



EVOLUTION AND MOLECULAR CHARACTERIZATION OF UROPLAKIN 3C, A NOVEL HUMAN TETRASPANIN ASSOCIATED UROPLAKIN PROTEIN

Javier Ugarte Chicote

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.

UNIVERSITAT ROVIRA I VIRGILI

EVOLUTION AND MOLECULAR CHARACTERIZATION OF UROPLAKIN 3C, A NOVEL HUMAN TETRASPANIN ASSOCIATED UROPLAKIN

Javier Ugarte Chicote

UNIVERSITAT ROVIRA I VIRGILI

EVOLUTION AND MOLECULAR CHARACTERIZATION OF UROPLAKIN 3C, A NOVEL HUMAN TETRASPANIN ASSOCIATED UROPLAKIN

Javier Ugarte Chicote

Javier Ugarte Chicote

**Evolution and molecular characterization of uroplakin 3c, a
novel human tetraspanin associated uroplakin protein**

Doctoral Thesis

Directed by Dr. Antonio García-España Monsonís

Department of Medicine and Surgery



UNIVERSITAT ROVIRA I VIRGILI

Tarragona 2016

UNIVERSITAT ROVIRA I VIRGILI

EVOLUTION AND MOLECULAR CHARACTERIZATION OF UROPLAKIN 3C, A NOVEL HUMAN TETRASPANIN ASSOCIATED UROPLAKIN

Javier Ugarte Chicote



Departament de Medicina i Cirurgia

C/ Sant Llorenç, 21

43203 – Reus (Tarragona)

Telèfon: 977 759 305

Fax: 977 759 322

FAIG CONSTAR que aquest treball, titulat “Evolution and molecular characterization of uroplakin 3c, a novel human tetraspanin associated uroplakin protein”, que presenta Javier Ugarte Chicote per a l’obtenció del títol de Doctor, ha estat realitzat sota la meva direcció al Departament de Medicina i Cirurgia d’aquesta universitat i que aconsegueix els requeriments per poder optar a Menció Internacional.

Tarragona, 29 de Novembre de 2015

El director de la tesi doctoral

Dr. Antonio García-España Monsonís

UNIVERSITAT ROVIRA I VIRGILI

EVOLUTION AND MOLECULAR CHARACTERIZATION OF UROPLAKIN 3C, A NOVEL HUMAN TETRASPANIN ASSOCIATED UROPLAKIN

Javier Ugarte Chicote

A Una Mancuerna y a Dos Terneros

UNIVERSITAT ROVIRA I VIRGILI

EVOLUTION AND MOLECULAR CHARACTERIZATION OF UROPLAKIN 3C, A NOVEL HUMAN TETRASPANIN ASSOCIATED UROPLAKIN

Javier Ugarte Chicote

AGRADECIMIENTOS

Agradecimiento es la acción de mostrar gratitud debido a un sentimiento que nos obligade a estimar el beneficio o favor que se nos ha hecho o ha querido hacer. De esta manera, a través de estas líneas, querría mostrar mi agradecimiento a todas las personas partícipes de alguna u otra manera de este ciclo que se cierra ahora, ya que no hay mayor beneficio que el haber aprendido algo de ellas.

Sin embargo, querría especialmente agradecerle a Antonio la oportunidad de brindarme el aprendizaje recibido por él durante estos años y, por lo tanto, haber sido capaz de sacar adelante este trabajo. Al igual, que siempre le estaré agradecido por la cercanía que ha demostrado dentro y fuera del laboratorio.

El número de horas invertidas en hipótesis disparatadas, errores de cálculo, obstáculos epistemológicos, prueba-error y déficits de atención afortunadamente no pueden ser reflejadas de manera alguna junto al formato compacto de resultados seleccionados como es el de una tesis. Aún así, pienso que es necesario mostrarles mi gratitud debido a que probablemente son de las que más uno puede llegar a aprender y a reírse.

Sin alegría e ilusión este trabajo no sería tal como es ahora y por tanto, quiero agradecer a todos los integrantes del Grupo B la alegría e ilusión que nos hemos contagiado durante todo este tiempo sonriendo, porque así, se ven las cosas de otra manera. Igualmente me gustaría agradecer estas sonrisas a las partículas elementales de cada día.

Finalmente, me gustaría mostrar mi más sincero agradecimiento a la técnica de la RT-PCR, dado que aunque tras una tortuosa larga relación sustentada en lo incomprensible, los dos fuimos capaces de llegar a un entendimiento de una manera adulta y convencional. Tú, diste pie al resto.

UNIVERSITAT ROVIRA I VIRGILI

EVOLUTION AND MOLECULAR CHARACTERIZATION OF UROPLAKIN 3C, A NOVEL HUMAN TETRASPANIN ASSOCIATED UROPLAKIN

Javier Ugarte Chicote

ÍNDICE

UNIVERSITAT ROVIRA I VIRGILI

EVOLUTION AND MOLECULAR CHARACTERIZATION OF UROPLAKIN 3C, A NOVEL HUMAN TETRASPANIN ASSOCIATED UROPLAKIN

Javier Ugarte Chicote

RESUMEN.....	1
INTRODUCCIÓN.....	7
HIPÓTESIS Y OBJETIVOS.....	19
MATERIALES Y MÉTODOS	23
RESULTADOS	39

ORIGEN Y EVOLUCIÓN DE LAS UROPLAQUINAS. IDENTIFICACIÓN DE TRES NUEVAS UROPLAQUINAS: UPK3c, UPK3d Y UPK2b..... 41

- 1.Búsqueda de secuencias en bases de datos de DNA y proteína.....41
- 2.Relaciones evolutivas de las uroplaquinas de la familia tetraspanina (UPK1a/1b)42
- 3.Relaciones evolutivas de las uroplaquinas de la familia UPK2/3: identificación de tres nuevas uroplakin UPK3c, UPK3d y UPK2b43
- 4.Origen de las uroplaquinas en el ancestro común de los vertebrados y desaparición específica en los peces óseos modernos.....48
- 5.La conservación de la pauta de exones codificantes e intrónes y de cisteínas extracelulares en UPK2s y UPK3s refuerza su origen evolutivo común.48

RELACIÓN EVOLUTIVA ENTRE LAS UROPLAQUINAS UPK2/3 Y LOS RECEPTORES SUBTIPO 3 DE TIROSINA FOSFATASAS (R3 PTPRS). CARACTERIZACIÓN DEL DOMINIO UROFIBRINA 51

- 1.Identificación de un fragmento de proteína similar a las uroplaquinas UPK2/3 en los receptores de proteína tirosina fosfatasa (R3 PTPRS)51
- 2.Caracterización del dominio urofibrina en UPK2/3 y R3 RPTPs de vertebrados e invertebrados.....54
- 3.El dominio urofibrina es probablemente un dominio FN3 divergente58
- 4.Relación evolutiva entre UPK3a y PTPRQ. Origen de las uroplaquinas UPK2/3.....60

UPK3c EN EL CONTEXTO DE LA HISTORIA EVOLUTIVA DEL CROMOSOMA 7 DE LOS GRANDES PRIMATES. POSIBLE MECANISMO DE TRASLOCACIÓN A TRAVÉS DE UN INTERMEDIARIO CIRCULAR.	63
1.Localización de UPK3c en el cromosoma 7 de los grandes primates	63
2.Caracterización de la duplicación segmental (DS2) translocada al locus genético 7p13.....	67
3.Señales de mecanismos de reparación de DNA de doble cadena en las zonas de unión entre la DS2 y el locus genético 7p13.....	68
4.Recombinación entre elementos Alu como posible mecanismo de la formación del intermediario circular.....	68
5.Análisis de la zona de inserción y apertura del intermediario circular	69
CARACTERIZACIÓN MOLECULAR DE UPK3c. DIVERGENCIA RESPECTO A UPK3a	73
1. Pauta de lectura de las tres copias de UPK3c	73
2. Imbricación de UPK3c con la subunidad J de la RNA Polimerasa II	73
3. Análisis transcripcional de UPK3c.....	75
4. Análisis de la secuencia proteica de UPK3c	84
5. Glicosilación en UPK3c	88
6. Immunolocalización de UPK3c en el urotelio de ratón	92
7. UPK3c y cáncer de vejiga	94
DISCUSIÓN	105
CONCLUSIONES	127
BIBLIOGRAFÍA	131
ANEXOS	151

RESUMEN

UNIVERSITAT ROVIRA I VIRGILI

EVOLUTION AND MOLECULAR CHARACTERIZATION OF UROPLAKIN 3C, A NOVEL HUMAN TETRASPANIN ASSOCIATED UROPLAKIN

Javier Ugarte Chicote

Uroplakins (UPKs) are a group of integral membrane proteins that form the protein subunits of the specific plaques (urothelial plaques) that cover the apical surface of mammalian bladder epithelium (urothelium) contributing to its remarkable permeability barrier function and mechanical stability. Uroplakins can be divided into two types. The first type comprises UPK1a and UPK1b, which belong to the tetraspanin family of proteins that cross the cell membrane four times. The second type which is called tetraspanin associated uroplakins (UPK2/3), comprises UPK2, UPK3a and UPK3b, spans the membrane only once and is not known to be related to any other protein family. Uroplakin defects underlie urinary tract anomalies, and one of them UPK1a, can serve as the receptor for the uropathogenic *E. coli* that causes over 85% of urinary tract infections.

Despite the fact that the urothelial plaques are unique to the urinary bladder of mammals, uroplakins also exist in lower vertebrates without a typical urothelium including amphibians where in *Xenopus*, UPK3a and UPK1b play a key role in sperm-egg fertilization, and in zebrafish where a UPK3-related gene product was recently found to play a role in epithelial polarization and morphogenesis of pronephric tubules. The evolutionary relationship among these fish, amphibian and mammalian uroplakins, that seem to be functionally divergent, remains unclear.

Taking advantage of the recent availability of greatly expanded genomic databases of chordates (vertebrates, cephalochordates and urochordates) to better understand the evolution of uroplakins we gather uroplakin-related sequences in a wide range chordates using bioinformatic tools and protein profiling techniques.

We have first pin-point the origin of uroplakins in the common ancestor of vertebrates since UPK-related sequences exists in lampreys (extant jawless basal vertebrate) but not in urochordates (*Ciona*) the closest animals to vertebrates, cephalochordates (*Amphioxus*), or in lower organisms. This observation places the origin of uroplakins over 500 million years ago when vertebrates radiated from cephalochordates and urochordates and most

likely underwent two rounds of whole genome duplication. Secondly, we have identified three novel UPK2/3 ortholog groups (UPK3c expressed in reptiles/birds and mammals; UPK2b in reptiles/birds and amphibians, and UPK3d in fish) that will possibly account with the four already known for all the existing UPK2/3s different groups in vertebrates. This data will help clarify the evolutionary relationships between fish, amphibian and mammalian uroplakins that may perform divergent functions specific to these different and physiologically distinct groups of vertebrates.

While the origin of uroplakin tetraspanins (UPK1a and UPK1b) could be explained by gene duplication from an ancestor invertebrate tetraspanin gene, the origins of UPK2/3 uroplakins remains unknown. In the present study, we show that UPK2/3s are related to the more ancient family of subtype R3 tyrosine phosphatase receptors (R3 PTPRs). Although the overall similarity of these proteins in low members of both families share a novel membrane-proximal FN3-like or “urofibrin” protein domain characterized by several amino acid signatures and conserved pairs of cysteine residues. In addition, phylogenetic analyses and the conservation of the intron/exon gene structure in their overlapping sequence regions (urofibrin and transmembrane regions), which in UPK2/3-comprises the whole protein except for the cytoplasmatic tail, indicates that UPK2/3 proteins have probably originated from a R3 PTPR related gene (PTPRQ) by gene duplication and domain loss in the common ancestor of vertebrates. This finding broadens our view on the structure and function of these two groups of integral membrane proteins with important roles in human disease. R3 PTPRs belong to the large family of phosphotyrosine phosphatases which together with protein tyrosine kinases regulate the levels of phosphotyrosine modifications in cells and its deregulation leads to various health disorders including cancer

Since UPK3c, in contrast to UPK2b and UPK3d, is expressed in mammals, including humans, we devoted our efforts to characterize this novel protein. First, by synteny analysis we have demonstrated that UPK3c originated by the duplication of the *UPK3b* gene in the common ancestor of reptiles/birds and mammals. Second, we have studied the expansion of the

UPK3c genetic loci that took place in gorillas, chimpanzees, and humans, which is linked to the duplications of a large segmental duplication originated after the split of the orangutan and human branches. We notice that an intrachromosomal transposition of one of these segmental duplications could have been produced by a circular intermediate mechanism, a phenomenon that to the best of our knowledge has only been reported once in mammals and never in primates. This suggestion is based on the presence of conserved sequence features compatible with a translocation through a circular intermediate: first in comparison with the putative ancestral sequence the inserted 80,6 Kb genomic block is precisely split in two fragments located in a head to tail orientation with the loss of only 4 bp of its sequence during process, second the existence of dsDNA breakage and repair fingerprints around the insertion point that could explain this sharp split of the translocated sequence, and third the signals left by an Alu Recombination Mediated Deletion (ARMD) event that may probably generated the circular intermediate.

Next using molecular biology techniques we have studied UPK3c expression (in a panel of 22 human tissues, and in a small sample of 31 human bladder tumors) and its localization in the mouse urothelium in transgenic models of superficial and invasive bladder tumors using confocal microscopy technique. We found that the expression of UPK3c was widespread in human tissues and it was particularly strong in corneal tissue. Additionally, UPK3c is N-glycosylated and contains up to 20 KDa of complex oligosaccharides in human normal corneal tissue, pterygium and bladder tumors.

In contrast with the localization of the uroplakins that form the urothelial plaques (UPK1a, UPK1b, UPK2a, UPK3a, and UPK3b), UPK3c was not present in the apical surface of the umbrella cells, and localizes predominately to the basal and intermediate cell layers of the mouse urothelium, maintaining this differential expression in mouse transgenic induced tumors. Finally in human bladder cancer the number of superficial

low grade tumors expressing UPK3c assessed by immunoblot was higher (9/11) than in superficial high grade tumors (6/9) or invasive tumors (1/11).

INTRODUCCIÓN

UNIVERSITAT ROVIRA I VIRGILI

EVOLUTION AND MOLECULAR CHARACTERIZATION OF UROPLAKIN 3C, A NOVEL HUMAN TETRASPANIN ASSOCIATED UROPLAKIN

Javier Ugarte Chicote

El uroepitelio o urotelio es el epitelio localizado en la pelvis renal, uréteres, vejiga, uretra superior y los conductos prostáticos de mamíferos. Está compuesto por tres capas celulares: basal, intermedia y apical. (Hatina and Schulz 2012) (Figura 1).

La capa basal se encuentra localizada sobre una lámina basal bajo la cual, en la vejiga, se encuentran la lámina propia y el músculo detrusor. Está compuesta por una única capa de células mononucleadas, de tamaño entre 5 y 10 μm , baja granularidad y un ratio núcleo:citoplasma elevado (Hatina y Schulz 2012). Aproximadamente, un 10% de las células basales podrían ser células madre uroteliales, tal como ha podido ser demostrado a través del marcaje del urotelio con bromodeoxiuridina (Kurzrock *et al.* 2008).

La capa intermedia está localizada entre la capa basal y la capa apical y ocasionalmente algunas de sus células están en contacto directo con la lámina basal (Hatina y Schulz 2012). El número de capas celulares que componen la capa intermedia varía según la especie, por ejemplo roedores y humanos tienen 2 y 5 capas respectivamente. No obstante, su grosor varía acorde al estado de distensión de la vejiga. Las células que componen esta capa son células mononucleadas, piriformes y de tamaño entre 10 y 15 μm (Jost *et al.* 1989). El urotelio posee un intervalo de renovación lento de entre 6 y 12 meses (Khandelwal *et al.* 2009). Sin embargo, en caso de que células de la capa apical sean dañadas por senescencia, infección bacteriana o manipulación experimental (Kanai *et al.* 2004, Lavelle *et al.* 2002; Mysorekar *et al.* 2002; Apodaca *et al.* 2003), las células de la capa intermedia en contacto directo con la capa apical están parcialmente diferenciadas y pueden rápidamente diferenciarse y ocupar su lugar.

Por último, la capa apical es la más superficial y está en contacto directo con el lumen. Es la principal responsable de formar la barrera que controla el libre intercambio de sustancias entre orina y sangre, y está

compuesta de una sola capa de células altamente especializadas y totalmente diferenciadas llamadas células “umbrella” (Figura 1).

De esta manera, las capas celulares del urotelio forman un gradiente de diferenciación. Las capas basales y apicales están compuestas de células no diferenciadas o totalmente diferenciadas respectivamente, mientras que la capa intermedia representa una capa de amplificación transitoria.

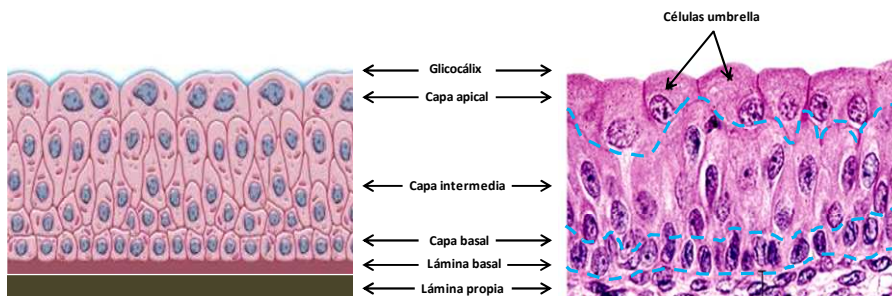


Figura 1: El urotelio está compuesto por tres capas celulares diferentes.

Representación e imagen de un corte de urotelio de primate teñido con hematoxilina – eosina (Jones y Cheng 2005) donde se pueden apreciar las diferentes partes que constituyen el urotelio. Desde la parte más luminal el urotelio está constituido por el glicocáliz de las células “umbrella” que forman la capa apical, inmediatamente por debajo se localizarían varias capas de células intermedias y finalmente una capa de células basales depositadas sobre una lámina basal. Las capas celulares del urotelio en la imagen están separadas por una capa azul discontinua.

Las células “umbrella” tienen un tamaño entre 25 y 250 μm , son mono- o multinucleadas según la especie, apareciendo en humanos frecuentemente como células binucleadas. Además, tienen forma poliedral aunque tanto su tamaño como su morfología varían según el estado de distensión de la vejiga. En el caso de que la vejiga esté vacía y el urotelio relajado, las células “umbrella” tienen forma cuboidal. En cambio, cuando el urotelio está distendido debido a que la vejiga está llena de orina, tienen un aspecto escamoso (Hicks *et al.* 1975; Truschel *et al.* 2002).

Las células “umbrella” son células polarizadas que poseen membranas basolaterales y apicales totalmente diferenciadas y delimitadas por uniones estrechas (Acharya *et al.* 2004; Varley *et al.* 2006; Lewis *et al.* 1976) (Figura 2a y 2b). Su membrana apical está tapizada por placas con regiones bisagra intermedias (Liang *et al.* 1999) (Figura 2a y 2b). Estas placas están formadas por unas estructuras cristalinas de 16 nm de diámetro compuestas de 6 subunidades formando un anillo interno y otro externo (Min *et al.* 2006; Walz *et al.* 1995) (Figura 2c y 2d). Los componentes de estas estructuras cristalinas de 16 nm son una familia de proteínas transmembrana llamadas uroplaquinas.

Las estructuras cristalinas de 16 nm hacen que la parte más externa de la membrana de las placas sea el doble de gruesa formando así lo que se conoce como la unidad de membrana asimétrica (Hicks 1965; Porter *et al.* 1967; Walz *et al.* 1995) (Figura 3a).

Las uroplaquinas no solo se encuentran en la unidad de membrana asimétrica ya que también se encuentran en vesículas discoidales o con forma fusiforme que en las células “umbrella” se acumulan debajo de la membrana y que se fusionan con la membrana apical cuando la vejiga se distiende (Hicks 1975, Truschel *et al.* 2002) (Figura 3b), aunque también se encuentran en las células de la capa intermedia que están parcialmente diferenciadas. Por lo tanto, las uroplaquinas son consideradas productos de diferenciación debido a que exclusivamente se localizan en las capas parcial o totalmente diferenciadas del urotelio (Wu *et al.* 1994).

Dentro de las uroplaquinas, se pueden diferenciar dos grupos: el grupo de las uroplaquinas de la familia de las tetraspaninas: uroplaquina 1a (UPK1a) y uroplaquina 1b (UPK1b) y el grupo de las uroplaquinas denominadas asociadas a tetraspaninas o UPKs 2/3: uroplaquina 2 (UPK2), uroplaquina 3a (UPK3a) y uroplaquina 3b (UPK3b) (Figura 4). Todas estas uroplaquinas son los componentes mayoritarios de las partículas cristalinas de 16nm de las placas excepto UPK3b, que es minoritaria (Wu *et al.* 1994; Deng *et al.* 2002).

