y ácido acético glacial (1% v/v) para comprobar la eficiencia de la transferencia. Las membranas se bloquearon en la solución de bloqueo (PBS 1X, 0.05% Tween 20 (Sigma) y 5% de leche en polvo desnatada) durante una hora en agitación. En el caso de los estudios

de fosforilación, la solución de bloqueo fue PBS 1X, 0.05% Tween 20 (Sigma) y 4% de albúmina de suero bovino (BSA), ya que la proteína caseína presente en la leche se encuentra fuertemente fosforilada y, por ello, reduciría la señal del anticuerpo empleado para detectar la fosforilación del resto de las proteínas, incluida nuestra proteína de interés. Posteriormente, las membranas se incubaron con los anticuerpos primarios correspondientes (Tabla 14) en solución de bloqueo durante una hora a RT o O/N a 4°C. Las membranas se lavaron con PBS 1X suplementado con 0.05% Tween 20 durante 20 minutos y, a continuación, se incubaron con el anticuerpo secundario (Tabla 14). Por último, se lavaron las membranas con PBS 1X y se procedieron a revelar. Se emplearon dos métodos diferentes de revelado, en el primero se trataron las membranas con la solución de revelado

Anticuerpos primarios	Huésped	Referencia	Casa comercial	Dilución
Anti V5	Ratón	V8012	Sigma	WB 1:5000, IP 1:100, IF 1:500, FACS 1:5000
Anti 5xHis	Ratón	34660	Qiagen	WB 1:5000, IP 1:10, FACS 1:1000
Anti-c-Myc (9B11)	Ratón	2276	Cell Signaling	WB: 1:2000, FACS 1:1000
Anti-Fosfotirosina (4G10 <sup>®</sup> Platinum)	Ratón	05-1050X	Millipore	WB 1:1000
Anti-RAGE humano	Conejo	GTX102797	GeneTex	WB 1:1000
Anti-RAGE ratón	Conejo	GTX23611	GeneTex	WB 1:800
Anti-manosidasa II	Conejo	AB3712	Chemicon	IF 1:100
Anti-calnexina	Conejo	SPA-860	Stressgen	IF 1:200
Anti-GFP	Ratón	G1546	Sigma	WB 1:2000, IP 1:200.
Anti-GAPDH	Ratón	G8795	Sigma	WB 1:5000
Anti-Tubulina	Ratón	MAB3408	Chemicon <sup>®</sup> International	WB 1:500
Anticuerpos secundarios	Reactividad	Referencia	Casa comercial	Dilución
IgG-Peroxidase	anti-ratón	A5906	Sigma	WB 1:5000
IgG-Peroxidase	anti-conejo	W4011	Promega	WB 1:2500
Li-Cor IRDye <sup>®</sup> 680 (680 nm)	anti- ratón	926-68020	Li-Cor	WB 1:10000
Li-Cor IRDye <sup>®</sup> 800 (800 nm)	anti- conejo	926-32219	Li-Cor	WB 1:50000
Alexa 488	anti- conejo	A-21206	Invitrogen	IF 1:500; FACS 1:500
Alexa 488	anti- ratón	A-21202	Invitrogen	IF 1:500; FACS 1:500
Alexa 555	anti- conejo	A-31572	Invitrogen	IF 1:500; FACS 1:500
Alexa 555	anti- ratón	A-31570	Invitrogen	IF 1:500; FACS 1:500
Alexa 647	anti- conejo	A-31571	Invitrogen	FACS 1:500
Alexa 647	anti- ratón	A-31573	Invitrogen	FACS 1:500

**Tabla 14. Listado de anticuerpos primarios y secundarios empleados en los análisis por Western blot (WB), inmunofluorescencias (IF) y citometría de flujo (FACS).** Se indica la concentración a la que fueron usados en cada tipo de análisis.

*Amersham™ ECL™ Western Blotting Detection Reagent* (GE Healthcare) según las indicaciones del fabricante y se revelaron las membranas en carcasas opacas (Fisher Biotech) y con autorradiografías (Curix RP2 Plus, Agfa). En el segundo método, cuando las membranas fueron incubadas con los anticuerpos secundarios Li-Cor IRDye® se procedió a revelar directamente las membranas en el Odyssey Infrared Imaging System (Li-Cor).

#### **3.3.4.4. Tinción de coomasie**

Una vez terminada la electroforesis, los geles de acrilamida se fijaron en una solución de metanol al 50% y ácido acético al 10% durante 20 minutos en agitación. Posteriormente, se realizó la tinción incubando entre 30 minutos y una hora en agitación con una solución de azul de Coomasie G250 (0.001% v/v), metanol (20% v/v) y ácido acético (10% v/v). Para la destinción del gel se utilizó una solución compuesta por un 7% (v/v) de metanol y un 10% (v/v) de ácido acético glacial durante alrededor de dos horas, incorporando una esponja que absorba el colorante. Todo el procedimiento se realiza a RT.

#### **3.3.4.5. Expresión de proteínas recombinantes mediante baculovirus**

##### **A) Purificación de proteínas recombinantes**

La purificación de proteínas recombinantes se realizó mediante cromatografía de afinidad gracias al epítipo de histidinas que se había incorporado previamente en su secuencia. Básicamente, se recoge el medio con las células y se centrifugan durante 10 minutos a 6000 x g en una centrifuga Beckman Avanti J-26 XP con el rotor JLA-16250. Se traspara el sobrenadante (alrededor de 150 a 200 mL totales por cada infección) a un Amicon Stirred Cells (Merck Millipore) donde se concentra hasta al menos un volumen de 5 mL. Para ello se requiere la utilización de una membrana cuyo tamaño de poro sea al menos 3 veces inferior al tamaño de las proteínas recombinantes a purificar. En nuestro estudio se emplearon tamaños de membrana de 3 kDa para las proteínas S100 y de 5 kDa para las diferentes isoformas de RAGE. Se toma una alícuota del sobrenadante concentrado o *input*. Una vez concentradas las proteínas, el sobrenadante se pasa por una columna PD-10 (GE Healthcare), previamente equilibrada con 25 mL de la solución 10 mM de Imidazole, y se eluye finalmente en 4 mL de esa misma solución 10 mM (Tabla 15). En paralelo se prepara la resina Ni-NTA-Agarose (Qiagen) empleando 1.2 mL de resina (600

$\mu\text{L}$  finales) en una columna de Poly-Prep y se lava con 30 mL de la solución con 10 mM de Imidazole. Posteriormente se incubó el eluido con la resina Ni-NTA-Agarose durante una hora a  $4^\circ\text{C}$  para permitir la unión de las proteínas recombinantes (mediante las His) a la resina. Tras la incubación se toma una alícuota del eluido que corresponderá con las proteínas no unidas a la resina. La resina se lava con 40 mL de la solución con 20 mM de Imidazole y posteriormente con 5 mL de la solución con 40 mM de Imidazole (Tabla 15). Se toman alícuotas de ambas fracciones. Por último se eluyen las proteínas recombinantes unidas a la resina con cuatro fracciones de 600  $\mu\text{L}$  de la solución con 250 mM de Imidazole (Tabla 15), tomándose alícuotas de cada una de las fracciones para posterior chequeo de la purificación.

Soluciones:	[ ] <sub>0</sub>	[ ] <sub>f</sub>	10 mM	20 mM	40 mM	250 mM
Tampón Fosfato*	100 mM	50 mM	30 mL	20 mL	2.5 mL	1.2 mL
NaCl	3 M	0.3 M	6 mL	4 mL	0.5 mL	0.24 mL
Imidazole	4M	---	0.15 mL	0.2 mL	0.05 mL	0.15 mL
H <sub>2</sub> O	---	---	23.85 mL	15.8 mL	1.95 mL	0.81 mL
Volumen final			60 mL	40 mL	5 mL	2.4 mL

**Tabla 15. Soluciones de Imidazole empleadas en la purificación de proteínas recombinantes con epítipo 6xHis mediante el sistema de baculovirus.** \*Tampón fosfato, pH 7.45: 405 mL de  $\text{Na}_2\text{HPO}_4$  0.2 M, 95 mL de  $\text{NaH}_2\text{PO}_4$  0.2 M y 500 mL de  $\text{H}_2\text{O}$ .

Para comprobar si la purificación funciona se cargan geles de acrilamida con las alícuotas obtenidas en cada paso y se tiñen con azul de Coomassie. Se observan las fracciones en las que se encuentra la proteína recombinante y se procede a dializar concentrar la proteína en un volumen menor. Para ello se emplearon columnas de Vivaspin 500 (Sartorius) con un tamaño de poro de 5 kDa para todos los casos (las columnas con un tamaño de poro inferior eran muy frágiles y no soportaban todos los lavados necesarios). Se lavaron las columnas con PBS + 0.05% Tween y centrifugaron a  $13300 \times g$  a  $4^\circ\text{C}$  hasta que quedasen unos 100  $\mu\text{L}$  de volumen residual. Posteriormente se lavaron con tres fracciones de 600  $\mu\text{L}$  de PBS 1X y centrifugaron. Las membranas se equilibraron con 600  $\mu\text{L}$  de la solución con 10 mM de Imidazole y se centrifugaron, posteriormente se añadieron las diferentes fracciones que contenían la proteína recombinante. Se lavaron las columnas que contenían la proteína con la solución de diálisis (20 mM Hepes ajustado a pH 7.4, 150 mM NaCl y 10% glicerol) en tres fracciones de 600  $\mu\text{L}$ , y se recuperaron 75  $\mu\text{L}$  tras el último lavado como la fracción con las proteínas dializadas, que se conservan en alícuotas a  $-20^\circ\text{C}$ . Las proteínas purificadas y dializadas se corrieron en un gel de poli-acrilamida con cantidades conocidas de BSA (Promega), y se tiñeron con azul de Coomassie. Las proteínas

se cuantificaron por densitometría tras escanear los geles y cuantificar todas las bandas con el programa ImageJ.

### B) Análisis MS-MALDI TOF de las proteínas purificadas

Al purificar la proteína S100A12 se observó una doble banda en el tamaño molecular esperado por la proteína de interés purificada, por lo que se procedió a su estudio. Para ello se corrió la muestra de la proteína purificada en un gel al 12% de Acrilamida/Bis, junto con estándares de peso molecular (sin preteñir ni colorear) con cantidades conocidas (1 a 2 µg de cada proteína), amablemente cedidos por el Servicio de Proteómica del CBMSO. Debido a la alta sensibilidad de las técnicas de espectrometría de masas, la polimerización del gel se realizó el día anterior a la carrera, empleando agua MQ y reactivos preparados expresamente para dicho experimento, manipulando con cuidado el extremo para evitar la contaminación de la muestra con queratinas. Posteriormente, se fijaron las proteínas al gel y se tiñeron con una solución de Coomassie (preparada expresamente para evitar la contaminación con queratinas). Primeramente se realizó un análisis del peso molecular de los péptidos/proteínas presentes en el gel mediante MALDI TOF-MS. Para ello se empleó el espectrómetro de masas tipo MALDI-TOF, modelo Autoflex de Bruker, equipado con reflector y sistema para PSD, disponible en el Servicio de Proteómica del CBMSO. Seguidamente las muestras se procesaron mediante desalación y concentración por Zip-Tip (Omix, Oasis), se realizó una digestión trípica y se procedió a su análisis por MALDI-TOF/MS obteniendo el mapa peptídico de ambas bandas. La identificación de las proteínas se realizó mediante el método denominado huella peptídica o “*Peptide Mass Fingerprinting*” donde mediante un análisis bioinformático se comparan las masas peptídicas experimentales con la lista de masas peptídicas calculadas de todas las secuencias de proteínas disponibles en las bases de datos. Estos análisis fueron realizados en el Servicio de Proteómica del CBMSO, servicio perteneciente al Instituto Nacional de Proteómica, ProteoRed.

### ***3.3.4.6. Estudio de modificaciones post-traduccionales***

#### A) Ensayos de deglicosilación

Se estudiaron los estados de glicosilación de las proteínas mediante el empleo de diversas enzimas como PNGase F (NEB) para determinar las N-glicosilaciones, O-

glycosidase (NEB) para determinar las O-glicosilaciones, y Neuraminidase (NEB) para determinar las ramificaciones por ácido siálico. Se emplearon dos pocillos de una placa de 24 pocillos de células trasfectadas con los plásmidos correspondientes a los genes *G6F*, *LY6G6D* y las quimeras. Se realizó el lisado de las células mediante el tampón desnaturalizante de glicoproteínas de NEB, junto con inhibidores de proteasas (Roche). Las muestras fueron tratadas en sus respectivos tampones durante 5 horas a 37°C.

### B) Ensayos de fosforilación

Los ensayos de fosforilación se realizaron en células K562 transfectadas transitoriamente por electroporación con las construcciones correspondientes a *G6F* y la quimera larga. Tras tres días de transfección, las células fueron estimuladas con o sin pervanadato durante quince minutos. El pervanadato fue preparado como se describe en De Vet et al. (2001), para reproducir resultados, con la amable ayuda de D. Antonio J. Montes y Dr. Juan M. Serrador (CBMSO), quienes también cedieron desinteresadamente los reactivos. Básicamente, se lisaron las células con el buffer de lisis Nonidet P40 (50 mM Tris-HCl pH 8.1, 1% NP40, 150 mM NaCl, 1 mM EDTA, e inhibidores de proteasa). Para las células tratadas con pervanadato, el buffer fue suplementado con 1 mM vanadato de sodio y 50 mM fluoruro de sodio. Los lisados fueron depurados por centrifugación y, posteriormente, se sometieron a una inmunoprecipitación (IP) con el anticuerpo anti-V5. Las proteínas se eluyeron con el buffer Laemmli y se analizaron por Western blot con el anticuerpo anti-V5 (Sigma) y el anticuerpo anti-fosfotirosina (Millipore) (Tabla 14), amablemente también cedido por el Dr. Juan M. Serrador.

### C) Estudios de anclaje por GPI a la membrana plasmática

El anclaje por GPI a la membrana plasmática se determinó mediante la acción enzimática de la enzima fosfolipasa C específica de fosfatidilinositol (PI-PLC, *phospholipase C - phosphatidylinositol specific*) (Sigma-Aldrich). Se empleó un pocillo de una placa de 24 pocillos para cada punto, que se transfectaron con los plásmidos correspondientes a los genes *LY6G6D* y la quimera corta. Se lavaron los pocillos con buffer PI-PLC (10 mM Tris-HCl, 144 mM NaCl, pH 7.4) y, posteriormente se añadieron 80 µl de dicho buffer con inhibidores de proteasas y unidades crecientes de enzima (A, sin enzima; B, 0.005U; C, 0.01U; D, 0.02U). Las células se incubaron con el enzima a 30°C durante una hora. Tras la incubación, se lavaron las células con PBS y se recogieron con 100 µl de buffer Laemmli para ser analizados mediante Western blot.



### 3.3.4.7. Ensayos de interacción proteica

#### A) Coinmunoprecipitaciones (CoIP)

Se realizaron ensayos de interacción proteica mediante experimentos de CoIP, así como también se realizaron IP directas para el estudio de la fosforilación de las proteínas. Se emplearon pocillos de placas de 24 pocillos transfectados (o cotransfectados) con las construcciones de interés, y se lisaron en el buffer de lisis IP (1X PBS, 1% NP40, 150 mM NaCl, e inhibidores de proteasas). Los lisados fueron depurados por centrifugación y se procedió a realizar las IPs a 4°C con los anticuerpos anti-V5 (Sigma), anti-His (Qiagen) ó anti-GFP (Sigma) (Tabla 14). Las IP se realizaron con resina *protein A-Sepharose* (Sigma) previamente bloqueadas con 1% BSA, ó con bolas magnéticas *Dynabeads® Protein G* (Life Technologies) siguiendo las indicaciones del fabricante. Posteriormente se lavaron las muestras con buffer de lisis IP 1X y 1/10X, y se eluyeron en buffer Laemmli para su estudio por Western blot.

#### B) Citometría de flujo

Para los ensayos de interacción proteica, así como para chequear la correcta transfección de las proteínas de interés en células (de aquellas proteínas que presentaban el epítipo extracelular), se realizaron ensayos de citometría de flujo. Las células en suspensión (electroporadas o no), y las células adherentes (transfectadas o no) levantadas por pipeteo se transfirieron a tubos de citometría. Se empleó un pocillo de una placa M24 ó 100.000 células por cada punto a medir. Se centrifugaron durante 5 minutos a 1000 xg eliminándose el sobrenadante y, posteriormente, se lavaron y bloquearon con PBS suplementado con 2% BSA (solución de lavado). En los ensayos de interacción proteica se incubaron las células con 5 µg/mL de las proteínas recombinantes purificadas por baculovirus en un volumen de 100 µL durante 15 minutos a RT, y se lavaron dos veces con 1 mL de la solución de lavado. Posteriormente, o inicialmente para el resto de los ensayos, se procedió a incubar las células con el anticuerpo primario correspondiente a la dilución requerida (Tabla 14) durante 15 minutos a RT, y se lavaron dos veces con 1 mL de la solución de lavado. A continuación, se incubaron las células con el anticuerpo secundario correspondiente a la dilución requerida (Tabla 14) durante 15 minutos a RT, y se lavaron de nuevo dos veces con 1 mL de la solución de lavado. Por último, las células se resuspendieron en 0.5-1 mL de PBS y se midieron en el citómetro FACSCalibur (Becton Dickinson) mediante el programa CellQuestPro, disponible en el Servicio de Citometría de Flujo (SCF) del CBMSO. Se

empleó la línea de láser azul (488 nm) con el detector FL1 para los anticuerpos secundarios Alexa 488, Alexa 555 o GFP y la línea de láser roja (633 nm) con el detector FL4 para el anticuerpo secundario Alexa 647 (Tabla 14). Los datos resultantes se analizaron posteriormente mediante el programa FlowJo (para Macintosh).

### 3.3.4.8. Inmunofluorescencias

#### A) Procesamiento de las muestras

Se cultivaron células HeLa en cubreobjetos circulares de vidrio depositados en placas de 24 y se transfectaron transitoriamente. Pasadas 16 horas de transfección, se fijaron las células en 2% (v/v) paraformaldehído (PFA) preparado en PBS durante 15 minutos a 4°C y se trataron con 50 mM de cloruro de amonio durante 30 minutos para reducir la autofluorescencia de las muestras. Se permeabilizaron las muestras con PBS suplementado con 0.1% tritón X-100 (solución de lavado) durante 20 minutos a RT. Posteriormente, se bloquearon las muestras con PBS suplementado con 0.1% Tritón X-100 y 5% BSA (solución de bloqueo) durante 30 minutos. Las inmunofluorescencias se realizaron con los anticuerpos primarios anti-V5 (Sigma), anti-manosidasa (Millipore) y anti-calnexina (Stressgen), y con los anticuerpos secundarios acoplados a Alexa 488 y 555 (Invitrogen), con reactividad anti-ratón o anti-conejo dependiendo de los anticuerpos primarios de cada experimento (Tabla 14), todos ellos preparados en la solución de bloqueo correspondiente. Cada anticuerpo se incubó durante una hora a RT y se realizaron dos lavados después de cada incubación con la solución de lavado. Para las muestras con condiciones no permeabilizantes no se realizó la etapa de permeabilización, las soluciones de lavado y bloqueo fueron realizadas sin detergente (Tritón X-100) y la incubación de los anticuerpos a 4°C, evitando de esta forma la permeabilización y la internalización de los anticuerpos. Finalmente, los ácidos nucleicos fueron teñidos con una dilución 1:5.000 del fluorocromo DAPI (*4',6-Diamidino-2-phenylindole*) (Calbiochem) tras una incubación de 10 minutos, y las preparaciones fueron montadas con *Prolong Gold Antifade* (Invitrogen) en portaobjetos para su visualización por microscopia de fluorescencia o confocal.

#### B) Microscopía de fluorescencia

Para la visualización y captura de imágenes se emplearon diversos microscopios de fluorescencia dependiendo de la procedencia de las preparaciones. Para la observación en

placas de 24 pocillos se empleó el microscopio invertido Axiovert 135 (Zeiss). Para el análisis de las preparaciones montadas en portaobjetos se empleó el microscopio vertical Axioskop 2 plus (Zeiss) acoplado a una cámara ccd color (Coolsnap FX color, Roper Scientific). Ambos microscopios se encuentran disponibles en el Servicio de Microscopía Óptica y Confocal (SMOC) del CBMSO. Las imágenes obtenidas se procesaron con el programa Metavue 5.07 (Universal Imaging).

### C) Microscopía confocal

Las imágenes se tomaron con el microscopio de barrido láser confocal LSM710 acoplado a un microscopio vertical AxioImager.M2 (Zeiss), disponible en el SMOC del CBMSO. Las imágenes obtenidas se procesaron con el programa Zeiss ZEN 2010. Se emplearon los objetivos de 63x y 100x, determinando el grosor de la sección con un Pinhole de 1 AU (*Airy Units*) que correspondería con la resolución máxima. Se realizaron secciones en el eje Z para el procesamiento y en algunos casos para su posterior deconvolución. Para las imágenes en microscopía óptica de fluorescencia se emplearon las líneas de láser Diodo (405) para el fluorocromo DAPI, Argón (488) para los anticuerpos secundarios Alexa 488 y Hene (543) para los anticuerpos secundarios Alexa 555. Además de las imágenes en microscopía óptica de fluorescencia, también se analizaron en microscopía óptica de transmisión (Nomarski). Esto se realizó colocando un polarizador (prisma de Wollafston) en el objetivo de 100x, cambiando el condensador a la posición DIC III y realizando un ajuste de Köhler.

### C) Tratamiento de imágenes y deconvolución

Las imágenes se trataron con el procesador de imágenes científicas Fiji, paquete de procesamiento de imágenes *Open Source* de ImageJ. La barra de medida se calculó escalando el número de píxeles al tamaño real de la imagen. En el caso de las imágenes que se iban a procesar para deconvolución se tomaron con 0.8 AU, 12 bits, secciones en el eje Z de 108 nm, un tamaño de pixel de 0.04  $\mu\text{m}$  y evitando la saturación en cualquiera de los canales. Las imágenes se deconvolucionaron con el programa Huygens a través del sistema operativo Linux Fedora. Se emplearon como parámetros de corrección por el aceite de inmersión (1.5150) y por el medio de montaje Prolong Gold (1.46).

## **4. RESULTADOS**

## 4.1. ESTUDIO DEL *SPLICING* EN EL GEN *RAGE*

Estudios previos realizados en nuestro grupo de investigación habían determinado la presencia de diferentes variantes de *splicing* del gen *RAGE* en diversas líneas celulares. Además, existían otros estudios publicados de dicho fenómeno en algún tejido de mamíferos, por lo que consideramos dicho gen como candidato para el estudio de diferentes eventos de *splicing*, la conservación de las variantes y eventos de *splicing* en diferentes especies de mamíferos y, sus posibles consecuencias a nivel transcriptómico y proteómico.

El estudio del *splicing* alternativo del gen *RAGE* se llevó a cabo mediante dos metodologías complementarias. El primer análisis fue realizado mediante la técnica de RT-PCR anidada, clonaje de cada una de las variantes observadas y secuenciación, así como cuantificación de las variantes de *splicing* obtenidas. El otro tipo de metodología consistió en análisis mediante datos de RNA-seq para estudiar la presencia de diferentes uniones de exones y su combinación en isoformas generadas “*in silico*”.

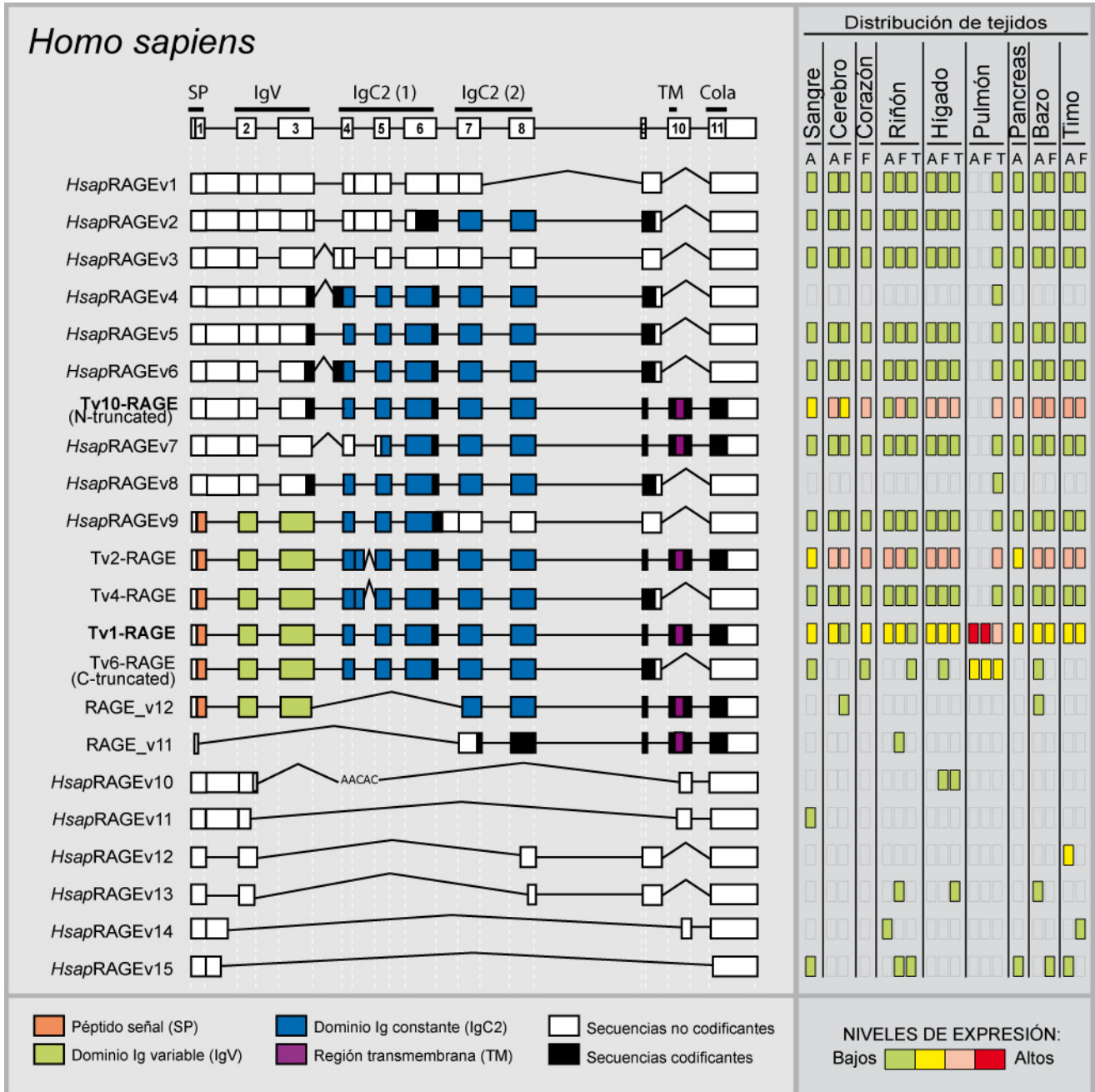
### 4.1.1. Análisis de las variantes de *splicing* del gen *RAGE* en diferentes especies de mamífero mediante RT-PCR

Las muestras empleadas en el estudio y caracterización de las variantes de *splicing* del gen *RAGE* en mamíferos fueron RNA total de muestras comerciales (ver Material y Métodos) de diversos tejidos en un total de seis especies diferentes: *Homo sapiens*, *Macaca mulatta*, *Sus scrofa*, *Bos taurus*, *Mus musculus* y *Rattus norvegicus*.

#### 4.1.1.1. Identificación y caracterización de las variantes de *splicing* de *RAGE* humanas

Para la completa caracterización de las variantes de *splicing* de *RAGE* en *Homo sapiens*, se realizó un ensayo RT-PCR específico desde el codón de inicio al codón de terminación de *RAGE* para dicha especie en 18 tejidos diferentes. El análisis detallado de estos resultados nos mostró un total de 22 variantes de *splicing* diferentes obtenidas en ocho tejidos adultos, siete fetales y tres tejidos tumorales (Figura 25). Interesantemente, quince de estas isoformas no habían sido anteriormente descritas (Figura suplementaria 1), y se nombraron como *HsapRAGEv1* a *v15* (Figura 25). La variante *Transcript variant 1* (o *full-length RAGE*) fue denominada Tv1-*RAGE* para simplificar. Esta isoforma canónica fue identificada en este estudio como una de las variantes de *splicing* más abundantes (Figura 25), en concordancia con lo que se había descrito previamente (Hudson et al. 2008b). Esta isoforma está compuesta por un péptido señal codificado en el exón 1, un dominio IgV

codificado en los exones 2 y 3, dos dominios IgC2 codificado en los exones 4 a 6 y 7 a 8 respectivamente, y una región transmembrana única en el exón 10 seguida de una corta cola citoplasmática codificada en el exón 11 (Hudson et al. 2008b).



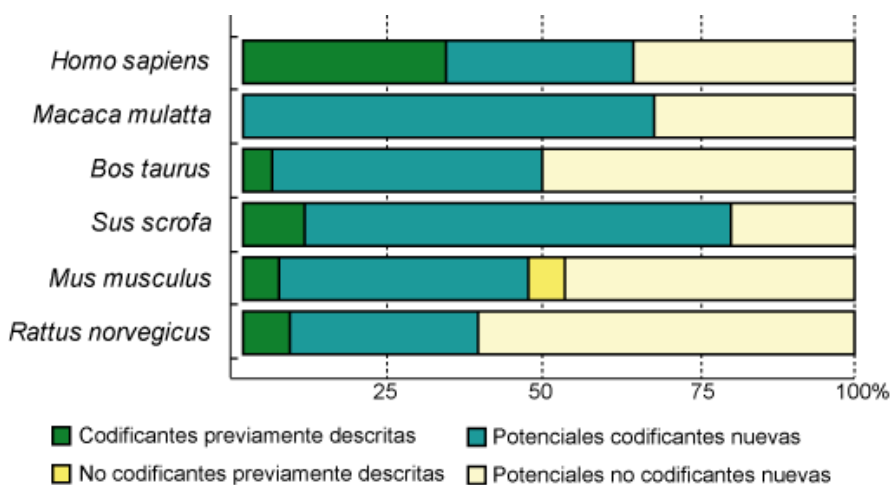
**Figura 25. Variantes de *splicing* de RAGE en *Homo sapiens*.** Representación de las variantes de *splicing* de RAGE humanas y sus niveles de expresión relativa en diferentes tejidos: adulto (A), fetal (F) y tumoral (T). Los dominios proteicos potenciales se muestran según los colores de la leyenda inferior, mientras que las secuencias no codificantes están mostradas en blanco. Las uniones de exones presentes en la variante canónica (Tv1-RAGE) están mostradas como líneas rectas, mientras que las uniones de *splicing* alternativas se muestran como triángulos abiertos. Los diferentes transcritos están organizados por tamaño, y representados a escala.

Otras variantes de AS encontradas en este estudio presentan variaciones en la presencia y composición de estos dominios. Un ejemplo de ello sería la presencia del intrón 1 retenido en la variante Tv10-RAGE (previamente denominada como N-truncada) y en las variantes *Hsap*RAGEv1 a v8 que provoca como resultado la aparición de codón de terminación prematuro de la traducción (PTC). Estas variantes presentan un codón de inicio potencial alternativo posterior a dicho PTC en el exón 3, lo que provocaría la exclusión tanto del péptido señal como del dominio IgV en las isoformas potencialmente codificantes (Figura 25). La variante Tv10-RAGE ha sido descrita recientemente por el NCBI como una variante no codificante, aunque múltiples autores la consideramos como una isoforma codificante con la pérdida del dominio IgV (Lopez-Diez et al. 2013; Sterenczak et al. 2013; Yonekura et al. 2003). La variante Tv10-RAGE solo presenta el evento del intrón 1 retenido, mientras que las otras variantes *Hsap*RAGEv1 a v8 presentan además una combinación de otros eventos de AS adicionales (Figura 25).

Otro evento de *splicing* común es la inclusión parcial de la porción 5' del intrón 9 y la exclusión del exón 10, lo que resulta en la pérdida de la region transmembrana y la cola citoplasmática generando isoformas proteicas sRAGE potencialmente solubles. Dicho evento se encuentra presente en las variantes Tv6-RAGE o C-truncada, Tv4-RAGE, y las variantes *Hsap*RAGEv1 a 6, v8 y v9. Además, hemos encontrado otros eventos de *splicing* diferentes, tales como el empleo de un sitio de *splicing* 5' donador diferente en el exón 4 (Tv2-RAGE y Tv4-RAGE, previamente descritos también por Hudson et al. (2008), o un nuevo sitio de *splicing* 3' acceptor en el exón 7 (RAGE\_v12) que causarían variaciones en la región codificante del dominio IgC y, por tanto, podría modificar la especificidad por ligando de RAGE y/o la afinidad de unión. Además, hemos encontrado múltiples variantes de *splicing* (RAGE\_v11 y las variantes *Hsap*RAGEv10 a v15) de pequeño tamaño que presentan la exclusión de múltiples exones (Figura 25). Tras estudios bioinformáticos basados en predicciones hemos considerado dichas variantes de pequeño tamaño como transcritos no codificantes, debido a la presencia de diferentes PTCs en su secuencia.

Hemos estudiado también la abundancia relativa de cada variante de *splicing* observando diferencias en los niveles de expresión entre los diferentes tejidos analizados. Las variantes Tv1-RAGE, Tv2-RAGE y Tv10-RAGE son las que presentaban unos niveles de expresión más elevados en la mayoría de los tejidos analizados (Figura 25). Las variantes de *splicing* de mayor tamaño (debidas a la retención de múltiples intrones) muestran unos niveles de expresión bajos pero uniformes a lo largo de los diferentes tejidos. Por el contrario, aquellas variantes de *splicing* de pequeño tamaño presentan también unos niveles

de expresión bajos pero con una distribución muy específica en tejidos concretos, incluso entre estados embrionarios o tumorales dentro de su tejido adulto normal equivalente. Todos los tejidos analizados mostraban un patrón de expresión similar, pero interesantemente hemos observado que los tejidos de pulmón tanto adulto como fetal solo expresaban las variantes Tv1-RAGE y Tv6-RAGE (C-truncada) a unos niveles de expresión muy elevados, mientras que en la muestra tumoral de pulmón se expresaban todas las variantes de mayor tamaño aunque a niveles bajos de expresión. Alrededor del 40% del total de las variantes descritas en humano son variantes no codificantes (Figura 26).



**Figura 26. Abundancia relativa de los transcritos analizados codificantes y no codificantes.** La barra de porcentajes refleja la proporción de variantes codificantes y no codificantes previamente descritas y aquellas que hemos encontrado en este estudio (codificantes y no codificantes) expresadas en las diferentes especies analizadas.

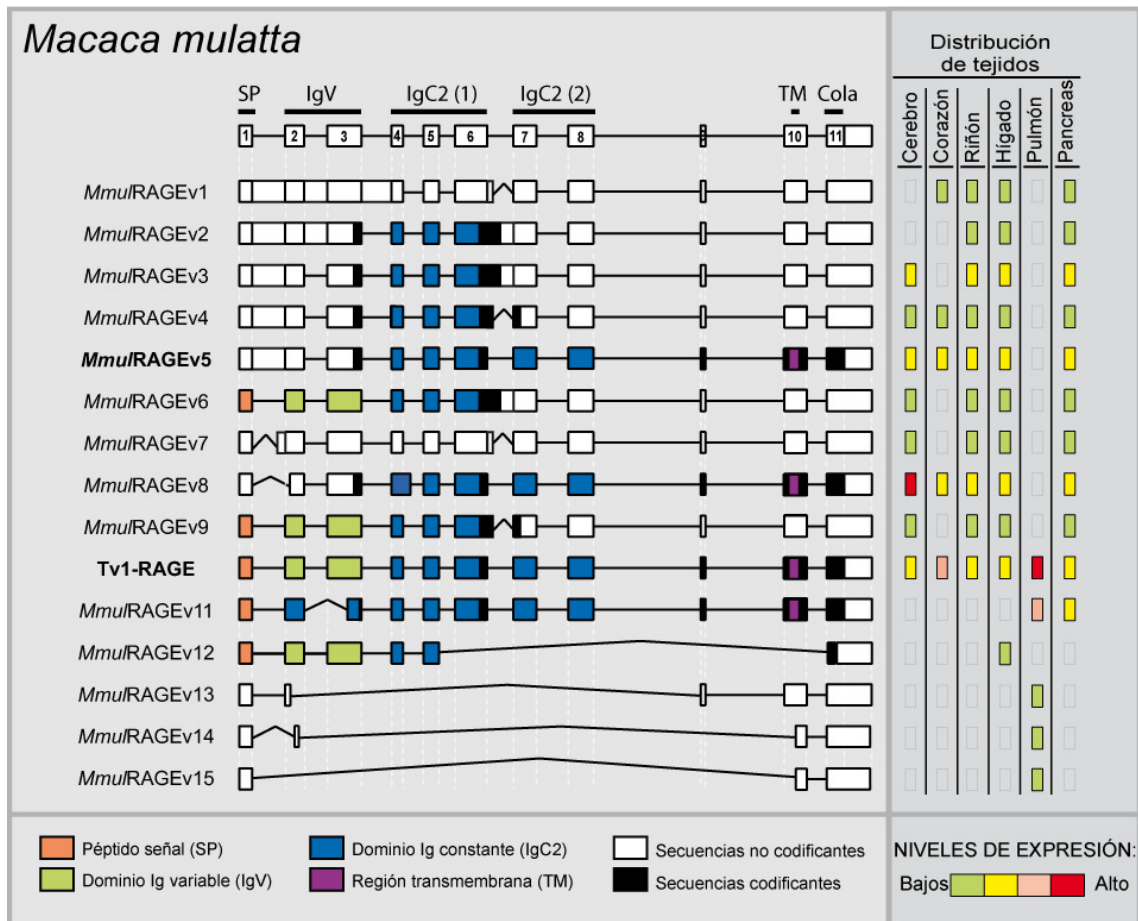
#### 4.1.1.2. Identificación y caracterización de las variantes de splicing de RAGE de el mono Rhesus

En *Macaca mulatta* (mono Rhesus) detectamos un total de 15 nuevas variantes de RAGE en los seis tejidos de adulto analizados (Figura 27), y fueron nombradas como *Mmul*RAGEv1 a v15. Solo la variante Tv1-RAGE había sido previamente anotada en las bases de datos (Figura suplementaria 1) mediante predicciones bioinformáticas basadas en su homología con la variante humana (NM\_001205117-GU139417). En este estudio mostramos la primera evidencia biológica de la expresión de esta isoforma en este primate.

Las variantes del mono Rhesus *Mmul*RAGEv5 y Tv1-RAGE están relacionadas directamente con las variantes Tv10-RAGE y Tv1-RAGE humanas, respectivamente, en relación a su estructura exónica/intrónica. Además, hemos encontrado una isoforma



(*Mmul*RAGEv3) que retiene los intrones 1 y 6, encontrándose directamente relacionada con una variante descrita previamente en humano (NtRAGE $\Delta$ ). En dos estudios previos se relacionó esta isoforma con la enfermedad de Alzheimer (Ding and Keller 2005a, 2005b), aunque nosotros no la hemos detectado en ninguno de los tejidos humanos analizados. Ninguna de las restantes variantes descritas en *Macaca mulatta* tienen homologías con variantes humanas (Tabla 16).



**Figura 27. Variantes de splicing de RAGE en *Macaca mulatta*.** La representación de las variantes de splicing de RAGE en el mono Rhesus y sus niveles de expresión relativa en diferentes tejidos de adulto. Ver Figura 25 para detalles. Tv1-RAGE y *Mmul*RAGEv5 (negrita) se corresponden a la forma canónica y a la variante humana Tv10-RAGE, respectivamente.

El evento de AS más común observado en las variantes encontradas es la inclusión del intrón 1, que se encuentra en las variantes *Mmul*RAGEv1 a v5, siendo la variante *Mmul*RAGEv1 una variante no codificante según predicciones bioinformáticas debido a la presencia de diferentes PTC en los intrones 1 y 6 y el exón 7. La completa inclusión del intrón 6 en *Mmul*RAGEv6 y su parcial inclusión en *Mmul*RAGEv9 genera también como resultado la aparición de PTCs en su secuencia. Sin embargo, en estos casos las variantes codificarían potencialmente un nuevo tipo de isoformas solubles que, además de perder la

	<i>Homo sapiens</i>	<i>Macaca mulatta</i>	<i>Bos taurus</i>	<i>Sus scrofa</i>	<i>Mus musculus</i>	<i>Rattus norvegicus</i>
	Tv1-RAGE	<i>Mmul</i> RAGEv10	Tv1-RAGE	Tv1-RAGE	Tv1-RAGE	Tv1-RAGE
	Tv10-RAGE	<i>Mmul</i> RAGEv5	<i>Btau</i> RAGEv7	<i>Sscr</i> RAGEv5	mRAGE_v2	<i>Rnor</i> RAGEv2
			<i>Btau</i> RAGEv1	<i>Sscr</i> RAGEv1		
		<i>Mmul</i> RAGEv3	<i>Btau</i> RAGEv3			
		<i>Mmul</i> RAGEv4	<i>Btau</i> RAGEv6			
	<i>Hsap</i> RAGEv14		<i>Btau</i> RAGEv22			
					mRAGE_v4	<i>Rnor</i> RAGEv3
				<i>Sscr</i> RAGEv8	<i>Mmus</i> RAGEv5	
					<i>Mmus</i> RAGEv11	<i>Rnor</i> RAGEv9
					<i>Mmus</i> RAGEv13	<i>Rnor</i> RAGEv12

**Tabla 16. Variantes de RAGE conservadas en los diferentes mamíferos analizados.** Conservación basada en la composición de exones e intrones. Las uniones de exones presentes en la variante canónica (Tv1-RAGE) están mostradas como líneas rectas, mientras que las uniones de *splicing* alternativas se muestran como triángulos abiertos. La retención de intrones se muestra como un cuadrado gris.

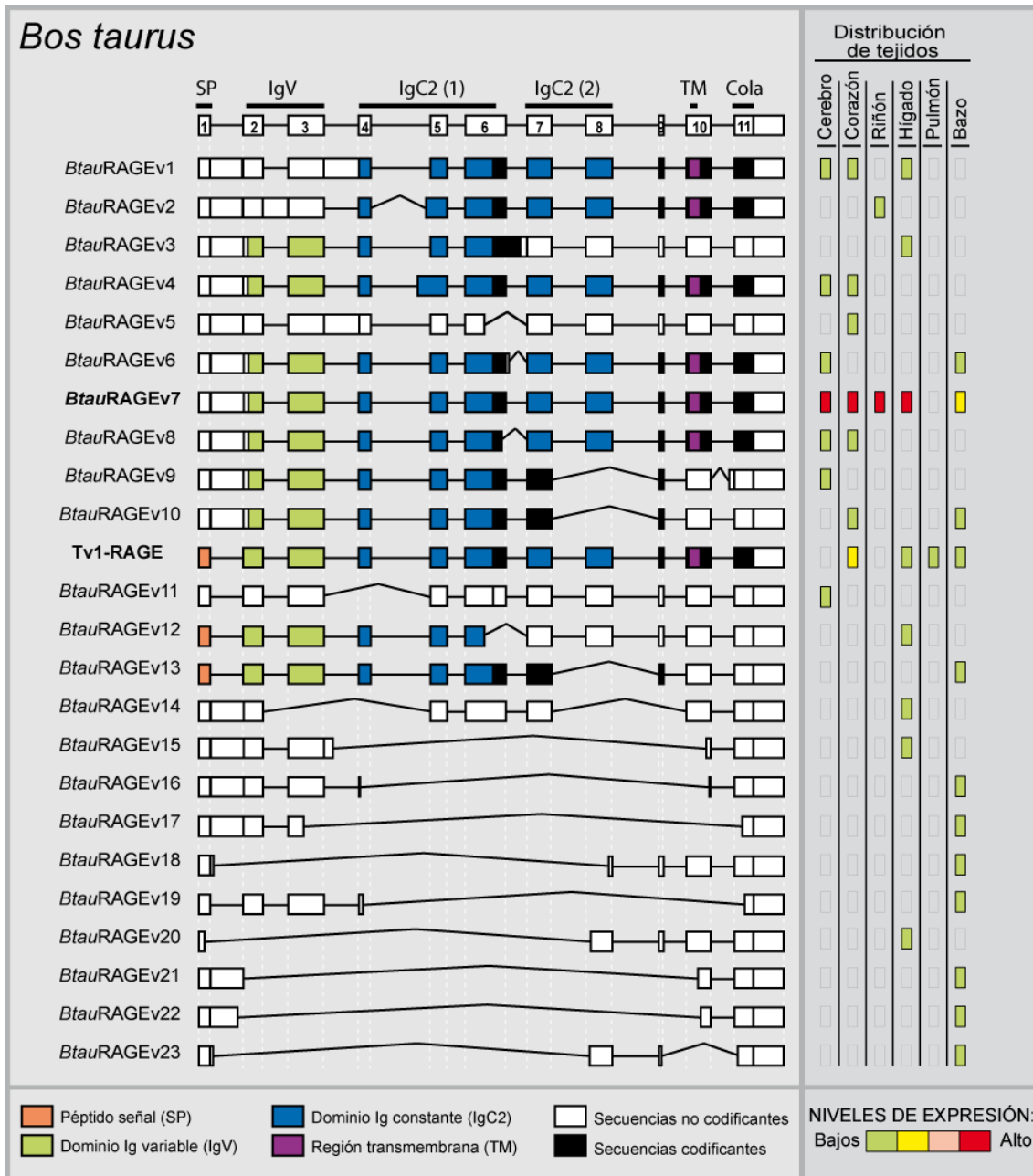
región transmembrana también carecerían del último dominio IgC (Figura 27). Hemos observado otro tipo de variación de los dominios proteicos en la variante *Mmul*RAGEv11 cuyo dominio IgV se convierte en un dominio IgC como resultado del uso de un sitio de *splicing* 5' donador alternativo en el exón 3. La pérdida del dominio IgV también se observaría en la variante *Mmul*RAGEv8 al presentar un sitio de *splicing* 3' aceptor alternativo en el exón 2 que implicaría cambios en la fase de lectura y la aparición de un PTC, por lo que posiblemente comenzase la traducción en un segundo codón de inicio de la traducción situado en el exón 3. Por último, es importante destacar la presencia de cuatro variantes de *splicing* de pequeño tamaño (*Mmul*RAGEv12 a v15) que se caracterizan por carecer de un número elevado de exones. Una de estas variantes podría potencialmente codificar una variante codificante soluble (*Mmul*RAGEv12) mientras que las otras tres las consideramos no codificantes según predicciones bioinformáticas. Las variantes no codificantes corresponden con aproximadamente el 30% de todas las variantes descritas en este estudio en el mono Rhesus (Figura 26).