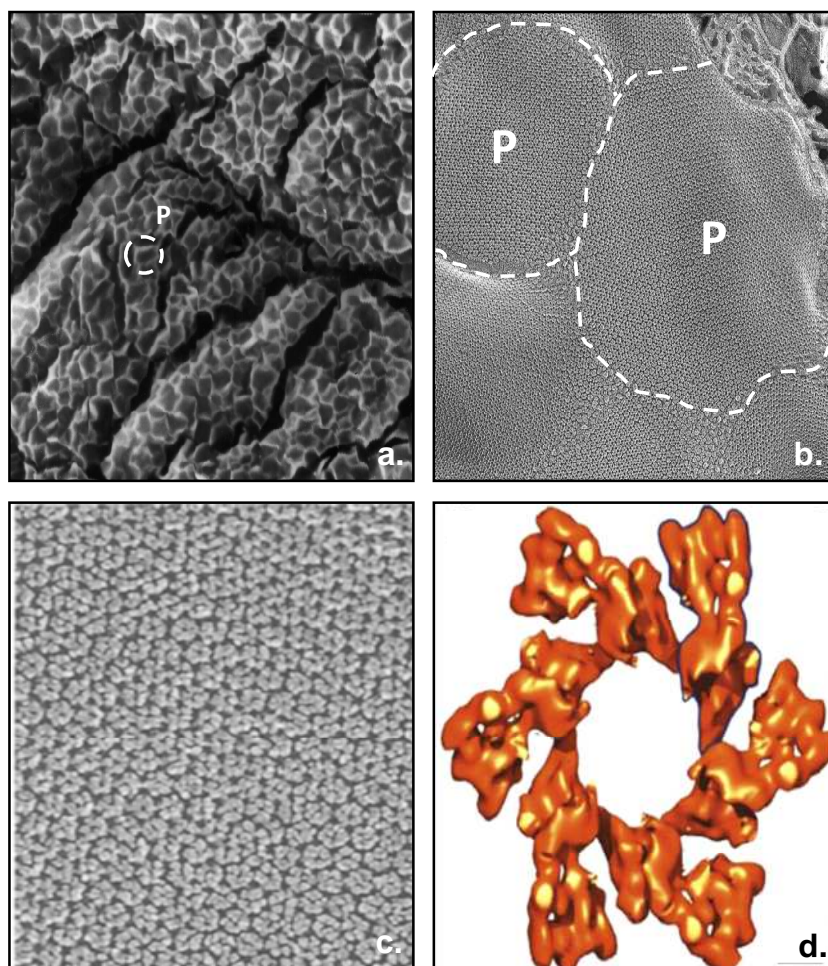


Figura 2: Las uroplaquinas forman las partículas de 16 nm de las placas que tapizan la membrana apical de las células “umbrella”. **(a.)** Imagen por microscopía electrónica de barrido de la superficie apical de las células “umbrella” de urotelio de ratón mostrando las placas del urotelio (P) (Adaptación de Sun *et al.* 2006). **(b.)** Imagen por microscopía electrónica de barrido usando la técnica de “quick-freeze deep-etch” de la superficie apical de células “umbrella” del urotelio de ratón donde se pueden apreciar las partículas de 16 nm empaquetadas hexagonalmente formando las placas uroteliales (P) que están interconectadas por áreas bisagra con partículas de 16 nm deformadas (Línea blanca discontinua). La parte superior derecha corresponde a un área de la parte apical de la membrana rota (Adaptación de Sun *et al.* 2006). **(c.)** Superficie apical de células “umbrella” del urotelio de ratón a mayor aumento donde se aprecia la forma hexagonal de las partículas de 16 nm. (Apodaca *et al.* 2004) **(d.)** Representación tridimensional de las partículas de 16 nm que están formadas por uroplaquinas (Sun *et al.* 2006).

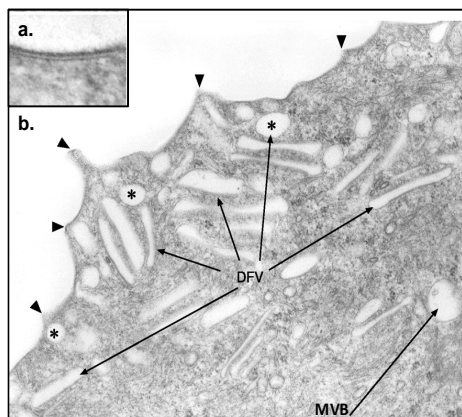


Figura 3: Unidad de membrana asimétrica y vesículas discoidales o fusiformes en la parte apical de una célula “umbrella”. (a.) Imagen por microscopía electrónica de barrido mostrando la unidad de membrana asimétrica de la membrana apical de una célula “umbrella” de rata. (b.) Imagen por microscopía electrónica de barrido de la parte más apical de una célula “umbrella” de rata. Las flechas marcan las regiones de bisagra entre las placas. Justo debajo de la membrana se pueden observar vesículas discoidales (*) o fusiformes (DFV) y cuerpos multivesiculares (MVB) más abajo en el citoplasma. (Khandelwal et al. 2009).

Modelos de ratones “knock-out” para uroplaquinas y estudios de transfección e interacción química han demostrado que las uroplaquinas forman heterodímeros (UPK1a/UPK2 y UPK1b/UPK3a o UPK3b), lo que les permite llegar a la parte apical sin quedar atrapadas en el retículo endoplasmático (Hu *et al.* 2005; Hu *et al.* 2000; Kong *et al.* 2004; Wu *et al.* 1995; Tu *et al.* 2002; Deng *et al.* 2002). No obstante, UPK1b es capaz de alcanzar la superficie apical de la célula por sí sola sin necesidad de formar heterodímeros con UPK3a o UPK3b (Tu *et al.* 2002). Atendiendo a esto, se ha postulado un modelo de ensamblaje de uroplaquinas que les permitiría llegar a la parte apical de las células “umbrella” (Wu *et al.* 2009). En principio la formación de los heterodímeros ocurriría en el retículo endoplasmático permitiéndoles así pasar al aparato de Golgi donde serían modificadas post-transduccionalmente. Esto permitiría la formación de las partículas de 16 nm mediante la unión de distintos heterodímeros. Desde el aparato de Golgi, las uroplaquinas oligomerizadas en partículas de 16nm (Hu *et al.* 2005) serían empaquetadas en las vesículas discoidales y fusiformes y liberadas en la membrana apical de las células “umbrella” (Alroy *et al.* 1982; Hicks 1966; Hudoklin *et al.* 2009).

Las uroplaquinas tetraspaninas pertenecen a la familia de las tetraspaninas y como tal poseen 4 hélices transmembrana (DTM) con dos asas extracelulares, mientras que las UPK2/3s son proteínas transmembrana tipo I con un solo DTM (Deng *et al.* 2002, Wu *et al.* 1994) (Figura 4). Las uroplaquinas en su estado maduro están glicosiladas mediante enlaces N-glicosídicos formando así parte de la capa de glicanos que se puede encontrar en la parte apical del urotelio (Parsons *et al.* 1990; Hurst *et al.* 1987; Parsons *et al.* 1977). La única uroplaquina en estado maduro que no está glicosilada es la UPK2 (Lin *et al.* 2004). Esto es debido a que es sintetizada como una pro-secuencia que es N-glicosilada pero que en el proceso de maduración para formar el heterodímero con UPK1a es cortada en una señal de corte para una endoproteasa furina que da lugar a la pérdida del fragmento N-glicosilado (Lin *et al.* 2004).

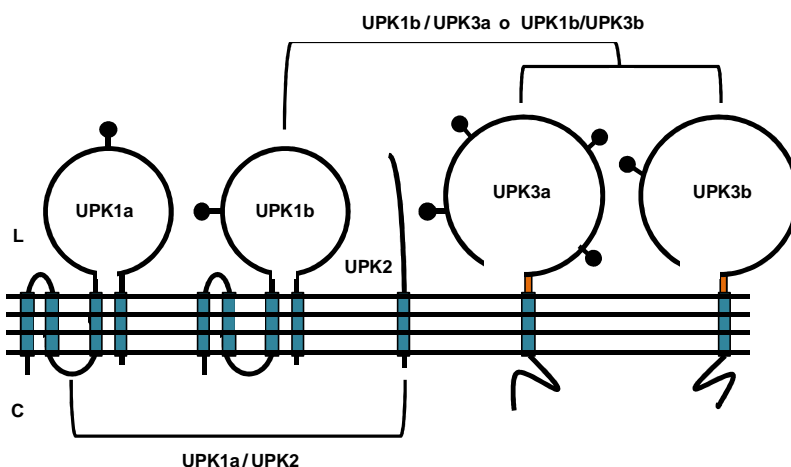


Figura 4: UPK1a/1b y UPKs 2/3 poseen características propias y forman heterodímeros específicos entre ellas. Representación de las UPKs tetraspaninas (UPK1a y UPK1b) y de las UPKs 2/3 (UPK2, UPK3a y UPK3b) donde se pueden apreciar las hélices transmembrana (DTM) (barras de color azul), los sitios potenciales de glicosilación (círculos negros), las regiones citoplasmáticas de las UPK3s y el fragmento de 19 aminoácidos (aa) yuxtapuesto a la DTM que está muy conservado en las UPK3s (barras color naranja). Los heterodímeros específicos que forman han sido reflejados mediante corchetes. L y C representan lumen y citoplasma respectivamente (Adaptación de Deng *et al.* 2002).

UPK3a y UPK3b tienen un fragmento de secuencia yuxtapuesto a su DTM altamente conservado, el cual ha sido sugerido como posible responsable de la formación de los heterodímeros UPK1b/UPK3a y UPK1b/UPK3b. Sin embargo, resultados previos no han demostrado esto aunque tampoco han descartado que esté implicado en el posterior ensamblaje de las partículas cristalinas de 16 nm (Deng *et al.* 2002).

El urotelio de mamíferos es el único que contiene placas tapizando su parte apical (Staehein *et al.* 1972; Hicks *et al.* 1975; Min *et al.* 2003). Por esta razón, las uroplaquinas, tras su descubrimiento como componentes mayoritarios de las placas, fueron consideradas mamífero y urotelio específicas (Wu *et al.* 1994). Posteriormente, se ha visto que no son mamífero específicas (Mitra *et al.* 2011, Sakakibara *et al.* 2005; García-España *et al.* 2006) ni urotelio específicas (Olsburgh *et al.* 2003, Adachi *et al.* 2001). Además, se ha podido establecer que el origen de las uroplaquinas tuvo que ocurrir al principio de la evolución de los vertebrados, dado que las secuencias más antiguas relacionadas con uroplaquinas se encontraron en peces cartilagosos (García-España *et al.* 2006).

Modelos de ratón “knock-out” para UPK2 y UPK3a han demostrado que su ausencia produce alteraciones en la formación de las placas, dando lugar al aumento de la permeabilidad a través de la capa de células “umbrella”, dejando claro su importancia en mantener la integridad de la membrana apical que actúa como barrera (Kong *et al.* 2004; Hu *et al.* 2000). Además, estos ratones “knock-out” presentan reflujo vesicoureteral que consiste en el paso de orina desde la vejiga a los riñones. Esto es una enfermedad hereditaria en humanos (Dillon *et al.* 1998), pero sin embargo, no se ha demostrado ninguna relación con mutaciones en UPK2 o UPK3a humanas (Kelly *et al.* 2005; Giltay *et al.* 2004). No obstante, mutaciones en UPK3a podrían causar hipodisplasia renal, caracterizada por riñones de menor tamaño y menor número de nefronas (Schönfelder *et al.* 2006; Jenkins *et al.* 2005).

Estudios de apoptosis celular inducida por la bacteria *E. coli* uropatogénica han mostrado señalización dependiente de UPK3a en las células uroepiteliales de mamíferos y que UPK3a tiene un papel importante en la respuesta del huésped frente a las infecciones del tracto urinario (Thumbikat *et al.* 2009) aunque se haya visto *in vitro* que el receptor para la unión del Fim H de *E. coli* sean los residuos de manosa de una arginina de UPK1a (Zhou *et al.* 2001).

El cáncer de vejiga fue el noveno cáncer con más incidencia a nivel mundial en el año 2012 (Lindsey *et al.* 2015). Dentro de los distintos tipos de cáncer de vejiga, el carcinoma urotelial (CU) es el más común de todos. El 70-80% de los casos son CUs que no invaden la capa muscular (NMICUs) y que están localizados en la mucosa (estadio Ta el 70% de los casos y carcinoma *in situ* (CIS) el 10%) o en la submucosa (estadio T1 el 20% de los casos). Estos NMICUs van desde tumores benignos de bajo grado y riesgo de progresión hasta tumores de elevado grado y riesgo de progresión, como los CIS que progresan en un 60% de los casos, y que casi con toda seguridad recidivan. El 20-30% restante representan la otra variante y son CUs que han invadido la capa muscular en el momento del diagnóstico, los cuales tienen una elevada capacidad de extenderse localmente y de progresar a metástasis (Falke and Witjes 2011).

Se han postulado dos rutas principales de carcinogénesis en los CUs con características moleculares exclusivas de cada una. Las dos rutas pierden heterocigosidad (LOH) en el cromosoma 9 y simplificándolas, por un lado los CUs de bajo grado presentarían un aumento de la activación de rutas de señalización de factores de crecimiento a través de mutaciones en el receptor del factor de crecimiento de fibroblasto subtipo 3 (FGFR3) (van Rhijn *et al.* 2002, Hernández *et al.* 2006; Kimura *et al.* 2001) o en el oncogén HRas (Bos 1989). Por otro lado, los tumores de alto grado se podrían caracterizar como una alteración de la regulación del ciclo celular con mutaciones que hacen perder la función de los genes supresores de

tumores TP53 y pRB1 como ejemplo de los cambios más característicos (Schulz 2006).

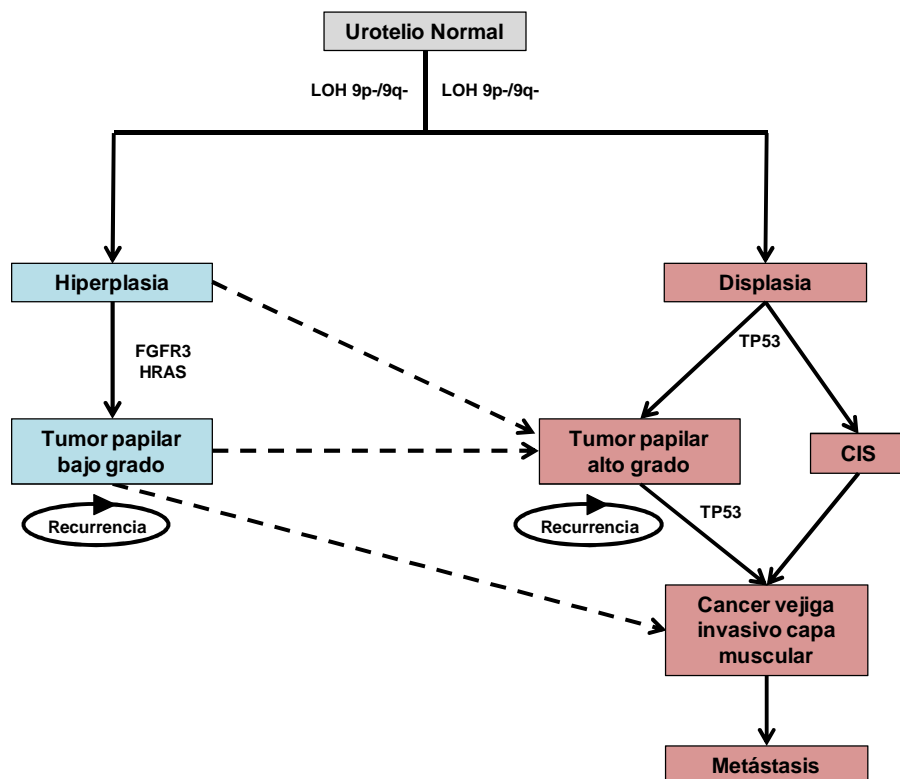


Figura 5: El carcinoma urotelial se desarrolla a través de dos rutas. La ruta de bajo grado (azul), se inicia por la LOH del cromosoma 9, dando lugar a una hiperplasia y a un tumor papilar de bajo grado. Esta ruta está asociada con mutaciones en FGFR3 y HRAS. Estos tumores tienen un elevado riesgo de recurrencia pero bajo riesgo de progresión. La ruta de alto grado (rojo) frecuentemente se inicia también por la LOH del cromosoma 9, dando lugar a displasia. Mutaciones en TP53 podrían dar lugar a CIS o a tumores papilares de alto grado. Mutaciones en TP53 están asociadas también con un potencial invasivo y de metástasis (Adaptación de Falke *et al.* 2011).

Estas dos rutas de carcinogénesis han podido ser recreadas en dos modelos de ratones transgénicos utilizando el promotor de UPK2 y a través de la activación constitutiva de HRAS en la variante papilar no invasiva (Zangh *et al.* 1999) y la inactivación de TP53 y pRB1 en la variante invasiva (Zangh *et al.* 2001) (Figura 5).

Las uroplaquinas son sintetizadas de manera dependiente a la diferenciación como se ha podido comprobar en su expresión diferencial en las distintas capas del urotelio (Moll *et al.* 1995; Olsburgh *et al.* 2003). Por tanto, cabría esperar que su expresión disminuyese durante la progresión y la tumorigénesis. Sin embargo, en CUs humanos no se ha podido demostrar una estricta relación entre la disminución de la expresión de las uroplaquinas y la progresión y tumorigénesis del urotelio (Huang *et al.* 2007; Olsburgh *et al.* 2003; Yuasa *et al.* 2003; Moll *et al.* 1995).

HIPÓTESIS Y OBJETIVOS

UNIVERSITAT ROVIRA I VIRGILI

EVOLUTION AND MOLECULAR CHARACTERIZATION OF UROPLAKIN 3C, A NOVEL HUMAN TETRASPANIN ASSOCIATED UROPLAKIN

Javier Ugarte Chicote

El trabajo realizado en esta tesis doctoral se ha centrado en una de las líneas de investigación del grupo de Biología Molecular del Urotelio Sano y Patológico perteneciente al Institut d'Investigació Sanitària Pere i Virgili. Desde hace más de 15 años el Dr. Antonio García-España Monsonís, responsable de este grupo, trabaja conjuntamente con el Dr. Tung-Tien Sun descubridor de las uroplaquinas. El descubrimiento por parte del Dr. Antonio García-España Monsonís de una nueva uroplaquina, UPK3c, motivó el planteamiento de unos objetivos preliminares que han dado lugar al trabajo presentado en esta tesis doctoral.

Aunque el Dr. Antonio García-España y colaboradores ya habían establecido una aproximación al origen de la familia de las uroplaquinas, el descubrimiento de UPK3c y la disponibilidad de un mayor número de genomas secuenciados de cordados marcó como objetivo estudiar de una manera más precisa la historia evolutiva de esta familia de proteínas al igual que comprobar si existían nuevas uroplaquinas además de la UPK3c.

Una vez establecida la historia evolutiva de la familia de las uroplaquinas, se marcó como un segundo objetivo poder establecer el origen de las UPK2/3 dado que, al contrario que UPK1a/UPK1b, que están relacionadas con la familia de las tetraspaninas, no se conocía que existiese relación con familias de proteínas más antiguas.

El gen de UPK3c ha sufrido cambios en su localización cromosómica y un aumento en el número de copias durante la especiación de los grandes primates africanos (gorila, chimpance y humanos) pasando de 1 copia en orangután a 3 copias en humanos. Como objetivo se intentó establecer las causas de este aumento en el número de copias y el de los cambios en la localización cromosómica.

Al no tener ninguna información previa sobre UPK3c, se marcó como objetivo su caracterización molecular y debido a que la familia de las uroplaquinas está vinculada al urotelio, se propuso estudiar su expresión en tumores de vejiga

MATERIALES Y MÉTODOS

UNIVERSITAT ROVIRA I VIRGILI

EVOLUTION AND MOLECULAR CHARACTERIZATION OF UROPLAKIN 3C, A NOVEL HUMAN TETRASPANIN ASSOCIATED UROPLAKIN

Javier Ugarte Chicote

1. Análisis *in silico*

a. Obtención de secuencias

Todas las secuencias que en el momento del estudio no se encontraban anotadas en ninguno de los proyectos de secuenciación del genoma NCBI (www.ncbi.nlm.nih.gov) y Ensembl (www.ensembl.org) o en artículos científicos (DeSalle *et al.* 2014; van Eekelen *et al.* 2010; García-España *et al.* 2006; Byrum *et al.* 2006; Ono-Koyanagi *et al.* 2000; Ono *et al.* 1999) se anotaron manualmente. Para anotar las secuencias manualmente se realizaron búsquedas Blast con el programa BLASTN en los genomas correspondientes utilizando las secuencias ortólogas de proteínas más cercanas evolutivamente al genoma en el que se realizaba la búsqueda. La delimitación de los extremos intrón-exón se realizaron con la función “Align two sequences” del NCBI Blast y las señales de splice de los intrones se anotaron manualmente. La traducción de la secuencia de DNA a proteína se realizó con el programa DNA club.

Para la obtención de secuencias de primates y el estudio de sus genomas se utilizó el servidor UCSC (www.genome.ucsc.edu). Se utilizaron los ensamblajes GRCH 37/hg19 para humano, CSAC 2.1.4/panTro4 para chimpancé, gorGor3.1/gorGor3 para gorila y WUGSC 2.0.2/ponAbe2 para orangután.

b. Análisis filogenético

En el análisis realizado se incluyeron todas las secuencias de proteínas transmembrana y dominios urofibrina de un grupo representativo de vertebrados, dos invertebrados deuterostomas (*Strongylocentrotus* y *Ciona*), tres protostomas (*Capitella*, *Drosophila* y *Caenorhabditis*) y un

porífero (*Amphimedon queenslandica*). Como secuencias no relacionadas se utilizaron las correspondientes regiones del representativo subtipo R2A de fosfatasas PTPRF y PTPRF de monosiga (*Monosiga brevicollis*). Los alineamientos se generaron con el servidor TranslatorX (Abascal *et al.* 2010). Para producir las matrices de secuencias de DNA y de proteína se llevaron a cabo alineamientos múltiples de secuencias de nucleótidos guiados por la traducción a aminoácidos (aa).

Al examinar las matrices de las secuencias de DNA se observó una saturación extrema de los datos. De esta manera, solamente se utilizaron las secuencias de aa para llevar a cabo la relación de los árboles filogenéticos de los miembros de la familia de genes. A los motivos moleculares y las fases intrónicas de estos genes se les otorgó un código de tal manera que se originaron seis caracteres binarios: presencia (G) y ausencia (T); cisteína 1-2 (C₁-C₂) y cisteína 3-4 (C₃-C₄); intrón 1, intrón 2, intrón 3 e intrón 4. Los caracteres de presencia/ausencia se codificaron como G y T respectivamente para permitir los análisis de secuencia de aa combinados y datos de motivos moleculares en el análisis de máxima parsimonia e inferencia Bayesiana. Los árboles Bayesianos se generaron utilizando MrBayes (Ronquist *et al.* 2012) y los de parsimonia utilizando PAUP para las secuencias de aa solas, los datos de motivos moleculares solo y para las dos particiones combinadas. Para los análisis de inferencia Bayesiana el modelo mezclado de "prset aamodelpr=mixed" comando con 500.000 generaciones (hasta el punto de que dos análisis de MrBayes convergiesen, desviación estándar = 0.0105) se utilizó. Los modelos de parsimonia fueron conducidos utilizando 1000 réplicas, TBR branch swapping. El árbol que se ha presentado es de un análisis de inferencia Bayesiana de matriz combinada.

c. Análisis de las secuencias y alineamientos

El servidor SMART (Letunic *et al.* 2014) y la base de datos CDD del NCBI se utilizaron para determinar la composición de los dominios de las secuencias de las proteínas analizadas.