Los tejidos de mono muestran niveles de expresión diferentes en cada variante. Tv1-RAGE, *Mmul*RAGEv8, *Mmul*RAGEv5 y *Mmul*RAGEv3 son las variantes más abundantes

en el estudio entre los diferentes tejidos analizados (Figura 27). *Mmul*RAGEv8 está sobre-expresado en cerebro, mientras que la variante Tv1-RAGE en mono está sobre-expresada en pulmón al igual que sucede con su homóloga humana. Este último tejido solo expresa la variante canónica y *Mmul*RAGEv11. Como *Mmul*RAGEv8 pierde el dominio IgV en relación con la variante canónica Tv1-RAGE, su sobre-expresión en cerebro podría indicar que la variante *Mmul*RAGEv8 podría estar uniéndose a otro ligando. Sin embargo, es importante destacar que esta variante no codificaría ningún péptido señal. Las pequeñas variantes no codificantes muestran niveles de expresión muy bajos y están solo presentes en pulmón, con la excepción de la variante codificante de pequeño tamaño *Mmul*RAGEv12 que se encuentra solo en hígado.

#### **4.1.1.3. Identificación y caracterización de las variantes de splicing de RAGE de la especie bovina**

En *Bos taurus* hemos detectado un total de 24 variantes de AS de RAGE en los seis tejidos adultos analizados (Figura 28), de las cuales 23 de ellas no estaban descritas previamente (Figura suplementaria 1) y fueron nombradas como *Btau*RAGEv1 a v23, siendo Tv1-RAGE la única isoforma descrita previamente por Neeper et al. (1992) (NM\_173982). Las isoformas Tv1-RAGE y *Btau*RAGEv7 están relacionadas estructuralmente a sus homólogas humanas y de macaco (Figura 26). Sin embargo, la variante de *splicing* *Btau*RAGEv7 presenta múltiples cambios a nivel proteico si lo comparamos con sus homólogos en primates. Esto es debido a la aparición de un codón de inicio de la traducción en el exón 2, en lugar de en el exón 3, lo que provoca la inclusión del dominio IgV en la secuencia de la proteína aunque careciendo del péptido señal. En relación con las proteínas N-truncadas, en la especie bovina solo nos encontramos con dos variantes, *Btau*RAGEv1 y *Btau*RAGEv2, que podrían generar este tipo de proteínas truncadas similares a las de humano, aunque para ello requiere la retención de dos intrones en lugar de únicamente el primero, como ocurre en otras especies. La variante *Btau*RAGEv22 es similar a la variante humana *Hsap*RAGEv14, siendo ambas variantes no codificantes. Además, las variantes *Btau*RAGEv3 y v6 tienen homólogos en las variantes de macacos (Tabla 16). En esta especie las isoformas sRAGE pueden ser generadas por diversos eventos de AS como son la pérdida del exón 8 (*Btau*RAGEv9, v10 y v13), el uso de sitios de *splicing* 5' donantes alternativos (*Btau*RAGEv12) o la retención de intrones (*Btau*RAGEv3), pero en todos los casos perdiendo el segundo dominio IgC además de la región transmembrana. Por último, destacar la presencia de múltiples variantes de *splicing*



**Figura 28. Variantes de *splicing* de RAGE en *Bos taurus*.** La representación de las variantes de *splicing* de RAGE en la especie bovina y sus niveles de expresión relativa en diferentes tejidos de adulto. Ver Figura 25 para detalles. Tv1-RAGE y *Btau*RAGEv7 (negrita) se corresponden a la forma canónica y a la variante humana Tv10-RAGE, respectivamente.

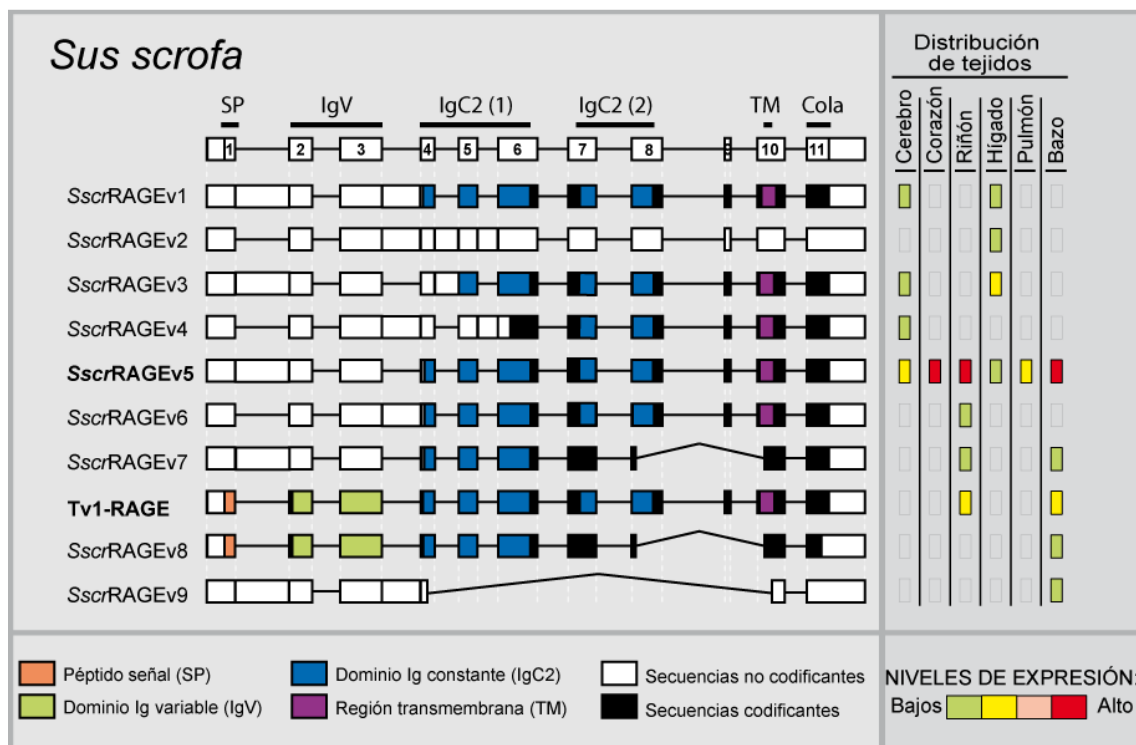
no codificantes consecuencia de largas “deleciones” (*Btau*RAGEv14 a v23), o de pequeñas “deleciones” y retención de intrones (*Btau*RAGEv5 y v11), representando alrededor del 50% del total de las variantes descritas en este trabajo (Figura 26).

En relación con la abundancia relativa de cada una de las variantes de *splicing*, la Figura 28 muestra *Btau*RAGEv7 con unos niveles de expresión altos (incluso más altos que la canónica) en la mayoría de los tejidos analizados, mientras que el resto de las variantes

muestran en general unos niveles de expresión bajos y específicos de tejido. En contraste con las variantes de humano, la isoforma Tv1-RAGE de la especie bovina presenta unos niveles de expresión bajos en pulmón, pero es la única isoforma que se expresa en este tejido.

#### 4.1.1.4. Identificación y caracterización de las variantes de splicing de RAGE de cerdo

*Sus scrofa* únicamente presenta diez variantes de AS de RAGE en los seis tejidos de adulto analizados (Figura 29), convirtiéndose en la especie con menos variantes incluso cuando comparamos con el mismo número y tipo de tejidos. Tras una búsqueda en las bases de datos nos encontramos con una predicción de la isoforma Tv1-RAGE, mientras que el resto de las variantes encontradas en este trabajo no habían sido descritas anteriormente (Figura suplementaria 1) y fueron denominadas como *SscrRAGEv1* a *v9*. Las variantes Tv1-RAGE y *SscrRAGEv5* son nuevamente las únicas dos variantes con homólogas humanas (Tv1-RAGE y Tv10-RAGE, respectivamente, Tabla 16), mostrando los mismos dominios que sus correspondientes humanas. La isoforma *SscrRAGEv1* está relacionada



**Figura 29. Variantes de splicing de RAGE en *Sus scrofa*.** La representación de las variantes de splicing de RAGE en cerdo y sus niveles de expresión relativa en diferentes tejidos de adulto. Ver Figura 25 para detalles. Tv1-RAGE y *SscrRAGEv5* (negrita) se corresponden a la forma canónica y a la variante humana Tv10-RAGE, respectivamente.

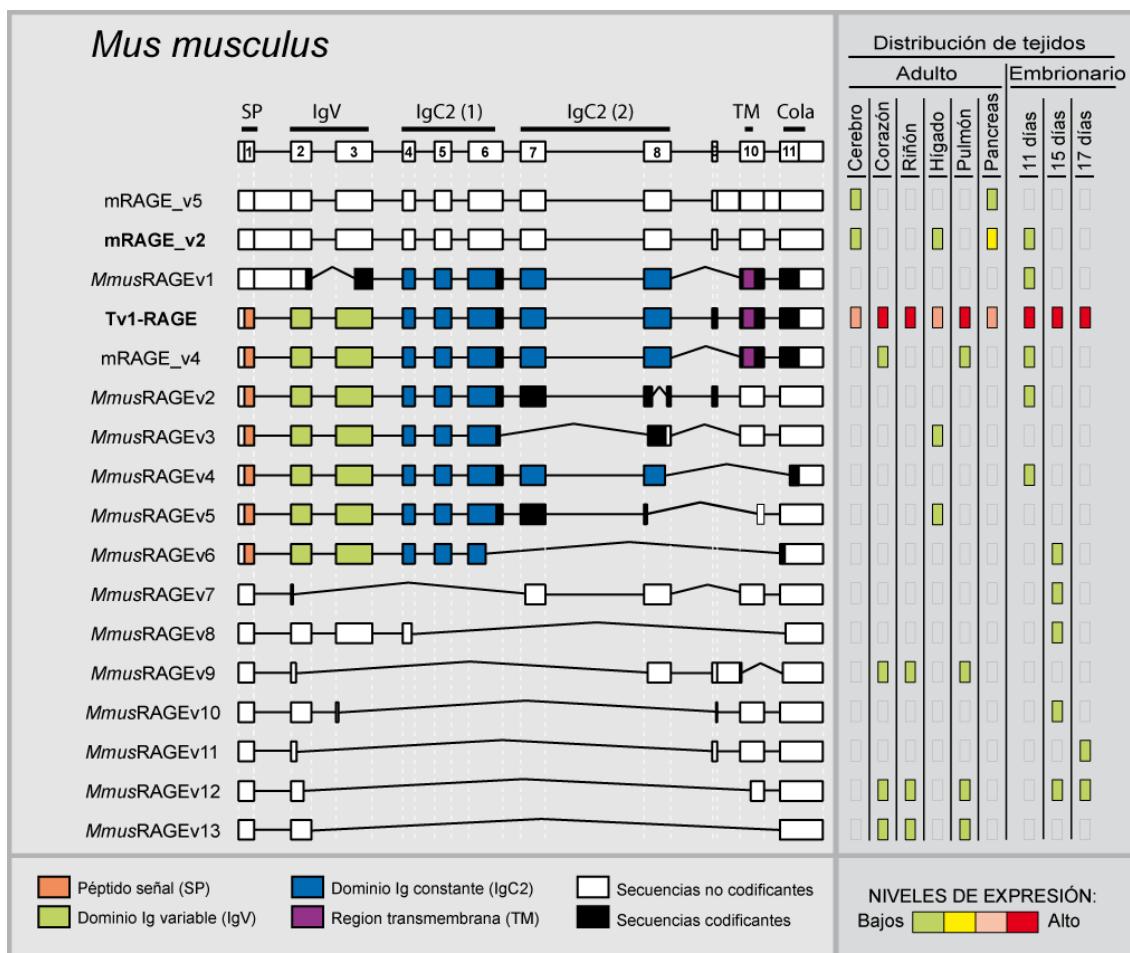
con la isoforma bovina *Btau*RAGEv1 (Tabla 16) y ambas presentan unos dominios proteicos similares. El evento de AS más característico de las variantes de cerdo es la retención de intrones, observándose como específicamente la retención de los intrones 1 y/o 3 son eventos comunes a varias isoformas (*Sscr*RAGEv1 a v7 y v9). Estos eventos resultan en la aparición de un PTC y un codón de inicio de la traducción localizado en el exón 4, en lugar de en el exón 3 presente en primates o en el exón 2 de la especie bovina. Este segundo codón de inicio de la traducción en cerdo provoca la potencial generación de proteínas que pierden el péptido señal y el dominio IgV. Las isoformas solubles de RAGE están también presentes en cerdo como consecuencia del uso de un sitio de *splicing* alternativo 3' acceptor en el exón 10 (*Sscr*RAGEv7 y v8) y, por lo tanto, perdiendo la región transmembrana. Por otro lado, el gen *RAGE* en cerdo solo genera dos variantes de *splicing* no codificantes (*Sscr*RAGEv2 y v9), representando únicamente el 20% de todas las variantes detectadas (Figura 26).

Los niveles de expresión de las diferentes variantes de *splicing* en cerdo (Figura 29) difieren bastante de los valores encontrados en las muestras de humano (Figura 25). La isoforma *Sscr*RAGEv5 expresada en cerdo, y equivalente a la isoforma Tv10-RAGE de humano, se encuentra altamente expresada en todos los tejidos analizados, incluso en pulmón donde su homólogo humano no se expresa. Sin embargo, la expresión de la isoforma Tv1-RAGE es moderada y únicamente se expresa en dos tejidos (riñón y bazo) y, lo que es más remarcable de esta isoforma es su ausencia de expresión en pulmón.

#### **4.1.1.5. Identificación y caracterización de las variantes de *splicing* de RAGE de ratón**

Hemos detectado un total de 17 variantes de AS de RAGE en *Mus musculus* dentro de los diferentes tejidos adultos y embrionarios analizados (Figura 30). La isoforma Tv1-RAGE se encontraba en las bases de datos, así como las variantes mRAGE\_v2, v4 y v5 habían sido previamente descritas (Kalea et al. 2009) (Figura Suplementaria 1). Las otras 13 variantes han sido descritas por primera vez en este estudio, y se nombraron como *Mmus*RAGEv1 a v13. Solo dos de estas variantes, Tv1-RAGE y mRAGE\_v2, tienen homólogos en términos de estructura exónica e intrónica al resto de las especies de mamífero analizadas (Tabla 16). Sin embargo, mRAGE\_v2 presenta otro PTC en el exón 5, no presente en el resto de sus homólogos, y como consecuencia de ello esta variante podría ser no codificante en ratón (Figura 30). Hay una variante potencialmente soluble en ratón (*Mmus*RAGEv5), que está relacionada con una variante de cerdo (*Sscr*RAGEv8). Además,

en ratón nos encontramos con otras variantes que podrían estar generando isoformas solubles, pero estas solo estarían formadas por el péptido señal y los dos primeros dominios Ig (*MmusRAGEv2* a *v6*). En un estudio previo se describió una isoforma (esRAGE) que podría generar una proteína soluble debido a retención del intrón 9, pero nosotros no hemos sido capaces de identificarla (Harashima et al. 2006) (Figura Suplementaria 1). Otro evento que podría producir alteraciones en los dominios proteicos de RAGE estaría presente en la variante *MmusRAGEv1* debido al uso de un sitio de *splicing* 3' aceptor en el exón 3, que provocaría la pérdida del dominio IgV y el péptido señal canónico. Hemos encontrado además múltiples variantes de *splicing* no codificantes, de las cuáles *mRAGE\_v5* y *mRAGE\_v2* habían sido previamente descritas (Kalea et al. 2009), y las variantes *MmusRAGEv7* a 13 fueron por primera vez descritas en este trabajo. En total, las variantes no codificantes representan más del 50% de las variantes de *splicing* murinas (Figura 26).



**Figura 30. Variantes de *splicing* de RAGE en *Mus musculus*.** La representación de las variantes de *splicing* de RAGE en ratón y sus niveles de expresión relativa en diferentes tejidos de adulto y embrionarios. Ver Figura 25 para detalles. Tv1-RAGE y *mRAGE\_v2* (negrita) se corresponden a la forma canónica y a la variante humana Tv10-RAGE, respectivamente.



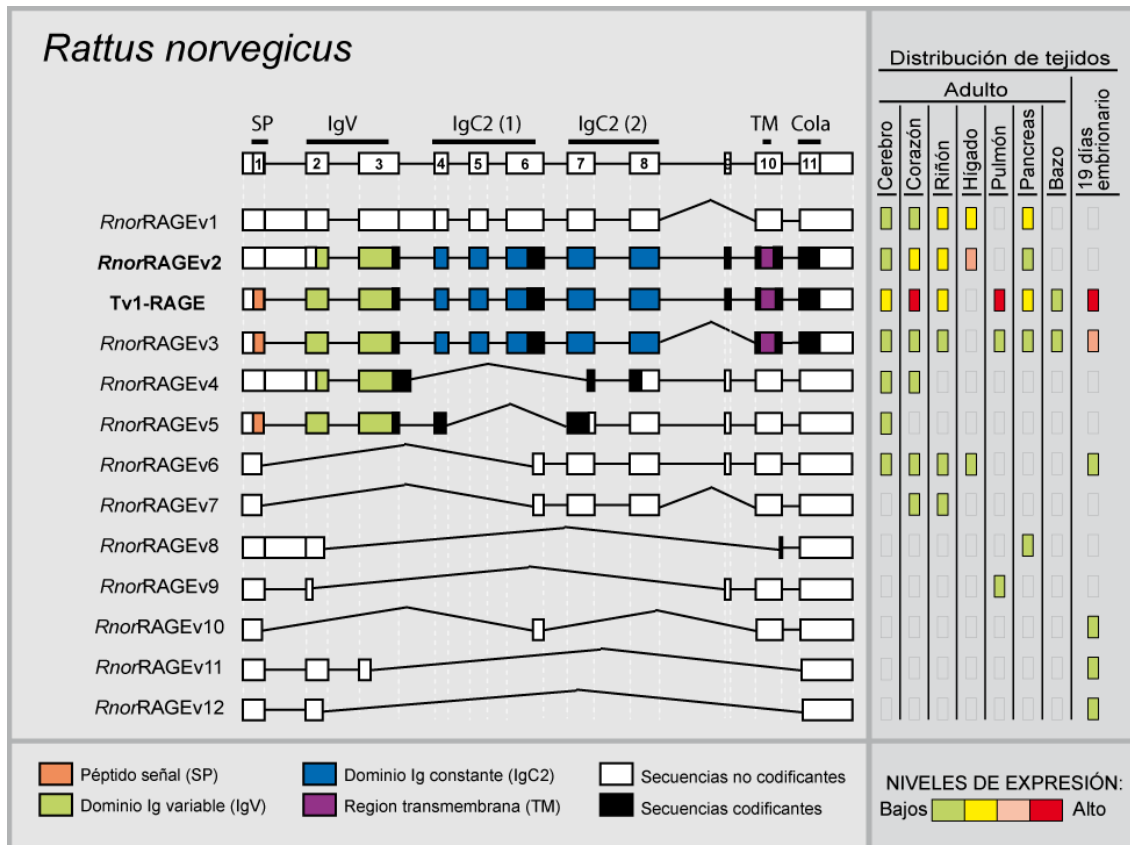
La abundancia relativa de cada variante de *splicing* muestra diferencias en los niveles de expresión en cada uno de los tejidos (Figura 30), así como si comparamos con las otras especies de mamíferos analizadas en este estudio. Tv1-RAGE presenta, al igual que en humano, unos niveles de expresión elevados a lo largo de todos los tejidos, tanto en tejidos embrionarios como en adultos. Sin embargo, la variante no codificante mRAGE\_v2 muestra unos niveles de expresión bajos en cerebro, hígado y en tejido embrionario de 11 días, y solo muestra niveles de expresión moderados en el tejido pancreático adulto, mientras que su homóloga humana se expresa en unos niveles de expresión elevados a lo largo de la mayoría de los tejidos. El resto de variantes de *splicing* murinas muestran unos niveles de expresión bajos y con una distribución específica de tejido y estados de desarrollo.

#### **4.1.1.6. Identificación y caracterización de las variantes de *splicing* de RAGE de rata**

En *Rattus norvegicus* detectamos un total de 13 variantes de *splicing* de RAGE en las diferentes muestras de tejido adulto y embrionario analizadas. La secuencia de la isoforma Tv1-RAGE se encontraba en las bases de datos pero basada en homologías, siendo esta la primera evidencia biológica. El resto de las doce variantes fueron descritas en este estudio por primera vez, y se nombraron como *RnorRAGEv1* a *v12* (Figura 31 y Figura suplementaria 1). Las isoformas Tv1-RAGE y *RnorRAGEv2* estaban relacionadas con las isoformas homólogas humanas, así como con sus homólogos en el resto de las especies analizadas (Tabla 16). La variante *RnorRAGEv2*, aunque está relacionada estructuralmente a la variante humana Tv10-RAGE, muestra cambios a nivel proteico siendo similar a la isoforma bovina *BtauRAGEv7* (Figura 28), ya que también conserva el dominio IgV en su secuencia (Figura 31).

La retención del intrón 1 está presente en varias variantes de *splicing* (*RnorRAGEv1*, *v2*, *v4* y *v8*), resultando en un PTC en la secuencia del intrón retenido y un codón de inicio alternativo localizado posteriormente en el exón 2. Este codón de inicio alternativo generaría variantes de *splicing* potencialmente codificantes, a no ser que presente también el intrón 3 retenido (*RnorRAGEv1*) o múltiples eventos de salto de exones (*RnorRAGEv8*), lo que generaría variantes de *splicing* no codificantes. Algunas variantes de rata están relacionadas con homólogos en ratón (Tabla 16), de las cuáles *RnorRAGEv3* (relacionada a la variante mRAGE\_v4 de ratón) se encuentra ampliamente expresada y presenta una estructura proteica muy similar a Tv1-RAGE, ya que solo pierde una pequeña región entre el segundo IgC y la región transmembrana. Las variantes no codificantes





**Figura 31. Variantes de *splicing* de RAGE en *Rattus norvegicus*.** La representación de las variantes de *splicing* de RAGE en rata y sus niveles de expresión relativa en diferentes tejidos de adulto y embrionario. Ver Figura 25 para detalles. Tv1-RAGE y *RnorRAGEv2* (negrita) se corresponden a la forma canónica y a la variante humana Tv10-RAGE, respectivamente.

(*RnorRAGEv1* y *RnorRAGEv6* a 12) representan más del 60% de las variantes totales descritas en *Rattus norvegicus* (Figura 26).

La abundancia relativa de cada variante de *splicing* en rata muestra diferencias en los niveles de expresión a lo largo de los diferentes tejidos analizados, especialmente las variantes no codificantes de pequeño tamaño que muestran una distribución muy específica de tejido y de estado de desarrollo. La isoforma Tv1-RAGE presenta los niveles de expresión más elevados en todos los tejidos, a excepción de en hígado donde no hemos encontrado expresión. La muestra de tejido de pulmón presenta unos niveles de expresión similares a la misma muestra en humano, con altos niveles de expresión de Tv1-RAGE y bajos niveles de *RnorRAGEv3* y pequeñas variantes de *splicing* no codificantes. Las variantes *RnorRAGEv1* y v2 muestran unos niveles de expresión moderados o moderados-altos en los tejidos de adulto, estando en concordancia la expresión de la variante *RnorRAGEv2* de rata y su homólogo humano Tv10-RAGE. En el tejido embrionario, Tv1-RAGE y *RnorRAGEv3* son los transcritos más abundantes (Figura 31).

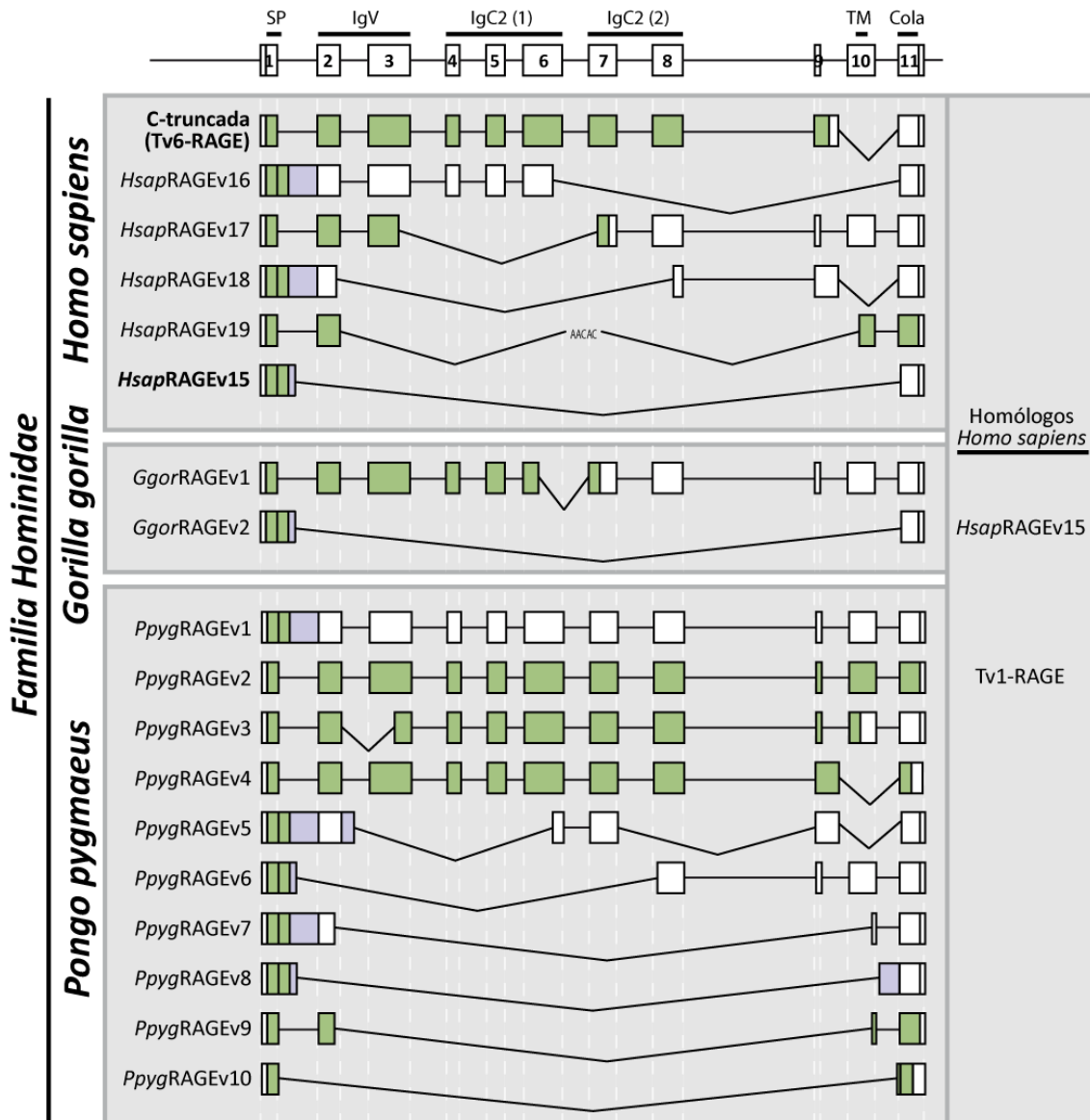
#### 4.1.2. Análisis de las variantes de *splicing* del gen *RAGE* en primates

A través de una colaboración con el Zoo-Aquarium de Madrid y el Servicio de Animalario del Instituto de Investigaciones Biológicas (IIB) se obtuvieron muestras de sangre de diferentes primates (ver Figura 21 de Materiales y Métodos). Se emplearon como controles muestras de sangre humanas procedentes de una colaboración con el Banco de Donantes del Hospital Universitario la Paz de Madrid. Todas las muestras fueron procesadas siguiendo el mismo procedimiento para la obtención del RNA presente en la fracción de PBMCs (ver Materiales y Métodos).

Para el análisis de las variantes de *splicing* se empleó el mismo procedimiento anterior de PCR anidada realizado para el estudio en las muestras de mamíferos, con la diferencia de que se tuvo que realizar un estudio previo de conservación de la secuencia para el diseño de cebadores. Esto es debido a que solo se encontraba disponible la secuencia de determinados primates en las bases de datos y, en función de estas secuencias, buscamos regiones de conservación con la finalidad de diseñar los cebadores para el estudio. Dichos cebadores fueron diseñados genéricos para todas las especies de primates dada la alta conservación existente en la secuencia exónica de este gen en el orden primates. Además, el empleo de cebadores comunes para todas las especies de primates reducía el riesgo de introducir sesgos en las amplificaciones pudiendo hacer una comparación más fiable.

En el estudio de la muestra de sangre de *Homo sapiens* se observa la expresión de seis variantes de *splicing*, de las cuales dos habían sido previamente descritas (Tv6-RAGE o C-truncada y *HsapRAGEv15*) y cuatro variantes nuevas que se denominaron como *HsapRAGEv16* a *v19* (Figura 32). Si comparamos esta muestra de sangre obtenida a partir del RNA de PBMCs con la muestra de RNA comercial estudiada anteriormente (Lopez-Diez et al. 2013), se observa como las dos isoformas Tv6-RAGE y *HsapRAGEv15* están expresadas en ambas muestras. Sin embargo, la muestra estudiada en este apartado presenta cuatro variantes nuevas con múltiples eventos de exclusión de exones que no están presentes en la muestra de RNA comercial. Igualmente, la muestra de RNA comercial presenta doce variantes de *splicing* adicionales, presentando la mayoría de las variantes múltiples eventos de retención de intrones. Las diferencias entre ambas muestras podían ser debidas a los métodos de extracción, por lo que se trató de minimizar esta variación igualando el método de procesamiento en todas las muestras de primates.

Las especies pertenecientes a la familia Hominidae (Humano u *Homo sapiens*, gorila o *Gorilla gorilla*, y orangután o *Pongo pygmeus*) presentan múltiples variantes de *splicing*, de las cuáles solo se comparte la variante *HsapRAGEv15* en humano y gorila



**Figura 32. Variantes de *splicing* de RAGE en sangre de primates, familia Hominidae.** Las uniones de exones presentes en la variante canónica (Tv1-RAGE) están mostradas como líneas rectas, mientras que las uniones de *splicing* alternativas se muestran como triángulos abiertos. Los diferentes transcritos están organizados por tamaño, y representados a escala. Se muestran en negrita las variantes conservadas en humano con la muestra de sangre comercial. A la derecha se indican los homólogos con *Homo sapiens*.

(nombrada como *GgorRAGEv2*). La isoforma canónica Tv1-RAGE solo está presente en la muestra de *P.pygmaeus* (*PpygRAGEv2*). Todas las variantes de *splicing* de *G. gorilla* y *P.pygmaeus* de RAGE encontradas en este estudio no se habían descrito con anterioridad y fueron nombradas como *GgorRAGEv1* a 2 para las variantes de gorila y *PpygRAGEv1* a v10 para las de orangután. Es interesante observar la presencia de *PpygRAGEv3* y v4 como dos nuevas variantes codificantes solubles en las muestras de *P.pygmaeus*. El evento de *splicing* que se encuentra más representado en estas especies es la exclusión de múltiples exones en la secuencia de los mensajeros produciendo variantes de pequeño tamaño.

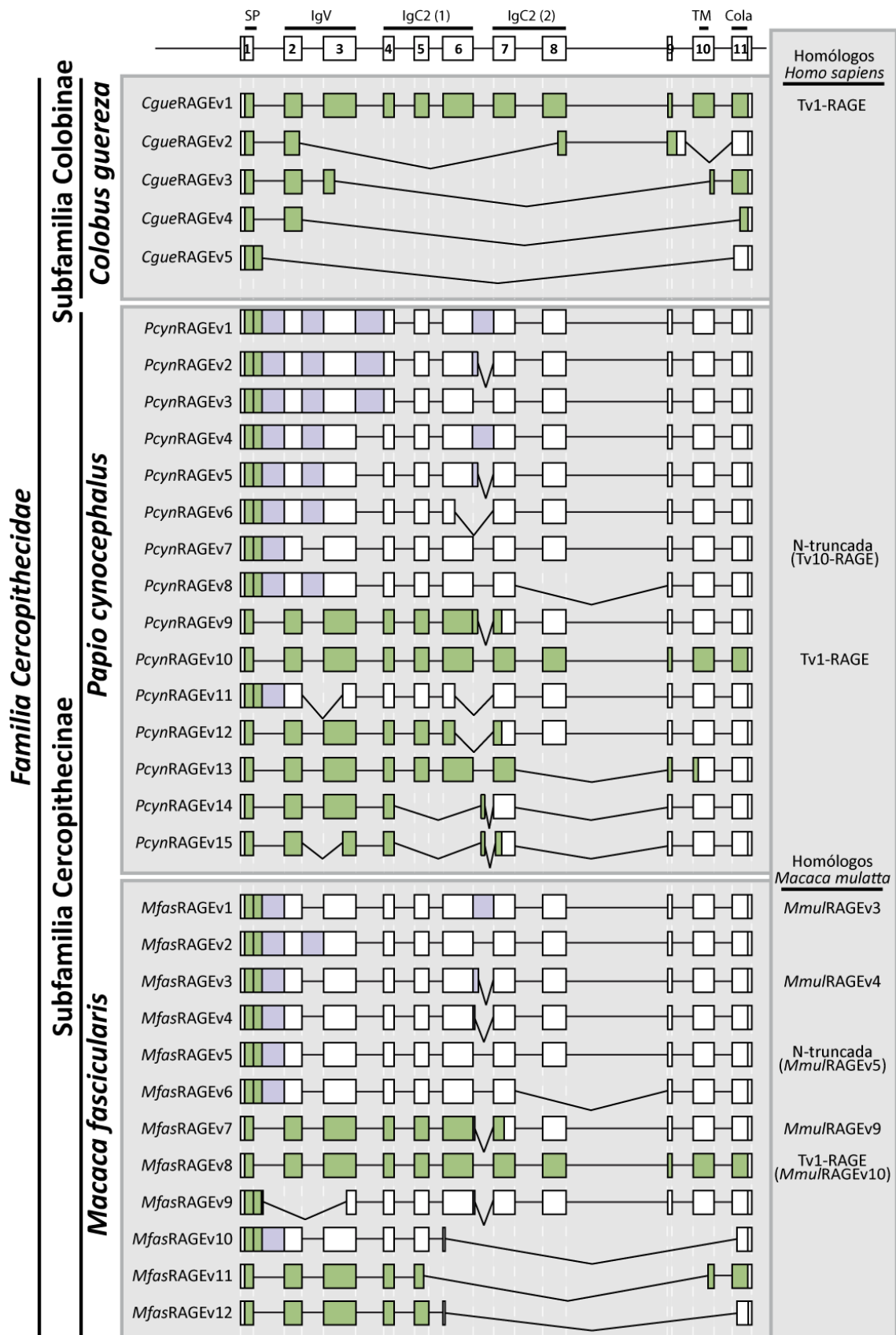


Figura 33. Variantes de *splicing* de RAGE en sangre de primates, familia Cercopithecidae. Para detalles, ver Figura 32.

En el caso de las especies pertenecientes a la familia Cercopithecidae (Colobo o *Colobus guereza*, papión o *Papio cynocephalus*, y macaco cangrejero o *Macaca fascicularis*) todas las variantes encontradas en este estudio no habían sido descritas previamente y se denominaron como *CgueRAGEv1* a *v5*, *PcynRAGEv1* a *v15* y *MfasRAGEv1* a *v12*, respectivamente (Figura 33). Se observa como las variantes de *splicing* de la subfamilia Colobinae (*C.guereza*) presentan como evento de *splicing* mayoritario la exclusión de múltiples exones en su secuencia, asemejándose a las variantes encontradas en la familia Hominidae. Sin embargo, en las dos especies de la subfamilia Cercopithecinae, *P.cynocephalus* y *M.fascicularis*, el evento más característico es la retención de uno o varios intrones en su secuencia. Ambas especies expresan la variante N-truncada como resultado de la retención del intrón 1 (*PcynRAGEv7* y *MfasRAGEv5*). Además, aunque las tres especies de esta familia presentan un patrón de expresión muy diferente entre ellas, es importante destacar la presencia de la variante Tv1-RAGE en todas las especies analizadas (*CgueRAGEv1*, *PcynRAGEv10* y *MfasRAGEv8*).

Si estudiamos en concreto la muestra del macaco cangrejero (*M.fascicularis*), observamos una alta similitud con aquellas variantes encontradas anteriormente en el mono Rhesus (*Macaca mulatta*), perteneciente ambos a un mismo género. Todas las variantes encontradas en *M.fascicularis* no se habían descrito previamente, solo existían predicciones bioinformáticas por homologías con otros primates estudiados (*H.sapiens* y *M.mulatta*). Las variantes que se encuentran conservadas entre ambos son *MmulRAGEv3*, *v5*, *v9* y *v10*, entre las que se encuentran la variante N-truncada (*MmulRAGEv5*) y la isoforma canónica Tv1-RAGE (*MmulRAGEv10*), como se detalla en la Figura 33.

De la familia Cebidae incluida en los monos del nuevo mundo solo tuvimos disponible una muestra de sangre de la especie *Cebus apella* (mono capuchino). Esta especie no había sido estudiada con anterioridad en relación a las variantes de *splicing* del gen RAGE, por lo que todas las variantes descritas son nuevas y se nombraron como *CapeRAGEv1* a *v6* (Figura 34). Basándonos en la homología con humano, encontramos la expresión de la variante N-truncada o Tv10-RAGE (*CapeRAGEv3*). Además, nos encontramos con la variante *CapeRAGEv6* que sería una variante potencialmente codificante y presentaría el dominio IgV modificado debido a un evento de AS en el exón 3. En global nos encontramos con un patrón de expresión de variantes similar a las especies más cercanas evolutivamente (*P. cynocephalus* y *M. fascicularis*) (Figura 33) con abundantes eventos de retención de intrones en sus variantes.

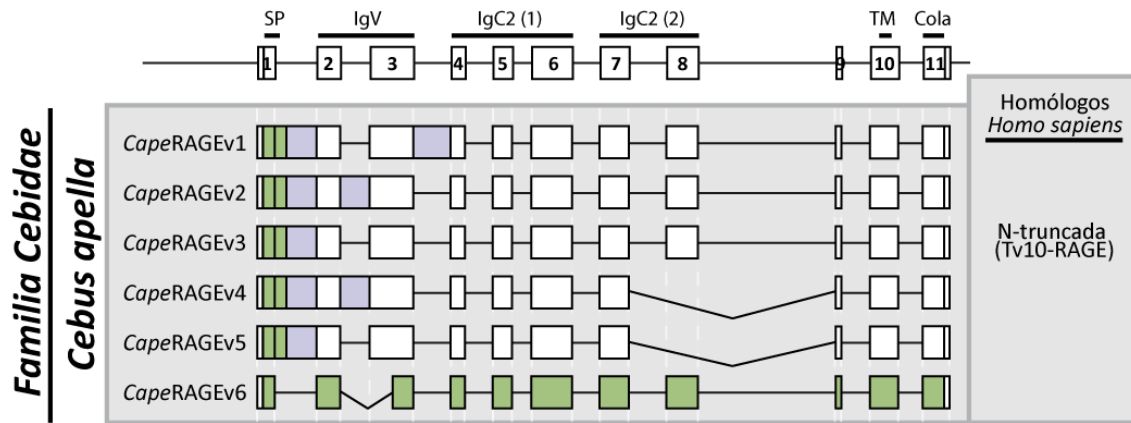


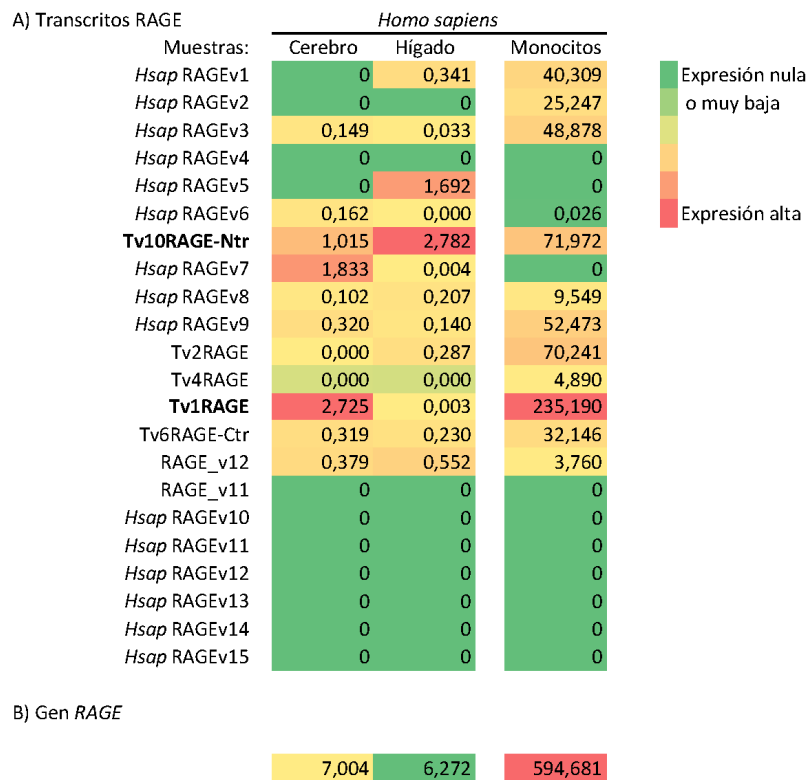
Figura 34. Variantes de *splicing* de *RAGE* en sangre de primates, familia Cebidae. Ver Figura 32 para detalles.

#### 4.1.3. Identificación de las variantes de *splicing* de *RAGE* mediante técnicas de secuenciación masiva

Actualmente, los datos de RNA-seq son una herramienta muy potente para el estudio y caracterización transcripcional, empleándose desde la descripción de nuevas isoformas hasta en la cuantificación de todos los transcritos expresados. Por este motivo, se decidió llevar a cabo el análisis de datos relacionados con los transcritos del gen *RAGE* procedentes de secuenciación masiva de RNA depositados y libremente accesibles en la base de datos SRA (Barbosa-Morais et al. 2012; Merkin et al. 2012; Pan et al. 2008) y compararlos con los resultados obtenidos con los análisis de RT-PCR anidadas. Para poder realizar esta comparativa, se emplearon datos de muestras de cerebro e hígado de humano, mono, bovino, ratón y rata. Posteriormente, se mapearon las lecturas de RNA-seq disponibles de las muestras de estos dos tejidos a cada uno de los genomas de referencia indicados y, posteriormente, se estimaron las abundancias relativas de cada uno de los transcritos detectados por RT-PCR anidada mediante el empleo del programa Cufflinks.

Los análisis confirmaron que *RAGE* es un gen de expresión muy baja ya que la cobertura que observamos en todas las muestras era muy baja o incluso nula en alguna de ellas (Figura 35 y Figura Suplementaria 2). En humano, la isoforma Tv1-RAGE muestra expresión moderada/alta en cerebro y la variante Tv10-RAGE en hígado, en conformidad con los resultados obtenidos por RT-PCR (Figura 35 y Figura 25).

Las variantes de macaco *MmulRAGEv2* y *MmulRAGEv6* se expresan mayoritariamente en cerebro, seguidas de niveles de expresión menores de las isoformas Tv1-RAGE y *MmulRAGEv1*, mientras que en los datos obtenidos por RT-PCR las variantes con mayores niveles de expresión son *MmulRAGEv8*, v5, v3 y Tv1-RAGE. En el



**Figura 35. Estimación de la expresión de los diferentes transcritos (A) y en global (B) del gen RAGE en muestras de RNA-seq humanas.** Se muestran el conjunto de todas las lecturas de cerebro e hígado, unidas por tejido [PMID:18978789 – Pan et al. (2008); PMID:23258890 – Barbosa-Morais et al. (2012); PMID:23258891 – Merkin et al. (2012)], y monocitos (datos procedentes del proyecto ENCODE SRX179434). Los niveles de expresión se muestran en una escala de gradiente mostrada a la derecha de la figura.

caso de las muestras de hígado ambos datos presentan cierta correlación con niveles de expresión elevados en *Mmul*RAGEv3 y v5 (Figura 36, Figura 27 y Figura Suplementaria 2). En los análisis de las muestras de bovino se observa como las variantes *Btau*RAGEv1, v2 y v4 se encuentran altamente expresadas en cerebro, mientras que las variantes con mayores niveles de expresión en hígado son *Btau*RAGEv1 y v3 (Figura 36, Figura 28 y Figura Suplementaria 2). Además, la isoforma bovina Tv1-RAGE está moderadamente expresada en cerebro pero no está expresada en hígado, mientras que la variante *Btau*RAGEv7, que se encuentra altamente expresada según los resultados de RT-PCR, no presenta expresión en las muestras de RNA-seq. Los datos de RNA-seq murinos indican como las variantes mayoritariamente expresadas son Tv1-RAGE (al igual que en los resultados de RT-PCR) y mRAGE\_v5 en ambos tejidos (mientras que esta variante no se encuentra expresada en hígado en los análisis por RT-PCR) (Figura 36, Figura 30 y Figura Suplementaria 2). Por último, los resultados obtenidos en rata por ambas tecnologías son muy similares, observándose la expresión de las variantes Tv1-RAGE, *Rnor*RAGEv1 y v2 (Figura 36, Figura 31 y Figura Suplementaria 2). Sin embargo, es importante destacar que todos los

<i>Macaca mulatta</i>			<i>Bos taurus</i>		
Muestras:	Cerebro	Hígado	Muestras:	Cerebro	Hígado
<i>Mmul</i> RAGEv1	6,275	3,427	<i>Btau</i> RAGEv1	4,735	8,371
<i>Mmul</i> RAGEv2	10,676	4,186	<i>Btau</i> RAGEv2	2,006	1,277
<i>Mmul</i> RAGEv3	3,051	4,297	<i>Btau</i> RAGEv3	1,650	7,160
<i>Mmul</i> RAGEv4	0	0	<i>Btau</i> RAGEv4	2,285	0
<b><i>Mmul</i> RAGEv5</b>	1,437	3,488	<i>Btau</i> RAGEv5	0,000	1,114
<i>Mmul</i> RAGEv6	8,224	2,159	<i>Btau</i> RAGEv6	0,000	0
<i>Mmul</i> RAGEv7	0	0	<b><i>Btau</i> RAGEv7</b>	0,000	0,000
<i>Mmul</i> RAGEv8	0	0,236	<i>Btau</i> RAGEv8	0,000	0,570
<i>Mmul</i> RAGEv9	0	0	<i>Btau</i> RAGEv9	0	0
<b>Tv1-RAGE</b>	4,185	0,717	<i>Btau</i> RAGEv10	0,000	0,000
<i>Mmul</i> RAGEv11	0,305	1,838	<b>Tv1-RAGE</b>	1,479	0,000
<i>Mmul</i> RAGEv12	0,937	0	<i>Btau</i> RAGEv11	0,000	0,000
<i>Mmul</i> RAGEv13	0	0	<i>Btau</i> RAGEv12	1,479	0,000
<i>Mmul</i> RAGEv14	0	0	<i>Btau</i> RAGEv13	1,622	1,753
<i>Mmul</i> RAGEv15	0	0	<i>Btau</i> RAGEv14	0	0
			<i>Btau</i> RAGEv15	0	0
			<i>Btau</i> RAGEv16	0	0
			<i>Btau</i> RAGEv17	0	0
			<i>Btau</i> RAGEv18	0	0
			<i>Btau</i> RAGEv19	0	0
			<i>Btau</i> RAGEv20	0	0
			<i>Btau</i> RAGEv21	0	0
			<i>Btau</i> RAGEv22	0	0
			<i>Btau</i> RAGEv23	0	0

<i>Mus musculus</i>			<i>Rattus norvegicus</i>		
Muestras:	Cerebro	Hígado	Muestras:	Cerebro	Hígado
mRAGE_v5	3,111	1,435	<i>Rnor</i> RAGEv1	32,580	2,488
<b>mRAGE_v2</b>	0,283	0,000	<b><i>Rnor</i> RAGEv2</b>	5,065	0,521
<i>Mmus</i> RAGEv1	0	0,246	<b>Tv1-RAGE</b>	23,267	2,952
<b>Tv1-RAGE</b>	4,251	1,014	<i>Rnor</i> RAGEv3	0,688	0,000
mRAGE_v4	2,379	0,000	<i>Rnor</i> RAGEv4	0,000	0,000
<i>Mmus</i> RAGEv2	0,068	0,251	<i>Rnor</i> RAGEv5	0,000	0,001
<i>Mmus</i> RAGEv3	1,321	0,233	<i>Rnor</i> RAGEv6	0	0
<i>Mmus</i> RAGEv4	0	0	<i>Rnor</i> RAGEv7	0	0
<i>Mmus</i> RAGEv5	0,151	0,172	<i>Rnor</i> RAGEv8	0	0
<i>Mmus</i> RAGEv6	0,898	0,836	<i>Rnor</i> RAGEv9	0	0
<i>Mmus</i> RAGEv7	0	0	<i>Rnor</i> RAGEv10	0	0
<i>Mmus</i> RAGEv8	0	0	<i>Rnor</i> RAGEv11	0	0
<i>Mmus</i> RAGEv9	0	0	<i>Rnor</i> RAGEv12	0	0
<i>Mmus</i> RAGEv10	0	0			
<i>Mmus</i> RAGEv11	0	0			
<i>Mmus</i> RAGEv12	0	0			
<i>Mmus</i> RAGEv13	0	0			

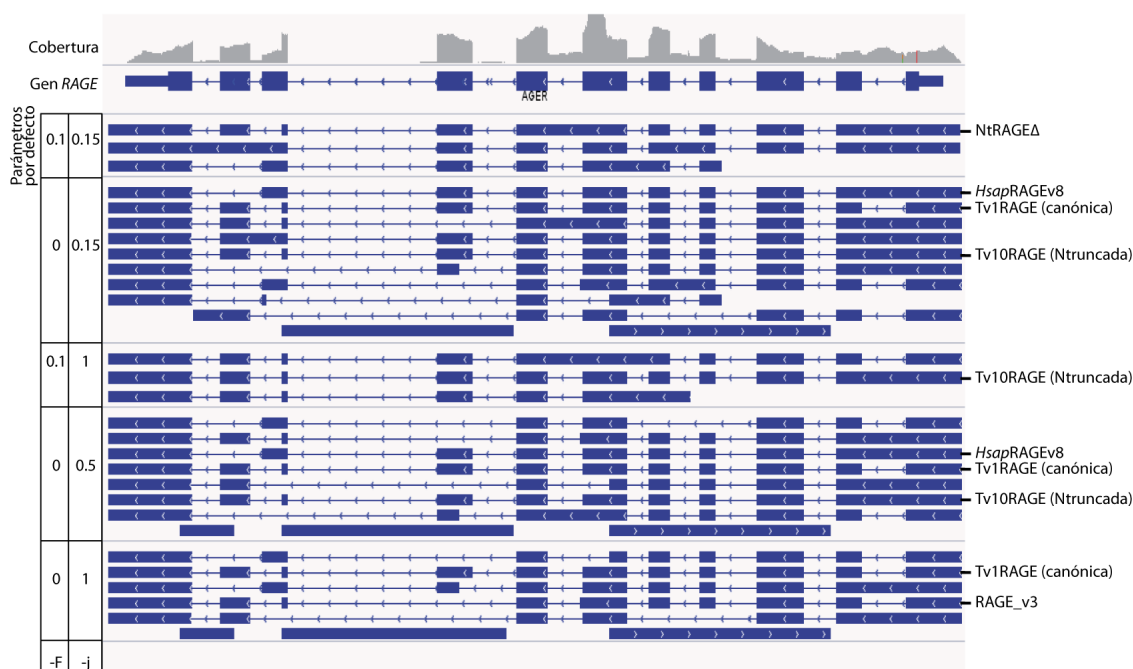
**Figura 36. Estimación por RNA-seq de la expresión de los diferentes transcritos de RAGE en diferentes especies y tejidos.** Se indica el conjunto de todas las lecturas de cerebro e hígado, unidas por tejido. Ver Figura 35 para detalles.

niveles de expresión observados por esta tecnología son extremadamente bajos. No se pudo hacer las comparaciones con cerdo pues no existían datos de RNA-seq disponibles de esta especie.

Además, hemos analizado los datos de una muestra humana de monocitos proveniente del proyecto ENCODE ya que disponía de un número de lecturas muy elevado. Usando los parámetros por defecto, Cufflinks solo era capaz de ensamblar tres transcritos (Figura 37) que no correlacionaban con aquellas variantes encontradas por RT-PCR



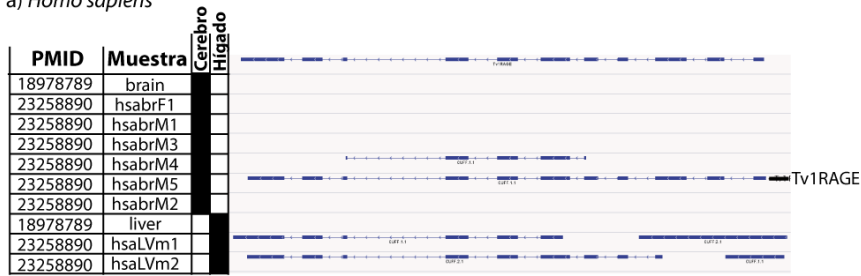
anidada, ni siquiera la isoforma canónica pudo ser ensamblada por este programa. Como Cufflinks suprime la fracción de variantes de *splicing* de menor expresión, realizamos un análisis alterando los parámetros que modificaban la fracción de isoformas mínima y la fracción de pre-mRNA. Como resultado, fuimos capaces de ensamblar las isoformas Tv1-RAGE, Tv10-RAGE (N-truncada) y *Hsap*RAGEv8 (Figura 37) en la muestra de monocitos humana, siendo las variantes más expresadas en dicha muestra (Figura 35) y, por tanto, condición indispensable para validar este tipo de análisis. Fijamos los parámetros en -F 0 y -j 0.15 para estudios en el resto de muestras.



**Figura 37. Optimización de los parámetros de Cufflinks para obtención *in silico* de las variantes de *splicing* de RAGE humanas según muestras de RNA-seq.** Se emplearon los parámetros por defecto de Cufflinks (-F 0.1 y -j 0.15), así como diferentes combinaciones de estos parámetros en la muestra de monocitos humanos (CD14+, proyecto ENCODE, SRX179434) para la predicción de las variantes de *splicing* de RAGE. Las variantes se representaron mediante el visualizador IGV.

Posteriormente, extendimos el abordaje a todas las muestras de cerebro e hígado previamente analizadas en las diferentes especies (Figura 38). La isoforma Tv1-RAGE se detectó únicamente en una muestra de cerebro humano. Las variantes relacionadas con la N-truncada humana se ensamblaron en cerebro de macaco (*Mmul*RAGEv5) y en ambos tejidos bovinos (*Btau*RAGEv7). La variante *Mmul*RAGEv6 se ensambló también en cerebro de macaco (Figura 38). Sin embargo, no se consiguió ensamblar ninguna isoforma en las muestras de roedores.

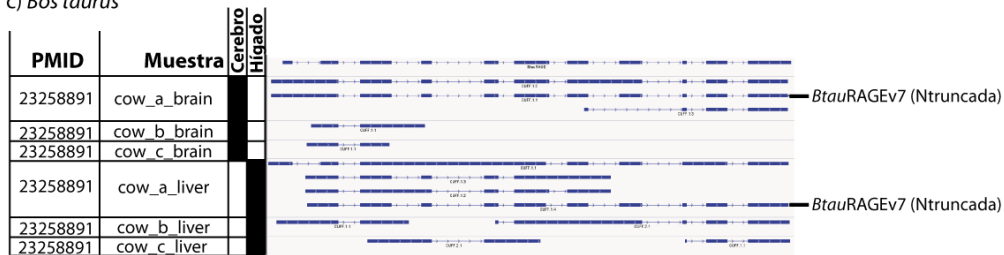
a) *Homo sapiens*



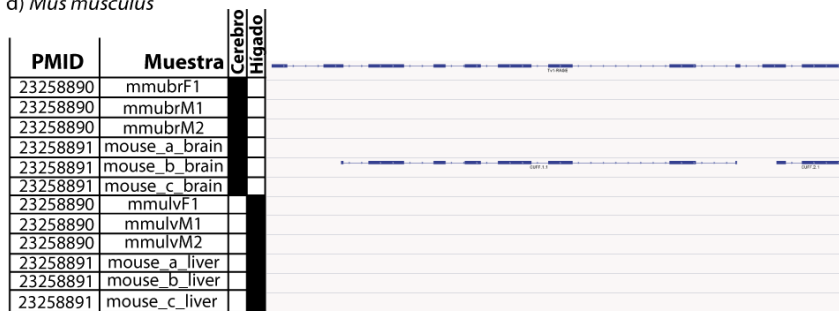
b) *Macaca mulatta*



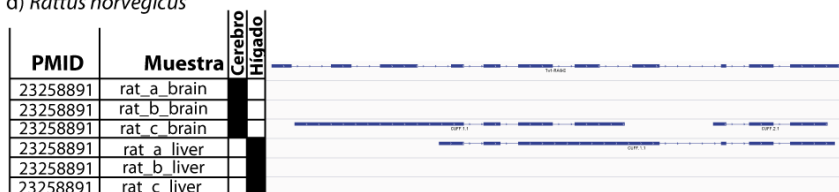
c) *Bos taurus*



d) *Mus musculus*



d) *Rattus norvegicus*

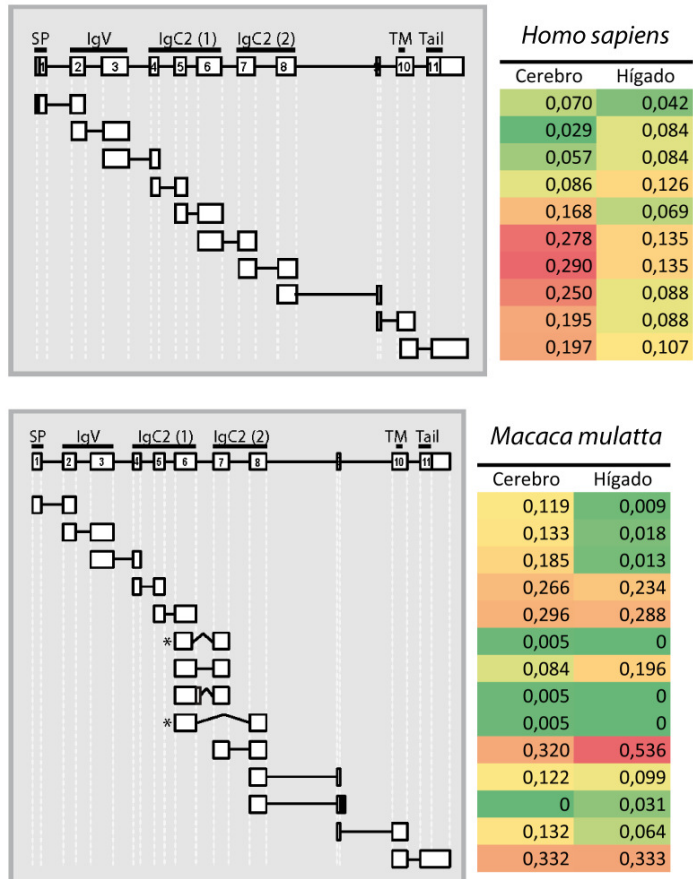


**Figura 38. Variantes de *splicing* de RAGE obtenidas *in silico* en las muestras de RNA-seq en las diferentes especies. Se emplearon los parámetros  $-F 0$  y  $-j 0.15$  de Cufflinks en muestras de cerebro e hígado de las diferentes especies a estudio [PMID:18978789 – Pan et al. (2008); PMID:23258890 – Barbosa-Morais et al. (2012); PMID:23258891 – Merkin et al. (2012)]. Las variantes se representaron mediante el visualizador IGV.**

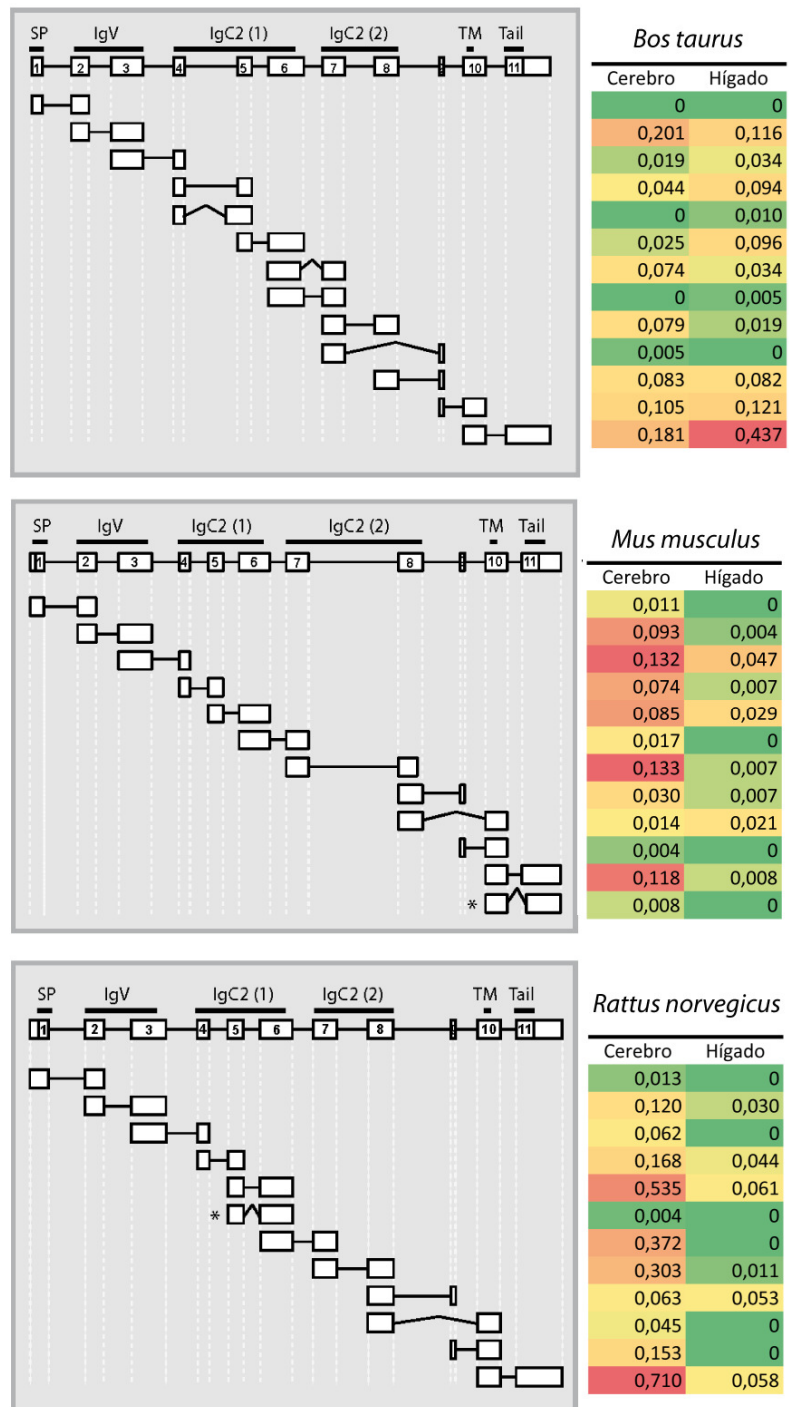
Dado que los análisis de Cufflinks no fueron capaces de ensamblar todas las isoformas detectadas por RT-PCR, ni siquiera las expresadas en mayor proporción, se decidió analizar el número de eventos de *splicing* detectados por TopHat gracias al archivo que genera de uniones de exones (Figura 39). En estas condiciones fue posible detectar la mayoría de los eventos de *splicing* canónicos (Figura Suplementaria 3), pero solo en aquellas muestras que tenían una cobertura de lecturas mayor. En macaco además detectamos algunos eventos no canónicos e incluso dos nuevas entre los exones 6-7 y 6-8 (Figura 39) pero con unos niveles de expresión muy bajos.

En bovino es interesante prestar atención sobre el evento de *splicing* entre el exón 1 y 2, pues no se encuentra presente en ninguna muestra de bovino, mientras que en los resultados de PCR aparecían dos variantes de *splicing* con dicho evento aunque con unos niveles de expresión muy bajos en comparación con el resto de las isoformas (Figura 28). Por esta razón, no se pudo ensamblar la isoforma Tv1-RAGE usando Cufflinks, y la isoforma equivalente a la N-truncada humana fue la única isoforma ensamblada (Figura 38). Además, hemos detectado pocos eventos de *splicing* alternativos en todas las especies (Figura 39 y Figura Suplementaria 3) y todos ellos con unos niveles de expresión extremadamente bajos. Algunos de estos eventos mostraron niveles entre 0,004 y 0,008

**Figura 39. Cuantificación de los diferentes eventos de *splicing* de RAGE en muestras de RNA-seq agrupadas por tejido.** Se cuantificaron el número de uniones de exones por millón de lecturas alineadas en las especies indicadas, agrupándose las cuantificaciones por tejido. Las uniones de exones presentes en la variante canónica Tv1-RAGE se muestran como líneas rectas, mientras que las uniones de exones alternativas se muestran como triángulos abiertos. El asterisco identifica los eventos de *splicing* no obtenidos por RT-PCR anidada. Los niveles de expresión se muestran en una escala de gradiente de verde (poca expresión) a roja (alta expresión).



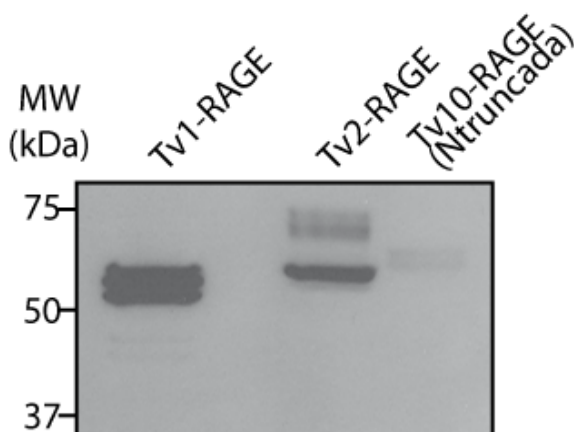
lecturas por millón de lecturas mapeadas, lo que corresponde con solo una o dos uniones encontradas en el total de la muestra. Sin embargo, estos eventos muestran unos patrones de expresión por tejido similares a los datos experimentales. Por ejemplo, la retención parcial del intrón 6 en las muestras de mono Rhesus es levemente detectado en cerebro, al igual que las variantes que contienen este evento de *splicing* (*Mmul*RAGEv1, v4, v7, y v9) (Figura 27). La misma consistencia en los resultados se observó en el evento de *splicing* que saltaba el exón 9 en las muestras de rata, con niveles bajos en cerebro observados por ambos abordajes (Figura 39 cont. y Figura 31).



**Figura 39 (cont.).** Cuantificación de los diferentes eventos de *splicing* de RAGE en muestras de RNA-seq agrupadas por tejido.

#### 4.1.4. Caracterización de las principales variantes de *splicing* de RAGE humanas a nivel proteico

Una vez realizados todos los estudios de expresión de las diferentes variantes en tejidos y especies, se procedió a continuar el estudio a nivel proteico y para ello se seleccionaron las tres variantes de *splicing* que se expresaban a mayores niveles en las muestras de humano. Estas variantes eran la isoforma canónica Tv1-RAGE, la variante Tv2-RAGE que presenta la incorporación de 16 aminoácidos adicionales en el segundo dominio IgC, así como la variante Tv10-RAGE o también denominada N-truncada debido a la retención del intrón 1 y un PTCs en su secuencia. Estas proteínas fueron etiquetadas en su N-terminal con el epítipo V5 o Cmyc en el caso de las variantes Tv1-RAGE y Tv2-RAGE, así como el epítipo V5 en el C-terminal de la variante Tv10-RAGE. Para el estudio de estos tres transcritos/proteínas se procedió a su sobre-expresión *in vitro* en la línea celular HEK293T y, posteriormente, al análisis por western blot con los extractos proteicos obtenidos (Figura 40).

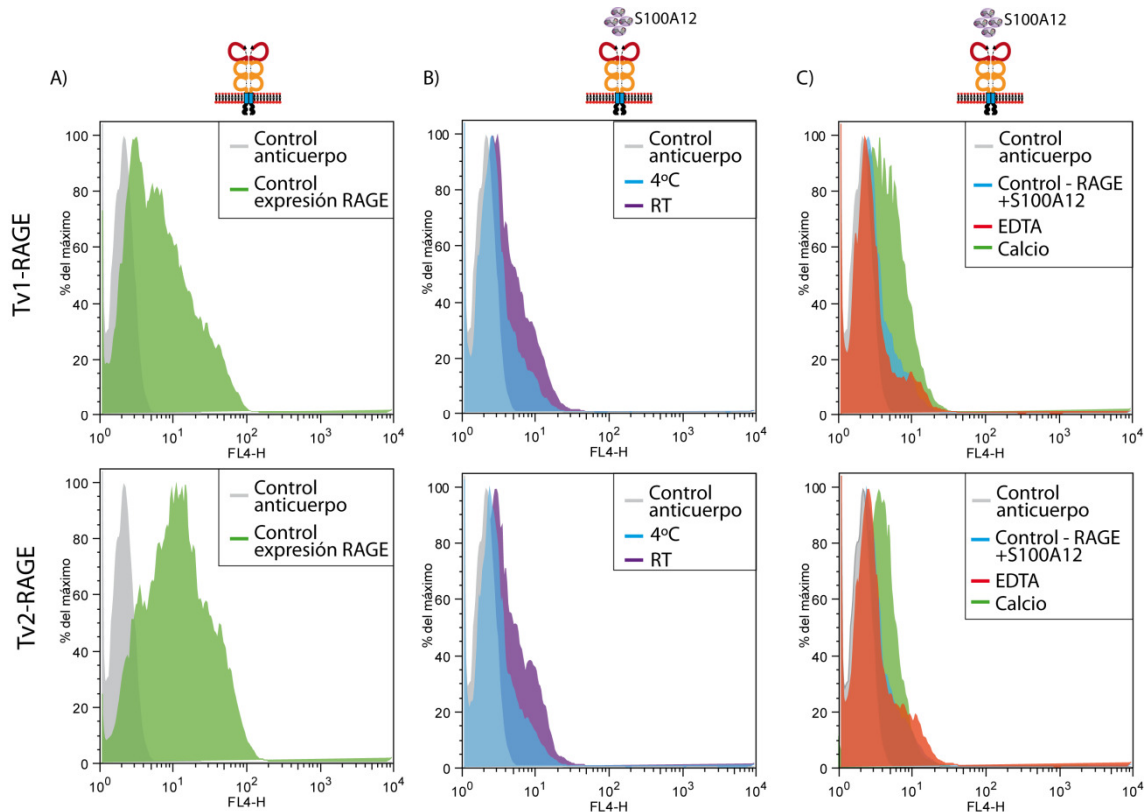


**Figura 40. Expresión de las isoformas del gen RAGE.** Análisis por Western blot en condiciones reductoras de la sobreexpresión de las isoformas Tv1-RAGE, Tv2-RAGE y Tv10-RAGE (N-truncada) en células HEK293T.

Los análisis por Western blot bajo condiciones desnaturalizantes nos muestran que la proteína Tv1-RAGE presenta tres bandas de pesos moleculares superiores a 50 kDa lo que aumenta ligeramente el tamaño esperado por las predicciones bioinformáticas de la proteína recombinante (44.2 kDa) (Figura 40). Además, la proteína Tv2-RAGE mostraba un tamaño superior a la isoforma canónica Tv1-RAGE en una única banda mayoritaria asociado a los 16 aminoácidos adicionales que dispone. Se observa también la presencia de dos pequeñas bandas de mayor peso molecular que podrían ser debidas a eventos de glicosilación, ya que según predicciones bioinformáticas estos aminoácidos presentan dos potenciales sitios de glicosilación. Respecto a la proteína N-truncada (Tv10-RAGE) se observa una ligera expresión de dos bandas pero con un tamaño superior a la proteína Tv1-RAGE. Esto podría ser debido a una diferente construcción con la etiqueta en el extremo C-terminal (Tv10-RAGE) que se añadió para no alterar la retención del intrón, en lugar de en

el extremo N-terminal (Tv1-RAGE). Los bajos niveles de expresión de esta isoforma podrían estar debidos a una supuesta regulación por NMD al presentar un PTC en la secuencia retenida del intrón 1. Sin embargo, esta proteína recombinante no se expresa según el tamaño obtenido por predicciones bioinformáticas (37 kDa) si la proteína se comenzase a traducir desde el segundo ATG localizado en el exón 3.

Estudios previos han indicado que la proteína RAGE es capaz de unirse a la familia S100 (Hofmann et al. 1999), entre otro tipo de ligandos proteicos. Esta interacción se realiza en los dos primeros dominios Ig, pero a través de estos estudios no es posible identificar las variantes a las que se está uniendo el ligando, o si dicha interacción es específica de una variante en concreto. Dado que las isoformas Tv1-RAGE y Tv2-RAGE muestran diferencias en el segundo dominio Ig, en el presente trabajo se quiso comparar su interacción con el ligando S100A12. Para ello se realizaron ensayos de interacción mediante citometría de flujo con células sin permeabilizar sobre-expresando las diferentes construcciones de las variantes y sus etiquetas. De esta forma se observaría la interacción directa únicamente con aquellos receptores que se encontrasen situados en la membrana

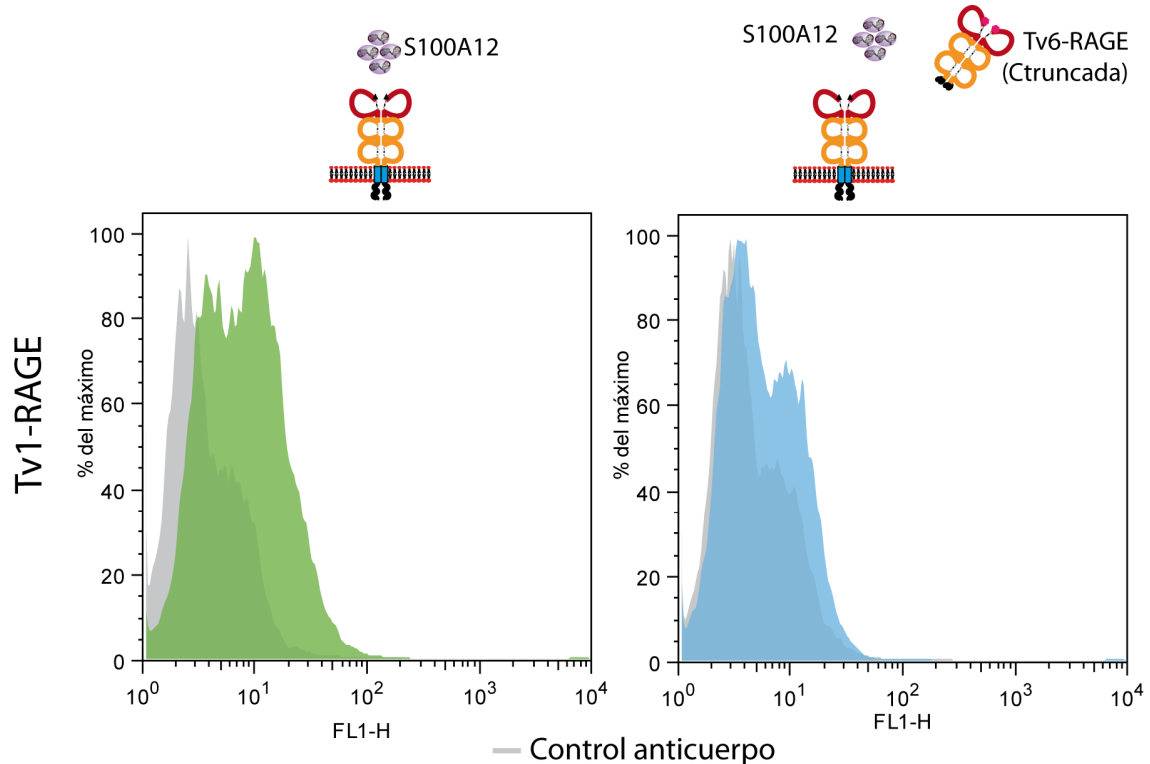


**Figura 41. Estudio de la interacción de la proteína recombinante S100A12 con los receptores Tv1-RAGE y Tv2-RAGE.** A) Control de la sobre-expresión de las proteínas Tv1-RAGE y Tv2-RAGE, respectivamente, con la etiqueta V5 en las células HEK293T. B) Interacción de las células sobre-expresando Tv1-RAGE y Tv2-RAGE con de la proteína recombinante S100A12 a distintas temperaturas. C) Implicación del catión calcio en la unión de la proteína recombinante al receptor a 4°C. Como control de anticuerpo se empleó anti-mIgG unido a Alexa 647.



celular. Se empleó el ligando S100A12 purificado a partir del sistema de baculovirus, incubándolo con células previo paso a la realización del ensayo de citometría.

Primeramente se comprobaron los niveles de sobre-expresión de las proteínas Tv1-RAGE y Tv2-RAGE (Figura 41A). Las células HEK293T sobre-expresando en su membrana las dos isoformas RAGE, respectivamente, se incubaron con la proteína recombinante S100A12 en dos condiciones de temperatura, 4°C y RT. Con ambos receptores Tv1-RAGE y Tv2-RAGE se encontraban unos niveles de interacción ligeramente superiores al aumentar la temperatura (Figura 41B). Dado que la interacción de familia de proteínas S100A12 con sus receptores parece estar modulada por la presencia de calcio soluble en el medio, realizamos dos puntos en el que primero se añadió calcio y, en el segundo se añadió un agente quelante de cationes divalentes (EDTA). Se observó como los niveles de interacción aumentaban considerablemente cuando se añadía calcio al medio y esta interacción era inhibida cuando se añadía el agente quelante durante el periodo de incubación del receptor con los ligandos (Figura 41C). Esta interacción era ligeramente superior en las células que sobre-expresaban la variante canónica Tv1-RAGE en detrimento de aquellas células que sobre-expresaban la variante Tv2-RAGE con el dominio IgC alterado.



**Figura 42. Análisis de la interacción de las proteínas S100A12 y Tv6-RAGE con el receptor RAGE.** Células HEK293 que sobre-expresaban Tv1-RAGE o un plásmido control (pZCV) fueron incubadas con la proteína S100A12 (A) y con ambas proteínas S100A12 y C-truncada (B) en presencia de calcio y a RT. Se empleó anti-mIgG unido a Alexa 488 como anticuerpo control.

Una vez establecidas las condiciones, realizamos el experimento de interacción incorporando los mejores parámetros obtenidos anteriormente. La incubación de las proteínas recombinantes S100A12 se realizó a RT y añadiendo calcio al medio utilizado. De esta forma obtuvimos unos niveles de interacción superiores de la proteína S100A12 con el receptor canónico Tv1-RAGE (Figura 42A). Además, se repitió el estudio añadiendo la isoforma soluble C-truncada (Tv6-RAGE), purificada igualmente a partir del sistema de baculovirus, con la proteína recombinante S100A12 en células que sobre-expresaban el receptor RAGE canónico. Se observa como la interacción del ligando S100A12 disminuye cuando se añade la proteína RAGE soluble por una posible competición del receptor soluble con el ligando (Figura 42B).

## **4.2. ESTUDIO DEL FENÓMENO DE QUIMERISMO ENTRE LOS GENES *G6F* Y *LY6G6D* COMO MODELO**

En estudios previos realizados en nuestro grupo de investigación se había determinado la presencia de quimeras entre los genes *G6F* y *LY6G6D* del MHC de clase III humano, así como la presencia de determinadas variantes de *splicing* tanto en la quimera como en el gen *LY6G6D* (Calvanese et al. 2008). Ambos genes se encuentran localizados muy próximos y en la misma orientación, características básicas de los fenómenos de quimerismo en cis. Sin embargo, consecuencia de la alta densidad génica presente en el MHC de clase III encontramos localizado otro gen (*LY6G6E*) entre los genes que forman la quimera, orientado en el sentido opuesto a estos. Por todo ello, consideramos este conjunto génico como buen candidato para el estudio de su quimerismo a nivel transcriptómico en diferentes especies de mamífero y así poder observar su nivel de conservación en las distintas especies, además de estudiar las isoformas a nivel proteómico para determinar la funcionalidad de dichas variantes quiméricas. Es importante destacar que esta región génica se encuentra en una zona asociada con susceptibilidad a padecer Artritis Reumatoide, tema de interés en esta Tesis.

### **4.2.1. Análisis de la presencia de quimeras en los genes *G6F* y *LY6G6D* en diferentes especies de mamífero.**

El estudio de la formación de quimeras así como el estudio del AS de los genes *G6F* y *LY6G6D* se realizó mediante la técnica de PCR anidada, clonaje de cada una de las variantes observadas y secuenciación. Las muestras empleadas en la caracterización de las quimeras en mamíferos fueron obtenidas de RNA total de muestras comerciales de los



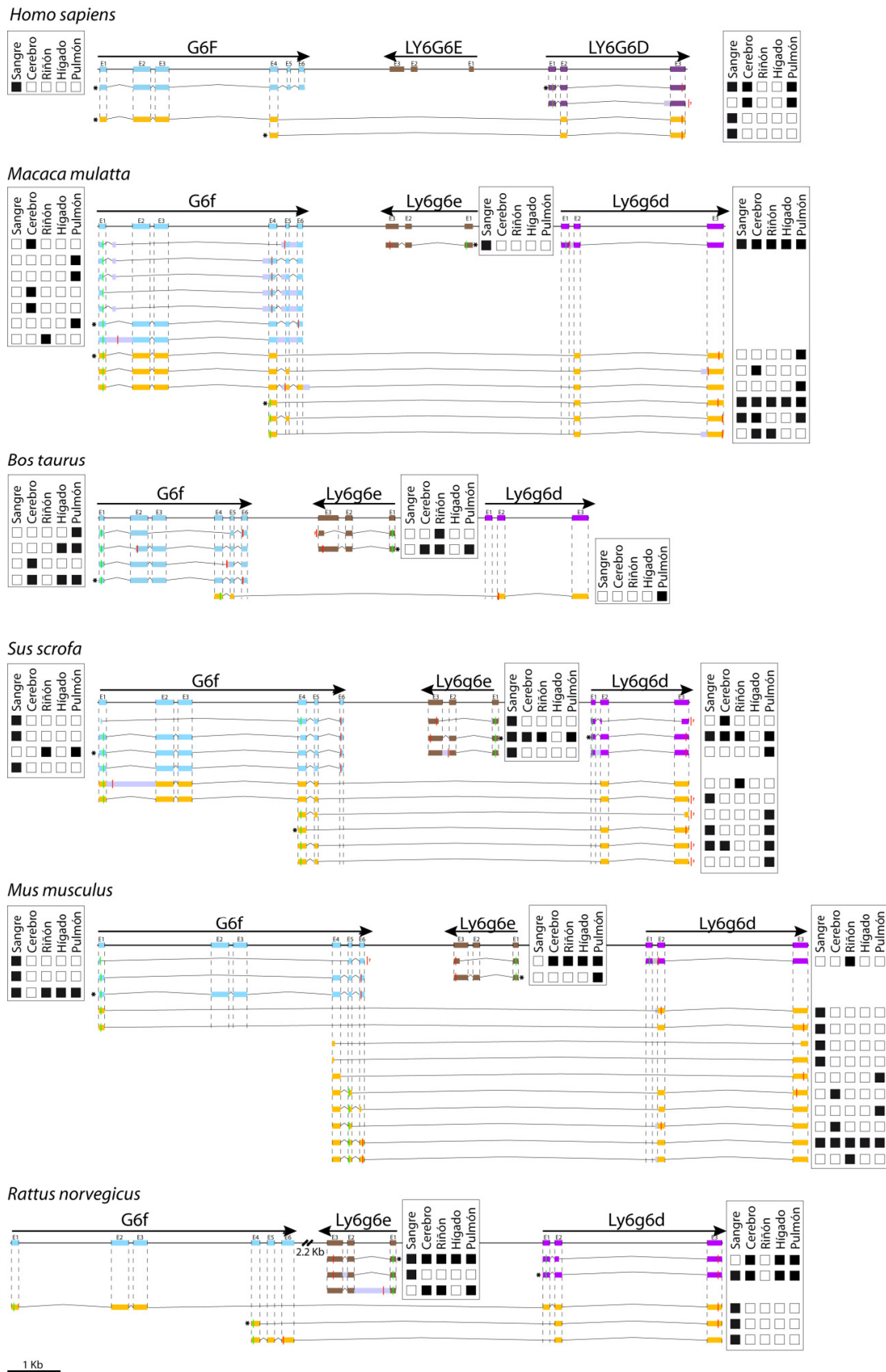


234), y una región transmembrana (aminoácidos 235-257) que genera una proteína transmembrana de tipo I y paso único, y una cola citoplasmática de 40 aminoácidos sin homología con ningún dominio conocido. Aunque carece de dominio conservado, esta cola citoplasmática contiene un residuo de tirosina (Y281) que se fosforila dependientemente de las proteínas GRB2 y GRB7 (De Vet et al. 2003) (Figura 43).

Tras análisis por RT-PCR fuimos capaces de detectar esta isoforma canónica en todas las especies analizadas excepto en rata, especie en la que *G6F* no presenta ningún transcrito en los tejidos analizados. En humanos solo se detectó expresión de la isoforma canónica y únicamente en sangre, mientras que en las demás especies había más variabilidad de variantes y expresión, encontrándose también la variante canónica en sangre solo en ratón. Se encontraron también variantes de AS en *Macaca mulatta*, *Sus scrofa*, *Bos taurus* y *Mus musculus* con especificidad de tejido (Figura 44). En *Macaca mulatta* observamos predominantemente retención de intrones que generan PTCs y la aparición de un nuevo exón en la secuencia del intrón 1. Sin embargo, en las otras tres especies estudiadas con expresión de *G6F*, encontramos preferentemente la exclusión de exones al completo o parciales, lo que genera principalmente la pérdida del dominio IgV o la región transmembrana en la proteína madura (Figura 44).

#### **4.2.1.2. Identificación y caracterización de *LY6G6E* y sus variantes de *splicing*.**

El gen *LY6G6E* fue previamente descrito como un pseudogen en humano (Mallya et al. 2002) careciendo de la presencia de ESTs. En consonancia, nosotros no hemos sido capaces de detectar la expresión de este gen en ninguno de los tejidos humanos analizados. Sin embargo, hemos detectado diferentes variantes de *splicing* en el resto de las especies estudiadas, encontrándose la variante canónica expresada ampliamente en la mayoría de los tejidos de las diferentes especies (excepto en ratón que únicamente se encuentra expresado en pulmón y en macaco que solo se expresa la canónica y en sangre). Como característica común a la familia de proteínas Ly6 previamente descrita por Mallya et al. (2002), y Calvanese et al. (2008), algunas variantes de AS de *Ly6g6e* en rata y cerdo presentan la retención de un intrón (Figura 44), lo que provoca la aparición de PTCs y, por tanto, la producción de variantes no codificantes. La retención del primer intrón se produce en varios tejidos de rata (cerebro, riñón y pulmón) mientras que la retención del segundo intrón se produce en ambas especies pero únicamente en la muestra de sangre.



**Figura 44. Representación esquemática de las variantes de *splicing* de *G6f*, *Ly6g6e*, *Ly6g6d* y quimeras en diferentes especies de mamífero. Las variantes canónicas se muestran con un asterisco. Los diferentes transcritos están organizados por tamaño, y representados a escala. Se muestra expresión en los tejidos en cuadrados negros y ausencia de expresión en blancos.**

#### **4.2.1.3. Identificación y caracterización de LY6G6D y sus variantes de splicing.**

La proteína LY6G6D pertenece a la superfamilia de proteínas Ly6, y presenta un péptido señal (aminoácidos 1-19) y un dominio u-PAR/Ly-6 rico en cisteínas (aminoácidos 24-103), según las herramientas bioinformáticas Pfam, SMART y SignalP (Figura 43). El dominio Ly-6 está caracterizado por la conservación de ocho o diez residuos de cisteína que forman un patrón definido de puentes disulfuro. La isoforma canónica de humano presenta ocho residuos de cisteína conservados mientras que curiosamente el resto de los ortólogos estudiados presentan diez residuos de cisteína (Mallya et al. 2006). Además, LY6G6D es una proteína de superficie celular anclada por GPI a partir del residuo de serina S104 (Mallya et al. 2006) (Figura 43). En este estudio hemos sido capaces de detectar la isoforma canónica únicamente en humano, cerdo y rata. Es importante destacar como todos los tejidos del mono Rhesus y la muestra de riñón en ratón presentan la retención del intrón 1 de Ly6g6d, característica común de las proteínas de la familia Ly6 (Calvanese et al. 2008; Mallya et al. 2002), provocando la aparición de un PTC en el intrón 1 ó en el exón 2. Ninguna transcrito de Ly6g6d fue encontrado en los tejidos analizados de *Bos taurus*.

#### **4.2.1.4. Identificación y caracterización de los transcritos quiméricos G6F-LY6G6D.**

Para la caracterización de los transcritos quiméricos G6f-Ly6g6d, inicialmente se realizó un estudio bioinformático buscando evidencias tanto a nivel de transcritos como a nivel proteico de la existencia de quimeras entre los genes *G6f* y *Ly6g6d*. Se observó la predicción de la secuencia de un transcrito, así como la presencia de una secuencia de proteína denominada MEGT1 que compartía exones de ambos genes a estudio. Estas predicciones podrían traducirse en una proteína quimérica (denominada en este estudio como Igch) que contenía la región extracelular de G6F incluyendo el péptido señal y los dominios Ig (exones 1-3), así como la región transmembrana (exón 4). Además, el transcrito quimérico también contaba con la secuencia de los exones 2 y 3 del gen *LY6G6D* que podría codificar un dominio Ly6 (Figura 43). Además, un estudio previo había mostrado la presencia de un transcrito quimérico de menor tamaño que se encontraba expresado en diferentes líneas celulares (Calvanese et al. 2008). En dicho estudio se realizaron ensayos de RT-PCR para amplificar tanto desde el exón 1 como desde el exón 4 de *G6F* hasta el exón 3 de *LY6G6D* en líneas celulares de humano. La amplificación a partir del exón 1 mostró la presencia de un transcrito quimérico largo (Igch) con una distribución completamente específica de tejido a lo largo de las muestras analizadas. En cambio, la

amplificación a partir del exón 4 y, por tanto, la potencial quimera corta (shch) estaba ampliamente expresada en la prácticamente todas las líneas celulares (Calvanese et al. 2008).

Esta distribución de la quimera larga y corta se ve corroborada en nuestros análisis por RT-PCR anidada (Figura 44). El estudio detallado de la quimera larga revela la expresión de esta isoforma en las muestras de sangre de humano y de pulmón de mono Rhesus. Además, se identificaron otros transcritos quiméricos que presentaban modificaciones tales como exclusión de determinados exones o inclusión de intrones. Una modificación relevante se observa en *Macaca mulatta* y *Sus scrofa*, y consiste en la inclusión también del exón 5 de G6f además del resto de exones del transcrito quimérico canónico. Esta nueva variante contendría el residuo de tirosina (Y281) responsable de las interacciones intracelulares en una manera dependiente de fosforilación, lo que podría estar variando la función de esta proteína quimérica en relación con la proteína quimérica canónica que no contiene este residuo potencialmente fosforilable. Existe otra isoforma en macaco que incluye también el exón 6 y parte de la región intergénica entre *G6f* y *Ly6g6d*, que generaría una isoforma proteica idéntica a G6f pero posiblemente regulada de manera diferente al tener un 3'UTR de mayor tamaño.

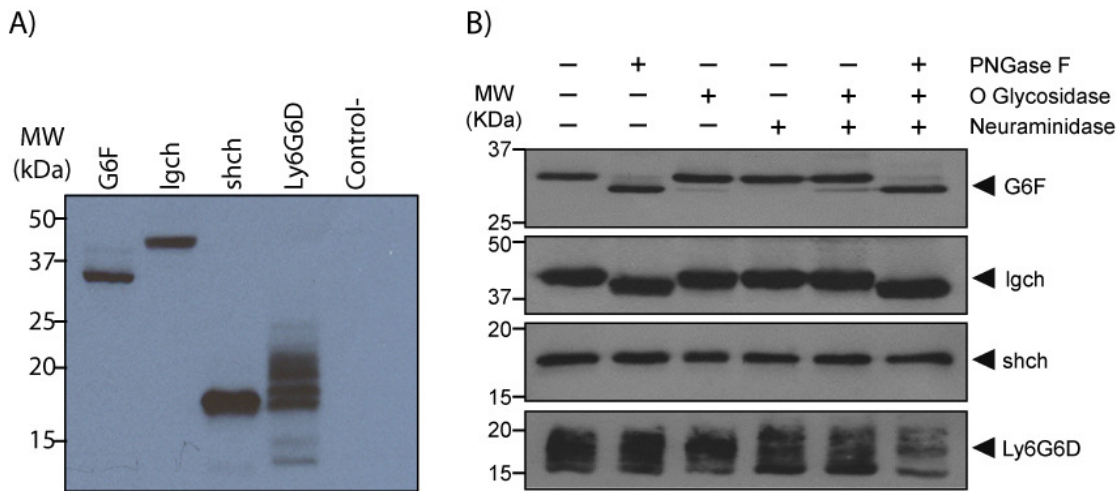
El estudio detallado de la quimera corta revela una expresión más amplia por la mayoría de los tejidos y especies analizadas si la comparamos con la quimera larga. La quimera shch contiene el exón 4 de G6f y los exones 2 y 3 de Ly6g6d, lo que resultaría en una posible región transmembrana o péptido señal seguida de un dominio Ly-6 y un residuo de serina que previsiblemente podría estar generando un anclaje de GPI a la superficie celular (Figura 43). Es importante destacar la presencia de este tipo de variante quimérica ampliamente expresada, mientras que la isoforma canónica de Ly6g6d no se encuentra expresada en determinadas especies (*Macaca mulatta*, *Bos taurus*, y *Mus musculus*), por lo que podría estar siendo reemplazada, ejerciendo ambas proteínas funciones similares ya que la proteína madura que codifican es teóricamente la misma. La isoforma shch canónica se expresa en humano, macaco, cerdo y ratón, y no hemos detectado isoformas que retengan el primer intrón. Además, hemos sido capaces de detectar otras doce variantes de AS generadas a partir del exón 4 de G6f, con eventos de *splicing* no presentes en las quimeras largas. Estos resultados parecen indicar que ambos transcritos quiméricos están regulados de una manera completamente diferente y, que la presencia de un inicio de transcripción alternativo podría estar implicado en esta expresión.