El gráfico hidropático se realizó con el servidor Kyte Doolittle Plot (<http://gcat.davidson.edu/DGPB/kd/kyte-doolittle.html>), la determinación del péptido señal con el servidor SignalP v4.1 (Petersen *et al.* 2011), la de los dominios transmembrana con el servidor TMHMM v2.0 (<http://www.cbs.dtu.dk/services/TMHMM>), los potenciales sitios de N-glicosilación con el servidor NetNGlyc v1.0 (Gupta *et al.* 2002), los posibles sitios de fosforilación con el servidor NetPhos v2.0 (Blom y Gammeltoft 1999) y los sitios de corte de la furina con el servidor ProP v1.0 (Duckert y Brunak 2004). Las posibles secuencias repetitivas en el genoma se delimitaron con el servidor Repeat Masker (A.F.A. Smit, R. Hubley & P. Green RepeatMasker at <http://repeatmasker.org>).

Los alineamientos se realizaron con el servidor MAFFT (Kato y Standley 2013) o el ClustalW (Thompson y Higgins 1994) y el Multialin (Corpet 1988) del servidor NPS@. La identidad y la similitud entre secuencias de proteínas se calcularon con el servidor DiAling2 (Morgenstern 1999)

2. Muestras

Los mRNA de vejiga fetal (540153), hígado (540017), páncreas (540023), tráquea (540145), pulmón (540019), estómago (540037), colon (54009), riñón (540013), útero (540043), testículo (540049), ovario (540071), corazón (540011), músculo esquelético (540029), bazo (540035), nódulo linfático (540021), placenta (540025), cerebro (540005), laringe (540047) y piel (540031) se obtuvieron de Stratagene – Agilent

Technologies (Estados Unidos), el de vejiga adulta (1100564F) se adquirió de Asterand Bioscience (Estados Unidos), los de próstata (636550) y retina (636579) de Nucliber (España) y el de cornea de ABS (Estados Unidos).

Las muestras humanas de cornea y de pterigion fueron cedidas por el Dr. Tung-Tien Sun del departamento de Biología Celular de la escuela de medicina en la Universidad de Nueva York (NYU Langome Medical Center, Nueva York, Estados Unidos). Las muestras de tumores de vejiga humanos se obtuvieron de pacientes intervenidos de CUs en el servicio de urología del Hospital Universitari Joan XXIII (Tarragona, España). El estudio fue aprobado por el comité ético del hospital y a todos los pacientes se les solicitó el consentimiento informado.

3.Síntesis de cDNA

Todas las reacciones de síntesis de cDNA se realizaron a partir de 200 ng totales de mRNA utilizando el kit Transcriptor First Strand cDNA Synthesis Kit (Roche, Alemania). Primeramente tanto el mRNA (20 ng/ μ l) junto con Random Hexamer Primer (60 pmol/ μ l) se desnaturalizaron durante 10' a 65°C. Posteriormente, se le añadió solución tampón (50 mM Tris/HCl, 30 mM KCl, 8 mM MgCl₂), junto con 20U de Inhibidores de protectores de RNase, dinucleotidos (dNTPS) (1mM) y 10U de la enzima Transcriptasa Reversa. La concentración final de mRNA fue de 10 ng/ μ l y los parámetros de incubación de 10' a 25°C, 30' a 55°C y 5' a 85°C. Todas las reacciones se almacenaron a -20°C hasta el momento de ser utilizadas.

4.RT-PCR

Para el cDNA de los tejidos y las muestras de tumores de vejiga, de cornea y pterigion se realizó la amplificación del transcrito de UPK3c utilizando dos primers que abarcan todo el CDS de las tres posibles copias de UPK3c en humanos (F: 5'-GACGGACGGACAGACAGATGGACA-3' y R: 5'-GCCCCCTCTGGAACCCCTCAG-3'). Una concentración final de 0.4 ng/μl de cDNA, junto solución tampón (16 mM (NH₄)SO₄ y 67 mM Tris/HCl) con dNTPs (Roche, Alemania) a 80 μM, MgCl₂ a 1.5 mM, 0.3 μM de cada primer y 2.5 U de BIOTAQ™ DNA polimerasa (Bioline, Reino Unido) primeramente se incubaron durante 5' a 95°C para luego realizar 35 ciclos de 30'' a 95°C, 30'' a 59°C y finalmente 60'' a 72°C. Finalmente las reacciones fueron incubadas otros 20' a 72°C.

5.Análisis de las copias de UPK3c

Para llevar a cabo el estudio de las copias de UPK3c se realizó una re-amplificación del producto de PCR (dilución 1:50) obtenido tras la amplificación de todo el CDS de las tres posibles copias para UPK3c en humanos. Para ello se diseñaron dos primers (F: 5'-TTCAGCAGCCACAACATCTC-3' y R: 5'-CGAGGTGCGTGGTGTATCTT-3') dentro de la secuencia del producto de PCR obtenido y común para las tres copias. Las condiciones para estos nuevos primers fueron las mismas que para la RT-PCR explicada anteriormente pero con una temperatura de alineamiento de 54°C. Dado que las muestras posteriormente se secuenciaron, previamente se realizó una purificación del producto de PCR con el kit GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare, Reino Unido) basado en un sistema de purificación a través de la adsorción y desorción del producto de PCR en columnas de sílice. Un total de dos reacciones de RT-PCR por muestra con un volumen final de 50 μl cada una se purificó de manera conjunta y se eluyó en un volumen de 30 μl de dH₂O.

Posteriormente, 10 µl de muestra a 50 ng/µl junto con 10 µl de 10 pmol/µl del primer interno diseñado dentro del producto de la reamplificación (5'-CTTCTCCCAGAGGGGCTACT-3') se secuenciaron por la empresa Macrogen (Corea). Varias de las muestras se secuenciaron por duplicado.

Con los resultados de los cromatogramas obtenidos se calculó el porcentaje relativo de cada copia de UPK3c. Para ello, se midió la altura de los picos para cada base nucleotídica en el lugar donde toca la sustitución UPK3c 531 C/A como ha sido realizado anteriormente (Kwok *et al.* 1994) y se calculó su porcentaje relativo a partir del tamaño de la suma de los dos picos.

6. Análisis relativo de la expresión transcripcional de uroplaquinas

La expresión relativa de las diferentes uroplaquinas y de POLR2J2/J3 se realizó con sondas TaqMan en un 7900HT Fast Real-Time PCR System (Applied Biosystems, Estados Unidos). Las sondas TaqMan para UPK1a (Hs01086736_m1), UPK1b (Hs00199583_m1), UPK2 (Hs00171854_m1), UPK3a (Hs00199590_m1) y UPK3b (Hs02385861_g1) fueron sondas ya diseñadas mientras que para POLR2J2/J3 y UPK3c se encargaron diseñar unas sondas específicas con un reporter en el extremo 5' FAM (6-carboxifluoresceína) y un quencher en el extremo 3' TAMRA (6-carboxitetrametilrodamina). En el caso de POLR2J2/J3 la sonda estuvo dirigida exclusivamente a su último exón dado que es la única parte de la secuencia que es capaz de diferenciarlas de POLR2J y de UPK3c a la vez. Mientras tanto, en el caso de UPK3c, la sonda estuvo dirigida solamente al primer exón de UPK3c porque así se podría llegar a diferenciar de POLR2J2/J3. Como genes normalizadores se utilizaron sondas TaqMan contra PPIa (Hs99999904_m1), RPLP0 (Hs99999902_m1) y GAPDH (Hs99999905_m1). Aproximadamente 0.5 µM de cada primer, 0.2 µM de la sonda y 25 ng de cDNA se mezclaron con TaqMan Universal PCR Master

Mix. Las condiciones de la PCR fueron 50°C durante 2', 95°C durante 10' y 40 ciclos de 95°C durante 15'' y 60°C durante 20''. Todas las muestras se corrieron por duplicado. Para calcular la expresión relativa se utilizó el método de $2^{-\Delta\Delta Ct}$ basado en el número del ciclo en el cual la amplificación del producto de PCR supera un valor umbral. La expresión se relativizó a los valores obtenidos en una mezcla de los mRNA de todos los tejidos.

7.Transformación y clonación de UPK3c

El producto de PCR obtenido con los primers que abarcan todo el CDS en vejiga normal fue clonado utilizando el kit TOPO TA Cloning® Kit Dual Promoter (Invitrogen, Estados Unidos). 4 µl de producto de PCR de vejiga normal se incubaron con 1 µl del pCR® II-TOPO® y con 1 µl solución salina (200 mM NaCl, 10 mM MgCl₂), para aumentar la eficiencia, durante 5' a temperatura ambiente y posteriormente se conservaron en hielo. Después, 2 µl de la reacción de ligación TOPO® Cloning con el producto de PCR de vejiga se incubó en hielo durante 30' junto con un vial de 50 µl de células One Shot® Chemically Competent *E. coli*. Cada 10' la reacción se mezcló suavemente. Tras la incubación en hielo, las células se sumergieron durante 30'' en un baño de agua a 42°C para permitir la internalización del DNA y posteriormente se conservaron en hielo. A estas células en hielo se le añadieron 250 µl de medio S.O.C. (2% Triptona, 0.5% Extracto de levadura, 10 mM de NaCl, 2.5 mM de KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM Glucosa) y se incubaron en agitación horizontal a 200 rpm durante 60' a 37°C. Finalmente, 3 volúmenes (25, 75 y 150 µl) de medio S.O.C con células competentes transformadas con UPK3c se sembraron en placas de LB (1% Triptona, 0.5% Extracto de Levadura, 1% NaCl, 1.5% de Agar Bacteriano). Las placas contuvieron además 100 µg/ml ampicilina y se le añadió horas previas a la siembra 100 µg/ml de X-gal. Las placas se incubaron en oscuridad a 37 °C durante toda la noche y a la

mañana siguiente se seleccionaron distintos clones transformados (color blanco) frente a los que no habían sido transformados (color azul).

Cada colonia transformada se incubó en medio LB líquido (1% Triptona, 0.5% Extracto de Levadura, 1% NaCl) con 50 µg/ml de ampicilina a 200 rpm en horizontal durante toda la noche en un volumen de 5 mL en tubos de un tamaño al menos 3 veces superior a ese volumen. Al día siguiente con 1 ml de células crecidas en LB líquido se prepararon stocks de glicerol (LB líquido 40% glicerol estéril) y almacenados a -80°C mientras que con los 4 ml restantes se extrajo el plásmido con el kit QIA filter™ Plasmid Mini Kit (Qiagen, Países Bajos) el cual tras una lisis celular el plásmido se purificó por un proceso de adsorción y desorción en columnas de sílice y eluido en un volumen final de 50 µl de dH₂O.

La presencia del transcrito de UPK3c en el plásmido pCR® II-TOPO® se realizó a través de la digestión de 100ng de plásmido con 10 U de la enzima de restricción EcoRI (New England Biolabs, Estados Unidos) a 37°C durante 60' y visualizado en un gel de agarosa al 2%. Los plásmidos purificados que contuvieron inserto se secuenciaron con los primers de la RT-PCR (5'- GACGGACGGACAGACAGATGGACA-3' y 5'-GCCCCTCTGGAACCCCTCAG-3'), otros dos primers internos (5'-GAGATGTTGTGGCTGCTGAA-3' y 5'-GTTGGCCCTCAGTGTGAGAT-3') y un primer fuera de la secuencia (5'-CAGGAAACAGCTATGAC-3').

8.Extracción y cuantificación de proteína

En el caso de la extracción de proteínas de la línea celular de CU T24, en el momento de recoger las células, el medio se desechó y las células se lavaron dos veces con PBS (0.01 M buffer fosfato, 0.137 M NaCl y 0.0027 M KCl). Posteriormente, las células se rasparon de la superficie

de cultivo, se recogieron y se centrifugaron a 200g durante 5'. El sobrenadante se desechó.

Tanto las células de la línea celular de CU T24 como las muestras de tumores de vejiga humanos, cornea y pterigion se sumergieron en una solución de lisis (25 mM Tris-HCl, 2 mM EDTA y 2% SDS). Posteriormente, las muestras se sonicaron a amplitud ultrasónica del 50% y de un ciclo de 1 durante 10 segundos 3 veces. Las muestras entonces se sometieron a 3 ciclos de congelación y descongelación para repetir finalmente de nuevo las 3 repeticiones de sonicación. El sobrenadante de las muestras se recogió tras haber sido centrifugadas a 28000 rpm durante 10'.

La cuantificación de proteína se realizó con el kit BCA Protein Assay Reagent Kit (ThermoFisher Scientific, Estados Unidos) basado en la detección colorimétrica de la reducción de Cu^{+2} a Cu^{+1} por parte de las proteínas en medios alcalinos. Ratio 50:1 de Reactivo A:Reactivo B se mezcló en ratio 1:20 muestra:reactivo AB estando la muestra diluida 1:10 con 0.1% SDS. La mezcla se incubó a temperatura ambiente durante 2h y la cuantificación se determinó leyendo la absorbancia a 560 nm en relación a los resultados obtenidos para una curva patrón realizada con BSA. La concentración de las muestras se ajustó con solución de lisis y con una solución de carga (Tris-HCl 225 mM pH: 7, 50% glicerol, 5% SDS, 0.05% de azul de bromofenol y 250 mM de DTT), se hirvieron durante 5' y se almacenaron a -80°C hasta el momento del experimento.

9. Western-Blot

Todas las muestras analizadas por WB se corrieron en Mini-PROTEAN® Precast Gels del 10% (Bio-Rad, Estados Unidos). En el caso de las transfecciones, se cargaron 5 o 15 μg de muestra mientras que en el caso de las muestras de tumores de vejiga, cornea y pterigion humanos se

cargaron 30 µg. La transferencia se realizó sobre Trans-blot® Turbo™ Mini Nitrocellulose Transfer Packs 0.2 µm (Bio-Rad, Estados Unidos) en un Trans-blot Turbo Transfer System (Bio-Rad, Estados Unidos) a 1.3 A y 25 V. Como control de carga para alguno de los WB y como control de transferencia, la membrana de nitrocelulosa se tiñó con Ponceau S solution (Sigma, Estados Unidos) durante 10' y posteriormente destañada con PBS-Tween 0.1%. El bloqueo de las membranas se realizó con una solución de PBS-Tween 0.1% y 3% de leche descremada. Tanto los anticuerpos primarios como los anticuerpos secundarios se prepararon en PBS 1% de leche descremada. Los anticuerpos primarios (Tabla 1) siempre se incubaron a 4°C durante la noche y los anticuerpos secundarios ECL Anti-ratón o conejo IgG unidos a HRP (Amershan Biosciences, Estados Unidos) siempre se incubaron durante 60' a temperatura ambiente y en agitación a una dilución 1:2000. La detección se realizó con SuperSignal™ West Femto Maximum Sensibility Substrate (ThermoFisher Scientific, Estados Unidos) que es un sustrato ECL para HRP.

PROTEÍNA	NOMBRE ANTICUERPO	ESPECIE	DILUCIÓN	PROVEEDOR
UPK3a	AU1	Ratón	1:100	Cedido por Dr. Sun (Liang et al. 2001)
UPK3b	UPK3B	Conejo	1:500	Atlas Antibodies (Suecia)
UPK3c	P3R2	Conejo	1:1000	Encargado diseñar a Biomedal (Sevilla)
UPK3c	UPK3BL	Conejo	1:1000	Origene (Estados Unidos)
FGFR3	FGFR3 (c51F2)	Conejo	1:1000	Cell Signaling Technologies (Estados Unidos)
p130	p130 Cas (E1L9H)	Conejo	1:1000	Cell Signaling Technologies (Estados Unidos)
Rb	Rb (4H1)	Ratón	1:500	Cell Signaling Technologies (Estados Unidos)

Tabla 1: Características, condiciones y proveedor de los anticuerpos utilizados para inmunodetección.

10. Transfección transitoria de UPK3c

UPK3c se transfectó de manera transitoria en la línea celular de CU T24 dado que por RT-PCR se observó que no expresa UPK3c de forma endógena (datos no mostrados). El plásmido de expresión utilizado para transfectar la línea celular T24 fue el plásmido CS2+ (pCS2+). De un clon con la construcción pCR® II-TOPO®-UPK3c con la secuencia correcta tras su secuenciación, se extrajo el transcrito de UPK3c con la enzima de restricción EcoRI (New England Biolabs, Estados Unidos) que posteriormente se extrajo del gel de agarosa con el kit GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare, Reino Unido) basado en un sistema de purificación a través de la adsorción y desorción del producto de PCR en columnas de sílice tras la disolución del fragmento de agarosa que lo contiene. El pCS2+ que contiene un sitio de restricción para EcoRI (New England Biolabs, Estados Unidos) se linearizó con esta enzima de restricción y posteriormente se desfosforiló para eliminar los fosfatos de los extremos 5' del plásmido y así evitar la religación y permitir la acción de las ligasas. Aproximadamente 400 ng de pCS2+, se incubó con 50 mM de Bis-Tris-propano-HCl, 1 mM MgCl₂, 0.1 mM ZnCl₂ y 5 U de fosfatasa Antartic (New England Biolabs, Estados Unidos) durante 15' a 37°C y posteriormente la enzima se inactivó durante 5' a 70°C. La ligación se realizó con la T4 DNA ligase (New England Biolabs, Estados Unidos) en un ratio molar Vector:Inserto 1:3 y con una concentración total de vector e inserto de 4,4 µg/ml incubándolo con 50 mM Tris-HCl, 10 mM MgCl₂, 1mM ATP, 10 mM DTT y 400 U de T4 DNA ligasa durante 5' a 25°C y posteriormente inactivada a 65°C durante 10'. 4 µl del producto de la ligación posteriormente se transformó en One Shot® Chemically Competent *E. coli* y purificado como ha sido explicado anteriormente. La presencia del inserto se comprobó también con la enzima de restricción EcoRI (New England Biolabs, Estados Unidos) como ha sido explicado anteriormente y su correcta orientación se estableció con los primers del plásmido Promotor SP6 (5'-CCCAAGCTTGATTTAGGTGAC-3') y el primer T7 (5'-AATACGACTCACTATAG-3') a través de su secuenciación.

Para llevar a cabo la transfección, la línea celular de CU T24 se sembró a una densidad de 150 células/mm² en medio Mc Coy's (Sigma, Estados Unidos), suplementado con 1.5 mM de L-Glutamina (Sigma, Estados Unidos) y 10% de Suero Fetal Bovino (Sigma, Estados Unidos), un cocktail de antibióticos (100 U de Penicilina , 25 mg/ml de Estreptomina) (Sigma, Estados Unidos) y 0.1 mg/ml de Plasmocin™ (Invivogen, Estados Unidos) y cuando alcanzaron una confluencia del 80-90% se transfectaron. Dos horas antes de realizar la transfección, el medio se cambió por medio nuevo sin antibióticos ni Plasmocin™. Para transfectar se utilizó Lipofectamine® LTX con Plus™ Reagent (Invitrogen, Estados Unidos). Plus™ Reagent con pCS2+-UPK3c se mezclaron en un ratio 1:1 de µl:µg en Opti-MEM® (Thermo Fisher Scientific, Estados Unidos) en tubos de poliestireno e incubados durante 10' a temperatura ambiente. Posteriormente, esta solución de Opti-MEM® se mezcló con otra solución de Opti-MEM® (1:1) que contenía Lipofectamine® LTX en un ratio final 1:4 o 1:2.5, plásmido:Lipofectamine® e incubada durante 30' a temperatura ambiente mezclando suavemente cada 10'. Finalmente el volumen final de Opti-MEM que es igual a 1:10 del volumen de medio en la placa se añadió al medio de cultivo de las células e incubado durante 6 horas. Después de la incubación, las células se lavaron dos veces con medio de cultivo Mc Coy's con antibiótico y Plasmocin™ y se le añadió medio fresco de nuevo dejándolas en incubación 24 y 48 horas. Los resultados de la transfección se analizaron por WB.

11. Deglicosilación de muestras

Los extractos proteicos de la línea celular de CU T24, como los de los tumores de vejiga humanos, cornea y pterigion se deglicosilaron con dos N-glicosidasas. Aproximadamente 30 µg de extracto proteico se incubó con un buffer desnaturante de glicoproteínas (5%SDS y 0.4 M DTT)

durante 10' a 100°C. Posteriormente el volumen de la reacción se dobló con 400 U de enzima y con 0.5M de citrato de sodio (pH: 5.5) en el caso de EndoH (New England Biolabs, Estados Unidos) o de 0.5 M fosfato de sodio (pH: 7.5) y 1% de NP40 en el caso de PNGase F (New England Biolabs, Estados Unidos). Ambas reacciones con la enzima se incubaron a 37°C durante 60' y se almacenaron a -80°C hasta el momento del experimento. A los controles para cada muestra se les expuso a las mismas condiciones que al resto pero sin contener enzima. La sensibilidad a ambas N-glicosidasas se examinaron a través de WB.

12. Immunolocalización de uroplaquinas en el urotelio de ratón

Las secciones de parafina de vejiga de ratón sano y de los modelos transgénicos de ratón HRAS y SV40T fueron cedidos por el Dr. Sun. La desparafinación y rehidratación de las secciones consistió en la incubación dos veces en xileno durante 15', otras dos en 100% etanol durante 5' y luego una vez durante 5' en un gradiente de distintos porcentajes de etanol (95, 80, 75, 50 y 30). La hidratación final se realizó en agua dos veces durante 3'. El desenmascaramiento y la recuperación del epítipo se realizó con una solución tampón de citrato a pH 6 (Sigma, Estados Unidos) en un microondas durante 5' a máxima potencia y otros 2' a potencia baja. Tras enfriar las muestras sumergidas en la solución tampón de citrato durante 60' a temperatura ambiente, las secciones se lavaron con PBS (pH 7.4) en agitación. El bloqueo de las secciones se realizó con una solución de PBS (pH 7.4) al 3% de BSA durante 30' a temperatura ambiente. Las secciones se incubaron con anticuerpos primarios (Tabla 2) a su respectiva concentración en una solución de PBS (pH 7.4) al 3% de BSA a 4°C durante toda la noche y los anticuerpos secundarios Alexa Fluor® 488, 594 y 633 de conejo, ratón y cabra respectivamente (ThermoFisher, Estados Unidos) se incubaron a temperatura ambiente durante 60' en oscuridad a

una dilución 1:100 de PBS (pH 7.4) al 3% de BSA. La tinción nuclear también realizada en oscuridad se llevó a cabo a una dilución 1:1000 durante 60'' en una solución de PBS (pH 7.4). Las secciones se cubrieron con el medio de montaje acuoso Fluoro-Gel (Electron Microscopy Sciences, Estados Unidos). Las imágenes se adquirieron con un microscopio confocal invertido Zeiss LSM₇₀₀ y las condiciones de adquisición de imágenes se llevaron a cabo con un objetivo de 100x y 1.4NA, el tamaño del pinhole fue de 0.7 μ M y el tamaño del pixel de 40 nm. Los fluoróforos para los anticuerpos secundarios emitieron a 520, 617 y 668 nm.

PROTEÍNA	NOMBRE ANTICUERPO	ESPECIE	DILUCIÓN	PROVEEDOR
UPK1a	2P22	Ratón	1:100	Cedido por Dr. Sun
UPK1b	7472	Conejo	1:100	Cedido por Dr. Sun
UPK2	N18	Cabra	1:50	Santa Cruz Biotechnology Inc. (Estados Unidos)
UPK2	S3045	Conejo	1:200	Cedido por Dr. Sun
UPK3a	M17	Cabra	1:50	Santa Cruz Biotechnology Inc. (Estados Unidos)
UPK3a	AU1	Ratón	1:100	Cedido por Dr. Sun
UPK3b	c362	Ratón	1:50	Cedido por Dr. Sun
UPK3c	m3c	Conejo	1:50	Cedido por Dr. Sun

Tabla 2: Características, condiciones y proveedor de los anticuerpos utilizados para inmunolocalización.

RESULTADOS

UNIVERSITAT ROVIRA I VIRGILI

EVOLUTION AND MOLECULAR CHARACTERIZATION OF UROPLAKIN 3C, A NOVEL HUMAN TETRASPANIN ASSOCIATED UROPLAKIN

Javier Ugarte Chicote

ORIGEN Y EVOLUCIÓN DE LAS UROPLAQUINAS. IDENTIFICACIÓN DE TRES NUEVAS UROPLAQUINAS: UPK3c, UPK3d Y UPK2b.

1. Búsqueda de secuencias en bases de datos de DNA y proteína.

La existencia de secuencias de las dos familias de uroplaquinas UPK1a/1b y UPK2/3 en vertebrados basales (peces óseos y cartilagosos, anfibios, reptiles/aves) y su ausencia en tunicados, que son los organismos evolutivamente más cercanos a los vertebrados, ha sido descrita en un estudio anterior en el que participó un investigador del grupo en el que he desarrollado mi trabajo de tesis (García-Espana *et al.* 2006). Los tunicados son animales marinos y aunque la mayoría de adultos son sésiles, las larvas son móviles y poseen una notocorda (médula espinal primitiva) que los clasifica como cordados, junto a los vertebrados y a los cefalocordados (Delsuc *et al.* 2006). La presencia de uroplaquinas en peces cartilagosos y su ausencia en tunicados es indicativa de que las uroplaquinas debieron aparecer durante las primeras fases de la evolución de los vertebrados. En ese estudio se clasificaron las nuevas uroplaquinas como pertenecientes a los distintos grupos ya conocidos (1a, 1b, 2, 3a y 3b) excepto las similares a UPK3 de peces, que se clasificaron como proto-UPK3s debido a la divergencia que presentaban con las UPK3a/3b de tetrápodos (anfibios, reptiles/aves y mamíferos) (García-Espana *et al.* 2006)

En el presente estudio hemos reexaminado la evolución y profundizado en el origen de las familias de las uroplaquinas debido al reciente aumento en el número de genomas parcial o totalmente secuenciados de cordados. Sobre todo, ha contribuido a ello, la disponibilidad de los genomas de: lamprea (*Petromizon marinus*), un pez sin mandíbula que puede considerarse como un fósil viviente de los vertebrados más basales (agnatos) (Kuratani 2012; Smith *et al.* 2012); catán pinto (*Lepisosteus oculatus*), un pez óseo primitivo (Amores *et al.*

2011); y el de un cordado basal (cefalocordado), lanceta o anfióxico (*Branchiostoma floridae*) (DeSalle *et al.* 2014; Putnam *et al.* 2008). Buscando en bases de datos de DNA y proteína, tal como se describe en materiales y métodos, mediante la utilización de secuencias ya descritas, se identificaron nuevas secuencias de uroplaquinas en vertebrados, incluyendo la lamprea, y no encontrándose uroplaquinas en cordados no vertebrados ni en organismos más antiguos. Las secuencias de proteína y DNA de las uroplaquinas de organismos representativos de los diferentes grupos de vertebrados mamíferos (placentarios, marsupiales y monotrema), reptiles/aves, anfibios y peces se enviaron a nuestro colaborador, el Dr. Rob DeSalle del *Natural History Museum* (Nueva York), para su estudio filogenético utilizando métodos de inferencia Bayesiana (Figura 6), máxima parsimonia, y máxima verisimilitud (Figura Suplementaria 3 y 4 de DeSalle *et al.* 2014).

2.Relaciones evolutivas de las uroplaquinas de la familia tetraspanina (UPK1a/1b)

Las tetraspaninas (tetraspan o TM4SF) constituyen una extensa superfamilia de proteínas con más de 33 miembros en humanos y con representantes en animales, hongos, plantas y protistas (eucariotas unicelulares) (García-España *et al.* 2008, Huang *et al.* 2005). Dentro de esta superfamilia, las uroplaquinas tetraspaninas (UPK1a/1b) constituyen un grupo bien definido con dos ramas que se corresponden con las secuencias similares a UPK1a y UPK1b (DeSalle *et al.* 2014; García-España *et al.* 2008; Huang *et al.* 2005; Hemler 2003; Boucheix *et al.* 2001). Al analizar filogenéticamente las uroplaquinas de la familia UPK1a/1b observamos que sus secuencias se mantienen claramente diferenciadas en

únicamente dos grupos, UPK1a y UPK1b, incluso con las secuencias de lamprea (Figura 6a y Figura 7).

La reconstrucción evolutiva de las UPK1a/1b concuerda con la historia de la diversificación de los grupos de vertebrados, excepto en el caso de los peces cartilagosos, que en el árbol filogenético se encuentran más cercanos a los tetrápodos (anfibios, reptiles/aves y mamíferos) que a los peces óseos (Figura 6a). Esto podría deberse a la gran diversificación y velocidad evolutiva que han experimentado los peces óseos en comparación con los peces cartilagosos y tetrápodos (Ravi *et al.* 2008). La diversificación de estas proteínas puede entenderse fácilmente como consecuencia de una duplicación génica en el ancestro común de vertebrados que dio lugar a UPK1a y UPK1b (Figura 7).

3.Relaciones evolutivas de las uroplaquinas de la familia UPK2/3: identificación de tres nuevas uroplakinas UPK3c, UPK3d y UPK2b

Los análisis de las uroplaquinas de la familia UPK2/3 revelaron la existencia de tres nuevos parálogos que no habían sido descritos con anterioridad y que hemos denominado UPK3c, UPK3d y UPK2b (Figura 6b).

UPK3c sería el más reciente evolutivamente y el único que se encuentra en mamíferos, incluyendo humanos. Otro de los nuevos parálogos, UPK3d, está relacionado con UPK3b y solo existe en peces (Figura 6b). UPK3d es la denominada “UPK3a-like protein” que recientemente se ha vinculado con la morfogénesis del conducto pronéfrico de pez cebra (*Danio rerio*) (Mitra *et al.* 2011).

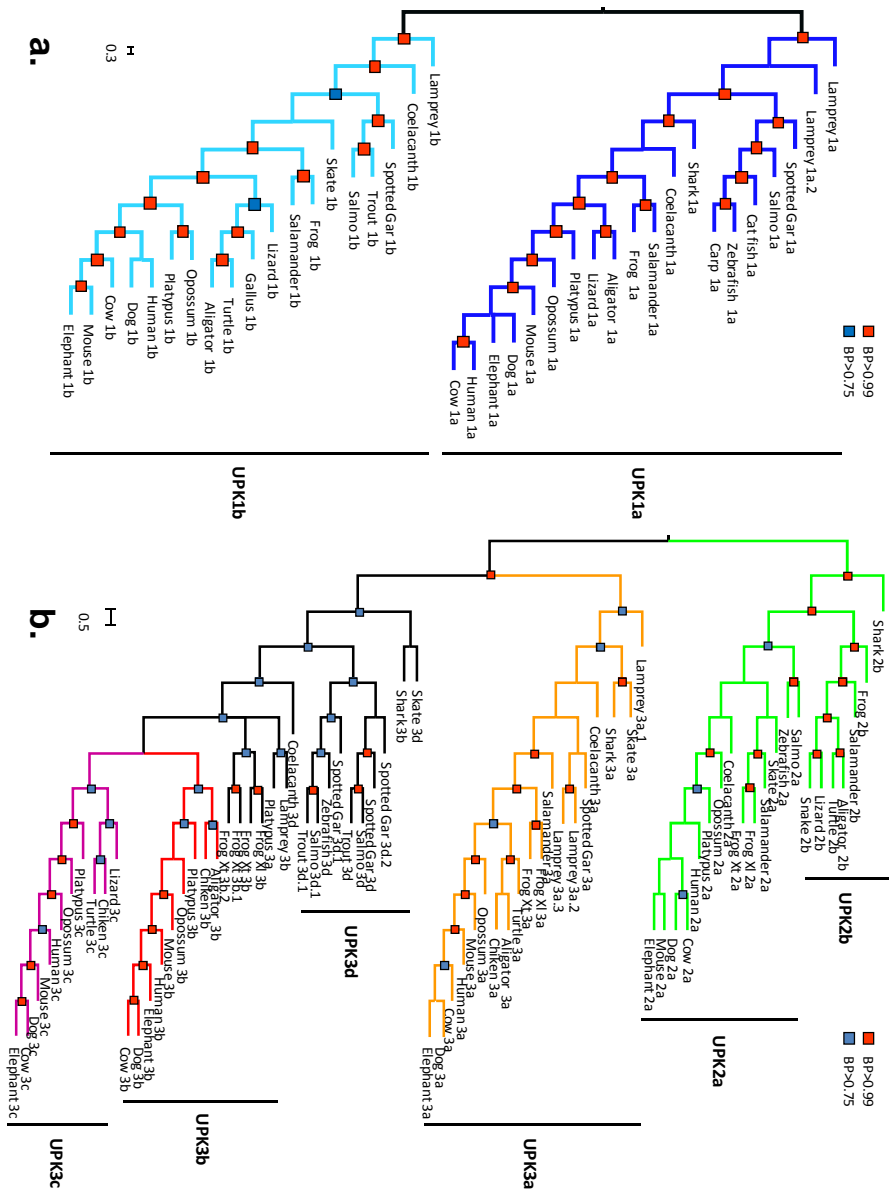


Figura 6: Relación evolutiva de las uroplaquinas ya conocidas y los nuevos miembros de la familia de las uroplaquinas. Árboles filogenéticos basándose en inferencia Bayesiana mostrando la relación evolutiva de (a.) las UPKs tetraspaninas (UPK1a y UPK1b), y de (b) las UPK2/3. Los principales grupos de uroplaquinas han sido indicados a través de líneas verticales. Cuadrados rojos BP>0.99. Cuadrados azules BP>0.75. Los colores en las ramas de los árboles delimitan la extensión de los distintos grupos de uroplaquinas. BP: Bayesian posterior

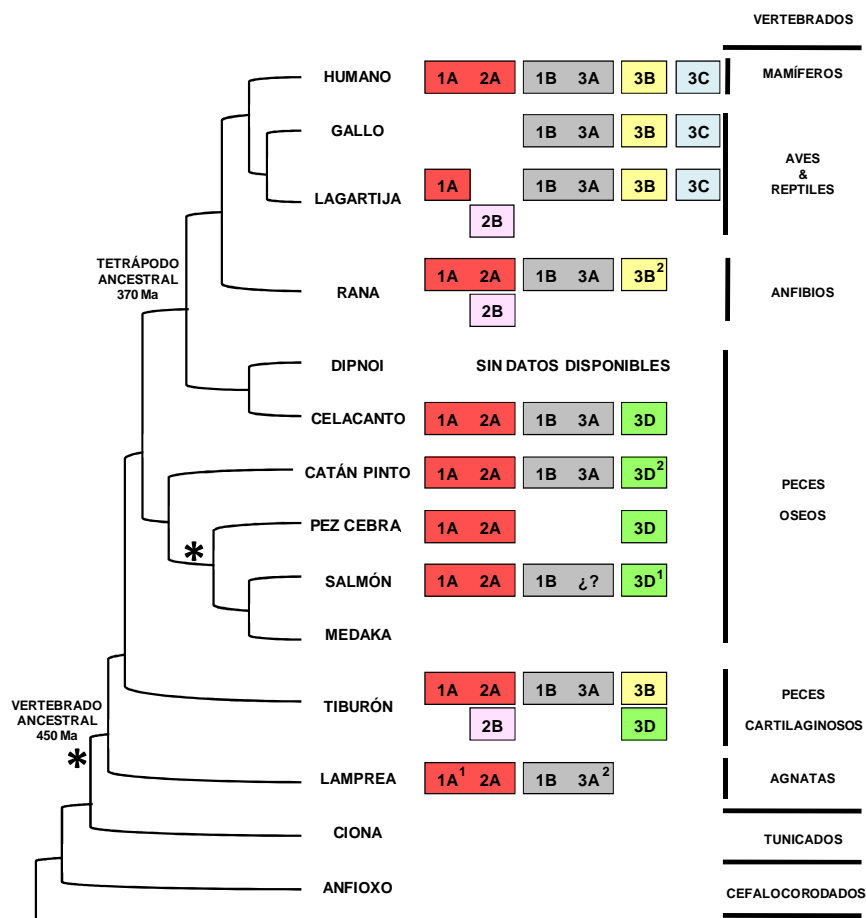


Figura 2: Árbol filogenético simplificado de la presencia y ausencia de uroplaquinas en los principales grupos de cordados. Esquema de la distribución de las uroplaquinas desde su origen en el ancestro común de vertebrados hasta humanos. 1A (UPK1a), 1B (UPK1b), 2A (UPK2a), 2B (UPK2b), 3A (UPK3a), 3B (UPK3b), 3c (UPK3c), 3D (UPK3d). Los superíndices 1 y 2 indican la existencia de 2 o 3 formas respectivamente de ese miembro de la familia de uroplaquinas. La falta de suficiente información para determinar la existencia o no de uroplaquinas ha sido indicada como ¿? y los * representan aquellos momentos en la historia donde tuvo lugar una duplicación completa del genoma. El árbol filogenético simplificado muestra los principales grupos taxonómicos con una especie como ejemplo representativo.

El último de los nuevos parálogos está relacionado con UPK2 y se le ha denominado UPK2b para distinguirlo de UPK2 que ha pasado a llamarse UPK2a. Solo se encuentra en aves/reptiles, anfibios y peces cartilagosos, a diferencia de UPK2a que tiene una distribución más amplia (Figura 6b y Figura 7).

Además de estos tres nuevos parálogos, en el caso específico de la lamprea, el catán pinto y la rana (*Xenopus tropicalis*) es posible encontrar tres copias de UPK3a, UPK3b y UPK3d respectivamente (Figura 7).

Al igual que en el caso de las UPK1a/1bs, la reconstrucción evolutiva de UPK2/3 está de acuerdo con la historia de la diversificación de los grupos de vertebrados menos por algunas excepciones (Figura 6b y Figura 7). Por ejemplo, la UPK2a, UPK3a.1 y UPK3a.2 en lamprea y la UPK3a tanto en celacanto como en ornitorrinco (*Ornithorhynchus anatinus*) no se encuentran en los grupos que le corresponderían. Esta incongruencia podría ser debida a "Long Branch Attraction", por la cual, secuencias que han evolucionado rápidamente se agrupan de forma conjunta indistintamente de su correcta relación (Felsenstein 2004).

La evolución de UPK2/3 es más compleja que la de UPK1a/UPK1b y varias rondas de duplicación serían necesarias para poder explicar su distribución en los distintos taxones animales (Conrad *et al.* 2007; Kleinjan *et al.* 2008; Rastogi *et al.* 2005; Dehal y Boore 2005; Lundin *et al.* 2003; Mc Lysaght *et al.* 2002) (Figura 7). En principio, podría pensarse que una duplicación habría coincidido con la primera duplicación del genoma que ocurrió en el ancestro común de vertebrados (Ohno *et al.* 1999) dando lugar a la separación de UPK2 y UPK3.

En el caso de UPK2, una duplicación adicional debió ocurrir que dio lugar a UPK2a y UPK2b. Esta duplicación podría haberse producido en el ancestro común de peces cartilagosos y óseos dado que la UPK2b más antigua encontrada es en peces cartilagosos (Figura 7). Sin embargo, dado que de vertebrados más primitivos solo se dispone del genoma de

lamprea, no se puede descartar la posibilidad de que UPK2a fuese duplicada ya en el ancestro común de vertebrados con la posterior pérdida de UPK2b en lamprea (Figure 7). En cuanto a UPK3s, a rasgos generales su evolución necesitaría al menos cuatro rondas de duplicación para poder llegar a explicar la actual distribución de los genes en el genoma de vertebrados. Sin embargo, no se puede precisar con exactitud cómo se produjeron estas duplicaciones excepto en el caso de UPK3c dado que su origen se explicaría como consecuencia de una duplicación del gen *UPK3b* en el ancestro común de reptiles/aves y mamíferos. Esto es demostrado a través de la sintenia o conservación de la localización de los genes en este caso adyacentes a *UPK3b* en los cromosomas de anfibios, reptiles/aves y mamíferos (Figura 8).

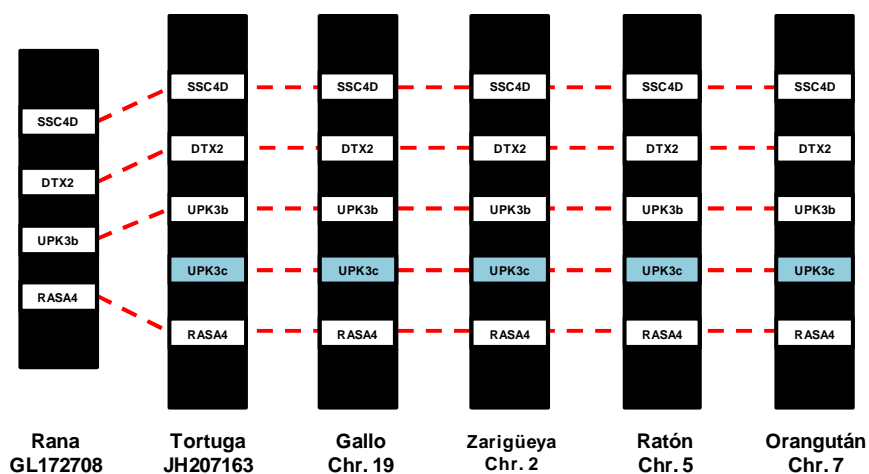


Figura 8: UPK3c se originó consecuencia de una duplicación del gen *UPK3b*. Análisis de sintenia donde la colocación preservada de genes adyacentes al gen *UPK3b* en cromosomas de diferentes especies muestra que el gen *UPK3c* se originó en el ancestro común de reptiles/aves y mamíferos a partir de una duplicación del gen de *UPK3b*.

4. Origen de las uroplaquinas en el ancestro común de los vertebrados y desaparición específica en los peces óseos modernos.

Las secuencias más antiguas relacionadas con las uroplaquinas se encontraron en lamprea, un vertebrado basal y no se detectaron en organismos más antiguos: tunicados y cefalocordados, equinodermos, insectos, nematodos, cnidarios, (medusas) o esponjas. Esta distribución sugiere que la familia de las uroplaquinas se originó hace 500 millones de años en el ancestro común de los vertebrados cuando estos se separaron de tunicados (urocordados) y cefalocordados (Holland *et al.* 2015).

Dado que las uroplaquinas de la familia UPK1a/1b pertenecen a la extensa y antigua familia de proteínas tetraspaninas probablemente se originaron a partir de la duplicación de un gen tetraspanina ancestral que no hemos podido identificar al no existir conservación de sintenia entre UPK1a o UPK1b en el genoma de los cordados basales. En el caso de la familia UPK2/3 no se conoce que estén relacionadas con otra familia de proteínas.

5. La conservación de la pauta de exones codificantes e intrones y de cisteínas extracelulares en UPK2s y UPK3s refuerzan su origen evolutivo común.

A pesar de que la similitud entre las secuencia de las distintas proteínas UPK2/3 es baja (12-44%), al alinear sus secuencias se observa que comparten un patrón génico idéntico. Todas, excepto las UPK2s, que carecen del último exón, están codificadas en seis exones con 5 intrones (1-5) en fases 1:1:2:1:2 (Figura 9). Los intrones en fase 1 y 2 interrumpen

un codón entre la primera y segunda base o entre la segunda y la tercera base respectivamente.