Para la identificación de un posible sitio de inicio de la traducción alternativo cercano al exón 4 de G6f, realizamos un estudio mediante la técnica de 5' RLM-RACE en la línea celular humana K562. Gracias a este método aseguramos la amplificación completa de los 5' terminales de los mensajeros para su estudio. En nuestro ensayo hemos identificado un posible inicio de la transcripción localizado en el intrón 3 de G6F humano, más detalladamente en la posición -30 empezando a contar desde el inicio del exón 4 de G6F (Figura 43).

#### 4.2.2. Caracterización de las principales variantes quiméricas a nivel proteico

Para analizar las proteínas G6F, LY6G6D, así como las proteínas quiméricas canónicas, y proceder a su caracterización, se han sobre-expresado *in vitro* en la línea celular HeLa. Estas proteínas fueron etiquetadas en su N-terminal con el epítipo V5 justo después de cada respectivo péptido señal. Los análisis por Western blot indicaron un peso molecular de aproximadamente 35 kDa para la proteína G6F cuando se corre un SDS/PAGE bajo condiciones desnaturalizantes (Figura 45A), tamaño similar a las predicciones bioinformáticas para la proteína G6F etiquetada con V5 (34.5 kDa) y a estudios previos (De Vet et al. 2003). Las proteínas quiméricas lgch y shch presentaron un peso molecular de 43 y 16.4 kDa, respectivamente, de manera análoga a las predicciones bioinformáticas de 42.9 y 17.6 kDa, respectivamente (Figura 45A). Asimismo, la proteína LY6G6D mostraba una variedad de bandas de entre 16 y 19 kDa, y alrededor de 23 kDa bajo condiciones desnaturalizantes (Figura 45A), con datos similares a los reportados por Mallya et al. (2006). *LY6G6E* está descrito como un pseudogen en humano (Mallya et al. 2002) y por ese motivo este transcrito no fue analizado en el estudio a nivel proteico.

Como las modificaciones post-traduccionales como glicosilaciones pueden ser de gran relevancia en la función de la proteína (Moller et al. 1993), se realizaron tratamientos de deglicosilación en las proteínas G6F, LY6G6D y las proteínas quiméricas para su estudio (Figura 45B). Se observó como la proteína G6F estaba claramente N-glicosilada, dato consistente con estudios previos (De Vet et al. 2003), así como con la presencia de sitios de N-glicosilación consensos (Asn-X-Ser/Thr) en las posiciones de aminoácidos 88-90. Ya que la proteína quimérica de mayor tamaño comparte estos sitios de N-glicosilación consenso en su secuencia, en el experimento de deglicosilación se observó el mismo patrón de N-glicosilación (Figura 45B2). En cambio, los tratamientos de deglicosilación en la proteína LY6G6D muestran modificaciones en residuos de ácidos siálico y O-glicosilaciones (Figura

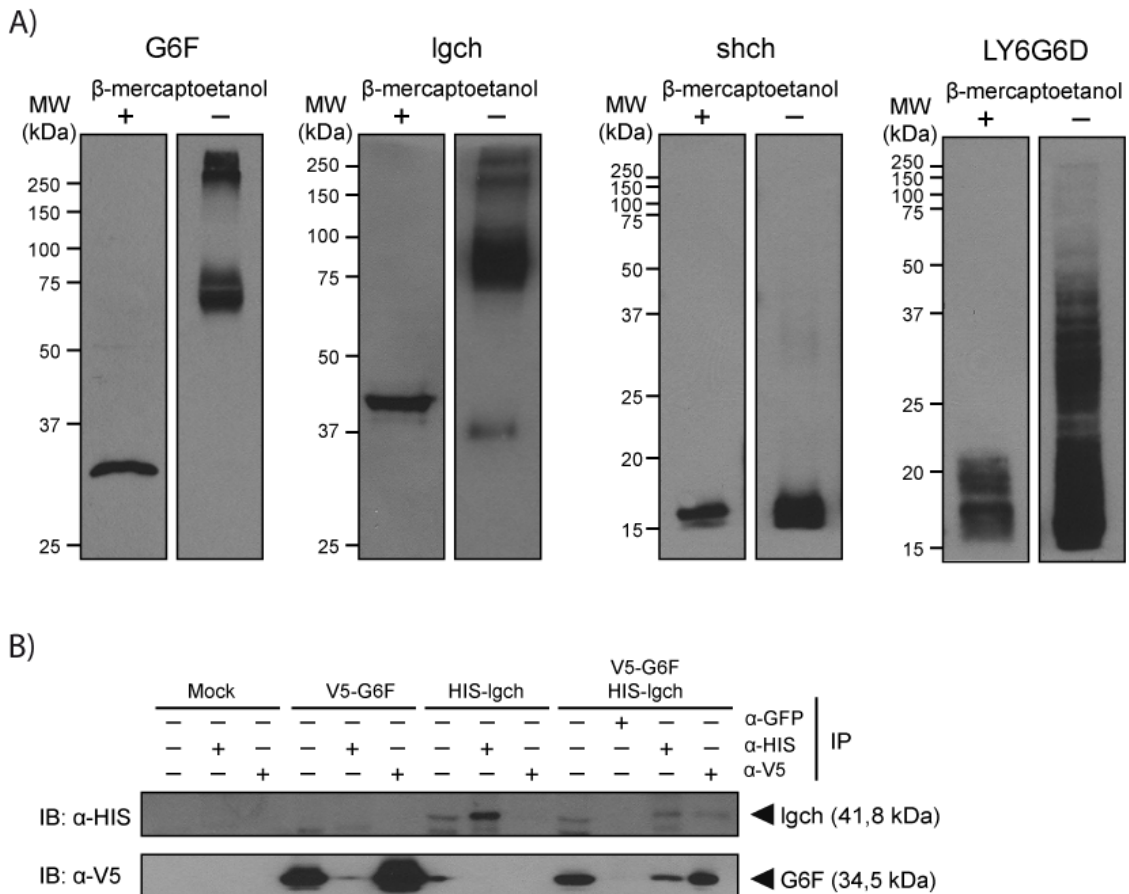


**Figura 45. Caracterización proteica de G6F, LY6G6D y quimeras humanas.** Las proteínas G6F, quimera larga (Igch), quimera corta (shch) y LY6G6D con la etiqueta V5 fueron sobre-expresadas en células HeLa. A) Análisis por western blot de lisados celulares que sobre-expresaban las proteínas. B) Tratamiento de deglicosilación de las proteínas con una enzima o combinación de ellas según se indica en la figura. Se empleó anti-V5 en todos los paneles. Los marcadores de peso molecular se indican a la izquierda de cada panel.

45B), encontrándose la mayor diferencia en peso molecular tras la deglicosilación con ambas enzimas, O-glicosilasa y neuraminidasa, tal y como se describía en estudios previos (Mallya et al. 2006). Sin embargo, no se encontró evidencias de la presencia de ningún tipo de glicosilación en la proteína quimérica de menor tamaño (Figura 45B).

Posteriormente, se procedió a estudiar la posible formación de dímeros por estas proteínas ya que estudios previos indicaban la presencia de homodímeros en la proteína G6F (De Vet et al. 2003), pudiendo aportar unas propiedades fundamentales para la función del receptor. En dicho estudio se mostraba la proteína en un peso molecular aproximado de 29 kDa bajo condiciones desnaturalizantes, mientras que bajo condiciones no desnaturalizantes su tamaño se doblaba (aproximadamente 60 kDa). Esto indicaría la formación de homodímeros mediante puentes disulfuro. Nuestros resultados son consistentes con dicho estudio mostrando el homodímero en la proteína G6F (Figura 46A), y mostrando además la formación de un multímero de alrededor de 250kDa. Para comparar si la quimera larga presentaba el mismo comportamiento realizamos el mismo análisis bajo condiciones no desnaturalizantes observando un homodímero con aproximadamente el doble de tamaño (85 kDa) de la proteína quimérica individual (Figura 46A). Los resultados con la quimera larga también nos permitieron observar la formación de un posible multímero. Además, para investigar si la quimera larga podría reconocer directamente e interactuar con la proteína G6F generando un heterodímero, realizamos co-inmunoprecipitaciones sobre-expresando ambas proteínas simultáneamente pero con





**Figura 46. Formación de dímeros entre las proteínas G6F, LY6G6D y quimeras.** A) Estudio de la formación de homodímeros y/o multímeros mediante el análisis de los lisados proteicos bajo condiciones reductoras (+) o no reductoras (-). Los marcadores de peso molecular se indican a la izquierda de cada panel. B) Estudio de la formación de heterodímeros entre G6F y la quimera larga. Se transfectaron células HeLa con el vector vacío (*mock*), G6F con la etiqueta V5 (V5-G6F), y la quimera larga con la etiqueta His (HIS-Igch), o estos dos últimos transfectados al mismo tiempo. La IP se realizó con los anticuerpos anti-V5 ( $\alpha$ -V5), anti-His ( $\alpha$ -HIS) y anti-sGFP ( $\alpha$ -GFP). Los inmunoprecipitados fueron analizados por western blot (WB) usando  $\alpha$ -HIS (panel superior) o  $\alpha$ -V5 (panel inferior).

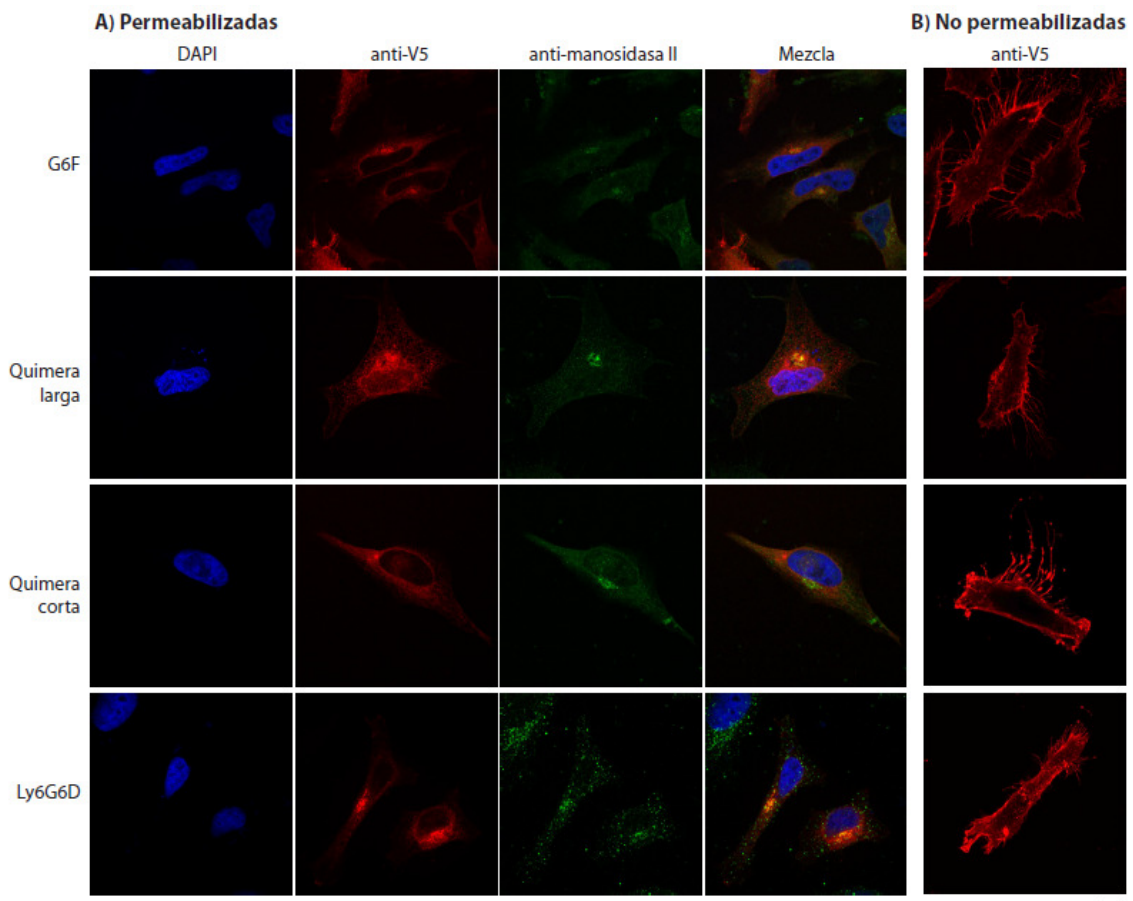
etiquetas diferentes (Figura 46B). Se observa como las dos proteínas interactúan entre ellas cuando ambas proteínas están sobre-expresadas al mismo tiempo, mientras como era de esperar no interactúan cuando solo una de ellas está sobre-expresada o cuando se realiza la co-inmunoprecipitación con un anticuerpo primario (anti-GFP) independiente a ambas construcciones (Figura 46B).

Sorprendentemente, al contrario del comportamiento observado por G6F y la proteína quimérica larga, observamos comportamientos muy diferentes entre LY6G6D y la quimera corta realizando el mismo ensayo. La proteína LY6G6D mostraba un aumento de su peso molecular en ausencia de BME aunque algunas proteínas se mantenían en su forma monomérica (Figura 46A), tal como se describió previamente (Mallya et al. 2006). Sin embargo, la proteína quimérica corta presentaba exactamente el mismo peso molecular en



ambas condiciones tanto desnaturalizantes como no desnaturalizantes (Figura 46A), indicando que esta proteína no forma dímeros.

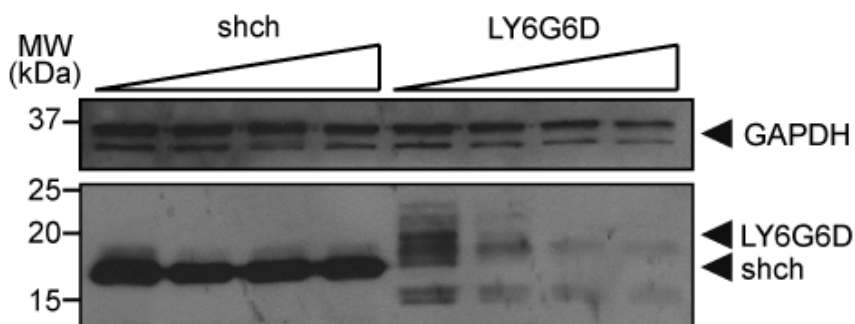
Otro estudio realizado para caracterizar estas proteínas fue el estudio de su localización subcelular. Los patrones de expresión en la célula se investigaron por microscopia de inmunofluorescencia confocal sobre-expresando las construcciones de las proteínas con la etiqueta V5 en su extremo N-terminal y sometiendo a las células a condiciones permeabilizantes y no permeabilizantes (Figura 47). Bajo condiciones permeabilizantes, todas las proteínas mostraban un patrón de tinción intracelular similar a una tinción de retículo (Figura 47A). Además, se observa como todas las proteínas presentan también un patrón de Complejo de Golgi al colocalizar con manosidasa II, a excepción de la proteína quimérica corta, la cual presenta únicamente un patrón de expresión similar a RE, evidenciándose de nuevo la diferencia de procesamiento con la proteína LY6G6D.



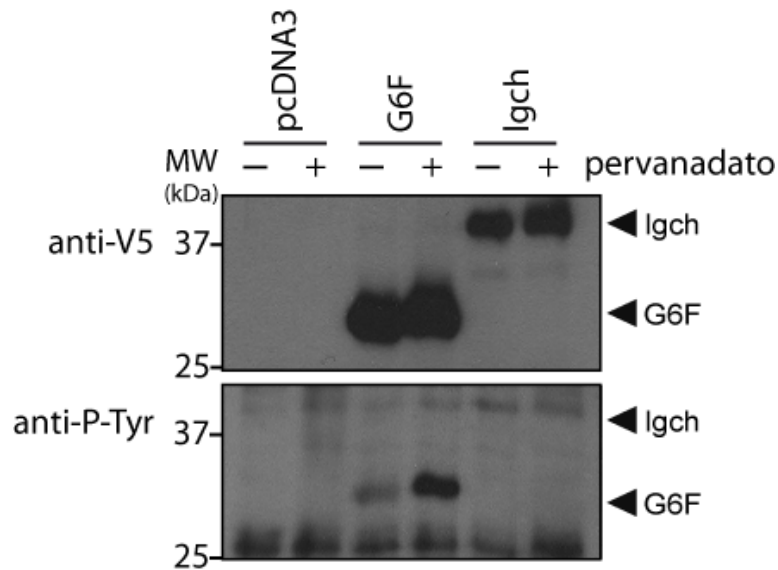
**Figura 47. Distribución celular de las proteínas G6F, LY6G6D, y quimeras Igch y shch.** Las proteínas fueron sobre-expresadas con la etiqueta V5 en células HeLa y la tinción fue realizada bajo condiciones permeabilizantes (A) y no permeabilizantes (B). Se muestra un estudio de localización del complejo de Golgi (anti-manosidasa II) y de colocalización de cada proteína con el complejo de Golgi (mezcla) bajo condiciones permeabilizantes. La barra de tamaño mostrada en la derecha inferior corresponde con 10 µm.

Bajo condiciones no permeabilizantes todas las proteínas analizadas presentaban un patrón de superficie celular (Figura 47B). Como G6F (De Vet et al. 2003), la proteína quimérica larga estaría anclada a la membrana debido a su región transmembrana, siendo ambas proteínas transmembrana de tipo I y paso único. Estos resultados son consistentes con nuestros análisis bioinformáticos previos. En el mismo sentido, la proteína LY6G6D mostraba un patrón de superficie celular, en consistencia con estudios previos donde se identificaba a LY6G6D como una proteína anclada por GPI (Mallya et al. 2006). La proteína quimérica corta presentaba también un patrón de superficie celular.

Para determinar si la proteína quimérica corta se encontraba también anclada por GPI a la membrana celular al igual que la proteína LY6G6D, ambas proteínas fueron transfectadas en células HeLa y, posteriormente tratadas con *Phosphatidylinositol Phospholipase C* (PI-PLC). Como era esperado, en LY6G6D según aumentábamos la dosis de PI-PLC disminuía la señal de detección de la proteína en el Western (Figura 48). Esto es debido a una liberación progresiva de la proteína de la membrana por eliminación de la unión por GPI. Sin embargo, la proteína quimérica corta no mostraba cambio alguno ni con la concentración más elevada de PI-PLC, mostrando que esta proteína no se encuentra anclada a la membrana celular por GPI (Figura 48). Teniendo en cuenta estos resultados y, dado que la proteína quimérica shch presentaba la etiqueta V5 situada inmediatamente posterior a la región transmembrana, podría indicar que la proteína quimérica corta estaría anclada a la membrana por un dominio transmembrana de paso único con la región C-terminal de la proteína en la región extracelular, es decir, podría caracterizarse como una proteína integral de tipo II y paso único.



**Figura 48. Tratamiento PI-PLC para el estudio de anclaje por GPI de la quimera corta (shch) y de la proteína LY6G6D.** Células HeLa que sobre-expresaban las proteínas con la etiqueta V5 fueron tratadas con dosis crecientes (0-0.02 U) de PI-PLC. Los lisados fueron analizados por Western blot. Los marcadores de peso molecular se indican a la izquierda y la posición de la banda esperada para cada proteína a la derecha de cada panel.



**Figura 49. Tratamiento de pervanadato para el estudio de fosforilación en la proteína G6F y la quimera larga (Igch).** Células K562 que sobre-expresaban las proteínas con la etiqueta V5 se trataron con (+) o sin (-) pervanadato, y los lisados se inmunoprecipitaron con el anticuerpo anti-V5. Las IP se analizaron por Western blot con los anticuerpos anti-V5 (panel superior) y anti-fosfotirosina (panel inferior). Como control negativo se emplearon células transfectadas con un vector pcDNA vacío y bajo los mismos tratamientos. Los marcadores de peso molecular se indican a la izquierda y la posición de la banda esperada para cada proteína a la derecha de cada panel.

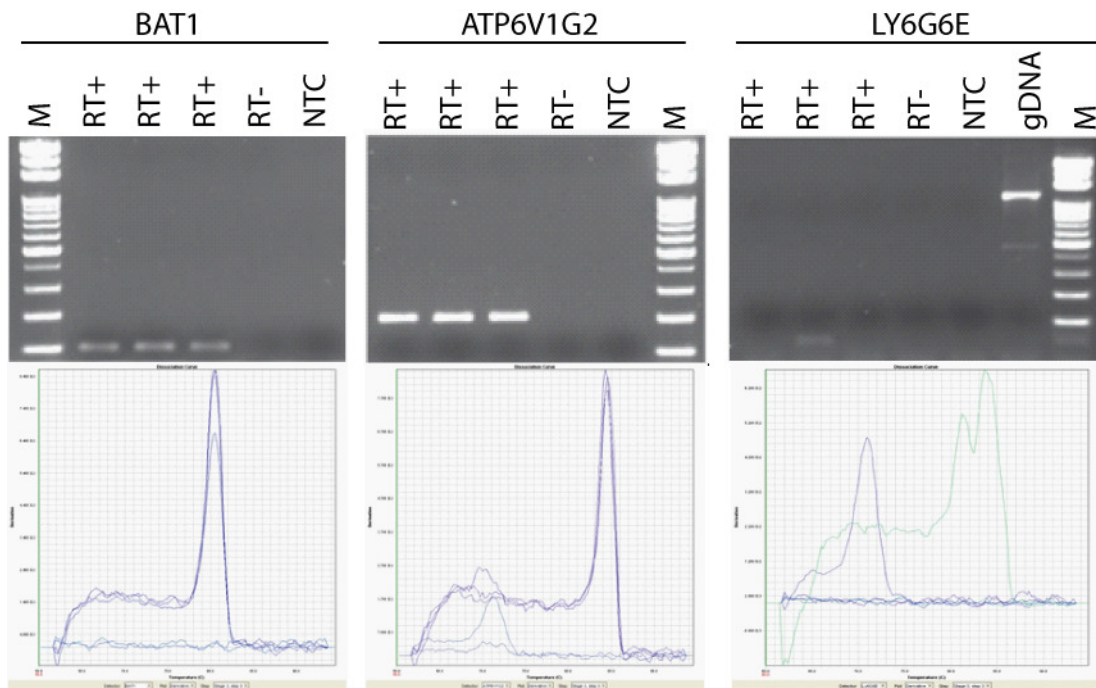
Como se estudió previamente, G6F contiene un residuo de tirosina (Y281) que es responsable de la señalización intracelular en una manera dependiente de fosforilación (De Vet et al. 2003). Sin embargo, la proteína quimérica larga difiere en la región intracelular de la proteína G6F. Para investigar si dicha quimera podría ser fosforilada de la misma manera que G6F, se sobre-expresaron células K562 con ambas construcciones independientemente y se trataron con pervanadato, un inhibidor de la actividad de la tirosina fosfatasa de tal forma que se estabilizaría la forma fosforilada de las proteínas diana. Como se muestra en la Figura 49, G6F se encuentra fosforilada en un residuo de tirosina en concordancia con el estudio previo (De Vet et al. 2003). Sin embargo, ninguna de las muestras que sobre-expresan la proteína quimérica larga detectaban expresión de tirosinas fosforiladas, ni de manera constitutiva (sin tratamiento con pervanadato), ni estimuladas con pervanadato (Figura 49).

#### 4.3. ESTUDIO DE LA ENFERMEDAD AR EN EL MHC DE CLASE III

Estudios previos realizados con pacientes de AR han indicado que esta enfermedad se encuentra relacionada con el MHC, observándose una fuerte asociación con regiones concretas del MHC de clase III. Asimismo, se ha descrito la presencia de diferentes polimorfismos situados en la región del MHC de clase III que parecen estar implicados en

la susceptibilidad a padecer esta enfermedad (Jenkins et al. 2000; Kilding et al. 2004; Kilding and Wilson 2005; Mulcahy et al. 1996). Por ello, se consideró esta región del MHC como candidata para el estudio de alteraciones en la expresión génica y su comparación entre grupos de individuos controles y pacientes con AR tanto con la enfermedad activa como en remisión.

El estudio de la expresión génica se llevó a cabo mediante la metodología de RT-PCR cuantitativa (qRT-PCR) combinando dos tecnologías diferentes: la tradicional qPCR mediante el termociclador ABI Prism 7900HT (Applied Biosystems) y la nanotecnología *BioMark™ HD System (Fluidigm Corporation)*. Todas las parejas de cebadores fueron diseñadas para la amplificación de los genes a estudio, o de unas variantes y eventos de *splicing* concretos. Estas parejas de cebadores fueron posteriormente chequeadas mediante ensayos de qRT-PCR tradicionales con el fin de observar su amplificación específica de producto y la ausencia de amplificación de un posible DNA genómico contaminante (Figura 48). Se muestran tres genes como ejemplo para detallar este chequeo: *BAT1* que presenta una amplificación modelo, *ATP6v1G2* que contiene dímeros de cebadores, y *LY6G6E* que es un pseudogen y, por tanto, no amplifica en las muestras de sangre (salvo un dímero de cebadores) y solamente amplifica en DNA genómico.



**Figura 48. Chequeo de tres parejas de cebadores por qRT-PCR.** Datos relativos a qRT-PCRs de *BAT1*, *ATP6V1G2* y *LY6G6E* con triplicados de cDNA de sangre comercial positivos (RT+) o negativo (RT -), y control sin muestra (NTC). En el caso de *LY6G6E* se muestra un control de DNA genómico (gDNA). Se muestran las curvas de disociación de los amplicones según la temperatura de melting (abajo) y la imagen de un gel con los productos finales obtenidos (arriba).

Una vez realizados todos los controles pertinentes, las parejas de cebadores se emplearon en dos experimentos utilizando la nanotecnología *BioMark<sup>TM</sup> HD System*. En el primer experimento se trataba de encontrar una pareja de normalizadores adecuado para el tipo y la naturaleza de nuestras muestras de sangre. En el segundo experimento se realizó el estudio de la expresión génica de los genes del MHC de clase III y otros genes e isoformas asociados con *RAGE* y las quimeras *G6F-LY6G6D*. Las muestras empleadas en el estudio proceden de donantes controles del Hospital Universitario La Paz y de pacientes de AR del Servicio de Reumatología del Hospital Universitario Ramón y Cajal. Los pacientes se clasificaron en dos grupos: pacientes con la enfermedad activa o en remisión. La actividad de la enfermedad se cuantificó según el criterio de un experto reumatólogo, observándose la presencia de al menos una articulación tumefacta o más de una articulación dolorosa y valores elevados de sedimentación globular. Ninguno de los pacientes de AR había sido tratado con fármacos modificadores de la respuesta biológica ya que podrían alterar los valores de expresión de diferentes genes del MHC de clase III, como por ejemplo del gen *TNF*, diana directa de la mayoría de estos tratamientos

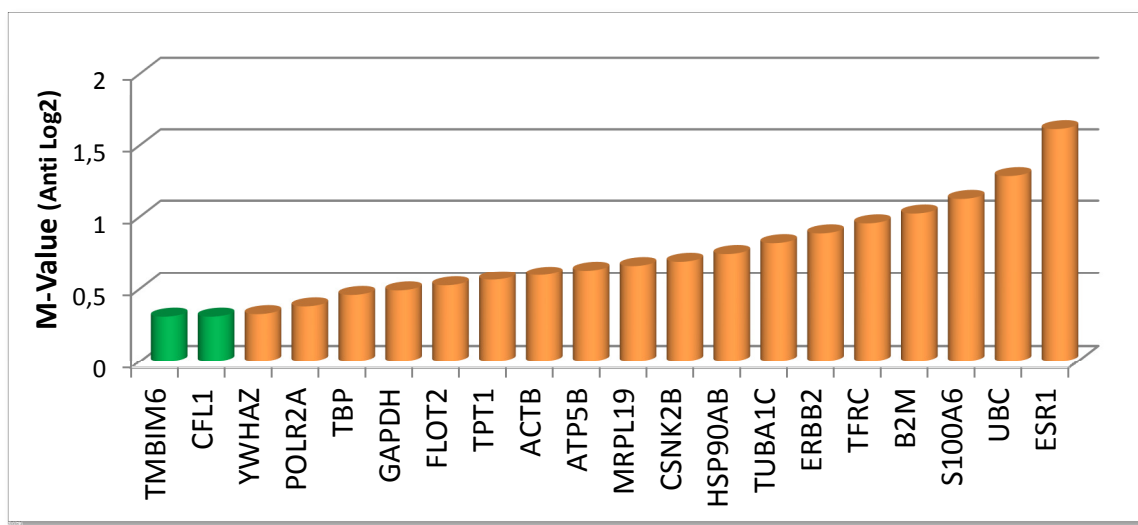
#### **4.3.1. Identificación de una pareja de normalizadores adecuados para el estudio.**

La cuantificación de la expresión génica con qRT-PCR se realiza generalmente mediante una cuantificación absoluta o relativa. Para emplear la técnica de cuantificación absoluta se requiere conocer el material de partida de las muestras mediante la realización de un patrón de diluciones estándar. En cambio, la cuantificación relativa no requiere de estándares con concentraciones determinadas, sino del empleo de unos genes normalizadores (o *housekeeping*) que nos permitan compensar variaciones en la cantidad y calidad del RNA de partida, que pueden afectar a las eficiencias de retrotranscripción y de PCR. De esta manera, midiendo los genes de interés y dividiéndolos por la expresión de los genes normalizadores seleccionados es posible comparar los primeros en diferentes muestras sin necesidad de conocer en términos absolutos su nivel de expresión.

No obstante, el aspecto crucial es que la estabilidad del gen normalizador sea real en todas las muestras analizadas. La estabilidad de su expresión se refleja por la desviación estándar en las muestras biológicas. Anteriormente se pensaba que cualquier gen implicado en tareas de organización de la célula, por ejemplo *ACTB*, así como genes involucrados en funciones básicas en la supervivencia celular, como *GAPDH*, presentaban una expresión constitutiva estable y podrían servir como genes normalizadores. Sin embargo, esta

expresión estable depende del tipo de muestra empleada en cada estudio. En nuestro caso estamos empleando el RNA extraído de PBMCs de sangre periférica. Por estudios previos realizados en este tipo de muestras se conoce que determinados normalizadores, como es el caso de *GAPDH*, pueden sufrir variaciones en su expresión en función del nivel de ayunas y del grado de obesidad de cada uno de los donantes/pacientes (Kloting et al. 2007). Por ello, no nos parecía apropiado emplear los tradicionales normalizadores y decidimos realizar un estudio previo buscando los dos normalizadores más apropiados para muestras de este tipo.

El estudio para la selección de los normalizadores adecuados para este tipo de muestras se realizó mediante un ensayo con un chip de 48\*48 de Biomark™ en el que se cuantificó la expresión de un conjunto de normalizadores: *CSNK2B*, *UBC*, *YWHAZ*, *B2M*, *ATP5B*, *HSP90AB1*, *FLOT2*, *TMBIM6*, *CFL1*, *TPT1*, *TUBA1C*, *ACTB*, *S100A6*, *MRPL19*, *TBP*, *TFRC*, *ESR1*, *ERBB2*, *GAPDH* y *POLR2A*. Estos normalizadores fueron seleccionados en base a su empleo previo en experimentos de similares características encontrados en la bibliografía [Vandesompele et al 2002; Andersen et al 2004; Spurgeon et al 2008]. Los resultados se analizaron empleando los algoritmos NormFinder (datos no mostrados) [Andersen et al 2004] y geNorm (Figura 49). Ambos algoritmos calculan la expresión media global de todos los genes en todas las muestras, y basado en esta comparación se calcula las desviaciones típicas (SD) de cada uno de los genes candidatos y se selecciona aquel/aquellos genes que presentan unas SD menores como los normalizadores más adecuados en las muestras a estudio.



**Figura 49. Clasificación de los genes de normalización candidatos empleando el algoritmo geNorm.** El eje de ordenadas muestra el M-Value, relacionado con las desviaciones típicas de los genes. Se indica el par de genes con menor variación (verde) y más adecuados para normalizar en las muestras empleadas.

Tras el análisis por ambos algoritmos, observamos que los genes *TMBIM6*, *CFL1* y *YWHAZ* eran los genes más apropiados para emplear como normalizadores en este estudio (Figura 49). Dado que el gen *YWHAZ* es un gen monoexónico y, por tanto, no se pueden diseñar cebadores *intron-spanning*, podría cuantificar erróneamente en el caso de una posible contaminación por DNA genómico. Sin embargo, los genes *TMBIM6*, *CFL1* si fueron diseñados de forma que presentan un intrón en su futuro amplicón, impidiendo la amplificación inespecífica de cualquier molécula de DNA genómico contaminante de la muestra. Por ello, seleccionamos los genes *TMBIM6* y *CFL1* como normalizadores para nuestro estudio en muestras de sangre.

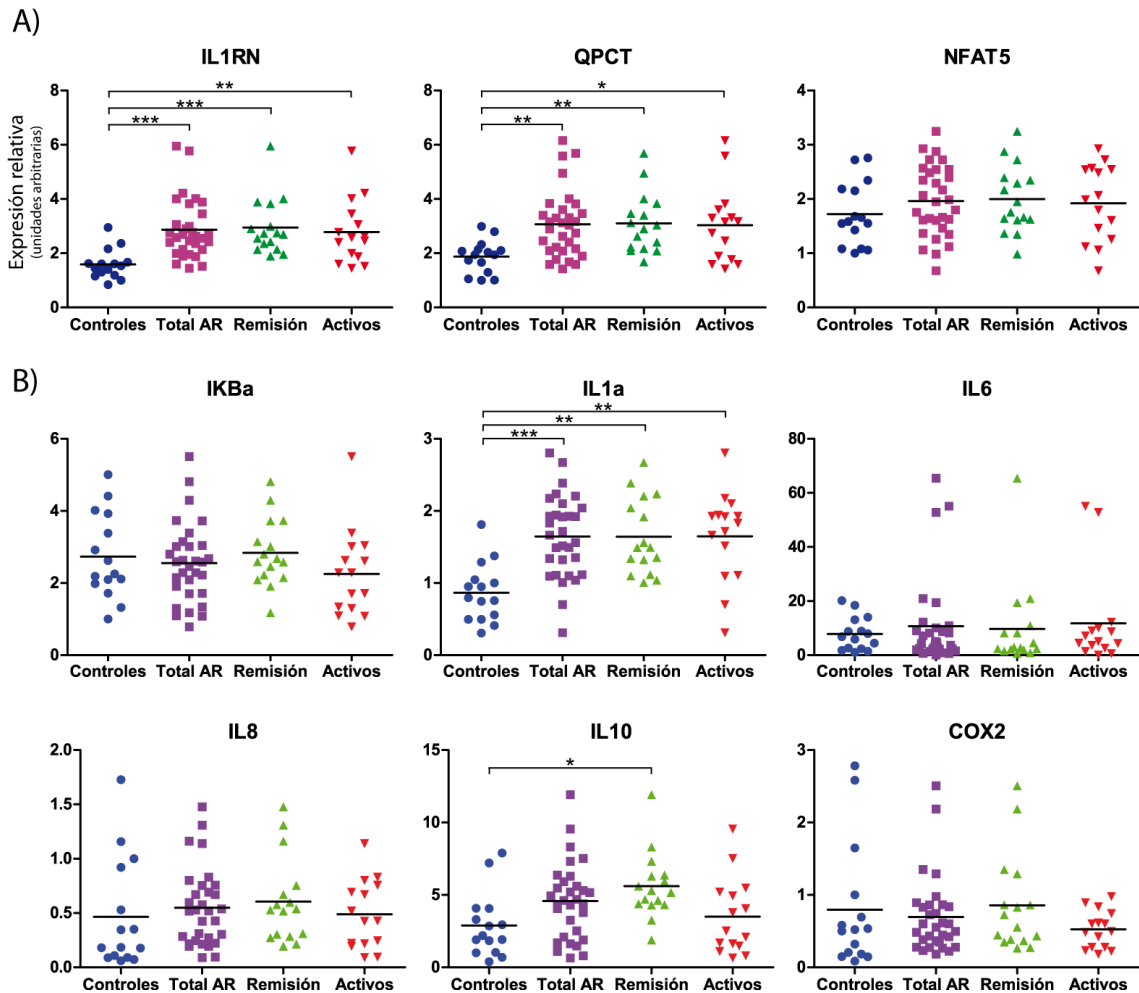
#### **4.3.2. Estudio de la expresión génica de los genes del MHC de clase III y otros genes involucrados en inflamación en pacientes con AR.**

Una vez seleccionados los normalizadores adecuados al tipo de muestra de nuestro análisis, procedimos a realizar el estudio en pacientes con AR cuantificando los niveles de expresión génica en el conjunto de los genes del MHC de clase III, así como en otros genes involucrados en procesos inflamatorios. Para dicho estudio se empleó un chip de 96\*96 de Biomark™ en el que se incluyeron 15 muestras de individuos controles y 31 muestras de pacientes con AR, de los cuales 16 muestras correspondían con pacientes que presentaban la enfermedad en remisión y 15 muestras a pacientes con la enfermedad activa en el momento de la extracción de la muestra de sangre. Para la comparación entre los tres diferentes grupos de muestras se realizó una cuantificación relativa de los genes a estudio empleando los dos normalizadores anteriores, *TMBIM6* y *CFL1* y un análisis estadístico no paramétrico mediante el test de Kruskal-Wallis, que no asume una distribución normal de las muestras a estudio. El nivel de significación establecido para este estudio fue de un p-valor 0,01.

En el estudio se incluyeron unos genes que habían sido previamente relacionados con la enfermedad de AR (Batliwalla et al. 2005; Teixeira et al. 2009). En dichos estudios previos se realizaron *arrays* de expresión y se observaron las diferencias significativas presentes entre un grupo de pacientes de AR y muestras de individuos controles. Además, se encuentran publicados otros estudios similares pero con un número de muestras inferior a 15 pacientes en todos los casos, por lo que no fueron considerados para este estudio (Bovin et al. 2004; Edwards et al. 2007; Meugnier et al. 2011). Los genes escogidos fueron *IL1RN* (*interleukin 1 receptor antagonist*, o anteriormente descrito IL1RA) y *QPCT* (*glutaminyl-peptide cyclotransferase*) pues se encontraban sobre-expresados tanto en los pacientes de



AR como en muestras enriquecidas en monocitos (Batliwalla et al. 2005), y *NFAT5* (*nuclear factor of activated T-cells 5, tonicity-responsive*) pues se encontraba sobreexpresado en muestras de PBMCs de pacientes de AR tratados únicamente con FAMES (Teixeira et al. 2009). Se diseñaron unos cebadores específicos para estos tres genes y se cuantificó la expresión relativa comparando tanto los controles con pacientes, como los controles con pacientes en remisión y pacientes activos por separado (Figura 50A). Al estudiar la cuantificación relativa se observaron diferencias significativas en la expresión de los genes *IL1RN* y *QPCT* en comparación con todos los pacientes de AR e, incluso con los pacientes desglosados por actividad de la enfermedad. En cambio, la expresión de *NFAT5* presentaba una ligera sobreexpresión en las muestras de los pacientes con AR, pero no generó diferencias significativas con el test estadístico empleado (Figura 50A).



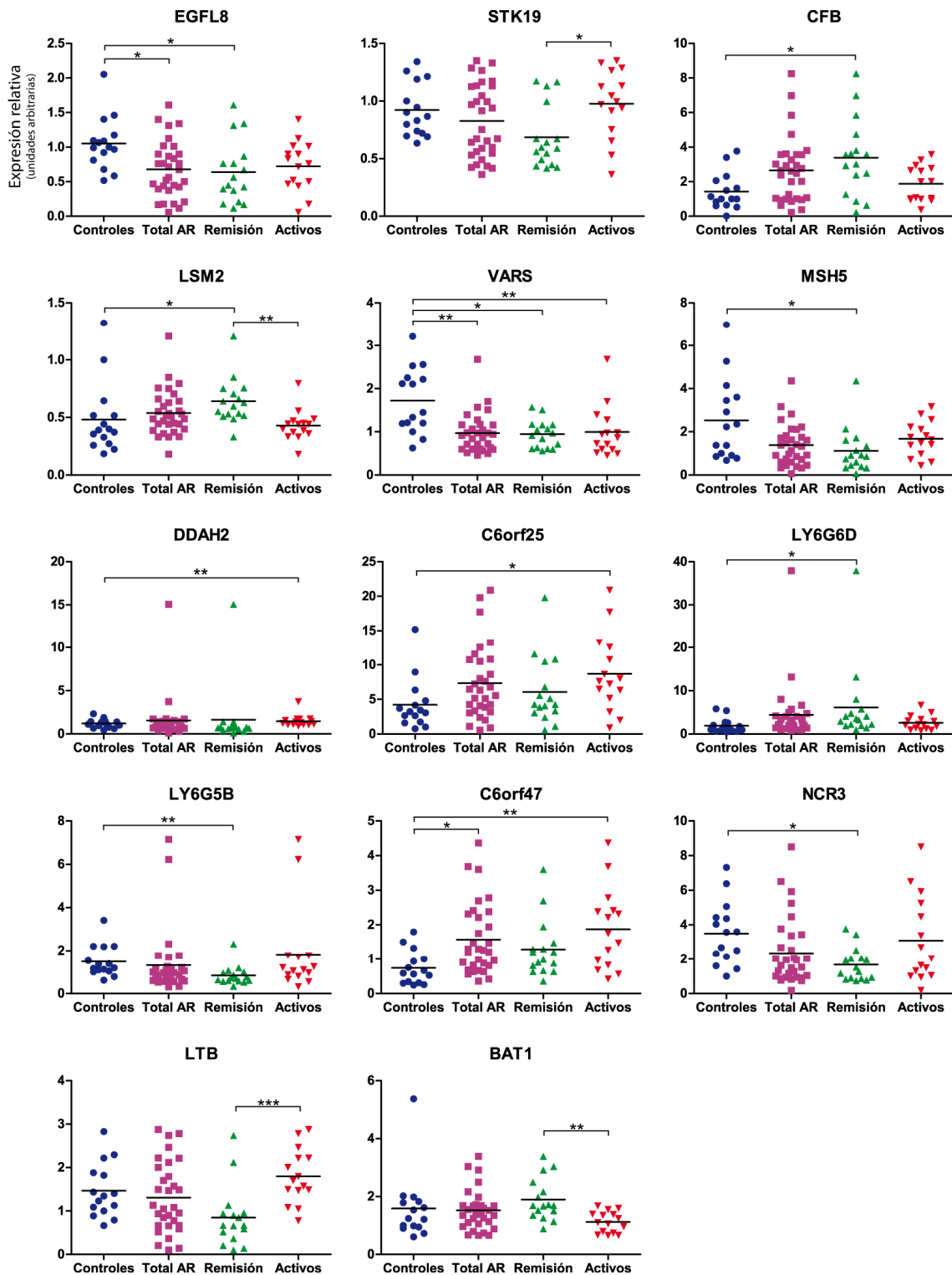
**Figura 50. Expresión relativa de genes relacionados con AR (A) o con procesos inflamatorios (B).** La expresión se distribuyó en 15 individuos controles (azul) y 31 pacientes de AR (morado), siendo 16 pacientes en remisión de la enfermedad (verde) y 15 con la enfermedad activa (rojo). La barra horizontal en cada nube de puntos representa la media. La significación estadística se detalla según: \*  $p < 0,01$ , \*\*  $p < 0,001$  y \*\*\*  $p < 0,0001$ .



Posteriormente, se estudiaron moléculas implicadas en procesos inflamatorios, que tradicionalmente se han visto involucrados con el desarrollo de la AR (Smolen et al. 2005), siendo las tradicionales moléculas diana para el tratamiento de esta enfermedad. Dichas moléculas fueron: IKBa (*nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor*), las interleuquinas IL1a, IL6, IL8 y IL10, así como COX2 (*cytochrome c oxidase subunit II*). Se observó como las interleuquinas IL1a y IL10 mostraban diferencias significativas en todos sus grupos o comparando controles con pacientes en remisión, respectivamente (Figura 50B). El resto de las moléculas inflamatorias estudiadas no mostraban diferencias significativas en su expresión a nivel de RNA en las muestras analizadas (Figura 50B), al igual que ocurría en los estudios descritos mediante la tecnología de *arrays* de expresión (Batliwalla et al. 2005; Bovin et al. 2004; Edwards et al. 2007; Meugnier et al. 2011; Teixeira et al. 2009).