Además en el alineamiento de estas secuencias se puede apreciar la conservación de cuatro cisteínas (C_1 - C_4) en la parte extracelular de la proteína con alguna excepción dado que UPK3c ha perdido las C_1 y C_2 en las secuencias de mamíferos y UPK2a no tiene las C_3 y C_4 (Figura 9 y Figura Suplementaria 1). Estas observaciones reforzarán los resultados de los análisis filogenéticos del origen común de las UPK2/3

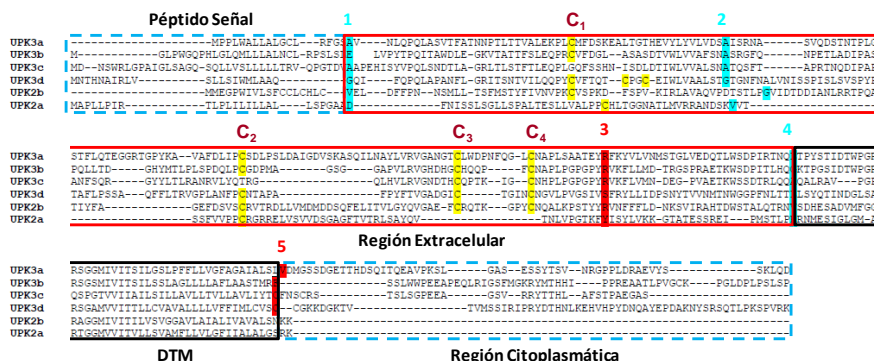


Figura 9: Conservación de los residuos de cisteína y la pauta de intrón exón en UPK2/3s. Alineamiento de las secuencias humanas de UPK2a, UPK3a, UPK3b y UPK3c, UPK2b de lagarto y UPK3d de pez zebra utilizando el servidor MAFFT con los parámetros estándar y ajustadas manualmente para poder alinear la posición del intrón 1. Los intrones del dominio urofibrina 1-4 (ver mas adelante y Figura 10) como el intrón 5 del DTM están señalados en la parte superior de la secuencia en color azul o rojo atendiendo si la fase es 1 o 2 respectivamente. De esta misma manera, los aa resultantes de un codón entre dos exones están resaltados con los mismos colores. Las cisteínas están resaltadas de color amarillo y numeradas del 1 al 4. UPK3d tiene dos cisteínas adicionales y los pares de cisteínas C_1 - C_2 y C_3 - C_4 no existen en UPK3c y UPK2a respectivamente. El péptido señal y la región citoplasmática están delimitados con una línea de color azul discontinua, el DTM está delimitado con una línea de color negro y la región extracelular con una línea de color rojo. Aunque la similitud entre las distintas UPK2/3 es baja, las fases de los intrones y los residuos de cisteína están muy conservados sugiriendo entonces un origen evolutivo común. Identidad de aa, idéntico (*); muy similar (:); poco similar (.) y sin similitud ().

RELACIÓN EVOLUTIVA ENTRE LAS UROPLAQUINAS UPK2/3 Y LOS RECEPTORES SUBTIPO 3 DE TIROSINA FOSFATASAS (R3 PTPRS). CARACTERIZACIÓN DEL DOMINIO UROFIBRINA.

1. Identificación de un fragmento de proteína similar a las uroplaquinas UPK2/3 en los receptores de proteína tirosina fosfatasa (R3 PTPRs)

Al contrario que las uroplaquinas UPK1a/1b que pertenecen a la familia de las tetraspaninas, las uroplaquinas UPK2/3 no se habían podido relacionar con otras familias de proteínas. Por este motivo utilizando como cebo diversas secuencias de UPK2/3s se iniciaron una serie de búsquedas en bases de datos de DNA y proteína.

En una de estas búsquedas, se observó que la proteína UPK3a humana reconocía, aunque con baja significancia ($E = 0.066$), un fragmento de 58 aa de la secuencia de proteína del receptor tirosina fosfatasa Q (PTPRQ) humana (Figura 10a).

Al alinear las secuencias de estas dos proteínas, la similitud entre ambas pudo extenderse a la región que abarca toda la secuencia extracelular y transmembrana de UPK3a y que corresponde en PTPRQ a la región entre los aa 1744 y 1928 (Figura 10b). Ambas secuencias comparten un 16% de aa idénticos, un 21% muy similares y las cuatro cisteínas (C_1 - C_4) identificadas anteriormente (Figura 9) como características de la familia UPK2/3 (Figura 10b).

La fosfatasa PTPRQ pertenece al subtipo 3 de receptores tirosina fosfatasa (R3 PTPRs). Los miembros de este subtipo 3 (R3 PTPRs) se caracterizan y distinguen de los otros 8 subtipos de receptores tirosina fosfatasas por tener un solo dominio catalítico y en la parte extracelular únicamente dominios fibronectina tipo III (FN3) (Andersen *et al.* 2001). Al igual que las UPK2/3, los R3 PTPRs son proteínas integrales de membrana

tipo I pero, a diferencia de estas (184-287 aa), son de gran tamaño (1216-2299 aa) (Figura 11). En vertebrados, existen cinco miembros de R3 PTPRs: PTPRB (VE-PTP), PTPRJ (DEP-1), PTPRH (SAP-1), PTPRO (GLEPP1) y PTPRQ (PTPS31) (Hendriks *et al.* 2013; van Eekelen *et al.* 2010; Andersen *et al.* 2001). Al comparar sus secuencias, la presencia de la región similar a UPK3a se observó también en PTPRB y PTPRJ pero no en PTPRO y PTPRH (Figura 11). Además, esta región, al igual que en las uroplaquinas UPK2/3, está codificada por cuatro exones comprendidos entre 5 intrones (1-5) en fases 1:1:2:1:2 (Figura 9 y Figura 11). Al igual que PTPRQ y las proteínas UPK2/3, las cisteínas C₁-C₄ también están en PTPRB (Figura 11) e incluso el tamaño de los exones es similar con los de UPK3a (Figura 11 y Figura Suplementaria 3).

- a.** PREDICTED: phosphatidylinositol phosphatase PTPRQ isoform X1 [Homo sapiens]
 Sequence ID: [ref|XP_011536592.1|](http://ref.xp_011536592.1) Length: 2321 Number of Matches: 1

Range 1: 1838 to 1894

Alignment statistics for match #1

Score	Expect	Method	Identities	Positives	Gaps
37.7 bits(86)	0.067	Compositional matrix adjust.	17/59(29%)	29/59(49%)	2/59(3%)

UPK3a	130	NAYLVRVGGANGTCLWDPNFQGLCNAPLSAATEYRFKYVLVNMSTGLVEDQTLWSDPIRT	188
		N + +GA+ C+ N +CN PL +Y FK+ N+ + +SDP++T	
PTPRQ	1838	NEEIYIIGADNACMIPGNEDKICNGPLPKPKQYLFKFRATNIMGQFTDSD--YSDPVKT	1894

b.

UPK3a	-AVNLQPQLASVTFATNNPTLT--TVALEKPLCMFDSKEALTGTHEVYLYVLVDSAISRNASVQDSTNTPLGSTFLOTEG
PTPRQ	APARPKTKPTPIYDATGKLLVTSTITIRMPICYYSDDHGPIKN-----VQVLVTEGTGAQHDGNVTKWYDAYFNKAR
	... : : : : : * : : : * : : : * : : : * : : : * : : : * : : : * : : : * : : : * : : : * : : : * : : : *
UPK3a	GRTGPYKAVAFDLIFCSDDLPSLDAIGDVSKASQILNAYLVRVGGANGTCLWDPNFQGLCNAPLSAATEYRFKYVLVNMSTG
PTPRQ	---PYFTNEGFPNFPCTEG-----KTKFSGNNEEIIYIIGADNACMIPGNEDKICNGPLPKPKQYLFKFRATNIMGQ
	: : : * : : : * : : : * : : : * : : : * : : : * : : : * : : : * : : : * : : : * : : : * : : : *
UPK3a	LVEDQTLWSDPIRTNQLTPYSTIDTWPGRRSGGMIVITSLGSLPFFLLVGFAGATAISLV
PTPRQ	FTDSD--YSDPVKT-----LG---EGLSERTVBIILSVTLCLLSIILLGTAIFA----
	: : : : : * : : : * : : : * : : : * : : : * : : : * : : : * : : : * : : : * : : : * : : : *

Figura 10: Región similar entre UPK3a y PTPRQ humanas. (a.) Fragmento de PTPRQ humana reconocida por UPK3a cuando esta es utilizada como cebo en la herramienta BLAST contra bases de datos de proteínas humanas. (b.) Alineamiento extendido de las regiones similares de UPK3a (17-235) y PTPRQ (1744-1928) humanas. Los DTM están subrayados en color rojo. Los residuos de cisteína están anotados como C₁-C₄ y resaltados de color amarillo. La identidad de los aa viene determinada como rojos o idénticos (*), verdes o alta similitud (:), azules o baja similitud (.) y negros o ningún tipo de similitud ().

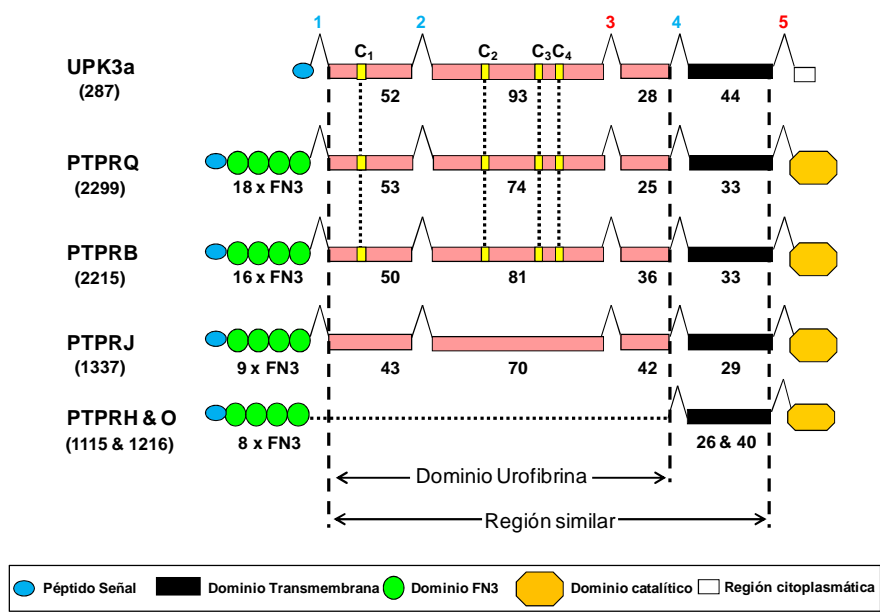


Figura 11: Identificación del dominio urofibrina dentro de la región similar entre uroplaquinas y receptores tirosina fosfatasa subtipo R3 (R3 PTPR) de humanos. Esquema representando los exones (rectángulos de color rosa) que constituyen el dominio urofibrina y con el DTM (rectángulo color negro), la región similar. Los números debajo de los rectángulos indican el número de aa. Los intrones son los números en la parte superior donde rojo y azul denotan fase 1 y fase 2 respectivamente. Aunque solo se represente UPK3a, todas las proteínas UPK2/3 humanas comparten la misma pauta de intrón-exón con la excepción del exón 5 que no existe en las UPK2s (Figura 9). C₁₋₄ indican las cuatro cisteínas de UPK3a que están conservadas en PTPR-Q, -B y -J pero no en -H y -O, la cual también ha perdido la región similar (línea punteada). Ver leyenda para información adicional de péptido señal, DTM, dominio catalítico PTP y dominios FN3 (cifras anotadas debajo indican el número de dominios FN3). El número de aa de cada proteína está reflejado entre paréntesis bajo el nombre de estas.

Por simplificar y para facilitar la lectura hemos llamado dominio “urofibrina” a la región comprendida entre los intrones 1 y 4 (Figura 11) dado que en general, los dominios en las proteínas están delimitados por intrones de la misma fase como en este caso (intrones 1 y 4 son de fase 1), presentan características comunes como el patrón génico y de cisteínas, y

porque, como se verá más adelante, esta región comparte cierta similitud con las uroplaquinas y con los dominios FN3 (Tabla 3).

2.Caracterización del dominio urofibrina en UPK2/3 y R3 RPTPs de vertebrados e invertebrados.

El dominio urofibrina está conservado en todas las proteínas UPK2/3 y es codificado por los exones del 2 al 4 con intrones en fases 1:1:2:1, con la excepción de la ausencia de las C₁-C₂ y C₃-C₄ en UPK3c y UPK2a respectivamente (Figura 9 y Figura Suplementaria 1). Para su caracterización se tuvieron que obtener secuencias de R3 RPTPs buscando en bases de datos genómicas como ha sido descrito en Materiales y Métodos y a través de artículos científicos (Ono *et al.* 1999; Ono-Kpyanagi *et al.* 2000; Byrum *et al.* 2006; van Ekelén *et al.* 2010). Para comprobar que solo pertenecían al subtipo R3, las secuencias debían de ser completas y poder así comprobar que tenían un solo dominio catalítico (PTP) y únicamente dominios fibronectina tipo III (FN3) en la parte extracelular. Las secuencias R3 RPTPs obtenidas (Material Suplementario 1) cubren un amplio rango evolutivo desde vertebrados a esponjas. No se encontraron secuencias de R3 PTPRs en genomas de organismos más antiguos que los animales (metazoos) como monosiga y capsaspora que son los protistas más cercanos al reino animal (King *et al.* 2008; Suga *et al.* 2013). Todas las secuencias de R3 PTPR contienen el dominio urofibrina excepto PTPRO, PTPRH en vertebrados y dos proteínas de invertebrados (ciona 1 y oikopleura 1) (Figura Suplementaria 2).

En el análisis comparativo de estos dominios urofibrina en UPK2/3 y R3 PTPRs (Figura 12 y Figura Suplementaria 3) se puede observar que los intrones localizados en los flancos del dominio (intrones 1 y 4) son de

origen más antiguo, estando ya presentes en esponja (porífera), mientras que los intrones 2 y 3 aparecieron a partir de deuterostomas (hemicordados y cordados) después de la separación de los protostomas (artrópodos, anélidos, moluscos etc) (Figura 12 y Figura Suplementaria 3).

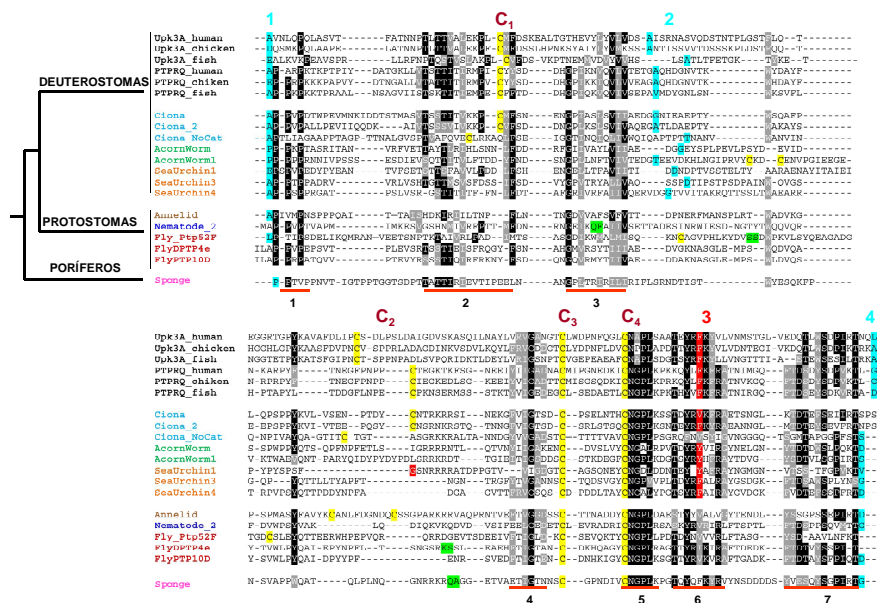


Figura 12: El dominio urofibrina se encuentra conservado a lo largo de la evolución de vertebrados e invertebrados. Alineamiento de los dominios urofibrina de secuencias representativas de UPK3a, PTPRQ y secuencias relacionadas con PTPR de invertebrados utilizando el servidor MAFFT con los parámetros de serie establecidos. El servidor BoxShade se utilizó para resaltar en diferentes tonalidades de grises la homología de los residuos. Los intrones del 1 al 4 están indicados en la parte superior de la secuencia y para las fases de los intrones, los aa están resaltados de azul (fase 1) o de rojo (fase 2). En el caso de fase 0, los dos aa que la flanquean están resaltados de verde. Las C₁-C₄ están indicadas en la parte superior y en el alineamiento, resaltadas de color amarillo. Las zonas de similitud 1-7 están indicadas con barras horizontales de color rojo. Las secuencias de arriba abajo son: vertebrados, negro; tunicados (ciona), azul; hemicordados (acornworm), verde; equinodermos (seaurchin), naranja; lophotrochozoa (annelid), marrón clarito; ecdysozoa (nematode), azul oscuro; insectos (fly), rojo; porífera (sponge), rosa.

Por otro lado, mientras que las C₃-C₄ están presentes en todos los organismos desde esponja hasta mamíferos, C₁-C₂ aparecieron a partir de los tunicados (ciona) (Figura 12 y Figura Suplementaria 3). Además, en el alineamiento de estos dominios urofibrina aparecen 7 secuencias cortas de aa de mayor similitud (1-7 en la Figura 12) y particularmente bien conservado aparece el motivo CNGPL en el fragmento 5 (Figura 12).

UPK2a en mamíferos es sintetizada como un precursor con una pro-secuencia que es cortada justo después de la C₂ por la endoproteasa furina en el motivo diana R_xR/KR para generar una isoforma madura capaz de unirse a UPK1a (Hu *et al.* 2008; Lin *et al.* 1994). Sitios putativos de esta endoproteasa furina son encontrados en posiciones parecidas además de en todos los dominios urofibrina de UPK2a, en algunos de UPK2b y en dominios urofibrina de R3 PTPRs de invertebrados (Figura 13 y Figura Suplementaria 3).

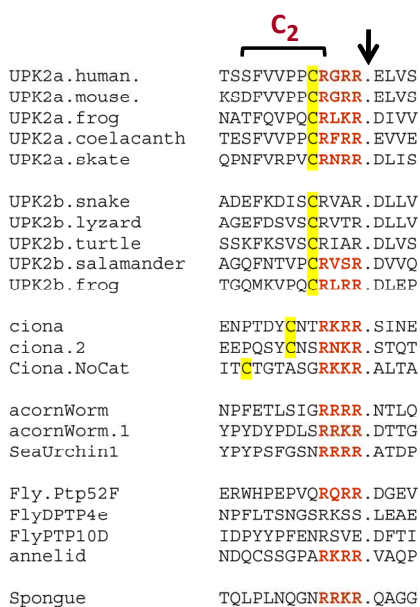


Figura 13: La señal de corte de la endoproteasa furina en UPK2a está conservada en UPK2b de anfibios y en R3 PTPRs de invertebrados. Alineamiento de un fragmento de secuencias representativas de UPK2a, UPK2b y de R3 PTPRs de invertebrados donde está incluida la señal de corte de la endoproteasa furina R_xR/KR. El sitio de corte está localizado posterior a la C₂ (resaltada de amarillo). La señal de corte (letras en rojo) está conservada en todas las UPK2a, las UPK2b de anfibios y en todas las R3 PTPRs de invertebrados excepto en las secuencias de mosca DPTP4e y PTP10D. El lugar donde ocurre el sitio de corte por parte de la endoproteasa está marcado con la flecha vertical de color negro.

Algunos dominios urofibrina de R3 RPTPs son reconocidos como pertenecientes a uroplaquinas y algunos dominios urofibrina de uroplaquinas como dominios FN3 utilizando el servidor SMART para el análisis de dominios FN3 y la base de datos de dominios conservados (CDD) que reconoce uroplaquinas (Tabla 3).

Protein	SMART Dominio FN3	CDD	
		Dominio UPK2/3	
UPK3a humano		cd09970	1.01e-108
UPK3a ratón		cd09970	5.40E-96
UPK3a gallo		cd09970	1.10e-71
UPK3a rana		cd09970	6.07e-56
UPK3a catán pinto		cd09970	1.19e-32
Lamprea		cd09970	1.86e-17
UPK3b humano		cd09969	2.62e-76
UPK3b ratón		cd09969	2.00e-65
UPK3b gallo		cd09968	2.63e-08
UPK3b rana		cd09969	1.12e-65
Lamprea		cd09968	8.29e-17
UPK2a humano	945	cd09967	5.0e-56
UPK2a ratón	129	cd09967	1.91e-57
UPK2a rana	945	cd09967	9.46e-13
UPK2a catán pinto	514	cd09967	9.31e-12
UPK3c humano		cd09969	9.23e-73
PTPRQ ratón		cd09968	2.97e-10
PTPRQ humano	1480	cd09968	2.63e-08
PTPRQ gallo	500	cd09968	1.40e-08
PTPRQ catán pinto	198	cd09966	1.01e-11
PTPRB humano	99		
PTPRB ratón	232		
PTPRB gallo	57		
PTPRB rana	63		
PTPRB catán pinto	474		
PTPRJ humano	30		
PTPRJ ratón	127		
PTPRJ gallo	450		
PTPRJ catán pinto			
Ciona	294	cd09968	3.53e-07
Ciona 2	2040	cd09968	3.27e-09
Ciona no catalítico			
Acornworm	374		
Acornworm 1	12		
Seaurchin 1	164		
Seaurchin 3	127		
Seaurchin 4	135		
Anélido	354	cd09966	9.89e-03
Fly Ptp52F	17		
Fly DPTP4e	5	cd09968	1.84e-03
Fly PTP10D	130	cd09966	8.62e-03
Nemátodo 2		cd09968	3.04e-03
Nemátodo 1			
Nemátodo 3	34		
Sponge	29	cd09966	6.10e-04

Tabla 3: Uroplaquinas y R3 PTPRs son reconocidas como dominios FN3 y como dominios UPK2/3.

Secuencias de R3 PTPRs y de uroplaquinas fueron analizadas en el servidor SMART o en la base de datos CDD para comprobar si eran reconocidas como dominios FN3 o dominios UPK2/3 respectivamente. En caso de ser reconocidas como tales, el valor adjudicado corresponde al e valor al cual son reconocidos como dominios FN3 o dominios UPK2/3. En los dominios UPK2/3 los códigos corresponden a cd09970 y UPK3a, cd09969 y UPK3b, cd09968 y UPK3, cd09967 y UPK2 y cd09966 y UPK2/3.

Finalmente no se encontraron dominios urofibrina fuera de las proteínas R3 RPTPs y UPK2/3 en una búsqueda específica de estos dominios utilizando los motivos más conservados como el C₃X_nC₄NGPL y variaciones de esta secuencia.

3.El dominio urofibrina es probablemente un dominio FN3 divergente

Al existir cierta similitud entre los dominios urofibrina y los dominios FN3, hemos comparado sus secuencias utilizando un alineamiento estructural descrito por Schroder *et al.* 2011.

La estructura tridimensional de los dominios canónicos de FN3, como el paradigmático décimo dominio de la proteína fibronectina, están formados por siete hebras beta antiparalelas denominadas desde el extremo amino terminal A, B, C, C', E F, y G que se distribuyen en dos laminas formando una estructura que permite una gran variabilidad a las cadenas que interconectan las hebras beta (Figura 14a) (Halaby *et al.* 1999; Koide *et al.* 1998). Comparando un grupo de dominios FN3 con estructura terciaria conocida con los dominios urofibrina de las secuencias de PTPRQ y UPK3a en diferentes organismos, se observa que las hebras beta de los dominios FN3 coinciden aproximadamente con las secuencias de aa 2 – 7 que mostraron una mayor conservación en el alineamiento de los dominios urofibrina (Figura 12 y Figura 14b). Estas mismas secuencias cortas 2-7 también coinciden con las predicciones de estructura secundaria de hebra beta (Figura 14b). Sin embargo, existen diferencias con los dominios FN3 canónicos como la pérdida en R3 PTPRs y UPK2/3s de un triptófano (W) muy característico de la hebra beta B de los dominios FN3 (Halaby *et al.* 1999) (Figura 14b), y que la longitud de conexiones entre las

hebras beta es mayor sobre todo entre C y C' en el dominio urofibrina (entre 36 y 53 aa más larga) (Figura 14b).

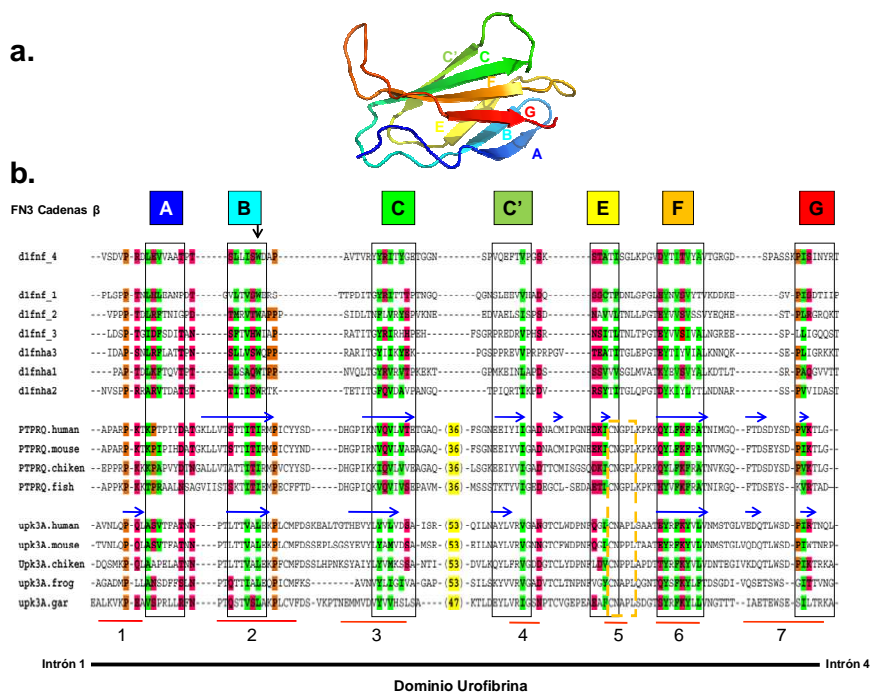


Figura 14: El dominio urofibrina está relacionado estructuralmente con los dominios fibronectina tipo 3 (FN3). (a). Representación 3D del dominio FN3 donde las láminas beta (A-G) están representadas con distintos colores (<http://scop.berkeley.edu/>). (b). Alineamiento de las secuencias de los dominios urofibrina de PTPRQ y UPK3a y dominios FN3 canónicos identificados en la figura con el código del *Protein Data Bank*. Las secuencias FN3 se obtuvieron del análisis realizado por Schroder *et al.* 2011. La conservación de los residuos está resaltada (verde, hidrofóbico; rosa, polar; naranja, prolina) acorde a Schroder *et al.* 2011. D1fnf_4 es el paradigmático décimo dominio FN3 de fibronectina. Las cajas con letras (A-G) representan las hebras beta en la estructura terciaria del décimo dominio FN3 de fibronectina (Koide *et al.* 1998 y Schroder *et al.* 2011). Las flechas azules indican las zonas de predicción de estructura secundaria beta predichas mientras que las secuencias de aa conservadas en los dominios urofibrina en la figura están subrayados con una línea roja. El motivo C₃X_nC₄NGPL está delimitado por una línea discontinua de color naranja. La flecha vertical indica la posición del triptófano (W) altamente conservado en los dominios FN3 canónicos. Los fragmentos entre 36 y 53 aa (resaltado en amarillo entre paréntesis) de los dominios urofibrina y que no se corresponden con los dominios FN3 canónicos se omitieron en la figura.

4.Relación evolutiva entre UPK3a y PTPRQ. Origen de las uroplakinas UPK2/3.

Para estudiar la relación evolutiva se generaron árboles filogenéticos con secuencias de los dominios urofibrina y DTM de PTPRB, PTPRQ, UPK3A, y UPK3B de vertebrados representativos y de todas las secuencias R3 RPTPs de invertebrados (Figura 15 y Figura Suplementaria 3).

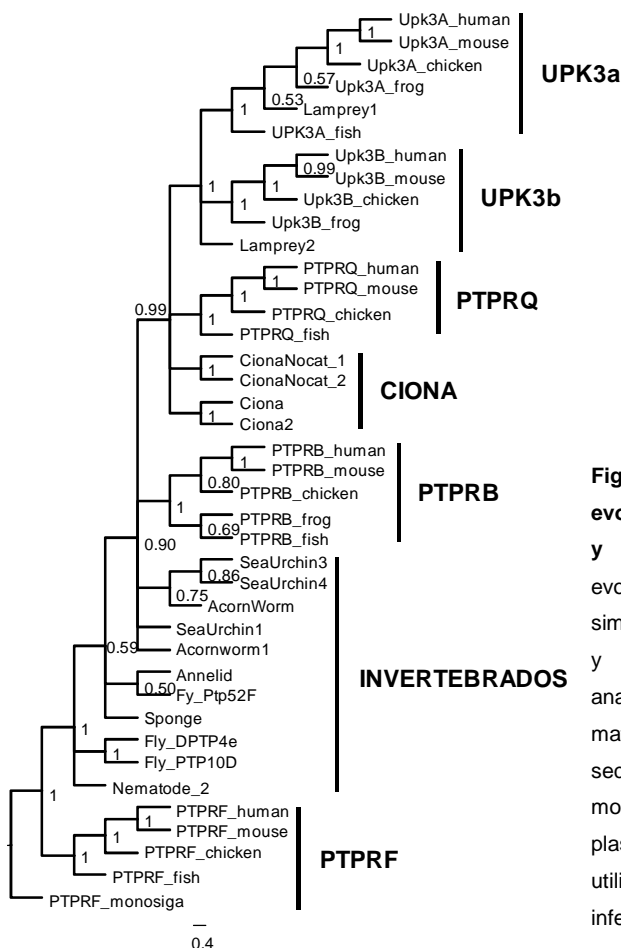


Figura 15: Relación evolutiva entre UPK3as y PTPRQs. La evolución de la región similar (dominio urofibrina y transmembrana) se analizó utilizando una matriz combinada de secuencia de proteína y morfología molecular y plasmada en un árbol utilizando métodos de inferencia Bayesiana.

Las secuencias de aa combinadas con caracteres de morfología molecular (presencia o ausencia de intrones y cisteínas) sitúan las secuencias de UPK3s en el mismo grupo que las PTPRQ en el análisis de inferencia Bayesiana (Figura 15 y Figura Suplementaria 3). Además muestran que UPK3a está estrechamente relacionada con las secuencias precursoras de PTPRQ en Ciona 1 (*C. intestinalis*) y Ciona (*C. savigny*) y también con dos secuencias de ciona relacionadas con PTPRQ que carecen de dominio catalítico. (Ciona Nocat.1 y Ciona Nocat.2) (Figura 16).

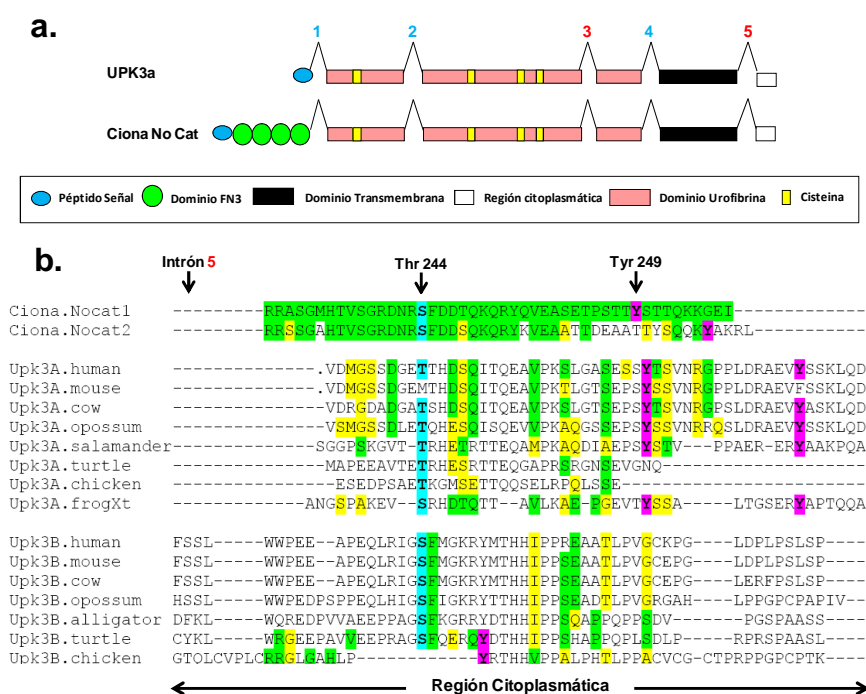


Figura 16: Similitud entre las regiones citoplasmáticas de UPK3a y las secuencias de ciona sin dominio catalítico PTP. (a) Representación de la estructura de UPK3a y Ciona No Cat. Intrones azules o rojos si la fase es 1 o 2 respectivamente. **(b)** Alineamiento de la región citoplasmática de las secuencias de ciona (CionaNocat1 y CionaNocat2) sin dominio catalítico con secuencias representativas de UPK3a y UPK3b. Residuos idénticos o similares a la secuencia de CionaNocat1 se resaltaron de verde o amarillo respectivamente. La Tyr de rana (Tyr₂₄₉) fosforilada durante la fertilización y los residuos de Tyr predichos para ser fosforilados están resaltados de magenta. Thr y Ser que coinciden con la Thr₂₄₄ humana fosforilada en respuesta a la adhesión de *E. coli* están resaltadas de azul.

Curiosamente, los fragmentos citoplasmicos de estas dos secuencias de invertebrados sin dominios catalíticos tienen cierta similitud con las secuencias citoplasmicas de UPK3a (Figura 16). En ello, se incluye un motivo putativo de fosforilación de tirosina (Tyr₂₄₉) que en *Xenopus* es transitoriamente fosforilado durante la fertilización de los ovocitos (Mahbub Hasan *et al.* 2011) y una treonina (Thr₂₄₄) que es fosforilada por una serina/treonina quinasa (caseína quinasa) en respuesta a la adhesión de la uropatogénica *E. coli* al urotelio humano (Thumbikat *et al.* 2009) (Figura 16).

Aunque la falta de genomas bien anotados impide la determinación precisa de la aparición de las UPK2/3, su relación evolutiva con las fosfatasa PTPRQ sugiere que la primera UPK2/3 se originó por duplicación y pérdida de dominios FN3 de un gen relacionado con PTPRQ durante las primeras fases de la evolución de los organismos vertebrados.

UPK3c DENTRO DE LA HISTORIA EVOLUTIVA DEL CROMOSOMA 7 DE LOS GRANDES PRIMATES. POSIBLE MECANISMO DE TRASLOCACIÓN A TRAVÉS DE UN INTERMEDIARIO CIRCULAR.

1. Localización de UPK3c en el cromosoma 7 de los grandes primates

Desde su origen, debido a una duplicación de UPK3b en el ancestro común de reptiles/aves y mamíferos (DeSalle *et al.* 2014) (Figura 8), hasta orangután solo existe una copia de UPK3c. Sin embargo, durante la especiación de los grandes primates (orangután, gorila, chimpancé, bonobo y humano) se produjo un aumento en el número de copias de tal manera que los genomas de gorila y chimpancé poseen dos copias de UPK3c mientras que en el genoma humano pueden ser localizadas tres copias.

Las copias de UPK3c en el genoma de los grandes primates están localizadas en su cromosoma 7. La historia evolutiva de este cromosoma en los grandes primates incluye dos inversiones cromosómicas que han sido estudiadas mediante la técnica de FISH utilizando BAC clones: primero una inversión pericéntrica tuvo lugar después de la separación de orangután del resto de grandes primates y luego una inversión paracéntrica en la rama que incluye chimpancé, bonobo y humano y que se produjo después de la separación de esta con la del gorila (Müller *et al.* 2004). El cromosoma 7 en humanos contiene un elevado número de duplicaciones segmentales derivadas (Jiang *et al.* 2007) que se definen como fragmentos de ≥ 1 Kb de longitud localizados más de una vez en el genoma con una identidad en su secuencia $\geq 90\%$ (She *et al.* 2004; Bailey *et al.* 2002).

El aumento en el número de copias de UPK3c estuvo relacionado con la generación de duplicaciones intracromosomales en el cromosoma 7 que dieron lugar a duplicaciones segmentales (DS) como se deduce de la información disponible en base de datos de DSs en Universidad de

California Santa Cruz (<http://genome.ucsc.edu>). En humanos, existen tres DSs que contienen alguna de las copias de UPK3c: dos localizadas en el locus genético 7q22.1: 102.178.864-102.184.083 pares de bases (pb) y 102.277.972-102.283.188 pb, y otra localizada en el locus genético 7p13: 44.024.594-44.029.806 pb. A estas DSs las hemos llamado DS1, DS3 y DS2 respectivamente. En chimpancé no existe una de las duplicaciones en 7q22.1.

Debido al aumento de las copias de UPK3c así como a sus traslocaciones durante la especiación de los grandes primates, se seleccionaron una serie de genes/pseudogenes que ayudasen a explicar lo que ocurrió. Todos los genes/pseudogenes utilizados son aquellos que aparecen anotados como más cercanos al punto donde ocurrieron las roturas para llevarse a cabo las inversiones cromosómicas. Estas regiones además fueron analizadas con el servidor GenScan por si algún gen/pseudogen no estuviese anotado. Los genes seleccionados se pueden agrupar en cinco grupos: (1) aquellos no afectados por ninguna de las dos inversiones: *CCZ1 homolog B*, *vacuolar protein trafficking and biogenesis associated (CCZ1B)* y *Family with sequence similarity 185, member A (FAM185A)*; (2) los afectados únicamente por la inversión pericéntrica: *Uroplakin 3b (UPK3b)*, *Ubiquitin-conjugating enzyme E2D4 (UBE2D4)*, *Drebrin-like (DBNL)* y *Core 1 synthase, glycoprotein-N-acetylgalactosamine 3-beta-galactosyl transferase 1 (C1GALT1)*; (3) los afectados únicamente por la inversión paracéntrica: *Fibrinogen-like 2 (FGL2)* y *Asparagine synthetase (glutamine-hydrolyzing) (ASNS)*; (4) los afectados por ambas inversiones: *Polymerase (RNA) II (DNA directed) polypeptide J (POLR2J/J2/J3/J4)*, *RAS p21 protein activator 4 (RASA4/4B/4DP/4CP)*, *HSPC047*, *Uroplakin 3c (UPK3c)* y *Lemur tyrosine kinase 2 (LMTK2)*; y (5) los que aparecieron de nuevo *Speedy/RINGO cell cycle regulator family member B2 (SPDYB2-L1/L2)*.

La región del cromosoma 7 afectada por ambas inversiones cromosómicas puede ser delimitada por los genes *CCZ1B* (7p22.1) y *FAM185A* (7q22.1) ya que conservan la misma orientación transcripcional y localización genómica desde orangután hasta humano (Figura 17). La inversión pericéntrica tuvo lugar entre los genes *CCZ1B* y *LMTK2* (7p22.1) y los genes *C1GALT1* y *ASNS* (7q11.23) debido a que aproximadamente la localización genómica de *LMTK2* en orangután ha sido ocupada por *C1GALT1* en gorila, chimpancé y humano. Sin embargo, *LMTK2* también se vio afectado por la inversión paracéntrica ya que en chimpancé y humano está localizado en el locus genético 7q22.1 mientras que en gorila se mantiene en 7q11.23 (Figura 17). La inversión paracéntrica ocurrió entre *UPK3b* y *UPK3c* y *FGL2* y *FAM185A* ya que las localizaciones cromosómicas de *UPK3c* y *FGL2* están prácticamente invertidas en gorila respecto a chimpancé y humano (Figura 17).

Las dos copias de *UPK3c* en 7q22.1 junto a *RASA4* o *4B*, *POLR2J3* o *J2*, dos copias de *SPDYB2-L1* y otras dos de *HSPC047* forman las DS1 y DS3 de humanos (Jiang *et al.* 2007) (Figura 17). En chimpancé, una copia de cada uno de estos genes/pseudogenes forman la DS1 (Figura 17). En gorila la región está mal anotada y no puede saberse con exactitud, por esta razón se ha reflejado únicamente la inversión (Figura 17). Una copia de cada uno de estos genes/pseudogenes constituye la DS2 de 7p13 (*UPK3c*, *SPDYB2-L2*, *POLR2J4*, *HSPC047* y *RASA4CP*) entre los genes/pseudogenes *DBNL* y *UBE2D4* en gorila, chimpancé y humano (Figura 17).

Todos los genes/pseudogenes incluidos en las DS se encuentran adyacentes en orangután excepto *SPDYB2-L1* o *SPDYB2-L2* que aparecen como componentes de las DS a partir de gorila (Figura 17).

El aumento de las copias de *UPK3c* durante la especiación de los grandes primates está asociado a la generación de las DS1-DS3. El cambio en la localización cromosómica en 7q11.23 de gorila se debe a la inversión cromosómica pericéntrica que tuvo lugar en la separación de orangután del

resto de grandes primates mientras que la localización en 7q22.1 de chimpancé y humano es consecuencia de la inversión paracéntrica que tuvo lugar en la separación de estos de gorila. Sin embargo, la localización de la copia de UPK3c que está localizada en la DS2 en el locus genético 7p13 de gorila, chimpancé y humano no puede ser explicada por ninguno de estos mecanismos.

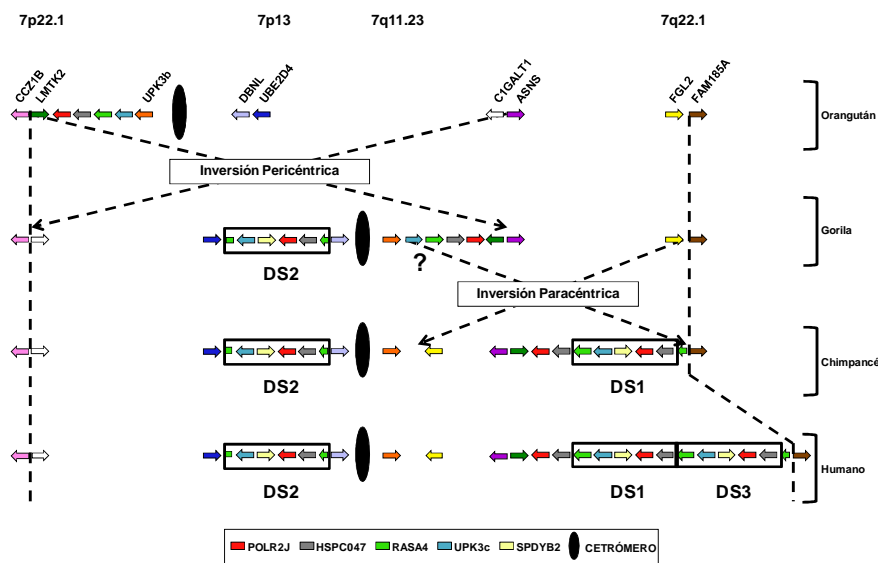


Figura 17: La actual localización cromosómica de UPK3c en gorila, chimpancé y humano no puede ser explicada solamente a través de las dos inversiones cromosómicas que ocurrieron durante la evolución de los grandes primates. Esquema de los cambios en la localización cromosómica de UPK3c durante la evolución de los grandes primates en los que se observa que la inversión pericéntrica determinó la localización de UPK3c en 7q11.23 en gorila y que una posterior inversión paracéntrica localiza a UPK3c en 7q22.1 de chimpancé y humano. Ambas inversiones no son capaces de explicar la localización cromosómica de UPK3c en 7p13. Los genes seleccionados para establecer los cambios en la localización cromosómica de UPK3c están representados mediante flechas de colores y en las que su dirección indica la orientación de la transcripción. Los genes que forman parte de las DS aparecen en la leyenda y el resto en la parte superior del esquema. Para poder simplificar el esquema, los genes de la misma familia o copias del mismo gen han sido representados del mismo color. Debido a la mala anotación que tiene el locus genético 7q11.23 en gorila (?), se ha representado solamente la situación como quedaría tras la inversión paracéntrica sin ningún aumento de copias.

2. Caracterización de la DS2 translocada al locus genético 7p13

La DS2 se encuentra en los genomas de gorila, chimpancé y humano de tal manera que su traslocación al locus genético 7p13 tuvo que ocurrir durante la especiación de los grandes primates en el momento de la separación del orangután del resto de ellos. En este momento también ocurrió la inversión pericéntrica y la DS1 fue traslocada del locus genético 7p22.1 al locus genético 7q22.11 y además también en algún momento debió de producirse la incorporación adyacente de SPDYB2-L1/L2 al resto de genes/pseudogenes de las DSs (Figura 17).

Las copias de genes/pseudogenes de la DS2 son las mismas que las de DS1 y DS3. Sin embargo, en este caso en ambos extremos de la DS se pueden encontrar dos fragmentos de una copia de RASA4. En el extremo más telomérico hay un fragmento de RASA4 que va desde el segundo exón (fragmento total perdido de 16.4 Kb) hasta el final del intrón 10-11 pero habiendo perdido 2,3 Kb que corresponde al exón 6 y parte de las secuencias intrónicas adyacentes (Fragmento A) (Figura 18a). En el extremo más telomérico existe otro fragmento de RASA4 (anotado en las bases de datos como pseudogen RASA4CP que comienza en el exón 11, habiendo perdido sus primeras 4 pb (TCAC), y acaba en su último exón (Fragmento B) (Figura 18a). En total, la DS2 tiene una longitud aproximada de 80.6 Kb y un fragmento muy similar con los fragmentos de DNA perdidos de 16.4 y 2.3 Kb pueden ser localizados en las DSs de 7q22.1 (Figura 17). Cuando los fragmentos A y B de la copia de RASA4 son unidos artificialmente, se observa que únicamente se han perdido 4 pb (TCAC) en la zona de unión en comparación con la secuencia completa del gen *RASA4* (NM_006989) sugiriendo de esta manera que la copia de RASA4 de DS2 podría haber sido partida durante el proceso de inserción. En el caso de que esto hubiese ocurrido así, la traslocación de la DS2 tendría que haber ocurrido a través de una estructura cerrada como podría ser un intermediario circular (IC).