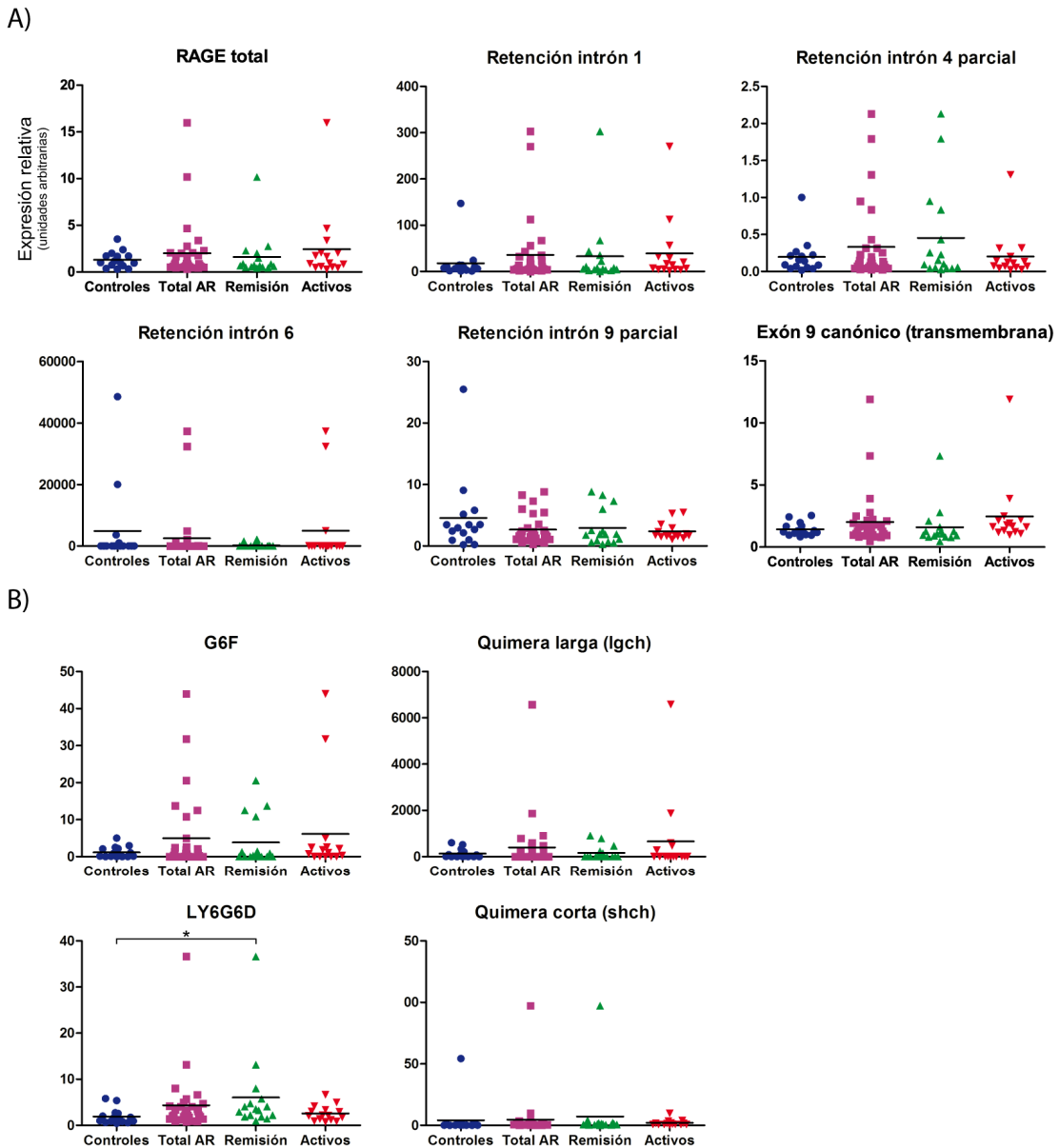
A continuación se realizó el estudio sobre los genes del MHC de clase III (Figura Suplementaria 4), observándose algunas diferencias significativas en la expresión de 14 genes en las muestras de AR (total, remisión o activos) en comparación con la expresión de las muestras de individuos controles (Figura 51). El único gen que presentaba variaciones significativas con todos los grupos de pacientes con AR en comparación con los individuos sanos fue el gen *VARS* (*valyl-tRNA synthetase*), que se encontraba inhibido en los pacientes de AR independientemente de la actividad de su enfermedad. Además, nos encontramos con los genes *DDAH2* (*dimethylarginine dimethylaminohydrolase 2*), *C6orf25* y *C6orf47* que se encuentran sobre-expresados en los pacientes con la enfermedad AR activa en comparación con los controles. Aparecen también diferencias entre los individuos controles y los pacientes con la enfermedad AR en remisión, con sobre-expresión de los genes *CFB* (*complement factor B*), *LSM2* (*LSM2 homolog, U6 small nuclear RNA associated (S. cerevisiae)*) y *LY6G6D* (*lymphocyte antigen 6 complex, locus G6D*), e inhibición de los genes *EGFL8* (*EGF-like-domain, multiple 8*), *MSH5* (*mutS homolog 5*), *LY6G5B* (*lymphocyte antigen 6 complex, locus G5B*) y *NCR3* (*natural cytotoxicity triggering receptor 3*) en los pacientes de AR en remisión. Por último, nos encontramos con un grupo de genes que presentan variaciones significativas en la expresión de los pacientes AR con la enfermedad activa y en remisión, no encontrándose diferencias con los individuos controles. En este grupo nos encontramos con los genes *STK19* (*serine/threonine kinase 19*) y *LTB* (*lymphotoxin beta (TNF superfamily, member 3)*) que se encuentran sobre-expresados en los pacientes con la enfermedad activa frente a los de remisión y los genes *LSM2* (*LSM2 homolog, U6 small nuclear RNA associated (S. cerevisiae)*) y *BAT1* (o también

denominado *DDX39B*, *DEAD (Asp-Glu-Ala-Asp) box polypeptide 39B*) que se encuentran sobre-expresados en los pacientes con la enfermedad en remisión frente a los activos.



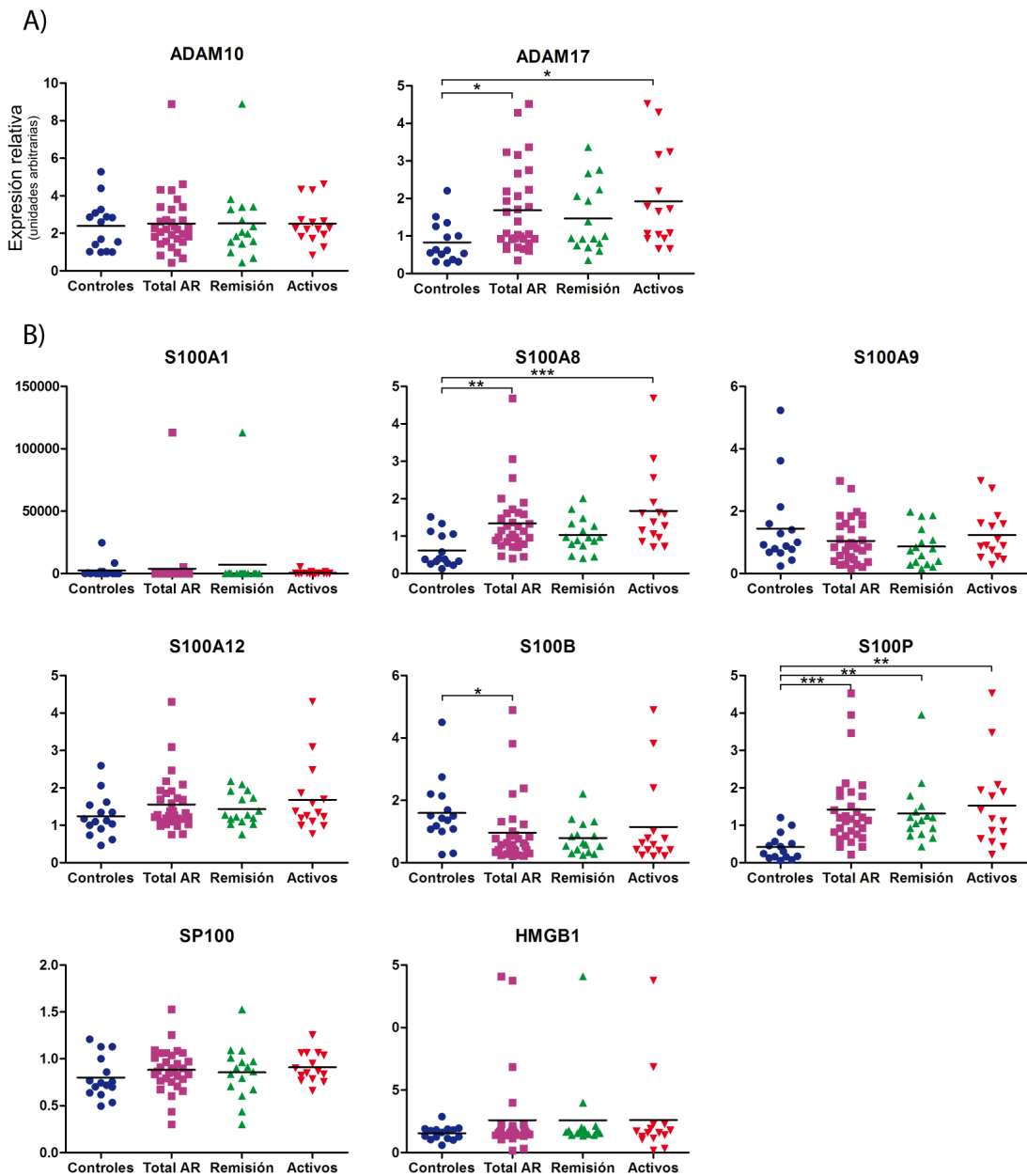
**Figura 51. Expresión relativa de los genes del MHC de clase III que presentan significación estadística.** Se indican aquellos genes que presentaban alguna diferencia significativa entre alguno de sus grupos. Ver Figura 50 para detalles. Para observar la expresión relativa del resto de los genes del MHC de clase III ver la Figura Suplementaria 4.

Además, aprovechamos este estudio para profundizar en los apartados centrales de esta Tesis, comparando los niveles de expresión de los eventos de *splicing* más relevantes que ocurren en el gen *RAGE*, así como las isoformas características de las quimeras entre *G6F* y *LY6G6D* (Figura 52). En el estudio de *RAGE*, se diseñó una pareja de cebadores (*RAGE* total) que amplifica presumiblemente la mayoría de las variantes, ya que los cebadores se diseñaron en la región común IgC que presentan la mayoría de las variantes (a excepción de las variantes *HsapRAGEv10* a *v15* que presentan la exclusión de múltiples exones) (Figura 25). El resto de las parejas de cebadores amplificaban específicamente un evento, que en la mayoría de los casos se encuentra compartido entre diferentes variantes de



**Figura 52.** Expresión relativa de los diferentes eventos de *splicing* de *RAGE* (A) y de las diferentes variantes quiméricas estudiadas en el apartado anterior (B). Ver Figura 50 para detalles.

*splicing*. Se comprobó la amplificación específica de los eventos deseados mediante qRT-PCR, añadiendo como material de partida plásmidos con cada uno de los eventos o la isoforma canónica (datos no mostrados). Únicamente se consideraron como específicos aquellas parejas de cebadores que no amplificasen otros productos inespecíficos o los amplificasen con una diferencia de al menos 10 ciclos (tres órdenes de magnitud en copias amplificadas). Ni el conjunto de las isoformas de RAGE ni ninguno de los eventos de *splicing* independientemente mostraron diferencias significativas presentes entre los donantes sanos y los pacientes con AR (Figura 52A).



**Figura 53. Expresión relativa de proteínas involucradas en la ruta biológica de RAGE.** A) Expresión relativa de proteínas shedasas implicadas en el corte proteolítico de RAGE y TNF. B) Expresión relativa de diferentes ligandos reconocidos por el receptor RAGE. Ver Figura 50 para detalles.

En el caso de los genes involucrados en la quimera, se diseñaron cebadores específicos para cada una de las isoformas y se comprobaron igualmente por qRT-PCR (datos no mostrados). Se encontraron diferencias significativas en la expresión del gen *LY6G6D*, encontrándose sobre-expresado en las muestras de pacientes con AR en remisión, en comparación con los individuos controles (Figura 52B). No se encontraron diferencias significativas en el resto de las isoformas a estudio.

Dado que no se encontraron variaciones significativas en la expresión de RAGE quisimos observar si otras proteínas involucradas en la ruta biológica de RAGE presentaba alteraciones, ya que estudios previos habían evidenciado su implicación en el desarrollo de esta enfermedad (Bovin et al. 2004; Edwards et al. 2007; Meugnier et al. 2011; Teixeira et al. 2009). Primeramente, se estudió los niveles de expresión de las proteínas shedasas implicadas en el corte proteolítico de RAGE (ADAM10) y de TNF $\alpha$  (ADAM17) (Raucci et al. 2008), lo que libera estos receptores a su forma soluble extracelular. Se observaron diferencias significativas en la expresión de ADAM17, mostrando sobre-expresión en las muestras de pacientes totales de AR y del grupo de pacientes que presentan la enfermedad activa (Figura 53A). Por último, se estudiaron los niveles de expresión de diversos ligandos de la proteína RAGE, entre los que se incluyen las proteínas SP100, HMGB1 y diversas proteínas de la familia S100. Los niveles de expresión de los genes S100A8, S100B y S100P mostraron diferencias significativas en pacientes con AR en comparación con su expresión en donantes sanos (Figura 53B). Además, estas variaciones significativas se extendían a los pacientes con la enfermedad AR activa en comparación con los controles en el caso del ligando S100A8, y a los pacientes con la enfermedad AR en remisión y activa en comparación con los controles en el caso del ligando S100P.

## **5. DISCUSIÓN**

### 5.1. Metodologías empleadas para el estudio de las variantes de *splicing* de un gen, el caso del gen *RAGE*.

En este estudio hemos caracterizado en profundidad un alto número de variantes de AS del gen *RAGE* en diferentes tejidos de mamíferos, y hemos comparado sus niveles de expresión relativos. La diversidad de funciones de *RAGE* había sido previamente predicha como resultado de la variedad de isoformas de *RAGE* generadas mediante eventos de AS (Hudson et al. 2008b; Kalea et al. 2009; Sterenczak et al. 2009; Yonekura et al. 2003). Estudios previos se centraban en las variantes de *splicing* de este gen en humanos (Hudson et al. 2008b) y ratones (Kalea et al. 2009) mostrando resultados muy similares en los niveles de expresión génica de cada variante de *splicing*. En humanos, *RAGE* había sido solo estudiado en detalle en pulmón y en células del músculo liso de la aorta (AoSMCs, *Aortic smooth muscle cells*) (Hudson et al. 2008b). Si realizamos una comparación entre este y nuestro estudio observamos unos niveles de expresión similar de la isoforma Tv1-*RAGE* (80% y altos niveles de expresión, respectivamente) y expresiones diferentes de la variante Tv6-*RAGE* (C-truncada) (7% y niveles de expresión moderada, respectivamente). Interesantemente, en nuestro estudio tanto el tejido adulto como el tejido fetal de pulmón presentaban únicamente estas dos isoformas, mientras que en el tejido de pulmón tumoral se expresaban un número muy alto de diferentes variantes de *splicing*.

En un estudio previo del AS del gen *RAGE* en muestras de tejidos de ratón (Harashima et al. 2006) se identificó una isoforma soluble en cerebro mediante el empleo de la tecnología de PCR anidada. Posteriormente, se cuantificó su expresión en los diferentes tejidos y en comparación con la isoforma canónica Tv1-*RAGE* mediante la realización de una RT-PCR entre los exones 9 y 11. Con este método se cuantificaba en conjunto a todas las isoformas solubles en comparación con todas las isoformas transmembranas, por lo que no se pudo comparar con nuestros resultados de variantes individuales. En ese trabajo, se describe también una isoforma soluble (es*RAGE*) no encontrada en nuestro estudio, pero esto puede ser debido a que esta isoforma fue generada a partir de unos cebadores específicos para el intrón 9 retenido durante la segunda ronda de PCR. En el estudio realizado por el grupo de la Dra. Schmidt, el tejido de pulmón murino muestra tanto expresiones de la isoforma canónica Tv1-*RAGE* (53%) como de la variante m*RAGE*\_v4 (41%) (Kalea et al. 2009), datos que correlacionan con los resultados obtenidos en esta Tesis con un alto y bajo nivel de expresión, respectivamente. Las variantes de menor tamaño encontradas en su estudio son diferentes a las descritas aquí, pero todas presentan unos niveles de expresión bajos (1% en cada variante de *splicing*).

En general nuestros resultados son generalmente consistentes con los estudios previos (Hudson et al. 2008b; Kalea et al. 2009) considerando los porcentajes mayores al 50% relacionados con nuestros niveles de expresión elevados, los porcentajes 50-40% con niveles de expresión moderado-alto, y los porcentajes 40-20% con niveles de expresión bajos, en la mayoría de los casos. Los porcentajes menores al 20% no son detectados a veces en nuestro estudio, mientras que, por el otro lado, las variantes con expresión baja detectadas aquí no fueron detectadas previamente en su estudio. En ambos estudios con muestras de humano y ratón (Hudson et al. 2008b; Kalea et al. 2009) las variantes de *splicing* fueron analizadas por RT-PCR y, posteriormente, realizando un análisis por enzimas de restricción según el número y el tamaño de los cortes en la secuencia de los transcritos. En nuestro caso, las variantes de *splicing* fueron analizadas mediante la separación electroforética de los productos de RT-PCR, posteriormente las bandas fueron clonadas y verificada la secuencia de cada transcrito por secuenciación. Ambos enfoques emplearon RT-PCR anidada debido a los niveles de expresión extremadamente bajos que presenta el gen *RAGE*, lo que hace imposible realizar el estudio dentro de un rango dinámico de la PCR. Por ello, esta tecnología puede influenciar ligeramente los datos limitando el poder de diferenciar la expresión entre las diferentes variantes de *splicing* y su abundancia relativa. Por esta razón, las variantes de baja expresión y, por tanto las más difíciles de detectar, mostraban a veces diferencias entre los diferentes estudios.

Además, es importante destacar que el mejor enfoque para determinar la expresión relativa de cada transcrito debería ser empleando la tecnología de RT-PCR cuantitativa (qRT-PCR). Sin embargo, debido a la presencia de múltiples eventos de AS comunes a varias isoformas entre sí, se vuelve completamente imposible diseñar un ensayo de qRT-PCR para discriminar entre las variantes. En cambio, esta técnica fue previamente empleada para el estudio de *RAGE* en muestras caninas (Sterenczak et al. 2009) para describir las variantes de *splicing* en tejidos sanos y tumorales, pero estos análisis fueron realizados únicamente para estudiar el evento de retención del intrón 1. En ese estudio se detectaron altos niveles de expresión del conjunto de las variantes que presentaban el intrón 1 retenido en las muestras de tejido neoplásico o tumoral. Por otro lado, la exclusión del intrón 1 fue designado al conjunto de estas variantes de *splicing* en muestras sanas y, por lo tanto, con niveles de expresión del evento de retención del intrón 1 muy bajos (Sterenczak et al. 2009). Este ensayo fue realizado para la observación del efecto global de retención del intrón 1, sin tratar de discriminar entre las diferentes variantes. La amplia diversidad de variantes de *splicing* con retención del intrón 1 que hemos encontrado en este estudio no podría ser



diferenciada con este ensayo. Por ello, en nuestro estudio encontramos variantes de *splicing* con retención del intrón 1 que están altamente expresadas en tejidos adultos y fetales sanos, así como encontramos variantes de *splicing* que presentan la exclusión del intrón 1 presentes también en tejidos tumorales con diferentes niveles de expresión.

Actualmente, la tecnología de secuenciación masiva está siendo empleada para la descripción de transcriptomas completos de diversos organismos (Barbosa-Morais et al. 2012; Merkin et al. 2012; Rastrojo et al. 2013). Esta técnica proporciona información de los niveles de expresión diferencial de los genes, mutaciones post-transcripcionales o fenómeno de *editing*, así como información de fusiones génicas, entre otras. Además, se pueden también observar los niveles de expresión de los diferentes transcritos de un gen, aunque para ello se requiere de una cobertura de lecturas suficiente. En el caso del gen *RAGE*, tanto los niveles de expresión bajos del gen como la elevada complejidad transcripcional hacen, como hemos visto, que sea extremadamente difícil abordar un estudio transcriptómico mediante esta tecnología. Además, en la muestra empleada con un nivel de cobertura elevado (monocitos humanos, con más de 200 millones de lecturas pareadas) solo fuimos capaces de detectar tres de las variantes de *splicing* detectadas por PCR anidada, y únicamente cuando los parámetros de Cufflinks fueron alterados. Por estas razones, los estudios transcripcionales detallados de genes de baja expresión como es el caso de *RAGE* pueden ser interpretados erróneamente cuando se estudian por tecnologías de secuenciación masiva, y deberían ser preferiblemente estudiadas por otras técnicas como la que proponemos y hemos empleado en este estudio, consistente en la amplificación específica del gen de interés mediante PCR anidada, clonaje y secuenciación por el método de Sanger.

## 5.2. Implicaciones funcionales del AS en la expresión del gen *RAGE*.

Las variantes de *splicing* encontradas en el gen *RAGE* y sus eventos de AS comunes podrían potencialmente tener un rol regulador sobre la función global de la isoforma canónica Tv1-*RAGE*, así como podrían estar implicadas en el desarrollo de diversas enfermedades como cáncer o Alzheimer, entre otras. Por ejemplo, los niveles bajos de la isoforma soluble s*RAGE* han sido asociados con un incremento en el riesgo a padecer Alzheimer (Emanuele et al. 2005), complicaciones diabéticas (Grossin et al. 2008), enfermedades cardiovasculares (Falcone et al. 2005), e hipertensión (Geroldi et al. 2005) en humanos.

Se ha descrito que aproximadamente un 40% de los genes que codifican proteínas transmembrana de paso único presentan isoformas que pierden completamente el dominio transmembrana mediante AS (Mittendorf et al. 2012; Xing et al. 2003). En su mayoría, esta pérdida se produce tras la exclusión de un exón único, ya que alrededor del 84,6% de las proteínas integrales de paso único presenta la región transmembrana localizada únicamente en un exón (Cline et al. 2004), como es el caso de la proteína RAGE. La eliminación de esta región transmembrana permite la generación de isoformas solubles sRAGE por diferentes procesos biológicos. En humanos, las variables solubles encontradas son debidas a un mecanismo de *splicing* alternativo (esRAGE) o a un corte proteolítico (Maillard-Lefebvre et al. 2009) generado por la shedasa ADAM10 (Raucci et al. 2008). Este corte proteolítico genera unas isoformas truncadas en su C-terminal a partir de las isoformas ancladas a la membrana plasmática (cRAGE) (Hanford et al. 2004). La isoforma sRAGE en rata adulta muestra unos niveles de expresión de la proteína elevados, dos veces superior al de ratas neonatales (Lizotte et al. 2007) y además, el corte proteolítico de las isoformas ancladas a la membrana plasmática parece ser la principal vía de generación de isoformas sRAGE (Hanford et al. 2004). Para concluir, cualquiera de estos mecanismos podría estar generando receptores solubles de RAGE que bloqueasen la unión de los ligandos al receptor por interacción directa con ellos y, de esta forma estar atenuando la señalización intracelular de RAGE. Nosotros hemos sido capaces de detectar isoformas solubles potenciales en todas las especies analizadas, aunque solo en humano (Tv4-RAGE y Tv6-RAGE) y en ratón (*Mmus*RAGEv4) contienen los tres dominios Ig y el péptido señal. Interessantemente, todas las especies excepto rata presentan isoformas solubles con solo los dos primeros dominios Ig, que son los dos dominios involucrados preferentemente en la unión a ligando.

Es importante destacar la presencia de un exón de tamaño muy reducido (únicamente de 27 nucleótidos), el exón 9, inmediatamente anterior al que codifica la región transmembrana. En un experimento previamente publicado se observa que cuando el tamaño de los exones es reducido artificialmente a menos de 50 bp se provoca la exclusión de dicho exón del mensajero final (De Conti et al. 2013; Dominski and Kole 1991). Sin embargo, únicamente las dos especies de roedores, ratón (*mRAGE\_v4*, *Mmus*RAGEv1 y v4) y rata (*Rnor*RAGEv1, v3 y v7), presentan este evento de exclusión del exón 9. En cambio, un experimento similar indica como la expansión artificial de exones de pequeño tamaño fomenta su inclusión (Berget 1995). Esto podría explicar la alta proporción de isoformas que presentan un sitio de AS 5' donador en el exón 9 provocando la retención parcial del intrón 9-10, aunque estas variantes solo las hemos encontrado en humano

(*Hsap*RAGEv1-6, v8-9, v12-13, Tv4-RAGE y Tv6-RAGE o C-truncada), orangután (*Ppyg*RAGEv4, *Ppyg*RAGEv5) y colobo (*Cgue*RAGEv2). Curiosamente, la elección del sitio de *splicing* 5' donador alternativo en el exón 9 provoca un fenómeno de exclusión del exón 10 en todas las variantes, siendo el principal evento implicado en la generación de las variantes solubles comentadas anteriormente.

Otra implicación funcional consecuencia de los fenómenos de AS estudiados en este trabajo es la modificación del péptido señal codificado por las diferentes variantes de *splicing* generadas. Los péptidos señal son importantes para la localización subcelular de las proteínas y, a menudo se encuentran alterados debido a procesos de AS. La presencia de estos eventos de AS pueden conllevar la localización de proteínas solubles en los diferentes compartimentos subcelulares (lumen del RE o del citosol, por ejemplo). Además, estos péptidos señal son de vital importancia para las proteínas transmembrana como RAGE, ya que juegan un papel muy relevante en la integración de las proteínas a la membrana plasmática (Mittendorf et al. 2012). Un ejemplo de eliminación de estas secuencias por AS se ha descrito en el receptor de la hormona paratiroidea (PTHrP). Una de las isoformas que expresa este gen presenta el péptido señal reemplazado por una secuencia hidrofílica y se expresa mínimamente en la membrana plasmática. En su lugar, esta isoforma se encuentra preferentemente expresada en compartimentos intracelulares, lo que podría significar un mecanismo de regulación decreciente de este receptor debido a la pérdida del péptido señal correspondiente (Joun et al. 1997).

En nuestro estudio nos encontramos con una amplia variedad de isoformas que carecen o presentan su péptido señal alterado. En la mayoría de las isoformas esta modificación es debida a la retención del primer (o primeros) intrón y a una señal de inicio de la traducción posterior. Las variantes resultantes suelen perder tanto el péptido señal como el primer dominio IgV, a excepción de aquellas isoformas expresadas en *Bos taurus*. En esta especie al aparecer un ATG iniciador alternativo en el exón 2 presentarían todos los dominios de la isoforma canónica y únicamente carecerían del péptido señal correspondiente. Por ello, consideramos que sería de gran relevancia estudiar en un futuro las consecuencias de la pérdida de este péptido señal en la posible localización de los receptores resultantes. Además, la mayoría de estas variantes de *splicing* que presentan la retención del intrón 1 (y otros) contienen en su secuencia un PTC y, por lo tanto, deberían ser degradadas rápidamente por el mecanismo NMD.

Además, hemos encontrado eventos de *splicing* que resultan en cambios en los dominios extracelulares que pueden afectar a los dominios de unión por ligando, por

inserción, delección o eliminación parcial del dominio IgV de RAGE. Se ha sugerido que la unión de los ligandos AGEs (Grossin et al. 2009) y los péptidos A $\beta$  (Ding and Keller 2005a) al dominio IgV en el receptor RAGE anclado a membrana podría contribuir a la patogénesis de complicaciones vasculares diabéticas y a la enfermedad del Alzheimer, respectivamente. Las modificaciones en el dominio IgV podría generar cambios en la afinidad por ligando de RAGE, de tal forma que las variantes de *splicing* con estas modificaciones en dichos dominios podrían ser factores importantes en la modulación de la enfermedad. Además, existen otras variantes que presentan todos los dominios Ig de la isoforma canónica pero, sin embargo, contienen una delección que reduciría la distancia entre la región transmembrana y el consecuente dominio IgC constante. Esta reducción podría estar impidiendo el plegamiento correcto de los dominios Ig y, por tanto, estar afectando la interacción de los ligandos. Un ejemplo de esta isoforma sería la variante *RnorRAGEv3*, que junto a la isoforma canónica Tv1-RAGE, son las dos variantes mayoritariamente expresadas en tejido embrionario de ratón. La expresión de ambas isoformas a niveles similares podría estar modulando la formación de los dímeros funcionales (homo y heterodímeros) y, modulando la señalización de RAGE en el tejido embrionario.

Sin embargo, estas variantes de *splicing* parecen ser estables, ya que se han identificado tanto en estudios previos (Hudson et al. 2008b; Kalea et al. 2009) como en nuestro estudio siendo relativamente abundantes. Las únicas dos variantes que se han encontrado en todas las especies son las isoformas equivalentes a las humanas Tv1-RAGE y Tv10-RAGE, siendo además las más abundantes en la mayoría de los tejidos. Además, la variante Tv10-RAGE (N-truncada) presenta un PTC en el intrón 1 retenido y, aún a pesar de ello, se encuentra incluso con mayores niveles de expresión que la isoforma Tv1-RAGE en algunas especies (grupo de artiodáctilos). Este dato es muy interesante desde el punto de vista evolutivo, ya que la conservación de la variante Tv10-RAGE (a la que le faltaría el dominio IgV) que se encuentra alta y ampliamente expresada en los diferentes tejidos y especies podría estar relacionado con la generación de una nueva función hasta ahora desconocida. En las muestras de humano, mono y cerdo, las predicciones proteicas parecen indicar la pérdida del primer dominio IgV, mientras que en las muestras de vaca y rata esta variante también codificaría el primer dominio IgV, aunque tiene que ser considerado que carecería del péptido señal correspondiente. Además, cuando se estudia esta isoforma en humano a nivel proteico se observan unos niveles de expresión muy bajos. Según

---

predicciones bioinformáticas esta variante Tv10-RAGE sería un transcrito no codificante en ratón por la presencia de múltiples PTC a lo largo de toda su secuencia.

Por otro lado, estudios previos mostraban otras variantes de *splicing* encontradas en muestras de tejido cerebral, como el dominante negativo de RAGE (dnRAGE o RAGE $\Delta$ ), que únicamente carecía de 16 aminoácidos en el dominio intracelular en comparación con la isoforma canónica (Ding and Keller 2005b). Esta variante de *splicing* puede interaccionar con el ligando A $\beta$  en los dominios Ig extracelulares pero sin generar la señalización intracelular de RAGE al carecer de dicha región intracelular y, de este modo, secuestrar los ligandos. Sin embargo, ni en nuestro estudio ni en otros estudios previos se han encontrado este tipo de variantes dnRAGE (Hudson et al. 2008b; Kalea et al. 2009; Lopez-Diez et al. 2013; Sterenczak et al. 2009; Yonekura et al. 2003), por lo que la variante dnRAGE podría ser una variante específica de tejido o de alguna patología como Alzheimer. En este estudio no hemos detectado ninguna variante que presentase alguna alteración en la cola citoplasmática en ninguna de las muestras de mamíferos analizadas.

En el presente trabajo hemos encontrado nuevos transcritos de función desconocida, que serían considerados principalmente como RNAs no codificantes (ncRNAs) de mayor y menor tamaño. La retención de intrones es un evento de *splicing* común en las variantes potenciales ncRNA mayores, mientras que la exclusión de múltiples exones genera potencialmente las variantes ncRNA de menor tamaño. Estudios previos relacionados con las variantes ncRNAs mostraban genes no codificantes como es el caso de XIST (Duret et al. 2006) o HOTAIR (Rinn et al. 2007). Pero poco se sabe de genes codificantes con variantes de AS no codificantes, los cuales podrían incrementar significativamente el número de ncRNAs. Además, estos ncRNAs podrían estar realizando funciones intracelulares vitales, como la regulación transcripcional o epigenética entre otros (Kikuchi et al. 2009; Marques and Ponting 2009; Ponting et al. 2009), siendo generados a partir de transcritos individuales o debidos a múltiples rondas de fragmentación de un único gen (Tuck and Tollervey 2011).

En este caso particular, las variantes de RAGE no codificantes podrían estar regulando la isoforma canónica de RAGE y/o sus funciones en cada tejido específico. Estas variantes no codificantes de mayor y menor tamaño parecen estar involucradas en diferentes procesos, ya que las variantes ncRNAs de mayor tamaño muestran una expresión amplia en los diferentes tejidos y, por el contrario, las variantes de menor tamaño solo presentan unos niveles de expresión bajos y con una distribución muy específica de tejidos. Además, hemos observado que la mayoría de los ncRNAs detectados en nuestro estudio parecen ser

variantes de *splicing* específicos de especie, encontrándose solo unos pocos conservados entre los diferentes mamíferos analizados. Estas variantes no conservadas podrían ser debidas al fenómeno denominado ruido transcripcional o *noisy splicing* (Baek and Green 2005; Melamud and Moulton 2009; Pickrell et al. 2010; C. Zhang et al. 2007; Z. Zhang et al. 2009), ya que este fenómeno está asociado a genes de baja expresión y uniones de *splicing* no conservadas en diferentes tejidos y especies. Algunos autores han mencionado la implicación del ruido transcripcional como una propiedad importante en la evolución del genoma (Pickrell et al. 2010). Sin embargo, para caracterizar esto en detalle sería totalmente imprescindible incrementar el número de tejidos analizados en cada especie, siendo la única forma para determinar verazmente la especificidad de especie de cada una de las variantes de *splicing* detectadas. Además, es importante destacar que tanto la distribución diferencial como la especificidad de especies de las variantes no codificantes de RAGE podrían generar una amplia diversidad funcional que hasta la fecha se mantiene desconocida.

### 5.3. El mecanismo de quimerismo entre dos genes. La frontera entre la definición de gen, fenómenos de *splicing* y la generación de quimeras.

En este estudio, hemos descrito la presencia de dos tipos de transcritos completamente diferentes generados como una combinación de los genes *G6F* y *LY6G6D*. El mecanismo denominado TIC o *Tandem Chimerism* es posiblemente el responsable de la generación de un único pre-mRNA entre estos dos genes. Sin embargo, otro mecanismo como es el *trans-splicing* podría también estar involucrado en la formación de este tipo de transcritos. Los genes a estudio son adyacentes y se encuentran igualmente orientados, siendo ambas características esenciales para definir un evento de TIC según los estudios previos que establecieron este mecanismo (Akiva et al. 2006; Parra et al. 2006). Por el contrario, estos transcritos quiméricos carecen de otras características posibles en los eventos de TIC que si se encuentran presentes en otras quimeras que se han estudiado previamente en el MHC de clase III. Estas características son la retención de la región intergénica completa situada entre los genes que forman la quimera en alguno de los transcritos quiméricos, como es el caso de la quimera entre *CSNK2B* y *LY6G5B* (Hernandez-Torres et al. 2013), o la formación de un nuevo exón en esa región intergénica, como es el caso de la quimera entre *PBX2* y *RAGE* (datos sin publicar). Sin embargo, los transcritos quiméricos de interés en este estudio no presentan ninguna de estas modificaciones. Por otro lado, para la generación de eventos de *trans-splicing* se requiere de las secuencias consenso de *splicing* que no se encontrarían en los extremos de los exones

terminales de ambos genes. Dado que estos exones se encuentran presentes en algunos de los transcritos a estudio, consideramos que la explicación más probable para su formación sea mediante el mecanismo TIC teniendo que considerar dichos transcritos como quiméricos.

Descartamos asimismo la posibilidad de que estos transcritos en realidad fuesen generados debido a reordenamientos génicos en lugar de a uniones transcripcionales fusionadas (Carninci et al. 2005; Kapranov et al. 2005) ya que encontramos ambos genes expresándose también de manera individual con transcritos independientes a cada gen. El hecho de que las quimeras se encuentren conservadas en tejidos y especies muy diferentes apoya igualmente la afirmación anterior. Además, por la naturaleza de las muestras es difícil la formación de este tipo de reordenamientos génicos en una zona del genoma tan conservada como el MHC de clase III, ya que se trata de RNA procedente de tejidos vivos y no de células en cultivo que tienden a sufrir este tipo de reordenamientos.

Además, si estudiamos en detalle la quimera corta shch, sería difícil determinar si este transcrito está formado por un mecanismo de quimerismo o, en su lugar, se trata únicamente de la presencia de un primer promotor y exón alternativo debido a un fenómeno de AS. Esta última opción se ve argumentada con diversos estudios donde se han analizado los sitios de inicio y terminación de la transcripción en diferentes especies, y se ha observado como en el 87% de los genes que presentaban extensiones en el 5' terminal de sus transcritos, dichas extensiones alcanzaban el gen situado anteriormente (Carninci et al. 2005; Denoeud et al. 2007; Kapranov et al. 2005), debido a eventos de *splicing* o a la presencia de toda la región intergénica. La presencia de estos exones iniciales alternativos múltiples fue también descrito en muchos genes humanos con diferencias en la regulación génica (Zhang et al 2004). Sin embargo, en los estudios de quimerismo se considera que el 12% de las quimeras contribuyen solo con el primer exón del primer gen (Akiva et al. 2006). Nosotros consideraríamos a este tipo de transcritos como quiméricos pues, ya que la quimera corta está a veces formada por el exón 4 de *G6F* pero en otras ocasiones también se incluye en el transcrito los exones 5 y/o 6.

Un detalle a considerar es el gen *LY6G6E* que se encuentra localizado entre los dos genes que forman la quimera y, aunque *LY6G6E* es un pseudogen en humano, este gen es codificante en el resto de las especies analizadas. Sin embargo, la orientación es la contraria a la de los genes contiguos *G6F* y *LY6G6D* y, por ello, se supone que no debería interferir este fenómeno de quimerismo. Además, las dos quimeras se encuentran conservadas entre la mayoría de las diferentes especies analizadas, siendo esta conservación específica de

tejido en el caso de la quimera larga lgch (conservación a nivel de primates) y ampliamente expresada en la mayoría de los tejidos y especies en el caso de la quimera corta shch (al menos podemos reportar su conservación a nivel de mamíferos). Además de esto, ambas quimeras muestran unos patrones de expresión completamente diferentes comparadas entre ellas y con las isoformas canónicas de *G6F* y *LY6G6D*. Sus niveles de conservación y expresión específica de tejido podrían indicar que estas quimeras juegan un papel relevante aún desconocido y debería ser investigado en un futuro.

Una característica que observamos en una gran parte de estos transcritos quiméricos es su formación mediante la unión entre el penúltimo exón del gen *G6F* y el segundo exón del gen *LY6G6D*, lo que es consistente con estudios previos (Tabla 2-Introducción) (Akiva et al. 2006; Nacu et al. 2011; Parra et al. 2006). Esta característica elimina el codón de terminación presente en el exón 6 de *G6F* permitiendo una traducción continuada con el siguiente gen en la quimera. Además, elimina la potencial región 3'UTR perteneciente al gen situado en el 5' y que podría incorporar una secuencia de aminoácidos inservible o, incluso la presencia de un PTC. De hecho, el primer exón de *LY6G6D* solo se retiene en un transcrito quimérico de rata, mientras que el último exón de *G6F* se encuentra presente en un transcrito quimérico del mono Rhesus y en tres transcritos de ambas especies de roedores analizadas, con inclusión de PTC en estos últimos. Sin embargo, en ninguna de las dos quimeras estudiadas, lgch y shch, nos encontramos con la presencia de un exón intergénico presente entre ambos genes. Aunque este dato es característico de únicamente el 12 o el 21% de las quimeras, según estudios (Akiva et al. 2006; Parra et al. 2006), este hecho podría ser debido a la localización de otro gen en la orientación contraria entre ambos genes. La presencia de un gen en la secuencia situada entre ambos genes podría haber impedido la generación y selección de un nuevo exón, ya que esto podría haber provocado la interrupción de secuencias importantes para su expresión.

Además, la cuestión que surge tras este estudio está relacionada con la estructura génica y las fronteras de su definición. Como es actualmente conocido, estamos describiendo dos genes completamente diferentes involucrados en dos procesos de quimerismo en tándem (Akiva et al. 2006; Parra et al. 2006). Con estos datos se podría proponer otro punto de vista sobre el mismo proceso, asumiendo que un gen puede tener diferentes promotores y primeros exones alternativos (Blencowe 2006), por lo tanto, se podría también considerar que estamos estudiando un gen único con tres exones de inicio alternativos (hasta ahora descritos como exón 1 de *G6F*, exón 4 de *G6F* y exón 1 de *LY6G6D*). Las variantes de AS podrían estar generando proteínas de funciones similares o



---

completamente diferentes (Kelemen et al. 2013). Además, cada una de las variantes descritas está diferencialmente expresada en los tejidos, e inclusive en las especies analizadas (con el mismo o diferente tejido en las diferentes especies), tal y como estaba descrito en otros genes con múltiples variantes de *splicing* (Lopez-Diez et al. 2013).

#### 5.4. Implicaciones funcionales del AS y quimerismo en los genes *G6F* y *LY6G6D*.

El fenómeno de quimerismo entre *G6F* y *LY6G6D* fue inicialmente descrito por Calvanesse et al. (2008) a partir del uso de diferentes líneas celulares. En esta Tesis se describen además unos patrones de conservación entre diferentes tejidos y especies de mamíferos, así como un estudio proteómico completo observando por primera vez diferencias entre dos isoformas quiméricas totalmente diferentes (lgch y shch) generadas a partir de los dos mismos genes originales. Aunque la función de las quimeras es todavía desconocida, algunos autores defienden que este tipo de fusiones podría generar proteínas bi-funcionales que podrían compartir propiedades de ambas proteínas originales (Akiva et al. 2006; Frenkel-Morgenstern et al. 2012). Para tratar de demostrar esto, hemos estudiado detalladamente ambas quimeras en comparación con las isoformas canónicas de los genes individuales. La quimera larga solo difiere en la región intracelular de la proteína G6F, mientras que conserva los mismos dominios extracelulares. Además, ambas proteínas son proteínas transmembrana de tipo I y parecen estar procesadas por los mismos mecanismos de glicosilación, por lo que los dominios IgV podrían ser idénticos en ambas proteínas. Debido a esto, ambas proteínas deberían interactuar con los mismos ligandos, aunque la transducción de señales podría ser diferente ya que hemos comprobado que la quimera larga lgch no puede ser fosforilada en su región intracelular de la misma manera que G6F. Sin embargo, ya que LY6G6D es una proteína extracelular anclada a la membrana a través de un anclaje por GPI (Mallya et al. 2006), la región intracelular de la quimera larga que corresponde con la proteína LY6G6D no conservaría las propiedades relacionadas con la isoforma canónica. Por ello, la quimera larga presumiblemente compite por la unión de los ligandos de G6F y previene la activación de la isoforma canónica, actuando como un posible dominante negativo. Nuevos estudios en relación a este tema son necesarios para entender la transducción de señales de G6F y, posteriormente, clarificar la validez de esta hipótesis.

En relación con la quimera corta, la presencia de un promotor alternativo podría estar generando esta isoforma debido a los fenómenos de AS (Blencowe 2006) y

---

quimerismo, lo que generaría una proteína quimérica que comparte solo el exón 4 de G6F y los dos últimos exones de LY6G6D. Esta quimera presenta un péptido señal diferente y, por lo tanto, las modificaciones post-traduccionales se encuentran alteradas. En este estudio hemos sido capaces de describir diferencias en la glicosilación de las proteínas, careciendo de cualquier tipo de glicosilación la quimera corta mientras que la proteína LY6G6D presenta O-glicosilaciones y modificaciones del residuo de ácido siálico (Mallya et al. 2006). Ya que es conocida la influencia de las modificaciones post-traduccionales en la estructura de la proteína y la estabilidad del centro activo (Rudd and Dwek 1997), así como en la habilidad de interacción a ligando del dominio Ly6 (Moller et al. 1993), la proteína quimérica podría estar uniéndose a los mismos ligandos con diferente afinidad, o uniéndose a diferentes ligandos, o inclusive perdiendo la capacidad de interacción con ningún ligando y, por tanto, teniendo una función aún desconocida.

Además, la presencia de sitios de inicio de transcripción alternativos se ha visto implicado en la organización de las proteínas en la membrana (M. J. Davis et al. 2006). Aunque esta quimera presenta los aminoácidos responsables para el anclaje por GPI, en este estudio hemos demostrado que la quimera no se ancla a la membrana por este tipo de modificaciones. En su lugar, se observa como la región C-terminal de la proteína quimérica es extracelular, de forma que la región transmembrana de la proteína G6F canónica parece estar también anclando la quimera corta a la membrana. En este caso particular, la región transmembrana de paso único localizada en el exón 4 de *G6F* parece estar generando una proteína transmembrana de tipo II (proteína G6F). Hay que tener en cuenta la incorporación de la etiqueta en la región extracelular próxima a la región transmembrana, lo que podría estar alterando el comportamiento de esta proteína quimérica. Esta proteína quimérica parece compartir propiedades de las dos proteínas originales, la región transmembrana y el dominio LY6 respectivamente, pero ambas propiedades se encontrarían modificadas. Por último, esta quimera corta se encontraría regulada tanto transcripcionalmente (por el promotor), como durante la traducción (5'UTR) por el gen *G6F* en contraste con la regulación de la isoforma canónica LY6G6D correspondiente al segundo gen.

En relación con el gen *LY6G6D* y sus variantes de *splicing*, se ha observado la presencia del evento de retención del intrón 1 conservado entre diferentes especies, como es el caso de mono Rhesus, cerdo y ratón. Este fenómeno ya se había descrito previamente tanto en este gen como en otros miembros de la familia Ly6 (Calvanese et al. 2008; Hernandez-Torres et al. 2013; Mallya et al. 2002), así como hemos observado que predominaba en las variantes descritas de RAGE en este estudio (Lopez-Diez et al. 2013).

Observamos como la retención del primer intrón es una característica importante que se encuentra conservada entre diferentes tejidos y especies. La conservación de este evento nos podría indicar un posible desacople de la maquinaria de transcripción y la maquinaria de *splicing*, pudiendo constituir un reservorio de mensajeros preparados para una posible respuesta rápida ante una señal determinada. Este tipo de desacople ya se había evidenciado en múltiples estudios donde se observaba el mecanismo de *splicing* en el citoplasma en lugar de en el núcleo celular en respuesta a estrés de retículo (Calfon et al. 2002; Gonzalez et al. 1999; Kawahara et al. 1998; Ruegsegger et al. 2001; Sidrauski et al. 1996; Sidrauski and Walter 1997; Uemura et al. 2009; Yoshida et al. 2001; Yoshida 2007).

Para concluir, es importante hacer énfasis en la relevancia tanto de los fenómenos de *splicing* como en el mecanismo de quimerismo para el aumento de la diversidad transcriptómica y proteómica de un individuo. Al igual que las implicaciones funcionales que hemos observado resultantes de la formación de transcritos quiméricos entre los genes *G6F* y *LY6G6D*, estos tipos de implicaciones se podrían ampliar a cualquier evento de quimerismo, ya que consideramos que es un mecanismo con enormes repercusiones a nivel proteico, combinando dominios proteicos que *a priori* no se podrían encontrar en las proteínas canónicas.

### 5.5. La enfermedad AR en la región del MHC de clase III

En este estudio hemos comparado los niveles de expresión de diferentes genes, entre los que se incluyen los 60 genes del MHC de clase III, entre pacientes de AR e individuos controles sanos. Los pacientes incorporados a nuestro estudio fueron seleccionados por características clínicas concretas para reducir la variabilidad de las muestras. En concreto se seleccionaron únicamente aquellas pacientes de sexo femenino, ya que esta enfermedad afecta con mayor predominancia a las mujeres frente a los varones, y que presentasen el factor reumatoide positivo (RF+). Esta última selección es debida a que aquellos pacientes con el factor reumatoide negativo (RF-) podrían estar erróneamente clasificados como pacientes de AR y, en realidad, estar padeciendo otro tipo de enfermedad reumática. Por lo tanto, aquellos estudios que incorporan pacientes tanto con RF+ como RF- deberían ser estudiados en detalle para evitar clasificaciones erróneas de genes asociados a enfermedades reumáticas y no a la enfermedad AR en concreto (Bovin et al. 2004).