3. Señales de mecanismos de reparación de DNA de doble cadena en las zonas de unión entre la DS2 y el locus genético 7p13

La falta de homología entre los extremos de la DS2 de 80.6 Kb insertada y la zona adyacente al punto de inserción en el locus genético 7p13 de orangután (DS2-7p13) descarta que el mecanismo de inserción tuviese lugar a través de un proceso de recombinación homóloga. Sin embargo, en estas zonas de unión se pueden encontrar señales de reparación de rotura de DNA de doble cadena (DSB): un fragmento de 13 pb en la zona de unión 5' y una microhomología de 5 pb en la zona de unión 3' típica del mecanismo de reparación de extremos no homólogos (NHEJ) (Lieber *et al.* 2008) (Figura 18a). Este mecanismo podría explicar la pérdida de 2 pb (TG) en el locus genético 7p13 en comparación con la secuencia de orangután y de 4 pb (TCAC) en la copia de RASA4 de la DS2 en comparación con la secuencia completa del gen *RASA4* (Figura 18a). Las zonas de unión con las marcas de NHEJ están conservadas tanto en humano como en gorila y chimpancé (Figura Suplementaria 4).

4. Recombinación entre elementos Alu como posible mecanismo de formación del intermediario circular

La recombinación homóloga intracromosomal entre elementos Alu localizados en la misma orientación puede dar lugar a la liberación de un IC a través de un mecanismo conocido como delección mediada a través de la recombinación homóloga de elementos Alu (ARMD) (Sen *et al.* 2006). Producto de esta recombinación en el cierre del IC y en el sitio de donde ha salido este habría un elemento Alu quimérico.

En el caso de que el IC de la DS2 se hubiese generado por una ARMD entre elementos Alu iguales pero de dos DS diferentes (ejemplo: DS1 y DS3 humanas), la identificación de los elementos Alu quiméricos sería complicada por la identidad que compartirían. Del total de 122 elementos Alu que posee la DS2, únicamente dos de ellos han podido ser identificados como elementos Alu quiméricos debido a que se han producido entre elementos Alu diferentes, lo que ha conllevado a la pérdida de los fragmento en la DS2 de 16.4 (dos primeros exones) y 2.3 Kb (exón 6) en la copia de RASA4.

Además de un elemento Alu quimérico, como señal de salida del círculo a través de alguna de estas ARMDs se deberían poder localizar los dos primeros exones de RASA4 aislados o una copia de RASA4 con dos exones 6 y en ninguno de los dos casos se han podido detectar. El elemento Alu quimérico resultante de la pérdida del fragmento de 16.4 Kb está bien anotado y se puede observar su conservación (Figura Suplementaria 5), por eso ha sido escogido para representar el posible mecanismo de cierre del intermediario circular conscientes de que es meramente especulativo (Figura 18b).

5. Análisis de la zona de inserción y apertura del IC

La región del locus genético 7p13 de orangután y la zona de apertura del IC en humano contienen un porcentaje de elementos genéticos transponibles contrapuesto: 93% de aproximadamente las 1000 pb adyacentes a cada lado del punto de inserción en 7p13 y 30% en las 1000 pb adyacentes a ambos lados de la zona de apertura del IC.

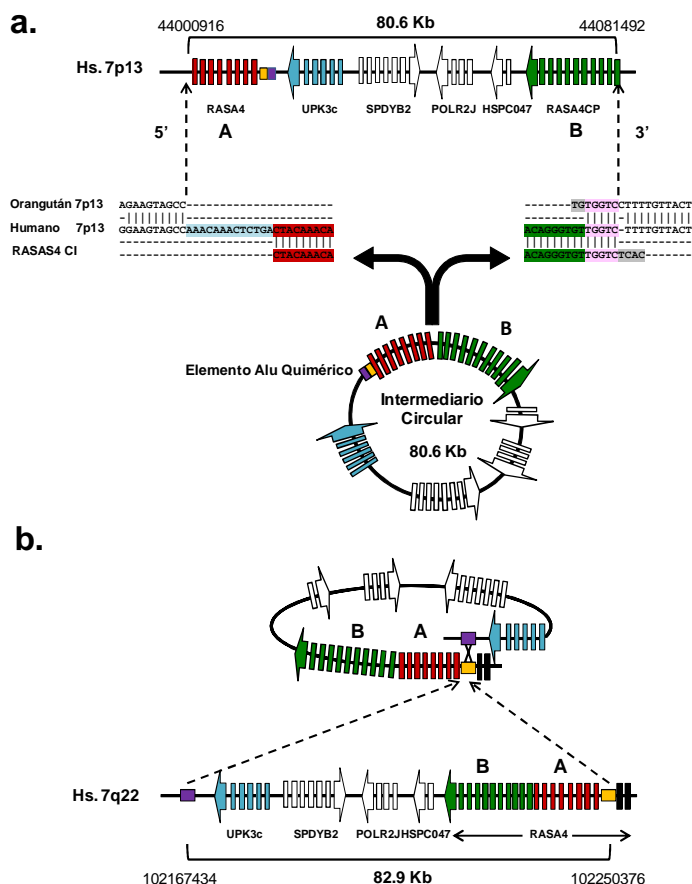


Figura 18: (a.) Marcas de mecanismos típicos de reparación de DSB en los extremos de 7p13 donde se ha insertado la DS2. Representación del locus genético 7p13 de humanos, el posible IC y el alineamiento de los extremos DS2-7p13. El alineamiento corresponde al locus genético de 7p13 en orangután, humano y una copia de RASA4. Las secuencias en color rojo y verde corresponden a los fragmentos A y B de la copia de RASA4, en color azul clarito las 13 pb insertadas, en rosa los 5 pb de microhomología y en gris los pb que han sido perdidos por la inserción. La identidad del alineamiento está indicada a través de barras verticales mientras que la falta de ella está con un guión. **(b.) Formación del IC a través de un ARMD.** Representación del locus genético 7q22.1 de humanos y la posible formación de un IC a través de una ARMD. Los genes están representados por cajas de colores y la cabeza de la flecha indica la orientación. Blanco: SPDYB2, POLR2J, HSPC047. Azul: UPK3c. Rojo: fragmento A de RASA4. Verde: fragmento B de RASA4. Negro: exones de RASA4 que quedan fuera del IC. Morado y amarillo: elementos AluS que llevan a cabo la ARMD que podría formar el IC. Los números encima o debajo de la representación de los loci genéticos corresponden a sus coordenadas.

La composición de elementos transponibles del locus genético de 7p13 es una secuencia repetitiva (AT), un elemento Alu, un DNA transposon MER46C, un retrotransposon MSTD fragmentado en tres trozos por un elemento Alu y un retrotransposon MER51A. El retrotransposon MER51A es un solo-Long Terminal Repeat (LTR) de un retrovirus endógeno humano (HERV) el cual se originaría como resultado de la recombinación de sus LTRs (Hughes y Coffin 2004). Aunque degenerados, en el solo-LTR se puede distinguir una caja TATA, una señal de poliadenilación y una región similar al fragmento típico de LTR12 típico de la familia de HERV MER51 (Jurka *et al.* 1996). La inserción del IC recae en el fragmento U3 dentro de la conformación característica de LTR (5'-U3-R-U5-3') (Figura Suplementaria 6).

Los extremos originados en el locus genético 7p13 tras la inserción del IC (DS2-7p13) poseen elementos Alu inversamente orientados y prácticamente equidistantes. En el extremo más centromérico donde se localiza la microhomología de 5 pb, existe un elemento Alu a cada extremo de ella a una distancia menor de 200 pb (133 y 189 pb). En el extremo más telomérico, a ambos lados del fragmento adicional de 13 pb también existen dos elementos Alu inversamente orientados y prácticamente equidistantes pero más alejados del punto de inserción (1774 y 2044 pb). En este último caso entre los dos elementos Alu existiría otro elemento Alu pero que comparte orientación con el del locus genético 7p13.

CARACTERIZACIÓN MOLECULAR DE UPK3c. DIVERGENCIA RESPECTO A UPK3a

1.Pauta de lectura de las tres copias de UPK3c

En el alineamiento del CDS de las tres copias de UPK3c se observa que las dos copias del locus genético 7q22.1 tienen una identidad superior al 99%, (Figura 19a) y su divergencia se basa únicamente en una sustitución por una transversión (C-A) en el pb 531 que no afecta a la secuencia de la proteína (Figura 19b).

En cambio, la copia de 7p13 presenta una identidad no superior al 97.1% y un total de 11 sustituciones por transversiones (5 sustituciones entre purinas y 6 sustituciones entre pirimidinas), 7 sustituciones por transiciones de purinas y pirimidinas y una inserción (Figura 19a) entre los pb 26 y 27. Esta inserción da lugar al desplazamiento del marco de lectura produciéndose la aparición de un codón de terminación en el pb 237 y dando lugar a un polipéptido de 78 aa sin identidad con UPK3c del locus genético 7q22.1.

2.Imbricación de UPK3c con la subunidad J de la RNA Polimerasa II

En humanos, la familia de genes de la subunidad J de la RNA polimerasa II (POLR2J) está compuesta por cuatro miembros: POLR2J, POLR2J2, POLR2J3 y POLR2J4 (Shpakovskii *et al.* 2004), siendo POLR2J el miembro más antiguo y el resto originados con las DS1-DS3.

a.

```

UPK3c_7q22.1_1 ATGGACAACAGCTGGAGGCTTGGCCC-GGCCATAGGGCTCTCTGCGGGACAGTCCCAGCTGTAGTGTGCTGTGCTACTGACCCG 90
UPK3c_7q22.1_2 ATGGACAACAGCTGGAGGCTTGGCCC-GGCCATAGGGCTCTCTGCGGGACAGTCCCAGCTGTAGTGTGCTGTGCTACTGACCCG 90
UPK3c_7p13 ATGGACAACAGCTGGAGGCTTGGCCC-GGCCATAGGGCTCTCTGCGGGACAGTCCCAGCTGTAGTGTGCTGTGCTACTGACCCG 90
*****

UPK3c_7q22.1_1 TGTCCAGCTGGGACAGAGCTGCTGCCCCAGAGCACATCAGCTATGTGCCAGCTCTCAAACGACACCTTGGCGGGGAGGCTCACCT 180
UPK3c_7q22.1_2 TGTCCAGCTGGGACAGAGCTGCTGCCCCAGAGCACATCAGCTATGTGCCAGCTCTCAAACGACACCTTGGCGGGGAGGCTCACCT 180
UPK3c_7p13 TGTCCAGCTGGGACAGAGCTGCTGCCCCAGAGCACATCAGCTATGTGCCAGCTCTCAAACGACACCTTGGCGGGGAGGCTCACCT 180
*****

UPK3c_7q22.1_1 GTCCACCTTCAGCTGGAGCAGCTCTAGGCCAGTTTACAGCAGCCACAACACTCTGACTTGGATACCATCTGGCTGTGGTGGCCCTCAG 270
UPK3c_7q22.1_2 GTCCACCTTCAGCTGGAGCAGCTCTAGGCCAGTTTACAGCAGCCACAACACTCTGACTTGGATACCATCTGGCTGTGGTGGCCCTCAG 270
UPK3c_7p13 GTCCACCTTCAGCTGGAGCAGCTCTAGGCCAGTTTACAGCAGCCACAACACTCTGACTTGGATACCATCTGGCTGTGGTGGCCCTCAG 270
*****

UPK3c_7q22.1_1 CAACGCCACCCAGAGCTTACCGGCCCAACGGAACAAACAGGACATCCCTGCTGCCAACTTCTCCAGAGGGGCTACTATCTCACACT 360
UPK3c_7q22.1_2 CAACGCCACCCAGAGCTTACCGGCCCAACGGAACAAACAGGACATCCCTGCTGCCAACTTCTCCAGAGGGGCTACTATCTCACACT 360
UPK3c_7p13 CAACGCCACCCAGAGCTTACCGGCCCAACGGAACAAACAGGACATCCCTGCTGCCAACTTCTCCAGAGGGGCTACTATCTCACACT 360
*****

UPK3c_7q22.1_1 GAGGGCCAACCGGCTGTGTACAGACAGAGGCGAGCTCCATGTCTCCGGCTCGGCAATGATACCCACTGCCAACCAACAAAATTTGG 450
UPK3c_7q22.1_2 GAGGGCCAACCGGCTGTGTACAGACAGAGGCGAGCTCCATGTCTCCGGCTCGGCAATGATACCCACTGCCAACCAACAAAATTTGG 450
UPK3c_7p13 GAGGGCCAACCGGCTGTGTACAGACAGAGGCGAGCTCCATGTCTCCGGCTCGGCAATGATACCCACTGCCAACCAACAAAATTTGG 450
*****

UPK3c_7q22.1_1 CTGCAACATCCCTTACAGGACCCGGCCCTTACAGGTTGAAGTTCCTGGTGATGAATGACGAAGGACCCGGTGGCTGAAACCAAGTGGTC 540
UPK3c_7q22.1_2 CTGCAACATCCCTTACAGGACCCGGCCCTTACAGGTTGAAGTTCCTGGTGATGAATGACGAAGGACCCGGTGGCTGAAACCAAGTGGTC 540
UPK3c_7p13 CTGCAACATCCCTTACAGGACCCGGCCCTTACAGGTTGAAGTTCCTGGTGATGAATGACGAAGGACCCGGTGGCTGAAACCAAGTGGTC 540
*****

UPK3c_7q22.1_1 CAGCGACACTCGCTGCAAGCAAGCCAGGCACTTGGGCTGTCCCGGCCCAAGAGCCGGGACCGTGTATCATCGCCACTCTGTCT 630
UPK3c_7q22.1_2 CAGCGACACTCGCTGCAAGCAAGCCAGGCACTTGGGCTGTCCCGGCCCAAGAGCCGGGACCGTGTATCATCGCCACTCTGTCT 630
UPK3c_7p13 CAGCGACACTCGCTGCAAGCAAGCCAGGCACTTGGGCTGTCCCGGCCCAAGAGCCGGGACCGTGTATCATCGCCACTCTGTCT 630
*****

UPK3c_7q22.1_1 TATCCTCTTGGCGCTCTCTCTCAGGCTCTCTGGCTGTGCTCATATACACTGCTTCAACAGCTGCAAGGACACTTCCCTATCAGGCC 720
UPK3c_7q22.1_2 TATCCTCTTGGCGCTCTCTCTCAGGCTCTCTGGCTGTGCTCATATACACTGCTTCAACAGCTGCAAGGACACTTCCCTATCAGGCC 720
UPK3c_7p13 TATCCTCTTGGCGCTCTCTCTCAGGCTCTCTGGCTGTGCTCATATACACTGCTTCAACAGCTGCAAGGACACTTCCCTATCAGGCC 720
*****

UPK3c_7q22.1_1 AGAGGAGGACAGGAGTGTGAGAAGATACACCAGCACCTCGCGTTGAGCAGCTCTGCCGAGGGGGTCTCTGTA 792
UPK3c_7q22.1_2 AGAGGAGGACAGGAGTGTGAGAAGATACACCAGCACCTCGCGTTGAGCAGCTCTGCCGAGGGGGTCTCTGTA 792
UPK3c_7p13 AGAGGAGGACAGGAGTGTGAGAAGATACACCAGCACCTCGCGTTGAGCAGCTCTGCCGAGGGGGTCTCTGTA 792
*****
    
```

b.

```

UPK3c_7q22.1_1 MNDNSRLGPAIGLSAGSQQLVSLLLLLTRVQPGTDVAPEHISYVPLSNDTLAAGRLTLSTFTLEQLPQFSSHNISDLDTIWLVLVLS 90
UPK3c_7q22.1_2 MNDNSRLGPAIGLSAGSQQLVSLLLLLTRVQPGTDVAPEHISYVPLSNDTLAAGRLTLSTFTLEQLPQFSSHNISDLDTIWLVLVLS 90
*****

UPK3c_7q22.1_1 NATQSFTAPRTNQDIPAPANFSRQYVILTRANRVLYQTRGQLHVLRVGNDTHCQPTKICGNHPLPGSPYRVKFLVMDNDEGVAETKWS 180
UPK3c_7q22.1_2 NATQSFTAPRTNQDIPAPANFSRQYVILTRANRVLYQTRGQLHVLRVGNDTHCQPTKICGNHPLPGSPYRVKFLVMDNDEGVAETKWS 180
*****

UPK3c_7q22.1_1 SDTRIQQAQALRAVPGPSGPTVLIAILSTLLAVLLTVLLAVLIYTCFNSCRSTSLSGPEEAGSVRRYTHLAFSTPAEAGS 263
UPK3c_7q22.1_2 SDTRIQQAQALRAVPGPSGPTVLIAILSTLLAVLLTVLLAVLIYTCFNSCRSTSLSGPEEAGSVRRYTHLAFSTPAEAGS 263
*****
    
```

Figura 19: Secuencias del CDS y proteína de las tres copias de UPK3c. (a.) Alineamiento de las secuencias del CDS de las tres copias de UPK3c. **(b.)** Alineamiento de las secuencias proteicas de las copias de UPK3c en 7q22.1. **(a. y b.)** Los alineamientos fueron realizados con el programa MAFFT. Las copias están nombradas acorde al locus genético dónde se sitúan y en el caso de las dos copias de 7q22.1, la copia con un 1 representa que está localizada más centromérica mientras que la secuencia con un 2 se sitúa más telomérica. Las letras azules y negras representan exones alternados, mientras que las letras rojas representan la fase de los intrones. Las letras resaltadas de verde representan las diferencias entre las 3 copias de UPK3c por transiciones, transversiones o inserciones. La última línea refleja identidad (*) y sin identidad ().

POLR2J2/J3/J4 están localizados aproximadamente a 24 Kb dirección 5' del ATG de la copia correspondiente de UPK3c. Acorde a lo anotado en el UCSC Genome Browser las copias más cercanas a POLR2J2 y POLR2J3 (7q22.1) forman parte de su 3' UTR mientras que en el caso de POLR2J4 (7p13) esto no ocurre (Figura 20).

Toda la secuencia del cDNA de UPK3c excepto el 5' UTR y su primer exón es la parte de su secuencia que constituye el 3' UTR de POLR2J2 y POLR2J3. Durante la evolución de los grandes primates, el gen ancestral que dio lugar a POLR2J2/J3/J4 perdió un fragmento de su secuencia que abarca desde parte del último intrón hasta el 3' UTR, perdiendo de esta manera el último exón. Su lugar ahora es ocupado por un fragmento de DNA que incluye a un gen SPDYB2-L1/L2 de la familia Speedy y un pseudogen de la familia PMS2. Parte de la secuencia del pseudogen de PMS2 actúa como la parte del último intrón y como último exón de POLR2J2/J3/J4. Posiblemente debido a que este fragmento no posee ninguna parte de su secuencia que pueda actuar como 3' UTR, ha ocurrido la imbricación de POLR2J2/J3 con las copias de UPK3c.

3. Análisis transcripcional de UPK3c

a. Análisis de ESTs

Como la mayor parte del cDNA de UPK3c está incluido en el 3' UTR de POLR2J2/J3 (Figura 20) se llevó a cabo la búsqueda de ESTs anotados en bases de datos y así poder tener una aproximación sobre la expresión de UPK3c en diferentes tejidos. La búsqueda se limitó para aquellos que tuviesen el primer exón en su secuencia (Figura 20) y poder asegurar de esta manera que corresponden a UPK3c.

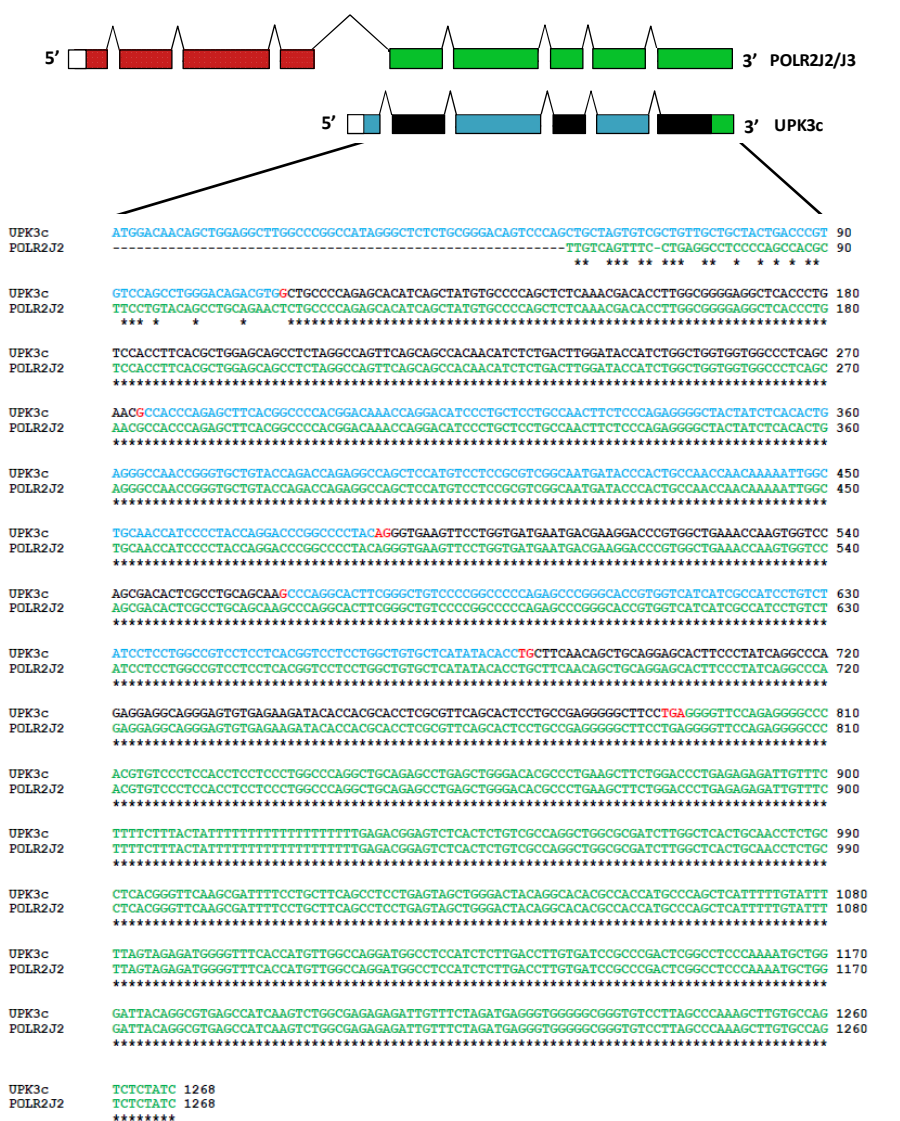


Figura 20: Imbricación de las copias de UPK3c con POLR2J2/J3 de la familia de POLR2J. **(Parte superior)** Esquema del cDNA de POLR2J2/J3 y UPK3c Los rectángulos blancos representan el 5' UTR; los rectángulos rojos el CDS de POLR2J2/J3; los rectángulos azules y negros los exones de UPK3c; las cajas verdes el 3' UTR de POLR2J2/J3 y de UPK3c. **(Parte inferior)** Alineamiento del CDS y 3' UTR de UPK3c con el 3' UTR de POLR2J2 mostrando su identidad. Las letras azules y negras representan exones alternados de UPK3c, las letras rojas demarcan la fase del intrón y el codón de terminación (TGA), las letras verdes muestran el 3'UTR de UPK3c y de POLR2J2. El alineamiento fue realizado con el servidor MAFFT con los parámetros de serie y muestra la total identidad (*) o la falta de identidad ().