Además, hemos seleccionado únicamente aquellos pacientes que habían sido tratados únicamente con AINEs y/o FAMEs, y nunca habían recibido un tratamiento de

segunda línea o biológico. Esta selección fue realizada ya que los tratamientos biológicos bloquean directamente la señalización de TNF (Infliximab, Remidace<sup>®</sup>; Etanercept, Enbrel<sup>®</sup>; Adalimumab, Humira<sup>®</sup>) o las diferentes citoquinas implicadas en inflamación (IL6: Tocilizumab, Actemra<sup>®</sup>; IL1: Anakinra, Kineret<sup>®</sup>; CD20: Rituximab o RTX, Mabthera<sup>®</sup>), y estarían modificando la expresión génica como consecuencia del tratamiento y no como consecuencia de la enfermedad, por lo que no valdría como método ni de detección ni de prognosis de la enfermedad. En un estudio donde se analizó la expresión de pacientes de AR con muestras de PBMCs previas y posteriores al tratamiento con Etanercept y Adalimumab se observó cómo, al igual que en nuestro estudio, no presentaba alteración de la expresión de TNF aún siendo la molécula inflamatoria clásica de esta enfermedad (Meugnier et al. 2011). Únicamente se observó una disminución en la expresión del receptor TNFR1 en las muestras de pacientes tras el tratamiento, siendo esta variación consecuencia directa del tratamiento biológico.

Al contrario de lo que ocurre con otras enfermedades, todos los estudios que se han realizado para identificar genes con niveles de expresión alterados en pacientes con AR se han obtenido a partir de un número muy limitado de muestras, lo que limita las conclusiones obtenidas por los mismos. Se han realizado estudios desde únicamente ocho (Meugnier et al. 2011) o nueve muestras (Edwards et al. 2007), hasta un máximo de 29 muestras (Batliwalla et al. 2005), encontrándose el resto en un número modesto de 14 (Bovin et al. 2004) y 18 muestras (Teixeira et al. 2009). En nuestro estudio hemos incorporado 31 muestras, ligeramente superior al estudio de Batliwalla et al. (2005), siendo esta Tesis el mayor estudio de expresión en pacientes con AR realizado hasta la fecha, lo que aporta unos análisis estadísticos y unas conclusiones más sólidas.

El tipo de tecnología empleada también podrían alterar las conclusiones obtenidas en el estudio. Estos estudios se han realizado preferentemente mediante la tecnología de *microarrays* cuantificando de esta forma la expresión relativa global de las muestras (Batliwalla et al. 2005; Bovin et al. 2004; Edwards et al. 2007; Teixeira et al. 2009), aunque únicamente dos de ellos han realizado la validación necesaria de dichos experimentos por RT-PCR cuantitativa (Bovin et al. 2004; Teixeira et al. 2009). De hecho, el único gen relacionado con AR en el que no se ha reproducido la significación estadística de su variación es el gen *NFAT5*, que se encontró alterado con unos valores muy elevados mediante la tecnología de *arrays* de expresión pero que no fue nunca validado mediante qRT-PCR (Teixeira et al. 2009). En el presente estudio no se ha podido encontrar alteraciones significativas de este gen en pacientes con AR. Los otros dos genes estudiados

que se habían relacionado previamente como alterados tanto en muestras de PBMCs de pacientes con AR como en muestras enriquecidas en monocitos son *IL1RN* y *QPCT* (Batliwalla et al. 2005). Ambos genes presentan variaciones significativas en nuestro estudio, dándole fortaleza a nuestro estudio.

El análisis estadístico de los resultados obtenidos por estas tecnologías cumplen un papel fundamental para la interpretación de los resultados, ya que todos los resultados de estudios previos se han analizado mediante análisis estadísticos paramétricos (t de Student o ANOVA) sin realizar un estudio de normalidad de las muestras en ninguno de los casos (Batliwalla et al. 2005; Bovin et al. 2004; Edwards et al. 2007; Teixeira et al. 2009). Tras realizar dicho estudio en nuestras muestras observamos cómo no presentan una distribución normal y, por lo tanto, se debería aplicar un análisis estadístico no paramétrico como el que empleamos en nuestro estudio. Desconocemos si las muestras empleadas en estos estudios previos presentaban una distribución normal o no, pero presumiblemente podríamos dudar dicha distribución por la naturaleza de las muestras. Si tenemos en cuenta, por ejemplo, los niveles de expresión del gen *S100A12*, observamos como en varios estudios previos presenta una variación significativa en los pacientes de AR (Bovin et al. 2004; Teixeira et al. 2009), mientras que en nuestro estudio no son significativas sus diferencias. Sin embargo, si modificamos el análisis estadístico realizando uno paramétrico (en lugar de no paramétrico) en nuestras muestras observamos que sí aparecen esas diferencias significativas. Además, el valor de significación empleado en dichos estudios es menos restrictivo (p-valor < 0.05) (Bovin et al. 2004; Edwards et al. 2007; Teixeira et al. 2009) que el empleado en nuestro estudio, a excepción del estudio realizado en el grupo de Batliwalla et al. (2005) donde el nivel de significación fue menor, con un p-value < 0.001. Esto implicaría que se han asignado diferencias significativas en la expresión de un mayor número de genes en comparación con nuestro estudio, al ser nuestro análisis más restrictivo (p-value < 0.01), pero al mismo tiempo más sólido.

Comparando de nuevo con nuestro estudio, en uno de los estudios previos realizados en *microarrays* se encontraron variaciones significativas en diez genes del MHC, de los cuales tres de ellos estaban situados en el MHC de clase III: los genes *AGPAT*, *MICB* y *STK19* (Edwards et al. 2007). Debido a la naturaleza de las muestras empleadas en ese estudio (pacientes de AR con la enfermedad activa y únicamente tratados con FAMES), podríamos comparar dichos resultados con nuestro grupo de pacientes con la enfermedad activa. De hecho, nosotros observamos también un aumento significativo en la expresión del gen *STK19* en los pacientes con la enfermedad activa en comparación con aquellos que

la presentaban en remisión. Sin embargo, no se observan diferencias significativas en la expresión de los genes *AGPAT* y *MICB*, aunque la falta de evidencias significativas podría ser debida a las opciones comentadas anteriormente de los análisis únicos por *microarrays* sin validación, el test estadístico o los niveles de significación empleados.

En conclusión, nuestro estudio ha identificado varios genes en PBMCs de pacientes con AR que podrían contribuir como marcadores de presencia de la enfermedad (*VAR5*, *DDAH2*, *C6orf25*, *C6orf47*, *CFB*, *LSM2*, *LY6G6D*, *EGFL8*, *MSH5*, *LY6G5B* y *NCR3*), así como de prognosis de la enfermedad (*STK19*, *LTB*, *LSM2* y *BATI*) dentro de la región del MHC de clase III. Sin embargo, sería necesario un análisis comparativo con otras enfermedades o procesos inflamatorios para poder determinar la relación directa de las variaciones en la expresión de dichos genes y los procesos patológicos específicos de la enfermedad AR.

## **6. CONCLUSIONES**

1. RAGE undergoes extensive AS in all the analysed mammalian species, generating different coding and non-coding splice variants.
2. Most of the found RAGE spliced variants seem to be species-specific and only two of them, Tv1-RAGE and Tv10-RAGE (N-truncated) are conserved among the analysed mammals.
3. The characterisation of splice variants by RNA-seq up to date is not possible for low expressed genes such as *RAGE*, *G6F* or *LY6G6D*.
4. Both Tv1-RAGE and Tv2-RAGE protein isoforms bind the S100A12 ligand.
5. Transcription induced chimerism (TIC) play an important role in *G6f* and *Ly6G6d* gene expression generating two completely different chimeric proteins: long (lgch) and short (shch) chimera, which also present AS and tissue and species specificity, as well as the independent genes.
6. The G6F and the long chimera (lgch) proteins can form homo and heterodimers, potentially binding to the same ligands, but their intracellular signaling should be different due to the lack of phosphorylation in the intracellular part of lgch.
7. LY6G6D and the short chimera (shch) proteins differ in their glycosylation pattern and their subcellular localisation, being LY6G6D a GPI anchorage protein and shch a transmembrane type II protein.
8. Some genes of the MHC class III region have significantly altered their expression levels in AR patients compared to controls, as well as between active and remission patients. However, *RAGE* and *G6F-LY6G6D* and some of their splicing events do not present any significant alteration of their expression.



1. *RAGE* sufre múltiples eventos de AS en todas las especies de mamíferos analizadas, generando diferentes variantes de *splicing* tanto codificantes como no codificantes.
2. La mayoría de las variantes de *splicing* de *RAGE* encontradas parecen ser específicas de especie y únicamente dos de ellas, Tv1-*RAGE* y Tv10-*RAGE* (N-truncada) se encuentran conservadas en las muestras de los seis mamíferos analizados.
3. La caracterización de las variantes de *splicing* por RNA-seq hasta la fecha no es posible para genes de baja expresión tales como *RAGE*, *G6F* y *LY6G6D*.
4. Ambas isoformas proteicas Tv1-*RAGE* y Tv2-*RAGE* se unen al ligando S100A12.
5. El mecanismo de quimerismo (TIC) juega un papel importante en la expresión de los genes *G6f* y *Ly6G6d* generando dos proteínas quiméricas completamente diferentes: la quimera larga (lgch) y corta (shch), que también presentan AS y especificidad de tejido y de especie, al igual que los genes independientes.
6. Las proteínas G6F y la quimera larga (lgch) pueden formar homo y heterodímeros, uniéndose potencialmente a los mismos ligandos, pero su señalización intracelular sería diferente debido a la pérdida de fosforilación en la región intracelular de la quimera lgch.
7. Las proteínas LY6G6D y la quimera corta (shch) difieren en sus patrones de glicosilación y su localización subcelular, siendo la proteína LY6G6D una proteína anclada por GPI a la membrana plasmática y la quimera corta una proteína transmembrana de tipo II.
8. Algunos genes de la región del MHC de clase III presentan significativamente alterados sus niveles de expresión en pacientes con AR comparados con los controles, así como entre pacientes con la enfermedad activa y en remisión. Sin embargo, tanto los genes *RAGE* y *G6F-LY6G6D* como sus eventos de *splicing* estudiados no presentan cambios significativos en su expresión.

## **7. BIBLIOGRAFÍA**

**A**

- Abelson, J., Trotta, C. R., and Li, H. (1998), 'tRNA splicing', *J Biol Chem*, 273 (21), 12685-8.
- Aguado, B. and Campbell, R. D. (1995), 'The novel gene G17, located in the human major histocompatibility complex, encodes PBX2, a homeodomain-containing protein', *Genomics*, 25 (3), 650-9.
- Aguado, B. and Campbell, R. D. (1998), 'Characterization of a human lysophosphatidic acid acyltransferase that is encoded by a gene located in the class III region of the human major histocompatibility complex', *J Biol Chem*, 273 (7), 4096-105.
- Akiva, P., et al. (2006), 'Transcription-mediated gene fusion in the human genome', *Genome Res*, 16 (1), 30-6.
- Andersen, C. L., Jensen, J. L., and Orntoft, T. F. (2004), 'Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets', *Cancer Res*, 64 (15), 5245-50.
- Ashburner, M., et al. (2000), 'Gene ontology: tool for the unification of biology. The Gene Ontology Consortium', *Nat Genet*, 25 (1), 25-9.
- Aznarez, I., et al. (2003), 'Characterization of disease-associated mutations affecting an exonic splicing enhancer and two cryptic splice sites in exon 13 of the cystic fibrosis transmembrane conductance regulator gene', *Hum Mol Genet*, 12 (16), 2031-40.

**B**

- Baek, D. and Green, P. (2005), 'Sequence conservation, relative isoform frequencies, and nonsense-mediated decay in evolutionarily conserved alternative splicing', *Proc Natl Acad Sci U S A*, 102 (36), 12813-8.
- Baker, K. E. and Parker, R. (2004), 'Nonsense-mediated mRNA decay: terminating erroneous gene expression', *Curr Opin Cell Biol*, 16 (3), 293-9.
- Bali, D., et al. (1999), 'Genetic analysis of multiplex rheumatoid arthritis families', *Genes Immun*, 1 (1), 28-36.
- Barash, Y., et al. (2010), 'Deciphering the splicing code', *Nature*, 465 (7294), 53-9.
- Barbosa-Morais, N. L., et al. (2012), 'The evolutionary landscape of alternative splicing in vertebrate species', *Science*, 338 (6114), 1587-93.
- Batliwalla, F. M., et al. (2005), 'Peripheral blood gene expression profiling in rheumatoid arthritis', *Genes Immun*, 6 (5), 388-97.
- Beckmann, J. S. and Trifonov, E. N. (1991), 'Splice junctions follow a 205-base ladder', *Proc Natl Acad Sci U S A*, 88 (6), 2380-3.
- Berget, S. M. (1995), 'Exon recognition in vertebrate splicing', *J Biol Chem*, 270 (6), 2411-4.
- Berget, S. M., Moore, C., and Sharp, P. A. (1977), 'Spliced segments at the 5' terminus of adenovirus 2 late mRNA', *Proc Natl Acad Sci U S A*, 74 (8), 3171-5.
- Blanchette, M., et al. (2005), 'Global analysis of positive and negative pre-mRNA splicing regulators in *Drosophila*', *Genes Dev*, 19 (11), 1306-14.
- Blencowe, B. J. (2000), 'Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases', *Trends Biochem Sci*, 25 (3), 106-10.
- Blencowe, B. J. (2006), 'Alternative splicing: new insights from global analyses', *Cell*, 126 (1), 37-47.
- Blencowe, B. J., et al. (1998), 'A coactivator of pre-mRNA splicing', *Genes Dev*, 12 (7), 996-1009.
- Blumenthal, T., et al. (2002), 'A global analysis of *Caenorhabditis elegans* operons', *Nature*, 417 (6891), 851-4.
- Boise, L. H., et al. (1993), 'bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death', *Cell*, 74 (4), 597-608.
- Bonen, L. (1993), 'Trans-splicing of pre-mRNA in plants, animals, and protists', *FASEB J*, 7 (1), 40-6.
- Bonnal, S. and Valcarcel, J. (2008), 'Molecular biology: spliceosome meets telomerase', *Nature*, 456 (7224), 879-80.
- Boue, S., Letunic, I., and Bork, P. (2003), 'Alternative splicing and evolution', *Bioessays*, 25 (11), 1031-4.
- Bovin, L. F., et al. (2004), 'Blood cell gene expression profiling in rheumatoid arthritis. Discriminative genes and effect of rheumatoid factor', *Immunol Lett*, 93 (2-3), 217-26.
- Box, J. A., et al. (2008), 'Spliceosomal cleavage generates the 3' end of telomerase RNA', *Nature*, 456 (7224), 910-4.
- Breen, M. A. and Ashcroft, S. J. (1997), 'A truncated isoform of Ca<sup>2+</sup>/calmodulin-dependent protein kinase II expressed in human islets of Langerhans may result from trans-splicing', *FEBS Lett*, 409 (3), 375-9.
- Brown, M. A., et al. (1998), 'The effect of HLA-DR genes on susceptibility to and severity of ankylosing spondylitis', *Arthritis Rheum*, 41 (3), 460-5.
- Buratti, E. and Baralle, F. E. (2004), 'Influence of RNA secondary structure on the pre-mRNA splicing process', *Mol Cell Biol*, 24 (24), 10505-14.
- Buratti, E., et al. (2007), 'SR protein-mediated inhibition of CFTR exon 9 inclusion: molecular characterization of the intronic splicing silencer', *Nucleic Acids Res*, 35 (13), 4359-68.
- Bustin, S. A. (2000), 'Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays', *J Mol Endocrinol*, 25 (2), 169-93.

Bustin, S. A., et al. (2013), 'The need for transparency and good practices in the qPCR literature', *Nat Methods*, 10 (11), 1063-7.

## C

- Caceres, J. F. and Kornblihtt, A. R. (2002), 'Alternative splicing: multiple control mechanisms and involvement in human disease', *Trends Genet*, 18 (4), 186-93.
- Calame, K., et al. (1980), 'Mouse Cmu heavy chain immunoglobulin gene segment contains three intervening sequences separating domains', *Nature*, 284 (5755), 452-5.
- Calarco, J. A., Zhen, M., and Blencowe, B. J. (2011), 'Networking in a global world: establishing functional connections between neural splicing regulators and their target transcripts', *RNA*, 17 (5), 775-91.
- Calarco, J. A., et al. (2007), 'Technologies for the global discovery and analysis of alternative splicing', *Adv Exp Med Biol*, 623, 64-84.
- Calfon, M., et al. (2002), 'IRE1 couples endoplasmic reticulum load to secretory capacity by processing the XBP-1 mRNA', *Nature*, 415 (6867), 92-6.
- Calvanese, V., et al. (2008), 'Regulation of expression of two LY-6 family genes by intron retention and transcription induced chimerism', *BMC Mol Biol*, 9, 81.
- Campbell, D. A., Thomas, S., and Sturm, N. R. (2003), 'Transcription in kinetoplastid protozoa: why be normal?', *Microbes Infect*, 5 (13), 1231-40.
- Carninci, P., et al. (2005), 'The transcriptional landscape of the mammalian genome', *Science*, 309 (5740), 1559-63.
- Caudevilla, C., et al. (1998), 'Natural trans-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver', *Proc Natl Acad Sci U S A*, 95 (21), 12185-90.
- Cech, T. R. (1990), 'Self-splicing of group I introns', *Annu Rev Biochem*, 59, 543-68.
- Claverie-Martin, F., et al. (2005), 'The Alu insertion in the CLCN5 gene of a patient with Dent's disease leads to exon 11 skipping', *J Hum Genet*, 50 (7), 370-4.
- Cline, M. S., et al. (2004), 'The effects of alternative splicing on transmembrane proteins in the mouse genome', *Pac Symp Biocomput*, 17-28.
- Colak, R., et al. (2013), 'Distinct types of disorder in the human proteome: functional implications for alternative splicing', *PLoS Comput Biol*, 9 (4), e1003030.
- Communi, D., et al. (2001), 'Cotranscription and intergenic splicing of human P2Y11 and SSF1 genes', *J Biol Chem*, 276 (19), 16561-6.
- Cooper, T. A., Wan, L., and Dreyfuss, G. (2009), 'RNA and disease', *Cell*, 136 (4), 777-93.
- Cox, P. R., Siddique, T., and Zoghbi, H. Y. (2001), 'Genomic organization of Tropomodulins 2 and 4 and unusual intergenic and intraexonic splicing of YL-1 and Tropomodulin 4', *BMC Genomics*, 2, 7.
- Cucca, F., et al. (1998), 'Evaluation of IgA deficiency in Sardinians indicates a susceptibility gene is encoded within the HLA class III region', *Clin Exp Immunol*, 111 (1), 76-80.
- Cuperlovic-Culf, M., et al. (2006), 'Microarray analysis of alternative splicing', *OMICS*, 10 (3), 344-57.
- Czaplinski, K., et al. (1998), 'The surveillance complex interacts with the translation release factors to enhance termination and degrade aberrant mRNAs', *Genes Dev*, 12 (11), 1665-77.
- Chang, Y. F., Imam, J. S., and Wilkinson, M. F. (2007), 'The nonsense-mediated decay RNA surveillance pathway', *Annu Rev Biochem*, 76, 51-74.
- Chavakis, T., et al. (2003), 'The pattern recognition receptor (RAGE) is a counterreceptor for leukocyte integrins: a novel pathway for inflammatory cell recruitment', *J Exp Med*, 198 (10), 1507-15.
- Chiba, T., et al. (2011), 'NFKBIL1 confers resistance to experimental autoimmune arthritis through the regulation of dendritic cell functions', *Scand J Immunol*, 73 (5), 478-85.
- Chikatsu, N., et al. (2000), 'Cloning and characterization of two promoters for the human calcium-sensing receptor (CaSR) and changes of CaSR expression in parathyroid adenomas', *J Biol Chem*, 275 (11), 7553-7.
- Choi, S. J., et al. (1995), 'Molecular cloning and sequence analysis of the rat liver carnitine octanoyltransferase cDNA, its natural gene and the gene promoter', *Biochim Biophys Acta*, 1264 (2), 215-22.
- Chow, L. T., et al. (1977), 'An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA', *Cell*, 12 (1), 1-8.
- Christopher, D. A. and Hallick, R. B. (1989), 'Euglena gracilis chloroplast ribosomal protein operon: a new chloroplast gene for ribosomal protein L5 and description of a novel organelle intron category designated group III', *Nucleic Acids Res*, 17 (19), 7591-608.

## D

- Daffu, G., et al. (2013), 'Radical roles for RAGE in the pathogenesis of oxidative stress in cardiovascular diseases and beyond', *Int J Mol Sci*, 14 (10), 19891-910.
- Dagan, T., et al. (2004), 'AluGene: a database of Alu elements incorporated within protein-coding genes', *Nucleic Acids Res*, 32 (Database issue), D489-92.
- Darzacq, X., et al. (2007), 'In vivo dynamics of RNA polymerase II transcription', *Nat Struct Mol Biol*, 14 (9), 796-806.
- Davies, A., et al. (1989), 'CD59, an LY-6-like protein expressed in human lymphoid cells, regulates the action of the complement membrane attack complex on homologous cells', *J Exp Med*, 170 (3), 637-54.

- Davis, M. J., et al. (2006), 'Differential use of signal peptides and membrane domains is a common occurrence in the protein output of transcriptional units', *PLoS Genet*, 2 (4), e46.
- Davis, R. E. (1996), 'Spliced leader RNA trans-splicing in metazoa', *Parasitol Today*, 12 (1), 33-40.
- Davis, R. E. and Hodgson, S. (1997), 'Gene linkage and steady state RNAs suggest trans-splicing may be associated with a polycistronic transcript in *Schistosoma mansoni*', *Mol Biochem Parasitol*, 89 (1), 25-39.
- De Conti, L., Baralle, M., and Buratti, E. (2013), 'Exon and intron definition in pre-mRNA splicing', *Wiley Interdiscip Rev RNA*, 4 (1), 49-60.
- De Vet, E. C., Aguado, B., and Campbell, R. D. (2003), 'Adaptor signalling proteins Grb2 and Grb7 are recruited by human G6f, a novel member of the immunoglobulin superfamily encoded in the MHC', *Biochem J*, 375 (Pt 1), 207-13.
- Delahaye, N. F., et al. (2007), 'Association analyses of NCR3 polymorphisms with *P. falciparum* mild malaria', *Microbes Infect*, 9 (2), 160-6.
- Denoeud, F., et al. (2007), 'Prominent use of distal 5' transcription start sites and discovery of a large number of additional exons in ENCODE regions', *Genome Res*, 17 (6), 746-59.
- Ding, Q. and Keller, J. N. (2005a), 'Evaluation of rage isoforms, ligands, and signaling in the brain', *Biochim Biophys Acta*, 1746 (1), 18-27.
- Ding, Q. and Keller, J. N. (2005b), 'Splice variants of the receptor for advanced glycosylation end products (RAGE) in human brain', *Neurosci Lett*, 373 (1), 67-72.
- Dominski, Z. and Kole, R. (1991), 'Selection of splice sites in pre-mRNAs with short internal exons', *Mol Cell Biol*, 11 (12), 6075-83.
- Douglas, K. B., et al. (2009), 'Complement receptor 2 polymorphisms associated with systemic lupus erythematosus modulate alternative splicing', *Genes Immun*, 10 (5), 457-69.
- Dreyfuss, G., et al. (1993), 'hnRNP proteins and the biogenesis of mRNA', *Annu Rev Biochem*, 62, 289-321.
- Dujardin, G., et al. (2013), 'Transcriptional elongation and alternative splicing', *Biochim Biophys Acta*, 1829 (1), 134-40.
- Duret, L., et al. (2006), 'The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene', *Science*, 312 (5780), 1653-5.

## E

- Early, P., et al. (1980), 'Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways', *Cell*, 20 (2), 313-9.
- Edwards, C. J., et al. (2007), 'Molecular profile of peripheral blood mononuclear cells from patients with rheumatoid arthritis', *Mol Med*, 13 (1-2), 40-58.
- Elkon, R., Ugalde, A. P., and Agami, R. (2013), 'Alternative cleavage and polyadenylation: extent, regulation and function', *Nat Rev Genet*, 14 (7), 496-506.
- Emanuele, E., et al. (2005), 'Circulating levels of soluble receptor for advanced glycation end products in Alzheimer disease and vascular dementia', *Arch Neurol*, 62 (11), 1734-6.
- Englert, M. and Beier, H. (2005), 'Plant tRNA ligases are multifunctional enzymes that have diverged in sequence and substrate specificity from RNA ligases of other phylogenetic origins', *Nucleic Acids Res*, 33 (1), 388-99.

## F

- Falcone, C., et al. (2005), '-374T/A polymorphism of the RAGE gene promoter in relation to severity of coronary atherosclerosis', *Clin Chim Acta*, 354 (1-2), 111-6.
- Faustino, N. A. and Cooper, T. A. (2003), 'Pre-mRNA splicing and human disease', *Genes Dev*, 17 (4), 419-37.
- Fears, S., et al. (1996), 'Intergenic splicing of MDS1 and EVI1 occurs in normal tissues as well as in myeloid leukemia and produces a new member of the PR domain family', *Proc Natl Acad Sci U S A*, 93 (4), 1642-7.
- Fehlbaum, P., et al. (2005), 'A microarray configuration to quantify expression levels and relative abundance of splice variants', *Nucleic Acids Res*, 33 (5), e47.
- Ferguson, M. A. (1992), 'Colworth Medal Lecture. Glycosyl-phosphatidylinositol membrane anchors: the tale of a tail', *Biochem Soc Trans*, 20 (2), 243-56.
- Fernando, M. M., et al. (2012), 'Transancestral mapping of the MHC region in systemic lupus erythematosus identifies new independent and interacting loci at MSH5, HLA-DPB1 and HLA-G', *Ann Rheum Dis*, 71 (5), 777-84.
- Ferrante, E., et al. (2010), 'Determinants of thromboxane biosynthesis in rheumatoid arthritis: Role of RAGE and oxidant stress', *Free Radic Biol Med*, 49 (5), 857-64.
- Filichkin, S. A., et al. (2010), 'Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*', *Genome Res*, 20 (1), 45-58.
- Flatz, L., et al. (2011), 'Single-cell gene-expression profiling reveals qualitatively distinct CD8 T cells elicited by different gene-based vaccines', *Proc Natl Acad Sci U S A*, 108 (14), 5724-9.
- Frenkel-Morgenstern, M. and Valencia, A. (2012), 'Novel domain combinations in proteins encoded by chimeric transcripts', *Bioinformatics*, 28 (12), i67-74.

- Frenkel-Morgenstern, M., et al. (2012), 'Chimeras taking shape: potential functions of proteins encoded by chimeric RNA transcripts', *Genome Res*, 22 (7), 1231-42.
- Fu, R. H., et al. (2013), 'Aberrant alternative splicing events in Parkinson's disease', *Cell Transplant*, 22 (4), 653-61.
- Fu, X. D. (1995), 'The superfamily of arginine/serine-rich splicing factors', *RNA*, 1 (7), 663-80.

**G**

- Garcia-Blanco, M. A., Baraniak, A. P., and Lasda, E. L. (2004), 'Alternative splicing in disease and therapy', *Nat Biotechnol*, 22 (5), 535-46.
- Garcia, A., et al. (2006), 'A global proteomics approach identifies novel phosphorylated signaling proteins in GPVI-activated platelets: involvement of G6f, a novel platelet Grb2-binding membrane adapter', *Proteomics*, 6 (19), 5332-43.
- Geroldi, D., et al. (2005), 'Decreased plasma levels of soluble receptor for advanced glycation end-products in patients with essential hypertension', *J Hypertens*, 23 (9), 1725-9.
- Gerstein, M. B., et al. (2010), 'Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project', *Science*, 330 (6012), 1775-87.
- Ghavami, S., et al. (2008), 'S100A8/A9 at low concentration promotes tumor cell growth via RAGE ligation and MAP kinase-dependent pathway', *J Leukoc Biol*, 83 (6), 1484-92.
- Gilbert, W. (1978), 'Why genes in pieces?', *Nature*, 271 (5645), 501.
- Gilles, A. M., et al. (1991), 'Nucleoside diphosphate kinase from human erythrocytes. Structural characterization of the two polypeptide chains responsible for heterogeneity of the hexameric enzyme', *J Biol Chem*, 266 (14), 8784-9.
- Gingeras, T. R. (2007), 'Origin of phenotypes: genes and transcripts', *Genome Res*, 17 (6), 682-90.
- Gingeras, T. R. (2009), 'Implications of chimaeric non-co-linear transcripts', *Nature*, 461 (7261), 206-11.
- Goh, S. Y. and Cooper, M. E. (2008), 'Clinical review: The role of advanced glycation end products in progression and complications of diabetes', *J Clin Endocrinol Metab*, 93 (4), 1143-52.
- Goncalves, V., Matos, P., and Jordan, P. (2009), 'Antagonistic SR proteins regulate alternative splicing of tumor-related Rac1b downstream of the PI3-kinase and Wnt pathways', *Hum Mol Genet*, 18 (19), 3696-707.
- Gonzalez, T. N., et al. (1999), 'Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway', *EMBO J*, 18 (11), 3119-32.
- Graveley, B. R. (2000), 'Sorting out the complexity of SR protein functions', *RNA*, 6 (9), 1197-211.
- (2008), 'The haplo-spliceo-transcriptome: common variations in alternative splicing in the human population', *Trends Genet*, 24 (1), 5-7.
- Graveley, B. R., et al. (2011), 'The developmental transcriptome of *Drosophila melanogaster*', *Nature*, 471 (7339), 473-9.
- Greetham, D., et al. (2007), 'Functional characterization of NF-kappaB inhibitor-like protein 1 (NFKBIL1), a candidate susceptibility gene for rheumatoid arthritis', *Hum Mol Genet*, 16 (24), 3027-36.
- Grossin, N., et al. (2009), 'Differential effect of plasma or erythrocyte AGE-ligands of RAGE on expression of transcripts for receptor isoforms', *Diabetes Metab*, 35 (5), 410-7.
- Grossin, N., et al. (2008), 'Severity of diabetic microvascular complications is associated with a low soluble RAGE level', *Diabetes Metab*, 34 (4 Pt 1), 392-5.
- Guergnon, J., et al. (2012), 'Single-nucleotide polymorphism-defined class I and class III major histocompatibility complex genetic subregions contribute to natural long-term nonprogression in HIV infection', *J Infect Dis*, 205 (5), 718-24.

**H**

- Hall, S. L. and Padgett, R. A. (1994), 'Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites', *J Mol Biol*, 239 (3), 357-65.
- Hanford, L. E., et al. (2004), 'Purification and characterization of mouse soluble receptor for advanced glycation end products (sRAGE)', *J Biol Chem*, 279 (48), 50019-24.
- Harashima, A., et al. (2006), 'Identification of mouse orthologue of endogenous secretory receptor for advanced glycation end-products: structure, function and expression', *Biochem J*, 396 (1), 109-15.
- Harja, E., et al. (2008), 'Vascular and inflammatory stresses mediate atherosclerosis via RAGE and its ligands in apoE<sup>-/-</sup> mice', *J Clin Invest*, 118 (1), 183-94.
- Hasler, J., Samuelsson, T., and Strub, K. (2007), 'Useful 'junk': Alu RNAs in the human transcriptome', *Cell Mol Life Sci*, 64 (14), 1793-800.
- Hastings, K. E. (2005), 'SL trans-splicing: easy come or easy go?', *Trends Genet*, 21 (4), 240-7.
- Heinrich, B., et al. (2009), 'Heterogeneous nuclear ribonucleoprotein G regulates splice site selection by binding to CC(A/C)-rich regions in pre-mRNA', *J Biol Chem*, 284 (21), 14303-15.
- Hendy, G. N., Canaff, L., and Cole, D. E. (2013), 'The CASR gene: alternative splicing and transcriptional control, and calcium-sensing receptor (CaSR) protein: structure and ligand binding sites', *Best Pract Res Clin Endocrinol Metab*, 27 (3), 285-301.
- Hentze, M. W. (2001), 'Protein synthesis. Believe it or not-translation in the nucleus', *Science*, 293 (5532), 1058-9.

- Heraï, R. H. and Yamagishi, M. E. (2010), 'Detection of human interchromosomal trans-splicing in sequence databanks', *Brief Bioinform*, 11 (2), 198-209.
- Hernandez-Torres, F., Rastrojo, A., and Aguado, B. (2013), 'Intron retention and transcript chimerism conserved across mammals: Ly6g5b and Csnk2b-Ly6g5b as examples', *BMC Genomics*, 14, 199.
- Hernandez, F., et al. (2004), 'Glycogen synthase kinase-3 plays a crucial role in tau exon 10 splicing and intranuclear distribution of SC35. Implications for Alzheimer's disease', *J Biol Chem*, 279 (5), 3801-6.
- Hillier, L. W., et al. (2009), 'Massively parallel sequencing of the polyadenylated transcriptome of *C. elegans*', *Genome Res*, 19 (4), 657-66.
- Hiwatashi, K., et al. (2008), 'A novel function of the receptor for advanced glycation end-products (RAGE) in association with tumorigenesis and tumor differentiation of HCC', *Ann Surg Oncol*, 15 (3), 923-33.
- Hofmann, M. A., et al. (1999), 'RAGE mediates a novel proinflammatory axis: a central cell surface receptor for S100/calgranulin polypeptides', *Cell*, 97 (7), 889-901.
- Holt, C. E. and Bullock, S. L. (2009), 'Subcellular mRNA localization in animal cells and why it matters', *Science*, 326 (5957), 1212-6.
- Hori, O., et al. (1995), 'The receptor for advanced glycation end products (RAGE) is a cellular binding site for amphoterin. Mediation of neurite outgrowth and co-expression of rage and amphoterin in the developing nervous system', *J Biol Chem*, 270 (43), 25752-61.
- Horton, R., et al. (2004), 'Gene map of the extended human MHC', *Nat Rev Genet*, 5 (12), 889-99.
- Hoskins, A. A. and Moore, M. J. 'The spliceosome: a flexible, reversible macromolecular machine', *Trends Biochem Sci*, 37 (5), 179-88.
- Hsieh, S. L. and Campbell, R. D. (1991), 'Evidence that gene G7a in the human major histocompatibility complex encodes valyl-tRNA synthetase', *Biochem J*, 278 ( Pt 3), 809-16.
- Hu, H. J., et al. (2011), 'Common variants at the promoter region of the APOM confer a risk of rheumatoid arthritis', *Exp Mol Med*, 43 (11), 613-21.
- Huang, C. H., et al. (2009), 'Rheumatoid arthritis-associated gene-gene interaction network for rheumatoid arthritis candidate genes', *BMC Proc*, 3 Suppl 7, S75.
- Hudson, B. I., et al. (2008a), 'Interaction of the RAGE cytoplasmic domain with diaphanous-1 is required for ligand-stimulated cellular migration through activation of Rac1 and Cdc42', *J Biol Chem*, 283 (49), 34457-68.
- Hudson, B. I., et al. (2008b), 'Identification, classification, and expression of RAGE gene splice variants', *FASEB J*, 22 (5), 1572-80.
- Hughes, C. E., et al. (2012), 'G6f-like is an ITAM-containing collagen receptor in thrombocytes', *PLoS One*, 7 (12), e52622.
- Hughes, T. A. (2006), 'Regulation of gene expression by alternative untranslated regions', *Trends Genet*, 22 (3), 119-22.
- Hull, J., et al. (2007), 'Identification of common genetic variation that modulates alternative splicing', *PLoS Genet*, 3 (6), e99.
- Hummel, T., et al. (2003), 'Axonal targeting of olfactory receptor neurons in *Drosophila* is controlled by Dscam', *Neuron*, 37 (2), 221-31.

## I

- Iborra, F. J., Jackson, D. A., and Cook, P. R. (2001), 'Coupled transcription and translation within nuclei of mammalian cells', *Science*, 293 (5532), 1139-42.
- Iborra, F. J., et al. (2004), 'Molecular cross-talk between the transcription, translation, and nonsense-mediated decay machineries', *J Cell Sci*, 117 (Pt 6), 899-906.
- Irimia, M. and Blencowe, B. J. (2012), 'Alternative splicing: decoding an expansive regulatory layer', *Curr Opin Cell Biol*, 24 (3), 323-32.
- Ishigaki, Y., et al. (2001), 'Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20', *Cell*, 106 (5), 607-17.
- Iwata, H., et al. (2009), 'Distinct C-terminus of the B subunit of factor XIII in a population-associated major phenotype: the first case of complete allele-specific alternative splicing products in the coagulation and fibrinolytic systems', *J Thromb Haemost*, 7 (7), 1084-91.
- Izquierdo, J. M. and Valcarcel, J. (2006), 'A simple principle to explain the evolution of pre-mRNA splicing', *Genes Dev*, 20 (13), 1679-84.

## J

- Jackson, I. J. (1991), 'A reappraisal of non-consensus mRNA splice sites', *Nucleic Acids Res*, 19 (14), 3795-8.
- Jacobson, K. and Dietrich, C. (1999), 'Looking at lipid rafts?', *Trends Cell Biol*, 9 (3), 87-91.
- Jawaheer, D. and Gregersen, P. K. (2002), 'The search for rheumatoid arthritis susceptibility genes: a call for global collaboration', *Arthritis Rheum*, 46 (3), 582-4.
- Jenkins, S. C., et al. (2000), 'A novel variant of the MHC-linked hsp70, hsp70-hom, is associated with rheumatoid arthritis', *Tissue Antigens*, 56 (1), 38-44.

- Johnson, B. D., et al. (2002), 'Photochemical treatment of donor lymphocytes inhibited their ability to facilitate donor engraftment or increase donor chimerism after nonmyeloablative conditioning or establishment of mixed chimerism', *Biol Blood Marrow Transplant*, 8 (11), 581-7.
- Johnson, J. M., et al. (2003), 'Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays', *Science*, 302 (5653), 2141-4.
- Joun, H., et al. (1997), 'Tissue-specific transcription start sites and alternative splicing of the parathyroid hormone (PTH)/PTH-related peptide (PTHrP) receptor gene: a new PTH/PTHrP receptor splice variant that lacks the signal peptide', *Endocrinology*, 138 (4), 1742-9.

## K

- Kadener, S., et al. (2001), 'Antagonistic effects of T-Ag and VP16 reveal a role for RNA pol II elongation on alternative splicing', *EMBO J*, 20 (20), 5759-68.
- Kalea, A. Z., Schmidt, A. M., and Hudson, B. I. (2011), 'Alternative splicing of RAGE: roles in biology and disease', *Front Biosci (Landmark Ed)*, 16, 2756-70.
- Kalea, A. Z., et al. (2009), 'Alternative splicing of the murine receptor for advanced glycation end-products (RAGE) gene', *FASEB J*, 23 (6), 1766-74.
- Kalsotra, A. and Cooper, T. A. (2011), 'Functional consequences of developmentally regulated alternative splicing', *Nat Rev Genet*, 12 (10), 715-29.
- Kallberg, H., et al. (2007), 'Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis', *Am J Hum Genet*, 80 (5), 867-75.
- Kanadia, R. N., et al. (2003), 'A muscleblind knockout model for myotonic dystrophy', *Science*, 302 (5652), 1978-80.
- Kapranov, P., et al. (2002), 'Large-scale transcriptional activity in chromosomes 21 and 22', *Science*, 296 (5569), 916-9.
- Kapranov, P., et al. (2005), 'Examples of the complex architecture of the human transcriptome revealed by RAGE and high-density tiling arrays', *Genome Res*, 15 (7), 987-97.
- Kataoka, N., et al. (2000), 'Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm', *Mol Cell*, 6 (3), 673-82.
- Kato, M., et al. (2003), 'Hodgkin's lymphoma cell lines express a fusion protein encoded by intergenically spliced mRNA for the multilectin receptor DEC-205 (CD205) and a novel C-type lectin receptor DCL-1', *J Biol Chem*, 278 (36), 34035-41.
- Kawahara, T., et al. (1998), 'Unconventional splicing of HAC1/ERN4 mRNA required for the unfolded protein response. Sequence-specific and non-sequential cleavage of the splice sites', *J Biol Chem*, 273 (3), 1802-7.
- Kelemen, O., et al. (2013), 'Function of alternative splicing', *Gene*, 514 (1), 1-30.
- Keren, H., Lev-Maor, G., and Ast, G. (2010), 'Alternative splicing and evolution: diversification, exon definition and function', *Nat Rev Genet*, 11 (5), 345-55.
- Kikuchi, K., et al. (2009), 'Increased inhibitory ability of conjugated RNA aptamers against the HCV IRES', *Biochem Biophys Res Commun*, 386 (1), 118-23.
- Kilding, R. and Wilson, A. G. (2005), 'Mapping of a novel susceptibility gene for rheumatoid arthritis in the telomeric MHC region', *Cytokine*, 32 (2), 71-5.
- Kilding, R., et al. (2004), 'Additional genetic susceptibility for rheumatoid arthritis telomeric of the DRB1 locus', *Arthritis Rheum*, 50 (3), 763-9.
- Kim, E., Magen, A., and Ast, G. (2007), 'Different levels of alternative splicing among eukaryotes', *Nucleic Acids Res*, 35 (1), 125-31.
- Kim, V. N., Kataoka, N., and Dreyfuss, G. (2001), 'Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex', *Science*, 293 (5536), 1832-6.
- Kloting, N., et al. (2007), 'Serum retinol-binding protein is more highly expressed in visceral than in subcutaneous adipose tissue and is a marker of intra-abdominal fat mass', *Cell Metab*, 6 (1), 79-87.
- Koch, M., et al. (2010), 'Structural basis for ligand recognition and activation of RAGE', *Structure*, 18 (10), 1342-52.
- Kole, R. and Weissman, S. M. (1982), 'Accurate in vitro splicing of human beta-globin RNA', *Nucleic Acids Res*, 10 (18), 5429-45.
- Kornblihtt, A. R., et al. (2009), 'When chromatin meets splicing', *Nat Struct Mol Biol*, 16 (9), 902-3.
- Kornblihtt, A. R., et al. (2013), 'Alternative splicing: a pivotal step between eukaryotic transcription and translation', *Nat Rev Mol Cell Biol*, 14 (3), 153-65.
- Kowalski, P. E., Freeman, J. D., and Mager, D. L. (1999), 'Intergenic splicing between a HERV-H endogenous retrovirus and two adjacent human genes', *Genomics*, 57 (3), 371-9.
- Kreivi, J. P. and Lamond, A. I. (1996), 'RNA splicing: unexpected spliceosome diversity', *Curr Biol*, 6 (7), 802-5.
- Kumar-Sinha, C., Kalyana-Sundaram, S., and Chinnaiyan, A. M. (2012), 'SLC45A3-ELK4 chimera in prostate cancer: spotlight on cis-splicing', *Cancer Discov*, 2 (7), 582-5.
- Kumar, N., et al. (2012), 'Major histocompatibility complex class I chain related gene-A microsatellite polymorphism shows secondary association with type 1 diabetes and celiac disease in North Indians', *Tissue Antigens*, 80 (4), 356-62.



**L**

- Langmead, B., et al. (2009), 'Ultrafast and memory-efficient alignment of short DNA sequences to the human genome', *Genome Biol*, 10 (3), R25.
- Le Hir, H., Moore, M. J., and Maquat, L. E. (2000a), 'Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions', *Genes Dev*, 14 (9), 1098-108.
- Le Hir, H., et al. (2000b), 'The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions', *EMBO J*, 19 (24), 6860-9.
- Le Hir, H., et al. (2001), 'The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay', *EMBO J*, 20 (17), 4987-97.
- Leff, S. E., Rosenfeld, M. G., and Evans, R. M. (1986), 'Complex transcriptional units: diversity in gene expression by alternative RNA processing', *Annu Rev Biochem*, 55, 1091-117.
- Lehmann, K. and Schmidt, U. (2003), 'Group II introns: structure and catalytic versatility of large natural ribozymes', *Crit Rev Biochem Mol Biol*, 38 (3), 249-303.
- Lei, H., Day, I. N., and Vorechovsky, I. (2005), 'Exonization of AluYa5 in the human ACE gene requires mutations in both 3' and 5' splice sites and is facilitated by a conserved splicing enhancer', *Nucleic Acids Res*, 33 (12), 3897-906.
- Lejeune, F., et al. (2002), 'The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling', *EMBO J*, 21 (13), 3536-45.
- Lev-Maor, G., et al. (2003), 'The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons', *Science*, 300 (5623), 1288-91.
- Lewandowski, U., et al. (2006), 'Elucidation of N-glycosylation sites on human platelet proteins: a glycoproteomic approach', *Mol Cell Proteomics*, 5 (2), 226-33.
- Li, B. L., et al. (1999), 'Human acyl-CoA:cholesterol acyltransferase-1 (ACAT-1) gene organization and evidence that the 4.3-kilobase ACAT-1 mRNA is produced from two different chromosomes', *J Biol Chem*, 274 (16), 11060-71.
- Li, H., et al. (2009), 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25 (16), 2078-9.
- Li, P., et al. (2010), 'The developmental dynamics of the maize leaf transcriptome', *Nat Genet*, 42 (12), 1060-7.
- Licatalosi, D. D. and Darnell, R. B. (2006), 'Splicing regulation in neurologic disease', *Neuron*, 52 (1), 93-101.
- Licatalosi, D. D. and Darnell, R. B. (2010), 'RNA processing and its regulation: global insights into biological networks', *Nat Rev Genet*, 11 (1), 75-87.
- Livak, K. J. and Schmittgen, T. D. (2001), 'Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method', *Methods*, 25 (4), 402-8.
- Lizotte, P. P., et al. (2007), 'Developmental expression of the receptor for advanced glycation end-products (RAGE) and its response to hyperoxia in the neonatal rat lung', *BMC Dev Biol*, 7, 15.
- Lopez-Bigas, N., et al. (2005), 'Are splicing mutations the most frequent cause of hereditary disease?', *FEBS Lett*, 579 (9), 1900-3.
- Lopez-Diez, R., et al. (2013), 'Complex tissue-specific patterns and distribution of multiple RAGE splice variants in different mammals', *Genome Biol Evol*, 5 (12), 2420-35.
- Luo, M. L., et al. (2001), 'Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly', *Nature*, 413 (6856), 644-7.
- Lykke-Andersen, J., Shu, M. D., and Steitz, J. A. (2000), 'Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon', *Cell*, 103 (7), 1121-31.
- Lykke-Andersen, J., Shu, M. D., and Steitz, J. A. (2001), 'Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1', *Science*, 293 (5536), 1836-9.

**M**

- Ma, W., et al. 'RAGE binds C1q and enhances C1q-mediated phagocytosis', *Cell Immunol*, 274 (1-2), 72-82.
- Macaulay, I. C., et al. (2007), 'Comparative gene expression profiling of in vitro differentiated megakaryocytes and erythroblasts identifies novel activatory and inhibitory platelet membrane proteins', *Blood*, 109 (8), 3260-9.
- Magen, A. and Ast, G. (2005), 'The importance of being divisible by three in alternative splicing', *Nucleic Acids Res*, 33 (17), 5574-82.
- Magrangeas, F., et al. (1998), 'Cotranscription and intergenic splicing of human galactose-1-phosphate uridylyltransferase and interleukin-11 receptor alpha-chain genes generate a fusion mRNA in normal cells. Implication for the production of multidomain proteins during evolution', *J Biol Chem*, 273 (26), 16005-10.
- Maher, C. A., et al. (2009), 'Chimeric transcript discovery by paired-end transcriptome sequencing', *Proc Natl Acad Sci U S A*, 106 (30), 12353-8.
- Maillard-Lefebvre, H., et al. (2009), 'Soluble receptor for advanced glycation end products: a new biomarker in diagnosis and prognosis of chronic inflammatory diseases', *Rheumatology (Oxford)*, 48 (10), 1190-6.

- Malherbe, P., et al. (1999), 'cDNA cloning of a novel secreted isoform of the human receptor for advanced glycation end products and characterization of cells co-expressing cell-surface scavenger receptors and Swedish mutant amyloid precursor protein', *Brain Res Mol Brain Res*, 71 (2), 159-70.
- Mallya, M., Campbell, R. D., and Aguado, B. (2002), 'Transcriptional analysis of a novel cluster of LY-6 family members in the human and mouse major histocompatibility complex: five genes with many splice forms', *Genomics*, 80 (1), 113-23.
- Mallya, M., Campbell, R. D., and Aguado, B. (2006), 'Characterization of the five novel Ly-6 superfamily members encoded in the MHC, and detection of cells expressing their potential ligands', *Protein Sci*, 15 (10), 2244-56.
- Marcotte, E. M., et al. (1999), 'Detecting protein function and protein-protein interactions from genome sequences', *Science*, 285 (5428), 751-3.
- Marques, A. C. and Ponting, C. P. (2009), 'Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness', *Genome Biol*, 10 (11), R124.
- Martin, K. C. and Ephrussi, A. (2009), 'mRNA localization: gene expression in the spatial dimension', *Cell*, 136 (4), 719-30.
- Martinez-Contreras, R., et al. (2006), 'Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing', *PLoS Biol*, 4 (2), e21.
- Matlin, A. J., Clark, F., and Smith, C. W. (2005), 'Understanding alternative splicing: towards a cellular code', *Nat Rev Mol Cell Biol*, 6 (5), 386-98.
- Mayeda, A., et al. (1999), 'Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements', *Mol Cell Biol*, 19 (3), 1853-63.
- Mayer, M. G. and Floeter-Winter, L. M. (2005), 'Pre-mRNA trans-splicing: from kinetoplasts to mammals, an easy language for life diversity', *Mem Inst Oswaldo Cruz*, 100 (5), 501-13.
- McGarvey, T., et al. (2000), 'The acute myeloid leukemia-associated protein, DEK, forms a splicing-dependent interaction with exon-product complexes', *J Cell Biol*, 150 (2), 309-20.
- McManus, C. J. and Graveley, B. R. (2011), 'RNA structure and the mechanisms of alternative splicing', *Curr Opin Genet Dev*, 21 (4), 373-9.
- Melamud, E. and Moul, J. (2009), 'Stochastic noise in splicing machinery', *Nucleic Acids Res*, 37 (14), 4873-86.
- Mercer, T. R., et al. (2011), 'Targeted RNA sequencing reveals the deep complexity of the human transcriptome', *Nat Biotechnol*, 30 (1), 99-104.
- Merkin, J., et al. (2012), 'Evolutionary dynamics of gene and isoform regulation in Mammalian tissues', *Science*, 338 (6114), 1593-9.
- Meugnier, E., et al. (2011), 'Gene expression profiling in peripheral blood cells of patients with rheumatoid arthritis in response to anti-TNF-alpha treatments', *Physiol Genomics*, 43 (7), 365-71.
- Michaeli, S. (2011), 'Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome', *Future Microbiol*, 6 (4), 459-74.
- Milicic, A., et al. (2000), 'Interethnic studies of TNF polymorphisms confirm the likely presence of a second MHC susceptibility locus in ankylosing spondylitis', *Genes Immun*, 1 (7), 418-22.
- Milner, C. M., et al. (1997), 'Identification of a sialidase encoded in the human major histocompatibility complex', *J Biol Chem*, 272 (7), 4549-58.
- Millar, J. K., et al. (2000), 'Chromosomal location and genomic structure of the human translin-associated factor X gene (TRAX; TSNAX) revealed by intergenic splicing to DISC1, a gene disrupted by a translocation segregating with schizophrenia', *Genomics*, 67 (1), 69-77.
- Miller, J. W., et al. (2000), 'Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy', *EMBO J*, 19 (17), 4439-48.
- Millot, P. and Dausset, J. (1978), '[The major OLA histocompatibility complex of sheep. Recent results in the genetic study of 11 factors previously defined by microlymphocytotoxicity]', *C R Acad Sci Hebd Seances Acad Sci D*, 286 (10), 801-5.
- Mittendorf, K. F., et al. (2012), 'Tailoring of membrane proteins by alternative splicing of pre-mRNA', *Biochemistry*, 51 (28), 5541-56.
- Mola, G., et al. (2007), 'Exonization of Alu-generated splice variants in the survivin gene of human and non-human primates', *J Mol Biol*, 366 (4), 1055-63.
- Moller, L. B., et al. (1993), 'N-linked glycosylation of the ligand-binding domain of the human urokinase receptor contributes to the affinity for its ligand', *J Biol Chem*, 268 (15), 11152-9.
- Montandon, P. E. and Stutz, E. (1984), 'The genes for the ribosomal proteins S12 and S7 are clustered with the gene for the EF-Tu protein on the chloroplast genome of *Euglena gracilis*', *Nucleic Acids Res*, 12 (6), 2851-9.
- Mudge, J. M., et al. (2011), 'The origins, evolution, and functional potential of alternative splicing in vertebrates', *Mol Biol Evol*, 28 (10), 2949-59.
- Mulcahy, B., et al. (1996), 'Genetic variability in the tumor necrosis factor-lymphotoxin region influences susceptibility to rheumatoid arthritis', *Am J Hum Genet*, 59 (3), 676-83.
- Muniz, M. and Riezman, H. (2000), 'Intracellular transport of GPI-anchored proteins', *EMBO J*, 19 (1), 10-5.

Munoz, M. J., de la Mata, M., and Kornblihtt, A. R. (2010), 'The carboxy terminal domain of RNA polymerase II and alternative splicing', *Trends Biochem Sci*, 35 (9), 497-504.

## N

Nacu, S., et al. (2011), 'Deep RNA sequencing analysis of readthrough gene fusions in human prostate adenocarcinoma and reference samples', *BMC Med Genomics*, 4, 11.

Neeper, M., et al. (1992), 'Cloning and expression of a cell surface receptor for advanced glycosylation end products of proteins', *J Biol Chem*, 267 (21), 14998-5004.

Newton, J. L., et al. (2004), 'Dissection of class III major histocompatibility complex haplotypes associated with rheumatoid arthritis', *Arthritis Rheum*, 50 (7), 2122-9.

Nilsen, T. W. and Graveley, B. R. (2010), 'Expansion of the eukaryotic proteome by alternative splicing', *Nature*, 463 (7280), 457-63.

Nurtdinov, R. N., et al. (2007), 'Conserved and species-specific alternative splicing in mammalian genomes', *BMC Evol Biol*, 7, 249.

## O

Ohashi, K., et al. (2010), 'A molecule in teleost fish, related with human MHC-encoded G6F, has a cytoplasmic tail with ITAM and marks the surface of thrombocytes and in some fishes also of erythrocytes', *Immunogenetics*, 62 (8), 543-59.

Okamoto, K., et al. (2003), 'Identification of I kappa BL as the second major histocompatibility complex-linked susceptibility locus for rheumatoid arthritis', *Am J Hum Genet*, 72 (2), 303-12.

## P

Pan, Q., et al. (2008), 'Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing', *Nat Genet*, 40 (12), 1413-5.

Pan, Q., et al. (2005), 'Alternative splicing of conserved exons is frequently species-specific in human and mouse', *Trends Genet*, 21 (2), 73-7.

Pardigol, A., et al. (1998), 'HCC-2, a human chemokine: gene structure, expression pattern, and biological activity', *Proc Natl Acad Sci U S A*, 95 (11), 6308-13.

Park, H., Adsit, F. G., and Boyington, J. C. (2010), 'The 1.5 Å crystal structure of human receptor for advanced glycation endproducts (RAGE) ectodomains reveals unique features determining ligand binding', *J Biol Chem*, 285 (52), 40762-70.

Parra, G., et al. (2006), 'Tandem chimerism as a means to increase protein complexity in the human genome', *Genome Res*, 16 (1), 37-44.

Pfaffl, M. W. (2001), 'A new mathematical model for relative quantification in real-time RT-PCR', *Nucleic Acids Res*, 29 (9), e45.

Pickrell, J. K., et al. (2010), 'Noisy splicing drives mRNA isoform diversity in human cells', *PLoS Genet*, 6 (12), e1001236.

Ponting, C. P., Oliver, P. L., and Reik, W. (2009), 'Evolution and functions of long noncoding RNAs', *Cell*, 136 (4), 629-41.

Poulin, F., Brueschke, A., and Sonenberg, N. (2003), 'Gene fusion and overlapping reading frames in the mammalian genes for 4E-BP3 and MASK', *J Biol Chem*, 278 (52), 52290-7.

Pozzoli, U. and Sironi, M. (2005), 'Silencers regulate both constitutive and alternative splicing events in mammals', *Cell Mol Life Sci*, 62 (14), 1579-604.

Pradet-Balade, B., et al. (2002), 'An endogenous hybrid mRNA encodes TWE-PRIL, a functional cell surface TWEAK-APRIL fusion protein', *EMBO J*, 21 (21), 5711-20.

## R

Rai, V., et al. 'Lysophosphatidic acid targets vascular and oncogenic pathways via RAGE signaling', *J Exp Med*, 209 (13), 2339-50.

Ramani, A. K., et al. (2011), 'Genome-wide analysis of alternative splicing in *Caenorhabditis elegans*', *Genome Res*, 21 (2), 342-8.

Ranum, L. P. and Cooper, T. A. (2006), 'RNA-mediated neuromuscular disorders', *Annu Rev Neurosci*, 29, 259-77.

Rastrojo, A., et al. (2013), 'The transcriptome of *Leishmania major* in the axenic promastigote stage: transcript annotation and relative expression levels by RNA-seq', *BMC Genomics*, 14, 223.

Rauci, A., et al. (2008), 'A soluble form of the receptor for advanced glycation endproducts (RAGE) is produced by proteolytic cleavage of the membrane-bound form by the sheddase a disintegrin and metalloprotease 10 (ADAM10)', *FASEB J*, 22 (10), 3716-27.

Reddy, M. A., et al. (2006), 'Key role of Src kinase in S100B-induced activation of the receptor for advanced glycation end products in vascular smooth muscle cells', *J Biol Chem*, 281 (19), 13685-93.

Reichert, V. L., et al. (2002), '5' exon interactions within the human spliceosome establish a framework for exon junction complex structure and assembly', *Genes Dev*, 16 (21), 2778-91.

- Rickman, D. S., et al. (2009), 'SLC45A3-ELK4 is a novel and frequent erythroblast transformation-specific fusion transcript in prostate cancer', *Cancer Res*, 69 (7), 2734-8.
- Rinn, J. L., et al. (2007), 'Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs', *Cell*, 129 (7), 1311-23.
- Robberson, B. L., Cote, G. J., and Berget, S. M. (1990), 'Exon definition may facilitate splice site selection in RNAs with multiple exons', *Mol Cell Biol*, 10 (1), 84-94.
- Robinson, J. T., et al. (2011), 'Integrative genomics viewer', *Nat Biotechnol*, 29 (1), 24-6.
- Rudd, P. M. and Dwek, R. A. (1997), 'Glycosylation: heterogeneity and the 3D structure of proteins', *Crit Rev Biochem Mol Biol*, 32 (1), 1-100.
- Ruegsegger, U., Leber, J. H., and Walter, P. (2001), 'Block of HAC1 mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response', *Cell*, 107 (1), 103-14.
- Russell, R. (2008), 'RNA misfolding and the action of chaperones', *Front Biosci*, 13, 1-20.

## S

- Sarkany, Z., et al. (2011), 'Solution structure of the soluble receptor for advanced glycation end products (sRAGE)', *J Biol Chem*, 286 (43), 37525-34.
- Scherer, S. (2010), 'Guide to the Human Genome', *Cold Spring Harbor*, (1st edn).
- Schmidt, A. M., et al. (1992), 'Isolation and characterization of two binding proteins for advanced glycosylation end products from bovine lung which are present on the endothelial cell surface', *J Biol Chem*, 267 (21), 14987-97.
- Schmucker, D., et al. (2000), 'Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity', *Cell*, 101 (6), 671-84.
- Schroeder, H. W., Jr., et al. (1998), 'Susceptibility locus for IgA deficiency and common variable immunodeficiency in the HLA-DR3, -B8, -A1 haplotypes', *Mol Med*, 4 (2), 72-86.
- Senis, Y. A., et al. (2009), 'The tyrosine phosphatase CD148 is an essential positive regulator of platelet activation and thrombosis', *Blood*, 113 (20), 4942-54.
- Shao, X., Shepelev, V., and Fedorov, A. (2006), 'Bioinformatic analysis of exon repetition, exon scrambling and trans-splicing in humans', *Bioinformatics*, 22 (6), 692-8.
- Shepard, P. J. and Hertel, K. J. (2008), 'Conserved RNA secondary structures promote alternative splicing', *RNA*, 14 (8), 1463-9.
- Shichi, D., et al. (2005), 'The haplotype block, NFKBIL1-ATP6V1G2-BAT1-MICB-MICA, within the class III-class I boundary region of the human major histocompatibility complex may control susceptibility to hepatitis C virus-associated dilated cardiomyopathy', *Tissue Antigens*, 66 (3), 200-8.
- Shklar, M., et al. (2005), 'GeneTide--Terra Incognita Discovery Endeavor: a new transcriptome focused member of the GeneCards/GeneNote suite of databases', *Nucleic Acids Res*, 33 (Database issue), D556-61.
- Sidrauski, C. and Walter, P. (1997), 'The transmembrane kinase Ire1p is a site-specific endonuclease that initiates mRNA splicing in the unfolded protein response', *Cell*, 90 (6), 1031-9.
- Sidrauski, C., Cox, J. S., and Walter, P. (1996), 'tRNA ligase is required for regulated mRNA splicing in the unfolded protein response', *Cell*, 87 (3), 405-13.
- Sigurdsson, S., et al. (2007), 'Association of a haplotype in the promoter region of the interferon regulatory factor 5 gene with rheumatoid arthritis', *Arthritis Rheum*, 56 (7), 2202-10.
- Simons, K. and Ikonen, E. (1997), 'Functional rafts in cell membranes', *Nature*, 387 (6633), 569-72.
- Smolen, J. S., et al. (2005), 'Pro-inflammatory cytokines in rheumatoid arthritis: pathogenetic and therapeutic aspects', *Clin Rev Allergy Immunol*, 28 (3), 239-48.
- Snoek, M., Teuscher, C., and van Vugt, H. (1998), 'Molecular analysis of the major MHC recombinational hot spot located within the G7c gene of the murine class III region that is involved in disease susceptibility', *J Immunol*, 160 (1), 266-72.
- Sorg, D., et al. (2013), 'Microfluidic high-throughput RT-qPCR measurements of the immune response of primary bovine mammary epithelial cells cultured from milk to mastitis pathogens', *Animal*, 7 (5), 799-805.
- Stamm, S., et al. (2005), 'Function of alternative splicing', *Gene*, 344, 1-20.
- Stephens, R., et al. (1999), 'Gene organisation, sequence variation and isochore structure at the centromeric boundary of the human MHC', *J Mol Biol*, 291 (4), 789-99.
- Sterenczak, K. A., Nolte, I., and Murua Escobar, H. (2013), 'RAGE splicing variants in mammals', *Methods Mol Biol*, 963, 265-76.
- Sterenczak, K. A., et al. (2009), 'Cloning, characterisation, and comparative quantitative expression analyses of receptor for advanced glycation end products (RAGE) transcript forms', *Gene*, 434 (1-2), 35-42.
- Sureau, A. and Perbal, B. (1994), 'Several mRNAs with variable 3' untranslated regions and different stability encode the human PR264/SC35 splicing factor', *Proc Natl Acad Sci U S A*, 91 (3), 932-6.
- Swanberg, M., et al. (2005), 'MHC2TA is associated with differential MHC molecule expression and susceptibility to rheumatoid arthritis, multiple sclerosis and myocardial infarction', *Nat Genet*, 37 (5), 486-94.

**T**

- Takahara, T., et al. (2005), 'Delay in synthesis of the 3' splice site promotes trans-splicing of the preceding 5' splice site', *Mol Cell*, 18 (2), 245-51.
- Tarn, W. Y. and Steitz, J. A. (1996), 'A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro', *Cell*, 84 (5), 801-11.
- Tarn, W. Y. and Steitz, J. A. (1997), 'Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge', *Trends Biochem Sci*, 22 (4), 132-7.
- Tasic, B., et al. (2002), 'Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing', *Mol Cell*, 10 (1), 21-33.
- Teixeira, V. H., et al. (2009), 'Transcriptome analysis describing new immunity and defense genes in peripheral blood mononuclear cells of rheumatoid arthritis patients', *PLoS One*, 4 (8), e6803.
- The MHC Consortium, The MHC Consortium (1999), 'Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium', *Nature*, 401 (6756), 921-3.
- Thomson, T. M., et al. (2000), 'Fusion of the human gene for the polyubiquitination coeffector UEV1 with Kua, a newly identified gene', *Genome Res*, 10 (11), 1743-56.
- Toure, F., et al. (2012), 'Formin mDia1 mediates vascular remodeling via integration of oxidative and signal transduction pathways', *Circ Res*, 110 (10), 1279-93.
- Trapnell, C., Pachter, L., and Salzberg, S. L. (2009), 'TopHat: discovering splice junctions with RNA-Seq', *Bioinformatics*, 25 (9), 1105-11.
- Trapnell, C., et al. (2010), 'Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation', *Nat Biotechnol*, 28 (5), 511-5.
- Traub, L. M. and Kornfeld, S. (1997), 'The trans-Golgi network: a late secretory sorting station', *Curr Opin Cell Biol*, 9 (4), 527-33.
- Treisman, R., Orkin, S. H., and Maniatis, T. (1983), 'Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes', *Nature*, 302 (5909), 591-6.
- Tuck, A. C. and Tollervey, D. (2011), 'RNA in pieces', *Trends Genet*, 27 (10), 422-32.

**U**

- Uemura, A., et al. (2009), 'Unconventional splicing of XBP1 mRNA occurs in the cytoplasm during the mammalian unfolded protein response', *J Cell Sci*, 122 (Pt 16), 2877-86.
- Ule, J., et al. (2005), 'Nova regulates brain-specific splicing to shape the synapse', *Nat Genet*, 37 (8), 844-52.
- Upadhyaya, A. B., Lee, S. H., and DeJong, J. (1999), 'Identification of a general transcription factor TFIIAalpha/beta homolog selectively expressed in testis', *J Biol Chem*, 274 (25), 18040-8.

**V**

- Valentijn, L. J., Koster, J., and Versteeg, R. (2006), 'Read-through transcript from NM23-H1 into the neighboring NM23-H2 gene encodes a novel protein, NM23-LV', *Genomics*, 87 (4), 483-9.
- Van der Ploeg, L. H. (1986), 'Discontinuous transcription and splicing in trypanosomes', *Cell*, 47 (4), 479-80.
- Vandesompele, J., et al. (2002), 'Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes', *Genome Biol*, 3 (7), RESEARCH0034.
- Vandiedonck, C. and Knight, J. C. (2009), 'The human Major Histocompatibility Complex as a paradigm in genomics research', *Brief Funct Genomic Proteomic*, 8 (5), 379-94.
- Vargas, D. Y., et al. (2011), 'Single-molecule imaging of transcriptionally coupled and uncoupled splicing', *Cell*, 147 (5), 1054-65.
- Villate, O., et al. (2008), 'Differential splicing, disease and drug targets', *Infect Disord Drug Targets*, 8 (4), 241-51.

**W**

- Wahl, M. C., Will, C. L., and Luhrmann, R. (2009), 'The spliceosome: design principles of a dynamic RNP machine', *Cell*, 136 (4), 701-18.
- Wang, et al. (2002a), 'Drosophila Dscam is required for divergent segregation of sister branches and suppresses ectopic bifurcation of axons', *Neuron*, 33 (4), 559-71.
- Wang, E. T., et al. (2008), 'Alternative isoform regulation in human tissue transcriptomes', *Nature*, 456 (7221), 470-6.
- Wang, G. S. and Cooper, T. A. (2007), 'Splicing in disease: disruption of the splicing code and the decoding machinery', *Nat Rev Genet*, 8 (10), 749-61.
- Wang, Z., et al. (2002b), 'Activation of astrocytes by advanced glycation end products: cytokines induction and nitric oxide release', *Acta Pharmacol Sin*, 23 (11), 974-80.
- Warf, M. B. and Berglund, J. A. (2010), 'Role of RNA structure in regulating pre-mRNA splicing', *Trends Biochem Sci*, 35 (3), 169-78.
- Wautier, M. P., et al. (2001), 'Activation of NADPH oxidase by AGE links oxidant stress to altered gene expression via RAGE', *Am J Physiol Endocrinol Metab*, 280 (5), E685-94.
- Wilhelm, O. G., et al. (1999), 'Cellular glycosylphosphatidylinositol-specific phospholipase D regulates urokinase receptor shedding and cell surface expression', *J Cell Physiol*, 180 (2), 225-35.

- Wilkie, S. E., et al. (2008), 'Disease mechanism for retinitis pigmentosa (RP11) caused by missense mutations in the splicing factor gene PRPF31', *Mol Vis*, 14, 683-90.
- Williams, A. F. and Barclay, A. N. (1988), 'The immunoglobulin superfamily--domains for cell surface recognition', *Annu Rev Immunol*, 6, 381-405.
- Wojtowicz, W. M., et al. (2004), 'Alternative splicing of Drosophila Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding', *Cell*, 118 (5), 619-33.
- Wu and Maniatis, T. (1993), 'Specific interactions between proteins implicated in splice site selection and regulated alternative splicing', *Cell*, 75 (6), 1061-70.
- Wu, Q. and Krainer, A. R. (1999), 'AT-AC pre-mRNA splicing mechanisms and conservation of minor introns in voltage-gated ion channel genes', *Mol Cell Biol*, 19 (5), 3225-36.
- Xie, T., et al. (2003), 'Analysis of the gene-dense major histocompatibility complex class III region and its comparison to mouse', *Genome Res*, 13 (12), 2621-36.
- Xing, Y. and Lee, C. (2005), 'Evidence of functional selection pressure for alternative splicing events that accelerate evolution of protein subsequences', *Proc Natl Acad Sci U S A*, 102 (38), 13526-31.
- Xing, Y. and Lee, C. (2007), 'Relating alternative splicing to proteome complexity and genome evolution', *Adv Exp Med Biol*, 623, 36-49.
- Xing, Y., Xu, Q., and Lee, C. (2003), 'Widespread production of novel soluble protein isoforms by alternative splicing removal of transmembrane anchoring domains', *FEBS Lett*, 555 (3), 572-8.
- Xu, Y., et al. 'Advanced glycation end product (AGE)-receptor for AGE (RAGE) signaling and up-regulation of Egr-1 in hypoxic macrophages', *J Biol Chem*, 285 (30), 23233-40.

**Y**

- Yan, S. D., et al. (1994), 'Enhanced cellular oxidant stress by the interaction of advanced glycation end products with their receptors/binding proteins', *J Biol Chem*, 269 (13), 9889-97.
- Yan, S. D., et al. (1996), 'RAGE and amyloid-beta peptide neurotoxicity in Alzheimer's disease', *Nature*, 382 (6593), 685-91.
- Yang, Y., et al. (2011), 'RNA secondary structure in mutually exclusive splicing', *Nat Struct Mol Biol*, 18 (2), 159-68.
- Yeo, G. W., et al. (2005), 'Identification and analysis of alternative splicing events conserved in human and mouse', *Proc Natl Acad Sci U S A*, 102 (8), 2850-5.
- Yonekura, H., et al. (2003), 'Novel splice variants of the receptor for advanced glycation end-products expressed in human vascular endothelial cells and pericytes, and their putative roles in diabetes-induced vascular injury', *Biochem J*, 370 (Pt 3), 1097-109.
- Yoshida, H. (2007), 'Unconventional splicing of XBP-1 mRNA in the unfolded protein response', *Antioxid Redox Signal*, 9 (12), 2323-33.
- Yoshida, H., et al. (2001), 'XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor', *Cell*, 107 (7), 881-91.

**Z**

- Zanelli, E., et al. (2001), 'The telomeric part of the HLA region predisposes to rheumatoid arthritis independently of the class II loci', *Hum Immunol*, 62 (1), 75-84.
- Zaphiropoulos, P. G. (1999), 'RNA molecules containing exons originating from different members of the cytochrome P450 2C gene subfamily (CYP2C) in human epidermis and liver', *Nucleic Acids Res*, 27 (13), 2585-90.
- Zhang, C., Krainer, A. R., and Zhang, M. Q. (2007), 'Evolutionary impact of limited splicing fidelity in mammalian genes', *Trends Genet*, 23 (10), 484-8.
- Zhang, Y., et al. (2012), 'Chimeric transcript generated by cis-splicing of adjacent genes regulates prostate cancer cell proliferation', *Cancer Discov*, 2 (7), 598-607.
- Zhang, Z., et al. (2009), 'Noisy splicing, more than expression regulation, explains why some exons are subject to nonsense-mediated mRNA decay', *BMC Biol*, 7, 23.
- Zhou, Z., et al. (2000), 'The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans', *Nature*, 407 (6802), 401-5.
- Zong, H., et al. (2010), 'Homodimerization is essential for the receptor for advanced glycation end products (RAGE)-mediated signal transduction', *J Biol Chem*, 285 (30), 23137-46.

## **8. DATOS SUPLEMENTARIOS**

**Figura Suplementaria 1. Representación de todas las variantes de *splicing* de RAGE previamente descritas anteriormente y en este estudio en *Homo sapiens*, *Macaca mulatta*, *Bos taurus*, *Sus scrofa*, *Mus musculus* y *Rattus norvegicus*.** Cada figura muestra la representación de las variantes de *splicing* de una especie particular. Se muestran los datos obtenidos experimentalmente como aquellos presentes en las bases de datos (únicamente aquellos descritos experimentalmente en un estudio previamente publicado). Cada variante de *splicing* presenta su nomenclatura actual y nombres previos (si han cambiado), además de su identificador de Genbank. Los potenciales dominios proteicos se muestran con diferentes colores según la leyenda, mientras que las secuencias no codificantes se muestran en blanco. Las uniones de exones presentes en la variante canónica se muestran como líneas rectas, mientras que las variantes de *splicing* alternativas se muestran como triángulos abiertos. Las variantes están organizadas por tamaño.

**Figura Suplementaria 2. Estimación de la expresión de los diferentes transcritos (A) y del gen (B) RAGE en las muestras de RNA-seq en *Homo sapiens*, *Macaca mulatta*, *Bos taurus*, *Mus musculus* y *Rattus norvegicus*.** En cada especie se muestra la abundancia relativa de cada variante de *splicing* de RAGE (A) descrita por nuestro estudio mediante el uso de un archivo GTF y la opción -G del programa Cufflinks. Posteriormente se muestra la expresión total del gen RAGE (B) calculado como la suma de la expresión de todos los transcritos. Se muestran las lecturas de cerebro e hígado (PMID:18978789 – Pan et al (2008); PMID:23258890 – Barbosa-Morais et al (2012); PMID:23258891 – Merkin et al (2012)), así como de monocitos (datos procedentes del proyecto ENCODE SRX179434) en el apartado de *Homo sapiens*. Los niveles de expresión se muestran en una escala de gradiente mostrada a la derecha de la figura. Las muestras se nombraron como originalmente: en el caso de las muestras de Barbosa-Morais et al (2012) con la abreviatura de la especie (por ejemplo “hsa para *Homo sapiens*), la abreviatura del tejido (br – cerebro y lv – hígado), el género (M o F) y replicados (1, 2, etc). Por el contrario, las muestras de Merkin et al (2012) solo abreviaron los diferentes replicados de cada muestra “a, b, o c”.



**Figura Suplementaria 3. Cuantificación de los diferentes eventos de *splicing* de RAGE en muestras de RNA-seq.** Se cuantificaron el número de uniones de exones por millón de lecturas alineadas en las diferentes especies a estudio (PMID:18978789 – Pan et al 2008; PMID:23258890 – Barbosa-Morais et al 2012; PMID:23258891 – Merkin et al 2012). Las uniones de exones presentes en la variante canónica Tv1-RAGE se muestran como líneas rectas, mientras que las uniones de exones alternativas se muestran como triángulos abiertos. El asterisco identifica los eventos de *splicing* no obtenidos por RT-PCR anidada. Los niveles de expresión se muestran en una escala de gradiente de verde (poca expresión) a roja (alta expresión).

**Figura Suplementaria 4. Expresión relativa de los genes del MHC de clase III en pacientes con AR.** La expresión se distribuyó en 15 individuos controles (azul) y 31 pacientes de AR (morado), siendo 16 pacientes en remisión de la enfermedad (verde) y 15 con la enfermedad activa (rojo). La barra horizontal en cada nube de puntos representa la media. La significación estadística se detalla según: \*  $p < 0,01$ , \*\*  $p < 0,001$  y \*\*\*  $p < 0,0001$ .

Figura Suplementaria 1.

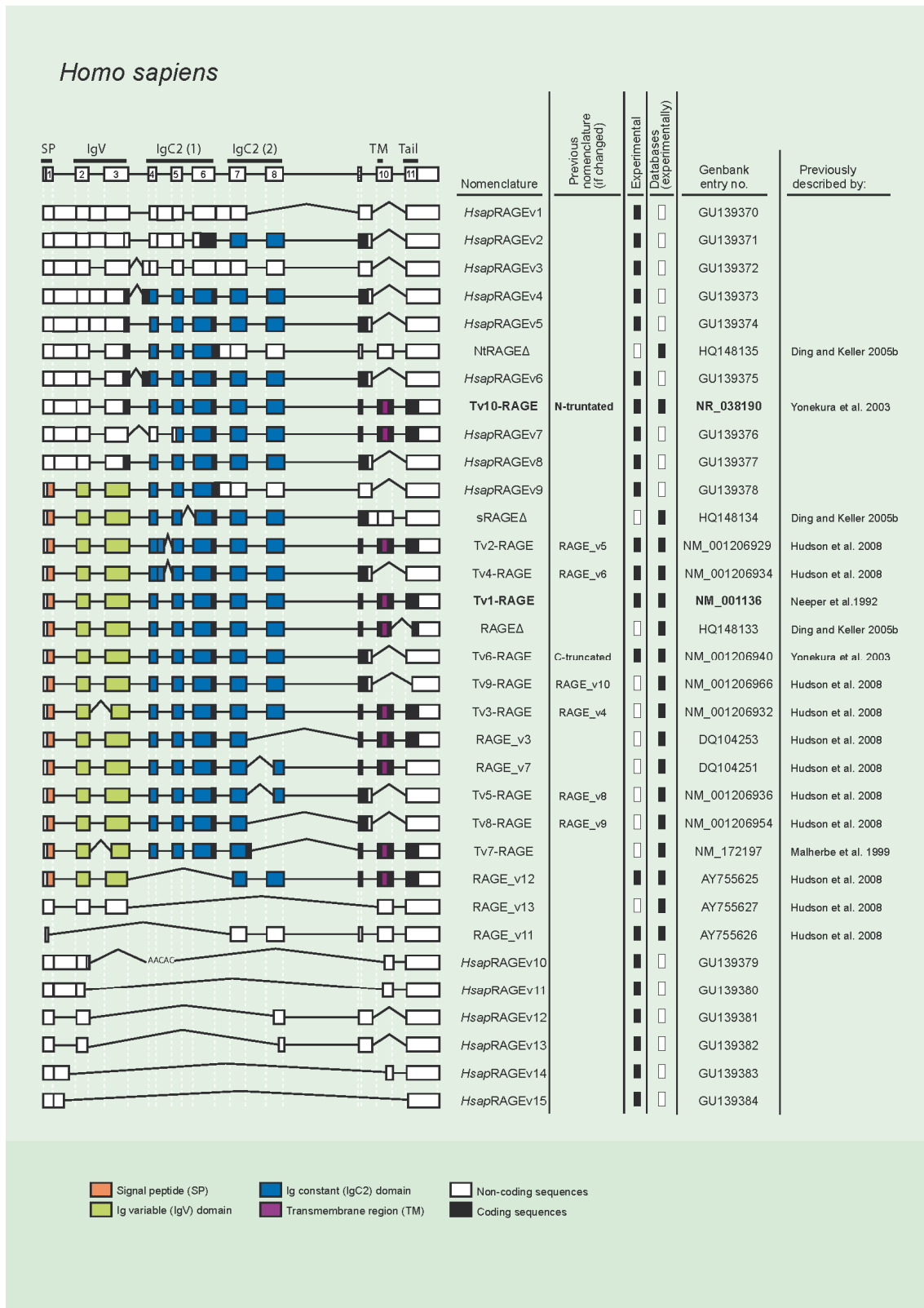


Figura Suplementaria 1 (continuación).

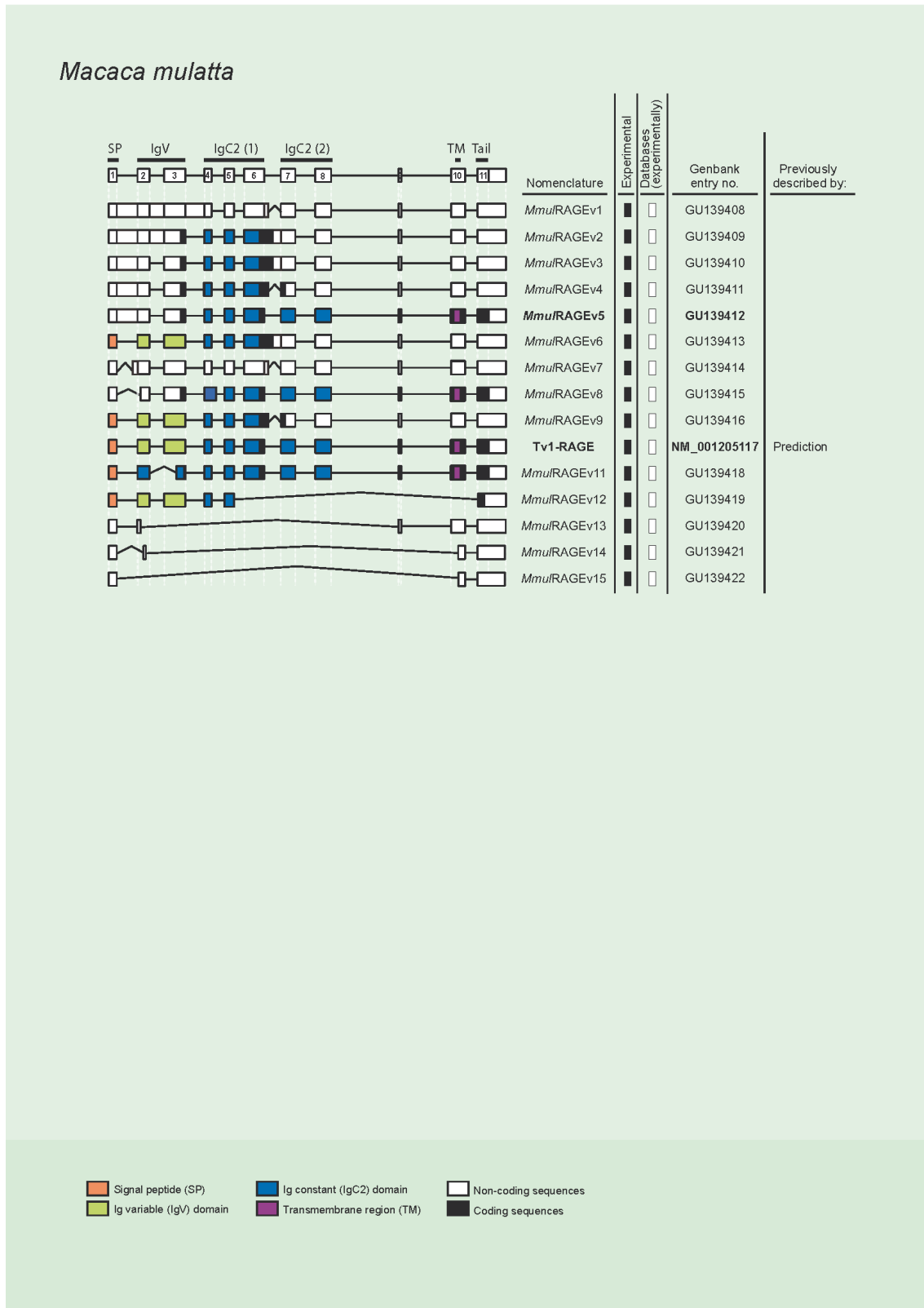


Figura Suplementaria 1 (continuación).

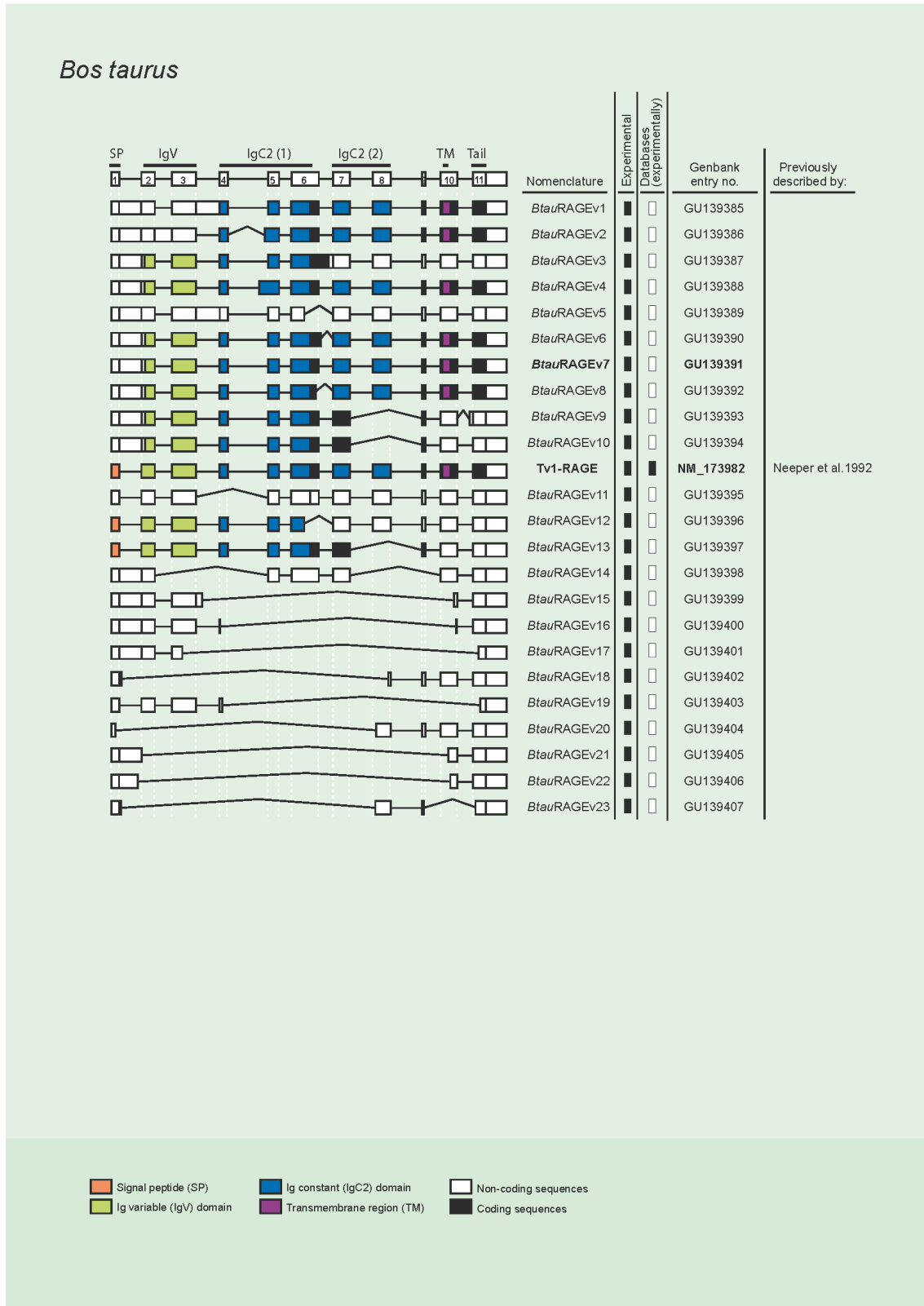


Figura Suplementaria 1 (continuación).

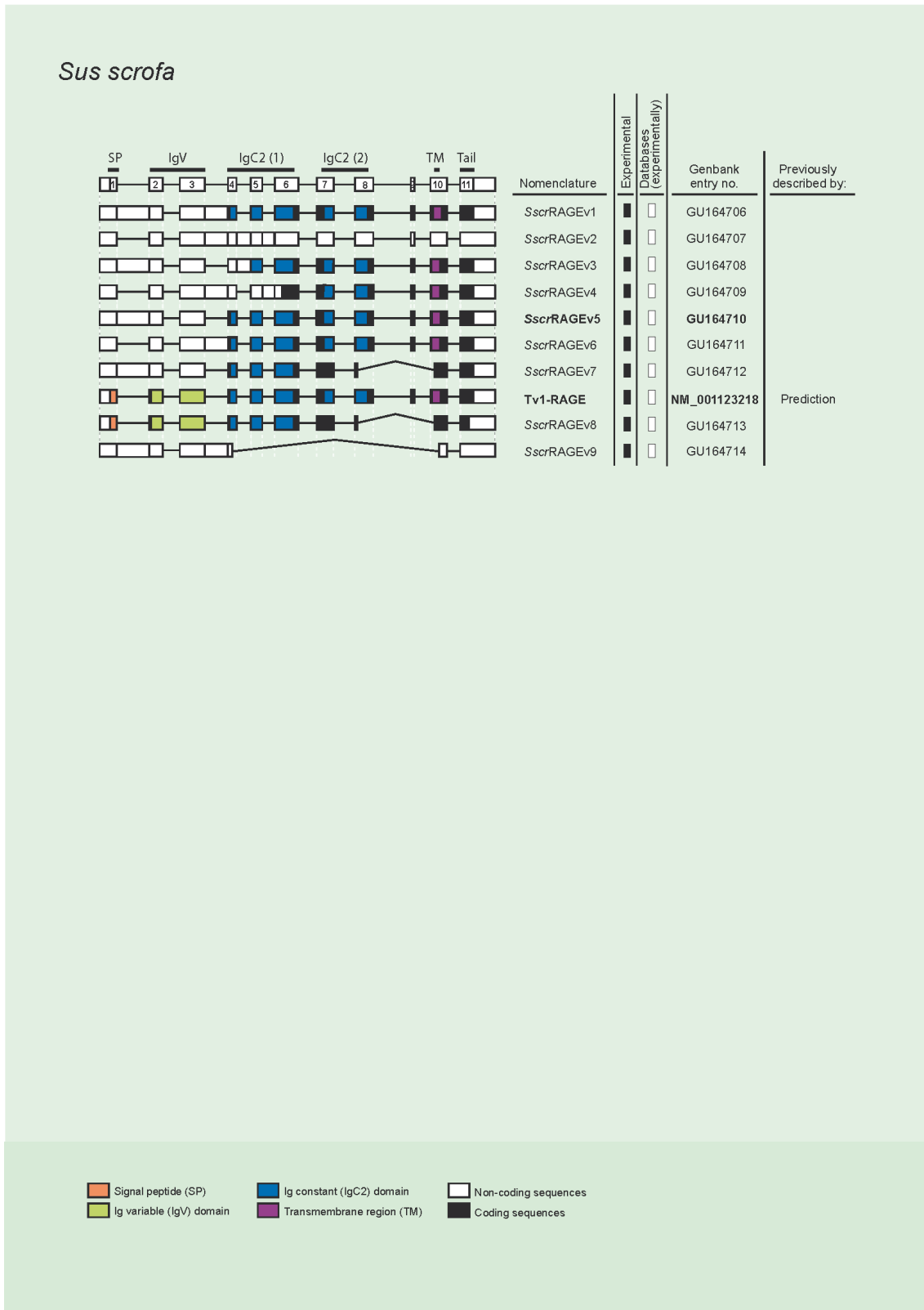


Figura Suplementaria 1 (continuación).

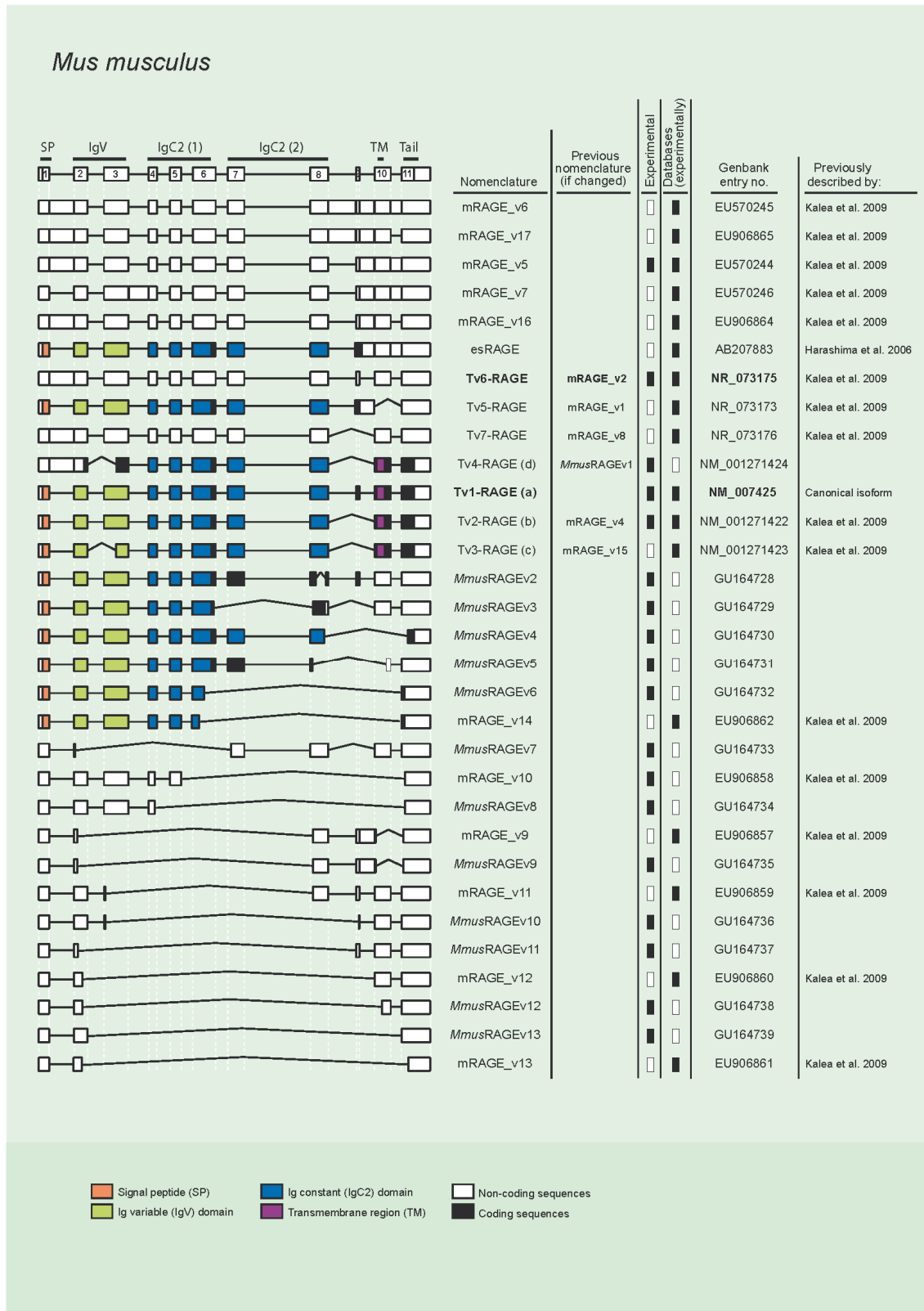


Figura Suplementaria 1 (continuación).

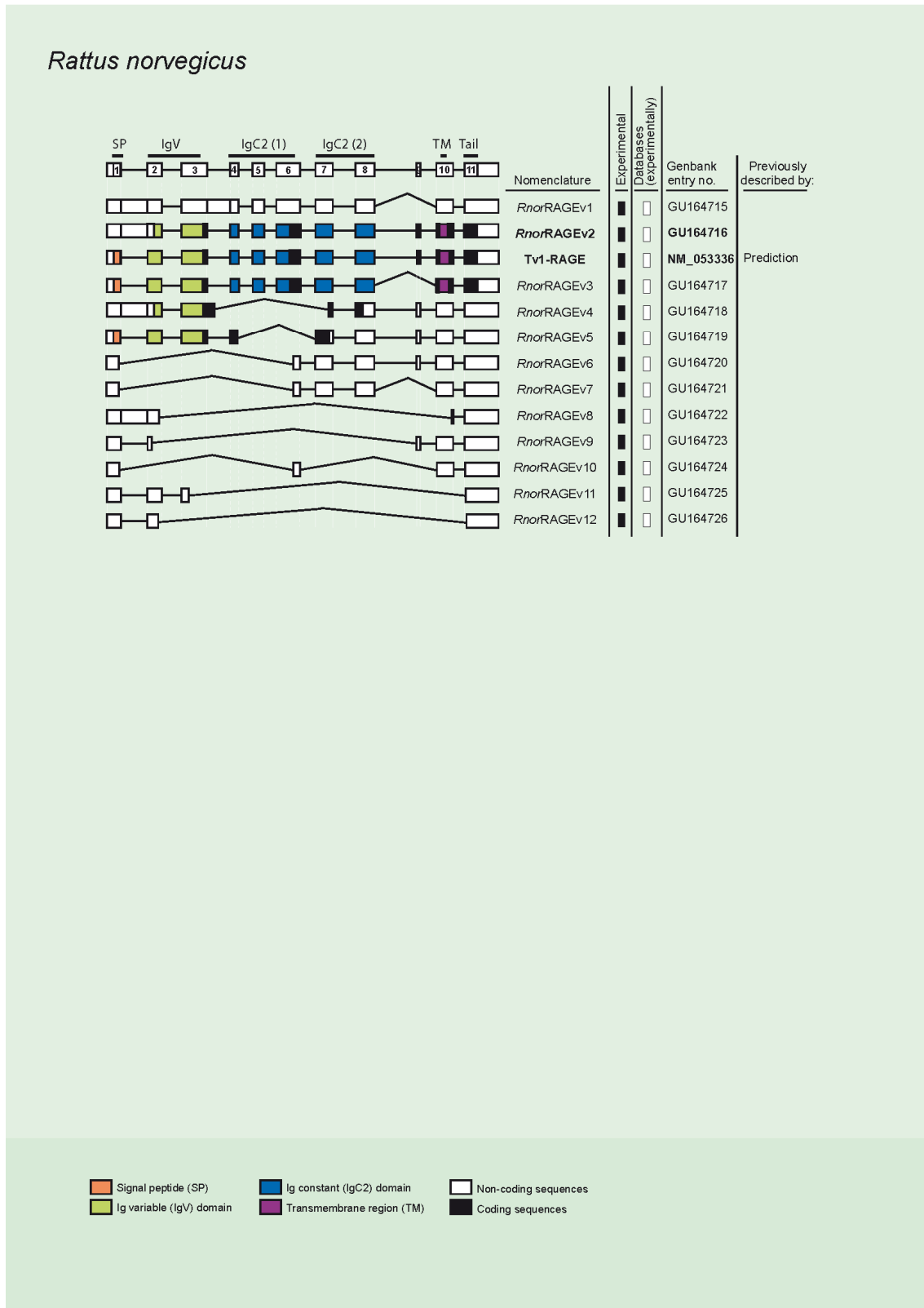


Figura Suplementaria 2.

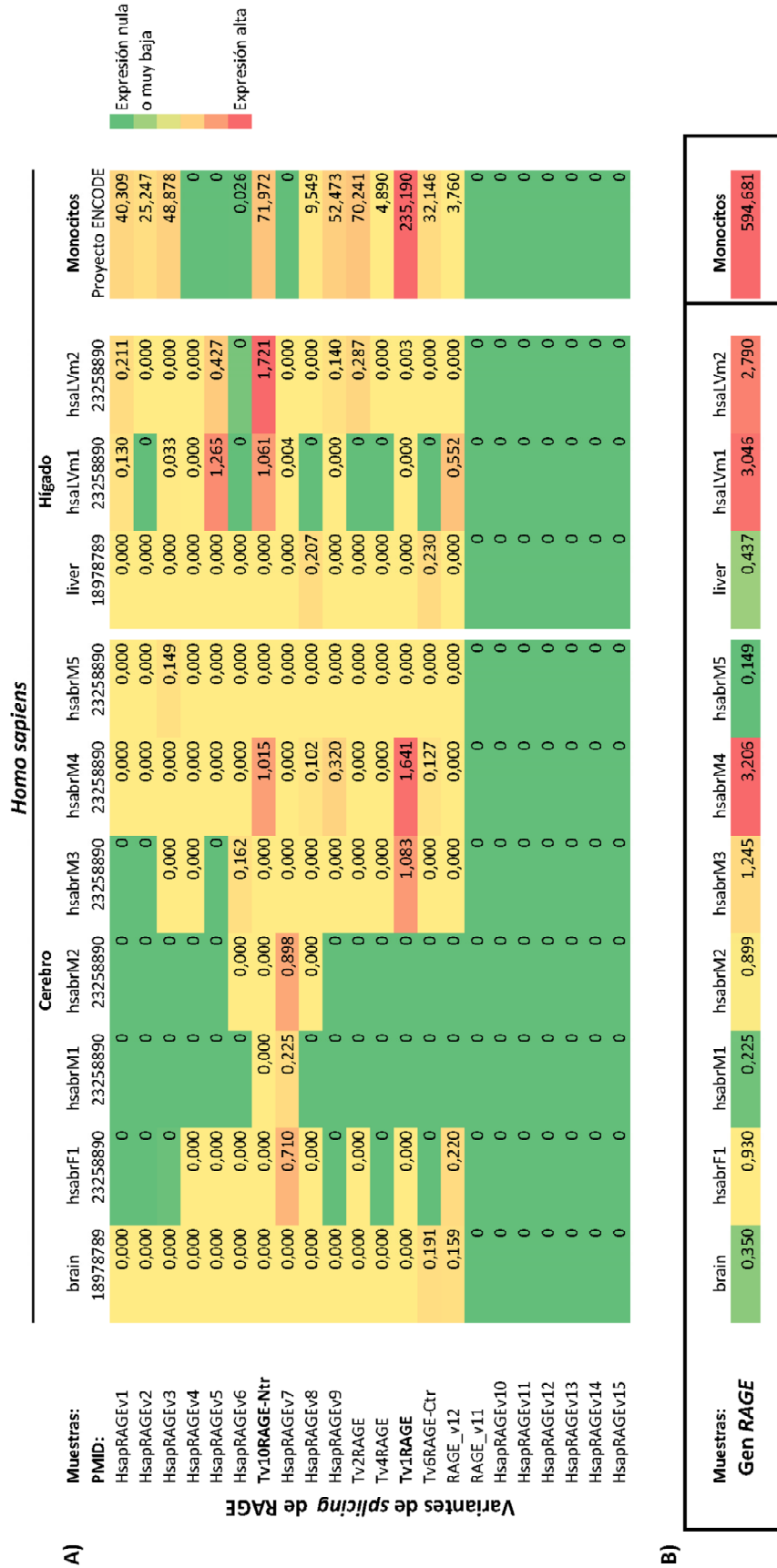
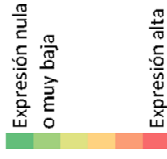




Figura Suplementaria 2 (continuación).

*Macaca mulatta*

Muestras: PMID:	Cerebro					Hígado					
	mm1brF1 23258890	mm1brM1 23258890	mm1brM2 23258890	rhesus_a_brain 23258891	rhesus_b_brain 23258891	rhesus_c_brain 23258891	mm1lvF1 23258890	mm1lvM1 23258890	rhesus_a_liver 23258891	rhesus_b_liver 23258891	rhesus_c_liver 23258891
MmuIRAGEV1	0	0	0	0,915	4,768	0,592	0	0,177	0,352	2,380	0,519
MmuIRAGEV2	0,177	0,000	0	0,546	8,775	1,178	0,244	0,000	0,233	3,624	0,085
MmuIRAGEV3	0	0,540	0,000	0,000	2,511	0,000	0,000	0,310	0,596	2,807	0,583
MmuIRAGEV4	0	0,000	0	0	0	0,000	0	0,000	0	0	0
MmuIRAGEV5	0	0,159	0	0	1,279	0,000	0	0,690	0,000	2,274	0,524
MmuIRAGEV6	0,363	0,000	0,721	0,801	6,197	0,142	1,005	0,000	0,001	1,012	0,142
MmuIRAGEV7	0	0	0,000	0	0	0,000	0,000	0	0	0	0
MmuIRAGEV8	0	0	0,000	0	0,000	0	0,000	0	0	0,236	0,000
MmuIRAGEV9	0,000	0,000	0,000	0	0	0,000	0,000	0	0	0	0
TV1-RAGE	0,848	0,000	0,308	0	3,029	0,000	0,000	0,000	0,000	0,649	0,068
MmuIRAGEV11	0,000	0,000	0,000	0	0,000	0,304	0,281	0,765	0,697	0,000	0,093
MmuIRAGEV12	0,085	0,852	0	0	0	0	0	0	0	0	0
MmuIRAGEV13	0	0	0	0	0	0	0	0	0	0	0
MmuIRAGEV14	0	0	0	0	0	0	0	0	0	0	0
MmuIRAGEV15	0	0	0	0	0	0	0	0	0	0	0



Muestras: Gen RAGE	mm1brF1 1,473	mm1brM1 1,551	mm1brM2 1,029	rhesus_a_brain 2,262	rhesus_b_brain 26,558	rhesus_c_brain 2,216	mm1lvF1 1,530	mm1lvM1 1,943	rhesus_a_liver 1,880	rhesus_b_liver 12,982	rhesus_c_liver 2,014
<b>Gen RAGE</b>	<b>1,473</b>	<b>1,551</b>	<b>1,029</b>	<b>2,262</b>	<b>26,558</b>	<b>2,216</b>	<b>1,530</b>	<b>1,943</b>	<b>1,880</b>	<b>12,982</b>	<b>2,014</b>

Figura Suplementaria 2 (continuación).

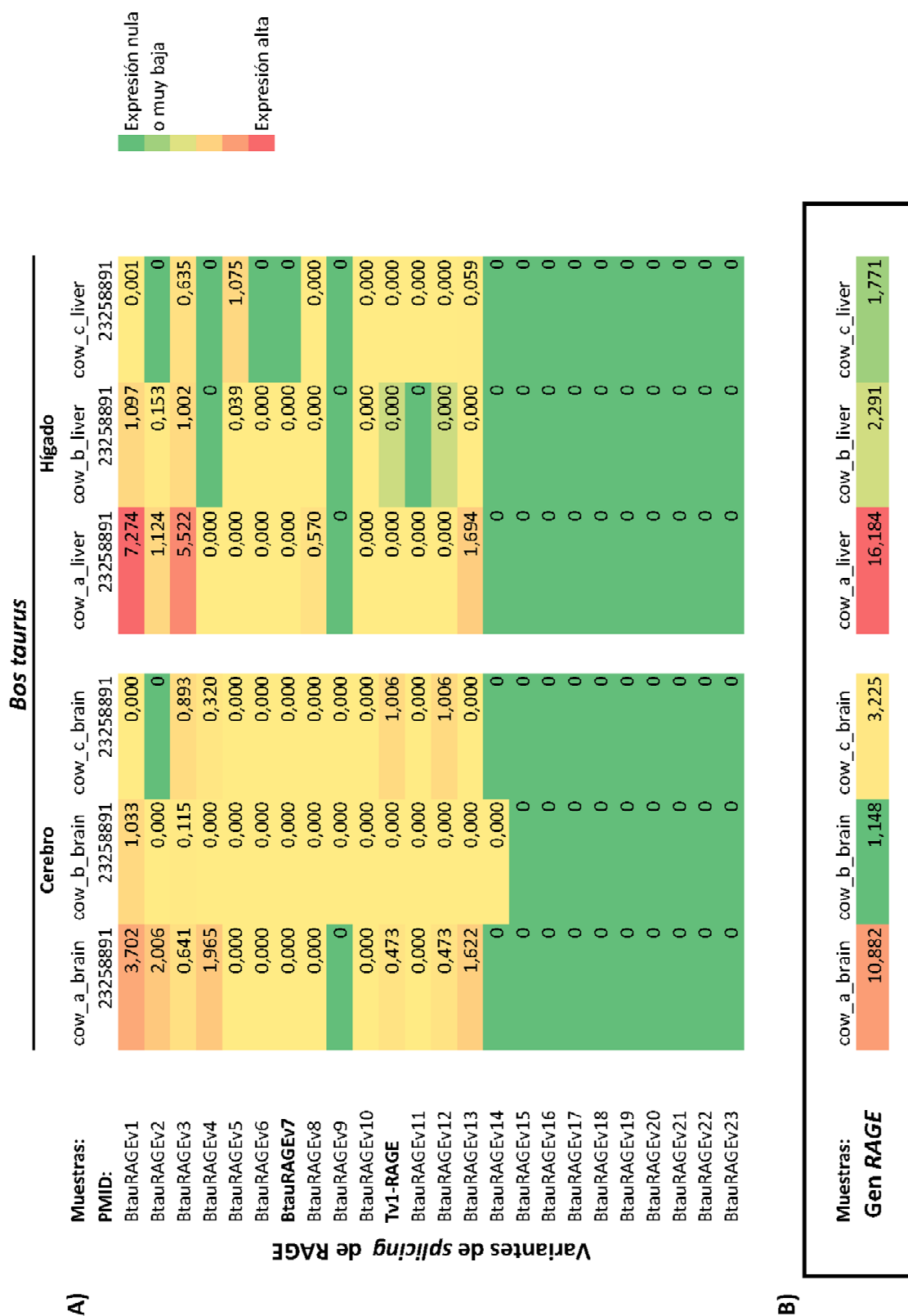


Figura Suplementaria 2 (continuación).

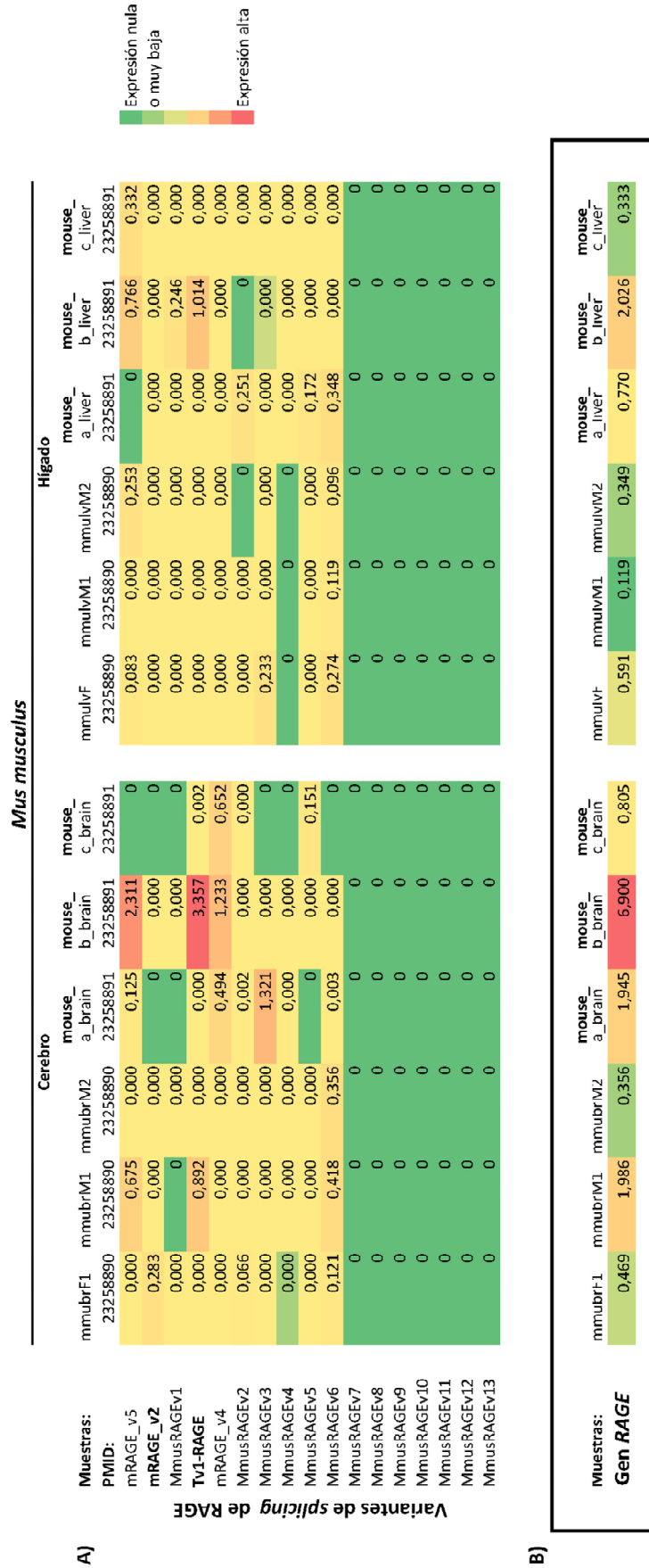


Figura Suplementaria 2 (continuación).

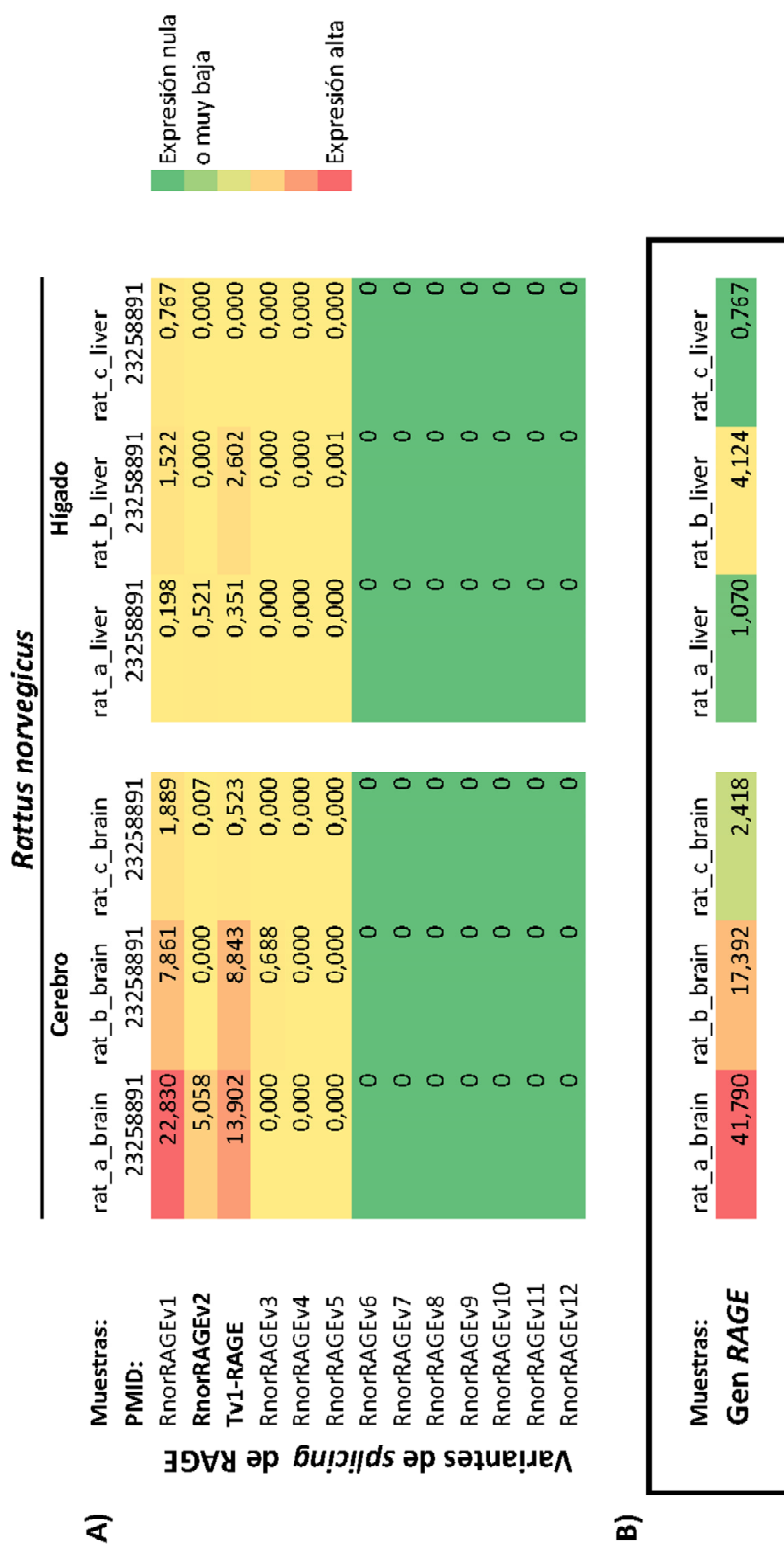
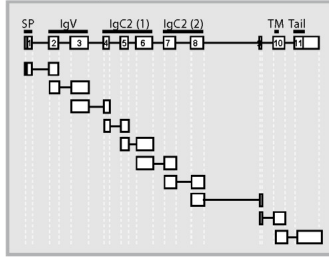


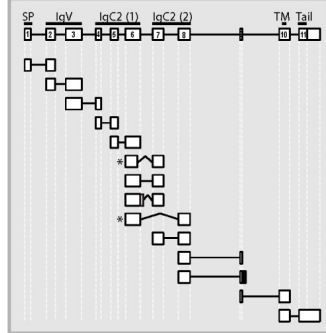
Figura Suplementaria 3.

*Homo sapiens*



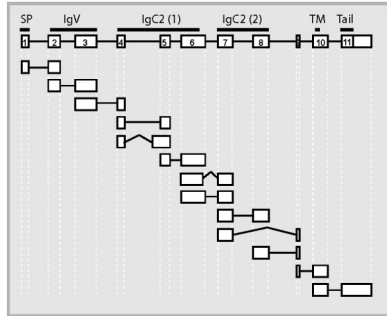
Tejido:	Cerebro						Hígado			
	brain	hsabrF1	hsabrM1	hsabrM2	hsabrM3	hsabrM4	hsabrM5	liver	hsaLvM1	hsaLvM2
Muestras:	18978789	23258890	23258890	23258890	23258890	23258890	23258890	18978789	23258890	23258890
PMID:										
	0,000	0,000	0,000	0,000	0,041	0,029	0,000	0,000	0,000	0,042
	0,000	0,000	0,000	0,000	0,000	0,029	0,000	0,000	0,000	0,084
	0,000	0,000	0,000	0,000	0,000	0,057	0,000	0,000	0,000	0,084
	0,000	0,000	0,000	0,000	0,000	0,086	0,000	0,000	0,000	0,126
	0,000	0,000	0,000	0,000	0,082	0,086	0,000	0,004	0,023	0,042
	0,000	0,000	0,053	0,000	0,082	0,143	0,000	0,000	0,093	0,042
	0,000	0,000	0,053	0,000	0,082	0,029	0,126	0,000	0,093	0,042
	0,000	0,000	0,053	0,000	0,082	0,115	0,000	0,000	0,046	0,042
	0,000	0,041	0,053	0,044	0,000	0,057	0,000	0,000	0,046	0,042
	0,000	0,041	0,000	0,000	0,041	0,115	0,000	0,000	0,023	0,084

*Macaca mulatta*



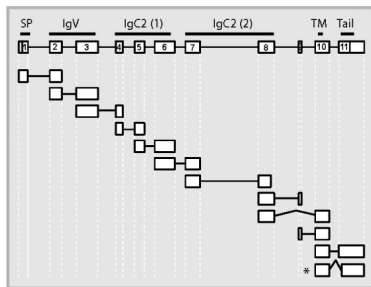
Tejido:	Cerebro						Hígado				
	mmubr	mmubr	mmubr	rhesus_a_brain	rhesus_b_brain	rhesus_c_brain	mmulv	mmulv	rhesus_a_liver	rhesus_b_liver	rhesus_c_liver
Muestras:	F1	M1	M2	a_brain	b_brain	c_brain	F1	M1	a_liver	b_liver	c_liver
PMID:	23258890	23258890	23258890	23258891	23258891	23258891	23258890	23258890	23258891	23258891	23258891
	0,052	0,044	0,000	0,000	0,023	0,000	0,000	0,000	0,000	0,009	0,000
	0,052	0,044	0,000	0,000	0,037	0,000	0,000	0,000	0,000	0,018	0,000
	0,052	0,044	0,000	0,029	0,060	0,000	0,000	0,000	0,000	0,013	0,000
	0,052	0,044	0,000	0,043	0,070	0,057	0,000	0,062	0,088	0,084	0,000
	0,052	0,000	0,000	0,100	0,144	0,000	0,092	0,093	0,018	0,066	0,019
	0,000	0,000	0,000	0,000	0,005	0,000	0,000	0,000	0,000	0,000	0,000
	0,000	0,000	0,047	0,000	0,037	0,000	0,000	0,062	0,070	0,027	0,037
	0,000	0,000	0,000	0,000	0,005	0,000	0,000	0,000	0,000	0,000	0,000
	0,000	0,000	0,000	0,000	0,005	0,000	0,000	0,000	0,000	0,000	0,000
	0,157	0,000	0,047	0,000	0,116	0,000	0,138	0,217	0,088	0,018	0,075
	0,000	0,000	0,047	0,014	0,042	0,019	0,000	0,062	0,000	0,000	0,037
	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,031	0,000
	0,000	0,000	0,000	0,000	0,056	0,076	0,000	0,000	0,000	0,027	0,037
	0,000	0,000	0,000	0,043	0,176	0,113	0,184	0,000	0,018	0,075	0,056

*Bos taurus*



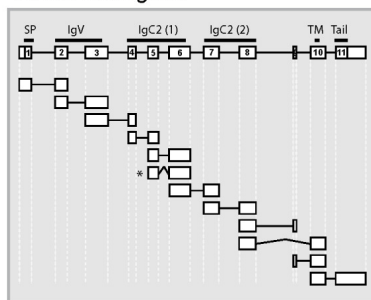
Tejido:	Cerebro			Hígado		
	cow_a_brain	cow_b_brain	cow_c_brain	cow_a_liver	cow_b_liver	cow_c_liver
Muestras:	a_brain	b_brain	c_brain	a_liver	b_liver	c_liver
PMID:	23258891	23258891	23258891	23258891	23258891	23258891
	0,000	0,000	0,000	0,000	0,000	0,000
	0,029	0,053	0,119	0,058	0,058	0,000
	0,019	0,000	0,000	0,034	0,000	0,000
	0,014	0,000	0,030	0,039	0,038	0,017
	0,000	0,000	0,000	0,010	0,000	0,000
	0,010	0,000	0,015	0,058	0,038	0,000
	0,024	0,035	0,015	0,034	0,000	0,000
	0,000	0,000	0,000	0,005	0,000	0,000
	0,014	0,035	0,030	0,019	0,000	0,000
	0,005	0,000	0,000	0,000	0,000	0,000
	0,053	0,000	0,030	0,063	0,019	0,000
	0,105	0,000	0,000	0,049	0,038	0,034
	0,163	0,018	0,000	0,180	0,154	0,103

*Mus musculus*



Tejido:	Cerebro						Hígado					
	mmubr	mmubr	mmubr	mouse_a_brain	mouse_b_brain	mouse_c_brain	mmulv	mmulv	mmulv	mouse_a_liver	mouse_b_liver	mouse_c_liver
Muestras:	F1	M1	M2	a_brain	b_brain	c_brain	F1	M1	M2	a_liver	b_liver	c_liver
PMID:	23258890	23258890	23258890	23258891	23258891	23258891	23258890	23258890	23258890	23258891	23258891	23258891
	0,000	0,000	0,000	0,011	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	0,040	0,013	0,000	0,017	0,008	0,015	0,000	0,000	0,000	0,004	0,000	0,000
	0,040	0,039	0,000	0,017	0,021	0,015	0,000	0,000	0,029	0,000	0,004	0,014
	0,000	0,000	0,000	0,017	0,042	0,015	0,000	0,000	0,000	0,000	0,007	0,000
	0,000	0,000	0,051	0,000	0,034	0,000	0,000	0,000	0,000	0,000	0,015	0,014
	0,000	0,000	0,000	0,000	0,017	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	0,040	0,013	0,000	0,006	0,059	0,015	0,000	0,000	0,000	0,000	0,007	0,000
	0,000	0,013	0,000	0,000	0,017	0,000	0,000	0,000	0,000	0,000	0,007	0,000
	0,000	0,000	0,000	0,006	0,008	0,000	0,021	0,000	0,000	0,000	0,000	0,000
	0,000	0,000	0,000	0,004	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	0,040	0,026	0,000	0,023	0,029	0,000	0,000	0,000	0,000	0,004	0,004	0,000
	0,000	0,000	0,000	0,000	0,008	0,000	0,000	0,000	0,000	0,000	0,000	0,000

*Rattus norvegicus*



Tejido:	Cerebro			Hígado		
	rat_a_brain	rat_b_brain	rat_c_brain	rat_a_liver	rat_b_liver	rat_c_liver
Muestras:	a_brain	b_brain	c_brain	a_liver	b_liver	c_liver
PMID:	23258891	23258891	23258891	23258891	23258891	23258891
	0,013	0,000	0,000	0,000	0,000	0,000
	0,084	0,005	0,031	0,019	0,011	0,000
	0,042	0,005	0,015	0,000	0,000	0,000
	0,080	0,057	0,031	0,000	0,008	0,036
	0,328	0,145	0,062	0,019	0,030	0,012
	0,004	0,000	0,000	0,000	0,000	0,000
	0,227	0,083	0,062	0,000	0,000	0,000
	0,189	0,083	0,031	0,000	0,011	0,000
	0,042	0,021	0,000	0,038	0,015	0,000
	0,029	0,016	0,000	0,000	0,000	0,000
	0,122	0,016	0,015	0,000	0,000	0,000
	0,332	0,239	0,139	0,000	0,034	0,024

Figura Suplementaria 4.

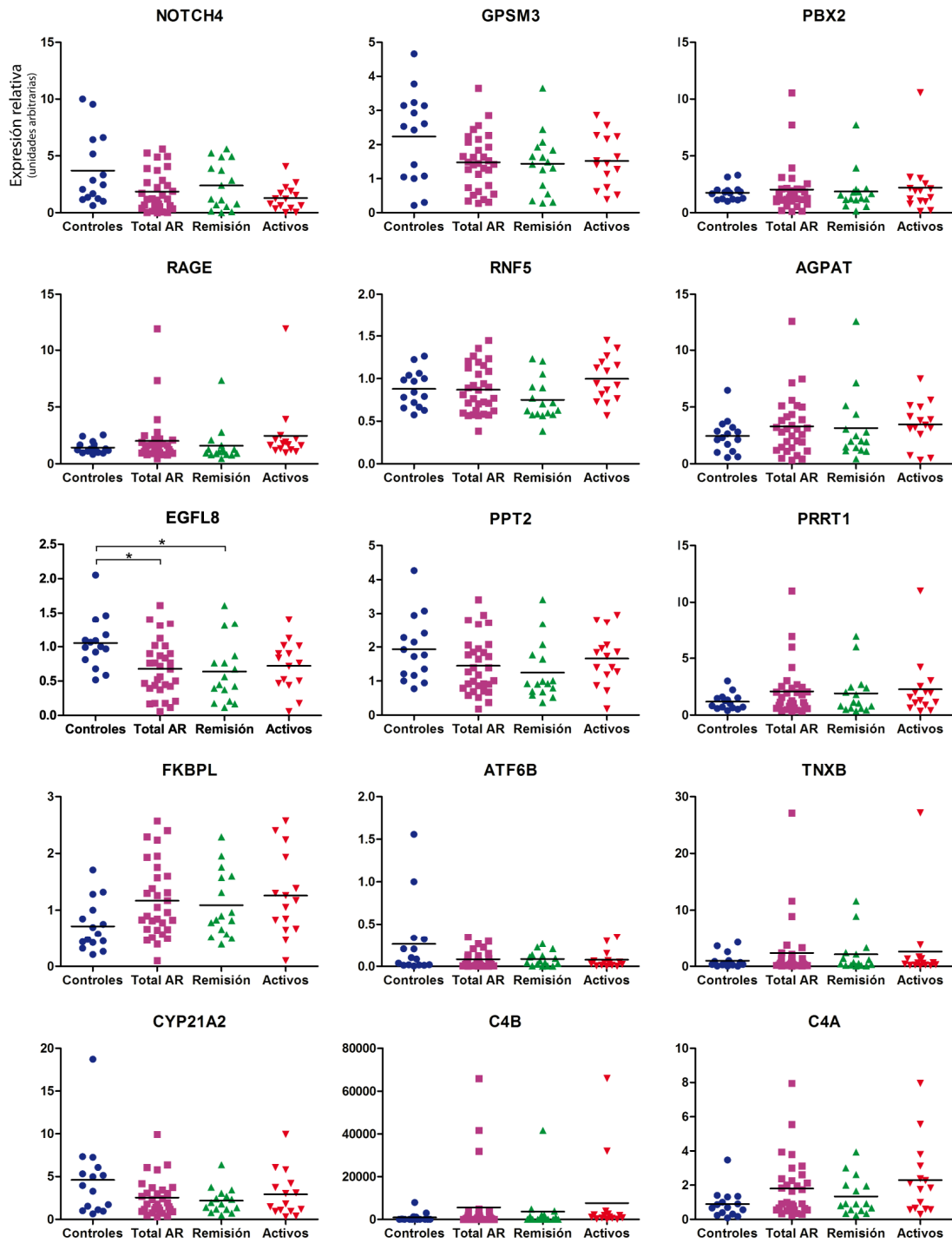


Figura Suplementaria 4 (continuación).

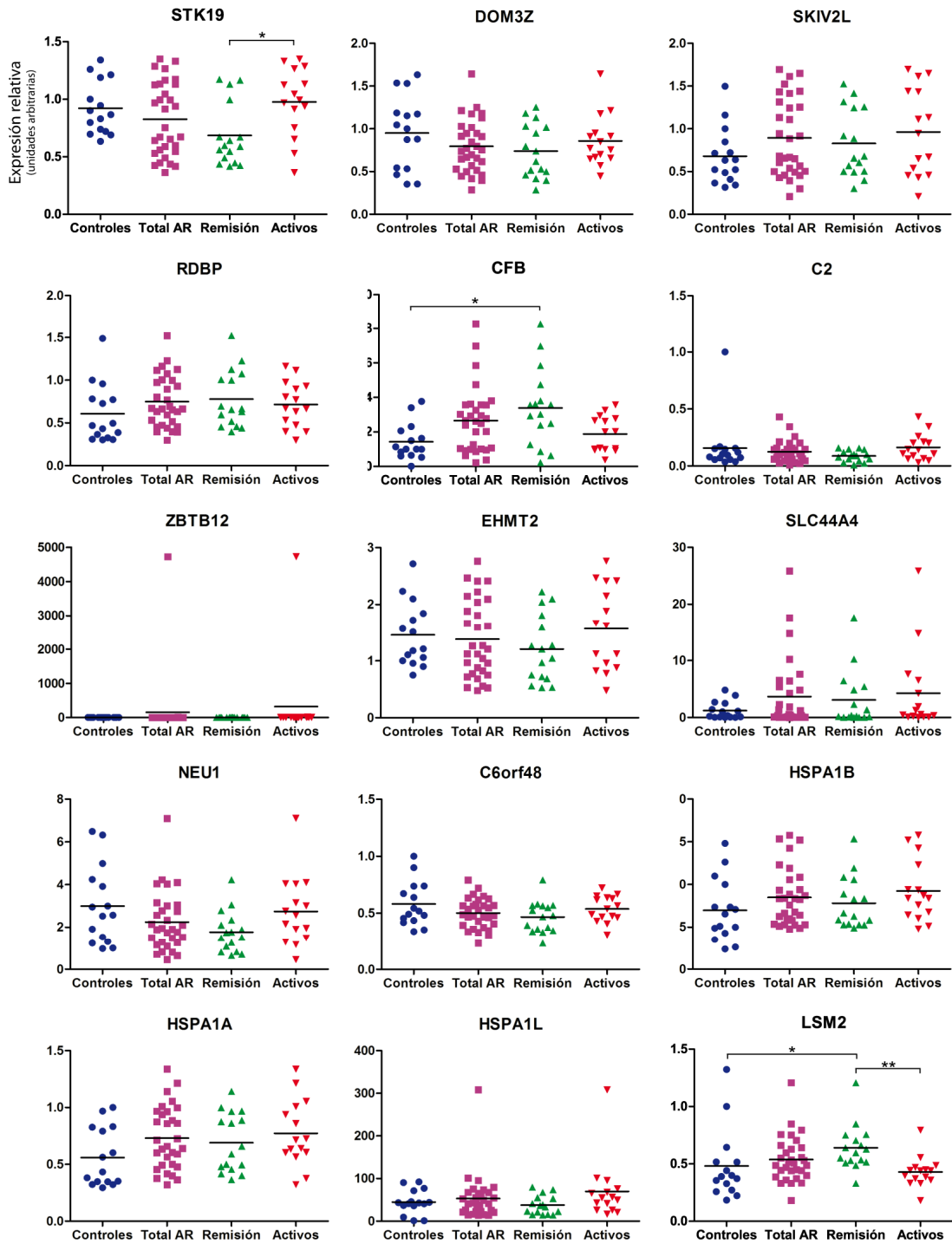




Figura Suplementaria 4 (continuación).

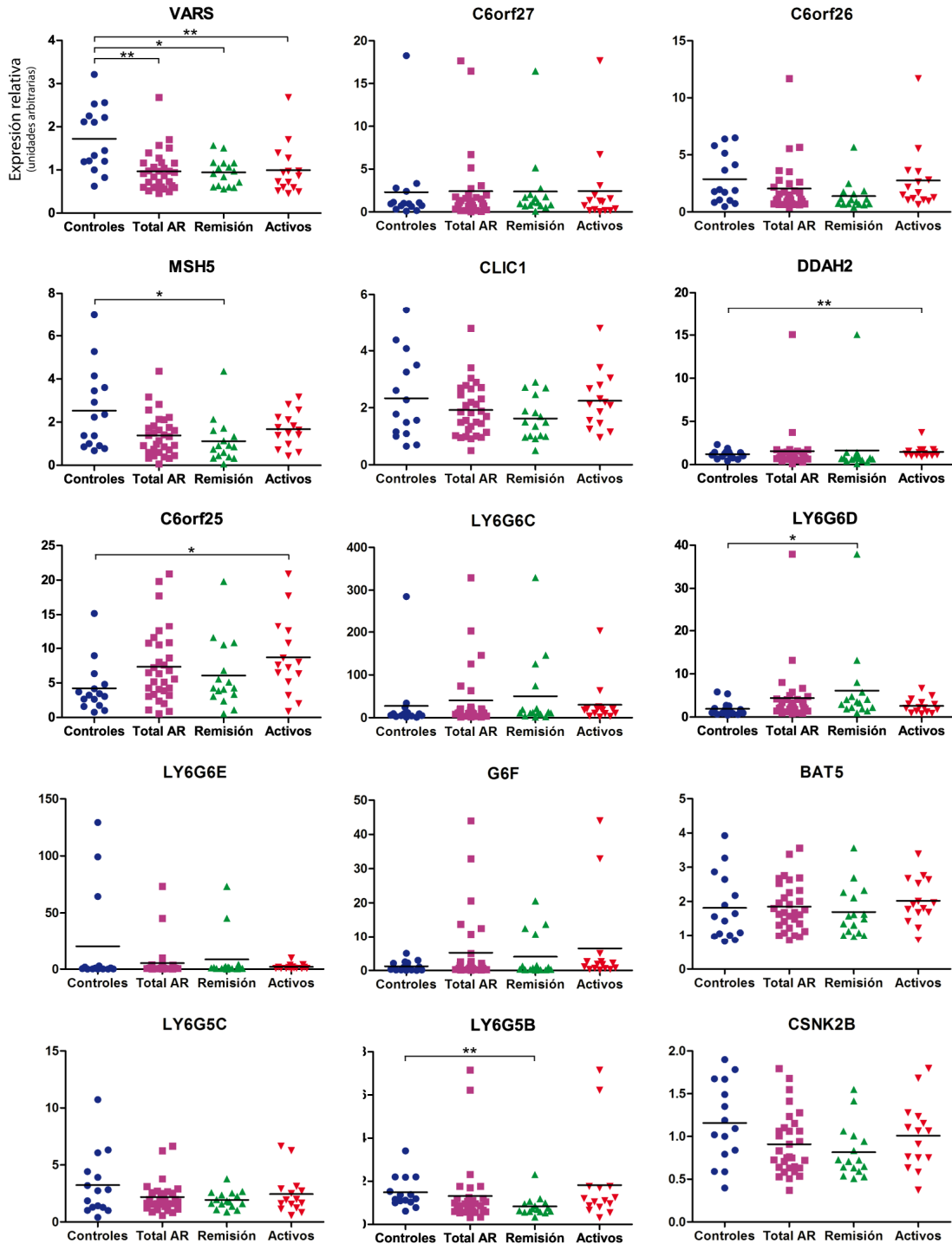
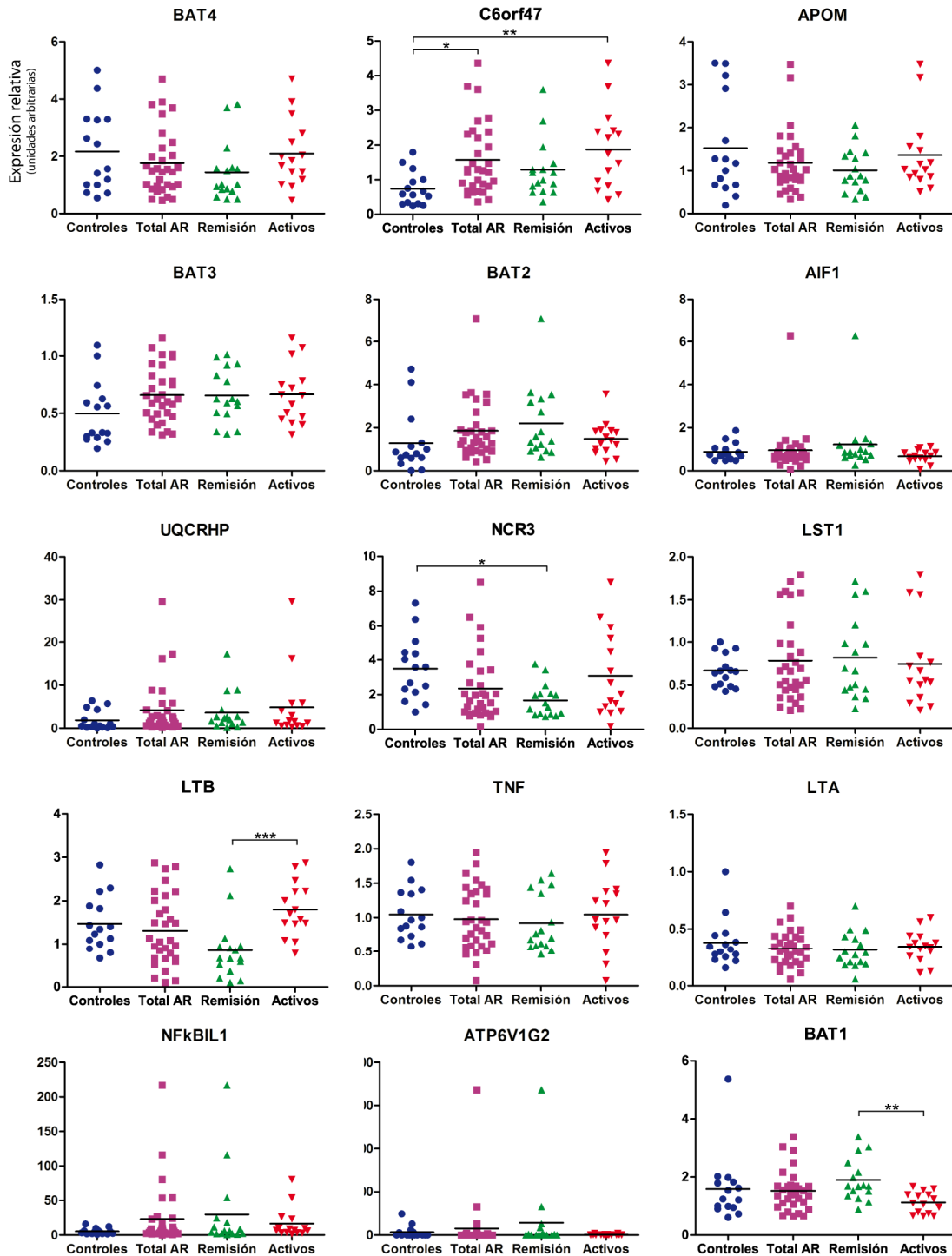




Figura Suplementaria 4 (continuación).



## **9. ANEXOS**