Se encontraron un total de 51 ESTs de los cuales 26 corresponden a tejido sano y 25 a tejido patológico (Tabla 4). Los ESTs de tejido sano se distribuyen en un amplio rango de tejidos y órganos siendo con un 21% el ojo, el órgano con mayor número de ESTs. El ojo con un 40% es también el órgano con mayor porcentaje de ESTs en tejido patológico seguido de vejiga con un 16% y de útero con un 12% y ovario y páncreas un 8%.

REFERENCIA	ORGANO/TEJIDO	C-A	REFERENCIA	ORGANO/TEJIDO	PATOLOGIA	C-A
DC424984.1	CARTILAGO	?	AA603194.1	ESTÓMAGO	TUMOR	?
DA093536.1	CEREBELO	?	AL520850.3	GLANDULA ADRENAL	NEUROBLASTOMA	?
B488409.1	CEREBRO	?	BE241615.1	NÓDULO LINFÁTICO	LEUCOFERESIS	?
AA077375.1	CEREBRO	?	BE783282.1	OJO	RETINOBLASTOMA	?
DN993120.1	CEREBRO	?	DR422125.1	OJO	PTERIGIÓN	C
BF816477.1	COLON	?	DR423533.1	OJO	PTERIGIÓN	C
BF758159.1	COLON	?	DR422867.1	OJO	PTERIGIÓN	C
R07794.1	HÍGADO Y BAZO	?	DR421655.1	OJO	PTERIGIÓN	C
B907605.1	LEUCOCITO	?	CV574749.1	OJO	QUERATOCONO	C
CR994138.1	LINFOCITO T	C	CV571894.1	OJO	QUERATOCONO	C
CR980593.1	LINFOCITO T	C	CV571740.1	OJO	QUERATOCONO	C
BM847378.1	NÓDULO LINFÁTICO	?	CV570877.1	OJO	QUERATOCONO	C
BM843730.1	NÓDULO LINFÁTICO	?	CV570067.1	OJO	QUERATOCONO	C
BM724442.1	OJO	?	BF038709.1	OVARIO	TUMOR	A
BG741801.1	PIEL	?	BF032261.1	OVARIO	TUMOR	?
BX404896.2	PLACENTA	?	BE735985.1	PÁNCREAS	TUMOR	C
DA855203.1	PLACENTA	C	BE736087.1	PÁNCREAS	TUMOR	A
AL544559.3	PLACENTA	A	BF923489.1	SISTEMA NERVIOSO	TUMOR	?
AL544513.3	PLACENTA	?	BF914307.1	ÚTERO	TUMOR	?
R28399.1	PLACENTA	?	BF912925.1	ÚTERO	TUMOR	?
DA878206.1	PRÓSTATA	?	BF912921.1	ÚTERO	TUMOR	?
BX457259.2	TIMO	?	BQ301028.1	VEJIGA	TUMOR	?
HY084937.1	TIMO	?	BQ300930.1	VEJIGA	TUMOR	?
HY092696.1	TIMO	?	BF738937.1	VEJIGA	TUMOR	?
HY063262.1	TIMO	?	BF737438.1	VEJIGA	TUMOR	?
DB249571.1	ÚTERO	?				

Tabla 4: El ojo es el órgano con más ESTs de UPK3c. Total de ESTs encontrados para las tres copias de UPK3c en tejidos sanos y patológicos en bases de datos de ESTs. Referencia se refiere al código del EST en la base de datos. C/A se refiere a la copia de UPK3c 531C o UPK3c 531A. El símbolo de interrogación se colocó en el caso de que no se pudiese diferenciar a cuál de las dos copias pertenecía el EST.

Las copias de UPK3c del locus genético 7q22.1 se encontraron entre los ESTs mientras que la copia del locus genético 7p13 no. Al estar la búsqueda limitada al primer exón y debido a que las dos secuencias en su CDS difieren solo en la transversión 531 C-A en el exón 4, en el 69% de los ESTs no se pudo diferenciar entre una y otra copia de UPK3c bien porque su longitud no era suficiente o porque no estaban bien anotados en esa región de la secuencia. De los restantes ESTs, el 25% correspondieron a UPK3c 531C y el 6% restante a UPK3c 531A. Aproximadamente el 69% de los ESTs de UPK3c 531C corresponden a ojo siendo cuatro ESTs de pterigion, el cual es un crecimiento anormal de la conjuntiva que tiende a dirigirse hacia la cornea (Jaworski *et al.* 2009), y cinco ESTs de queratocono, el cual es un adelgazamiento de la cornea que hace que protruya hacia delante (Chaerkady *et al.* 2013). De los cinco ESTs de placenta, mientras que en tres no se puede distinguir entre las dos copias de UPK3c del locus genético 7q22.1, de los dos restantes existe uno de cada UPK3c. En cáncer de páncreas, de los dos ESTs uno corresponde a cada copia de UPK3c. Esta observación indica la existencia a nivel transcripcional de las dos UPK3c de 7q22.1 y su coexistencia en tejidos sanos y patológicos.

b. Estudio del porcentaje relativo de expresión de las tres copias de UPK3c

Aunque la existencia de las copias de UPK3c del locus genético 7q22.1 se comprobase mediante el análisis de ESTs, en la mayoría de ellos no se podía diferenciar entre una y otra copia de UPK3c. Por lo tanto, para poder obtener una idea más precisa sobre la transcripción de estas dos copias de UPK3c y comprobar si la copia de UPK3c en el locus genético

7p13 se transcribe, se secuenciaron el producto de PCR de 23 tejidos con un primer común para las tres copias de UPK3c.

En ninguno de los 23 tejidos analizados se detectó la UPK3c del locus genético de 7p13 de tal manera que probablemente no se transcriba. Por otro lado, se detectaron transcritos de las dos copias de UPK3c del locus genético 7q22.1 (531C y 531A) en todos los tejidos excepto en corazón donde solo se detectó la copia UPK3c 531C (Figura 21).

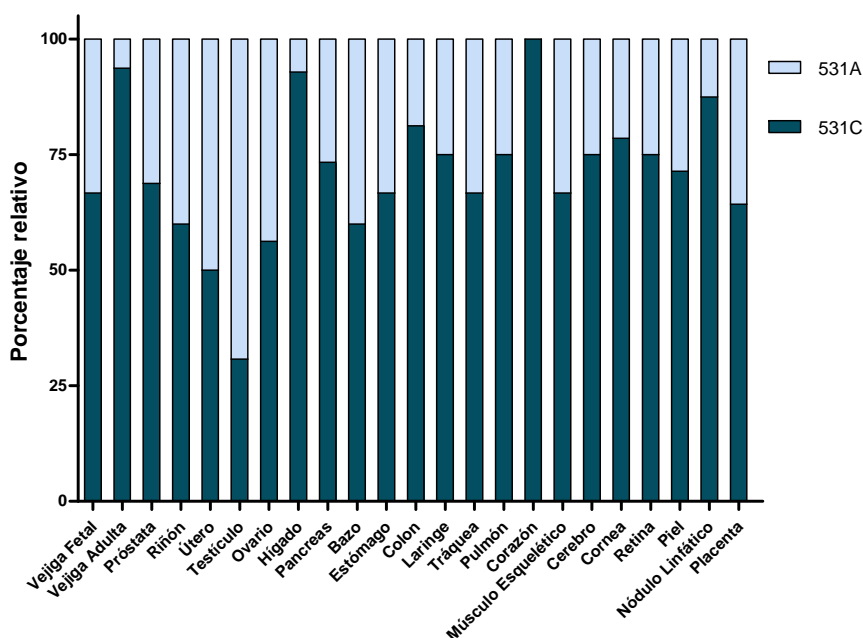


Figura 21: UPK3c 531C posee un porcentaje de expresión relativa mayor que UPK3c 531A Porcentaje relativo de expresión en tejidos humanos de las dos copias de UPK3c en el locus genético 7q22.1 humano: UPK3c 531C (azul oscuro) y UPK3c 531A (azul claro). El porcentaje relativo se calculó respecto a la suma total de los picos para ambas copias en cada cromatograma.

En vejiga adulta, hígado, colon, laringe, pulmón, nódulo linfático, cerebro, cornea y retina el porcentaje relativo de la copia UPK3c 531C fue superior al 75% (Figura 21). En vejiga fetal, próstata, tráquea, estómago, riñón, músculo esquelético, bazo, placenta y piel el porcentaje relativo de la

copia UPK3c 531C fue superior o igual al 60% e inferior al 75% (Figura 21). Útero, ovario y testículo fueron los únicos tejidos que tuvieron un porcentaje relativo de la copia UPK3c 531C similar o inferior a la copia UPK3c 531A (Figura 21). Resaltar la diferencia en el porcentaje relativo de la copia UPK3c 531C entre vejiga fetal y adulta dado que en la vejiga fetal es cercano al 75% mientras que en la adulta es del 90% (Figura 21).

Debido a la disponibilidad de tejidos sano, en el caso de cornea, tres muestras pudieron ser analizadas teniendo un porcentaje relativo con un error estándar de 72 ± 3 para la copia UPK3c 531C y de 27 ± 3 para la copia UPK3c 531A. Aunque el tamaño muestral es bajo, aparentemente en tejido sano parece que el porcentaje relativo de cada copia de UPK3c del locus genético 7q22.1 es constante.

c.Estudio de la expresión de las uroplaquinas mayoritarias de las placas (UPK1a, UPK1b, UPK2a y UPK3a) en tejidos humanos.

De los 22 tejidos humanos analizados, solamente en vejiga fetal, vejiga adulta y riñón se expresaron las cuatro uroplaquinas mayoritarias a la vez. El heterodímero UPK1a-UPK2a en vejiga fetal es en el que se observó una expresión más elevada entre estos órganos (Figura 22).

En próstata y útero se observó la expresión de 3 uroplaquinas mayoritarias: UPK1a/2a/3a y UPK1a/2a/1b respectivamente. La expresión observada en UPK3a de próstata fue claramente superior a la del resto de uroplaquinas en este órgano y no se expresó su pareja de heterodímero UPK1b. Útero es el único de los órganos analizados que no presenta urotelio en los que se ha observado expresión de varias uroplaquinas (Figura 22).

En tráquea se observó la expresión de UPK1a y UPK1b y en ovario, estómago, corazón, cornea nódulo linfático y placenta solamente de UPK1b. De todos los tejidos y uroplaquinas analizadas, la expresión más elevada se observó para la UPK1b en cornea. UPK1b es la única de las uroplaquinas mayoritarias que se expresó en varios órganos que carecen de urotelio (Figura 22).

En testículo, hígado, páncreas, bazo, colon, pulmón, músculo esquelético, cerebro, retina y piel no se observó expresión de ninguna de las uroplaquinas mayoritarias de las placas (Figura 22).

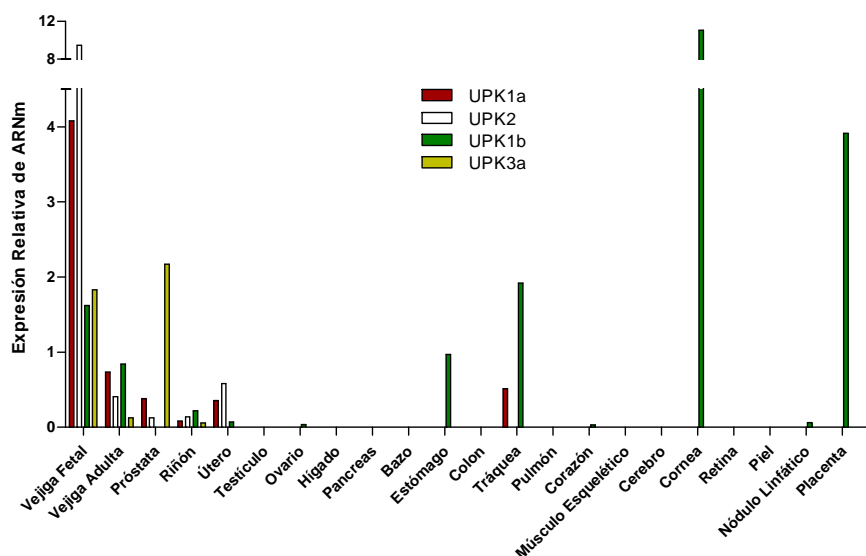


Figura 22: Las uroplaquinas mayoritarias de las placas se expresan principalmente en el urotelio. Expresión relativa de mRNA en distintos tejidos humanos de las uroplaquinas mayoritarias de las placas. Las uroplaquinas mayoritarias de las placas se expresan primordialmente en órganos con urotelio y UPK1b es la única de ellas que se expresa de forma no aislada fuera de este. UPK1a (rojo), UPK2 (blanco), UPK1b (verde) y UPK3a (amarillo).

d. Estudio de la expresión de UPK3c y POLR2J2/J3 en tejidos humanos

Debido a la existencia de una imbricación entre las copias de UPK3c que se transcriben y POLR2J2 y POLR2J3, se examinó la expresión de ambos genes en una serie de tejidos humanos con el fin de comparar sus niveles de expresión (Figura 23). Para ello se diseñaron sondas específicas, la primera dirigida exclusivamente al primer exón de las copias de UPK3c quedando excluidos de esta manera POLR2J2/J3 y la segunda dirigida al último exón de POLR2J2/J3 evitando así la detección de POLR2J y de UPK3c.

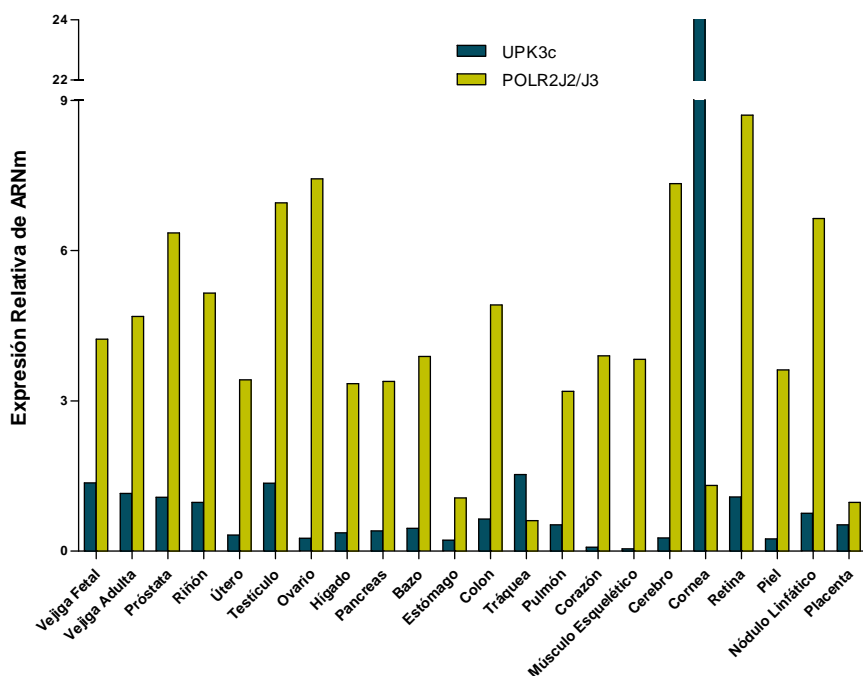


Figura 23: La imbricación de UPK3c y POLR2J2/J3 no se refleja en su patrón de expresión. UPK3c y POLR2J2/J3 tienen pautas de expresión transcripcional independientes. Expresión relativa del mRNA de POLR2J2/J3 y UPK3c en diferentes tejidos y órganos humanos. UPK3c (azul oscuro) y POLR2J2/J3 (amarillo).

En todos los tejidos excepto tráquea y cornea la expresión de POLR2J2/J3 es superior que la de UPK3c. POLR2J2/J3 muestran una expresión ubicua clara en los tejidos analizados mientras que UPK3c en corazón y músculo esquelético es prácticamente indetectable y en cornea es 21 veces superior que la expresión relativa media de las copias de UPK3c en el resto de tejidos. Esto sugiere que la imbricación no afecta a la transcripción de ambos genes (Figura 23).

e. Estudio de la expresión de UPK3c y UPK3b en tejidos humanos

UPK3c se originó como consecuencia de una duplicación de UPK3b en el ancestro común de reptiles/aves y mamíferos. El origen de UPK3b se sitúa en el ancestro común de peces óseos/cartilaginosos (De Salle *et al.* 2014). Debido a esta relación evolutiva entre UPK3c y UPK3b se compararon sus expresiones relativas en una serie de tejidos humanos (Figura 24).

Excepto UPK1b el resto de uroplaquinas mayoritarias de las placas en humanos se expresan principalmente en el urotelio (Figura 22). Sin embargo, la expresión de UPK3b y UPK3c no está tan restringida e incluso la de UPK3c es más generalizada que la de UPK3b (Figura 24). Todos los tejidos mostraron expresión para UPK3c excepto corazón y músculo esquelético cuyos niveles prácticamente son indetectables. La expresión relativa de cornea fue 21 veces superior a la de la media de la expresión relativa observada en el resto de tejidos. UPK3b no se expresó en páncreas, tráquea, estómago, placenta, cerebro y cornea. Pulmón, útero y testículo mostraron la mayor expresión relativa entre los tejidos analizados siendo en el primero de estos más de 3 veces superior al de los otros dos. Al contrario que en el resto de las uroplaquinas donde la expresión relativa entre vejiga adulta y fetal es claramente superior en esta última, en el caso

de UPK3c la expresión entre ambos órganos es prácticamente equiparable (Figura 22, 23 y 24).

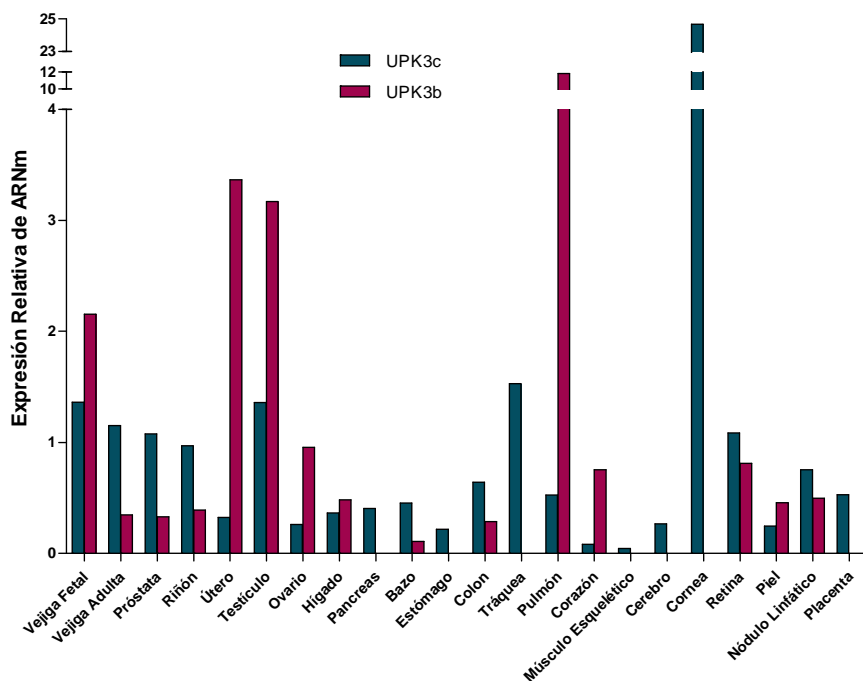


Figura 24: UPK3c y UPK3b se expresan en un amplio número de tejidos. Análisis de la expresión relativa de mRNA de UPK3c y UPK3b en diferentes tejidos humanos muestra que además de expresarse en varios tejidos, su expresión no es mayoritaria en el urotelio. UPK3c (azul oscuro) y de UPK3b (magenta).

4. Análisis de la secuencia proteica de UPK3c

a. Características asociadas a la secuencia de proteína de UPK3c

La traducción de los mRNAs de las copias de UPK3c 531C y UPK3c 531A dan lugar a la misma secuencia de una proteína de 263 aa de

longitud codificados por seis exones en fases 1:1:2:1:2 (Figura 25). Su perfil hidropático muestra tres regiones de carácter hidrofóbico localizadas en los 33 primeros aa, entre los aa 83 y 89 y desde el aa 203 hasta el 225 (Figura 25b). El servidor para péptidos señales SignalP v4.1 detecta un sitio de corte de péptido señal entre los aa 33 y 34 coincidiendo con la primera región hidrofóbica y el servidor para hélices transmembrana TMHMM v2.0 detecta un posible DTM entre los aa 203 y 225 coincidiendo con el tercer fragmento hidrofóbico. De esta manera la secuencia de aa de la proteína de UPK3c tiene un extremo extracelular amino terminal de 202 aa y un extremo citosólico carboxilo terminal de 38.

Acorde al servidor de ExpASy que calcula el peso molecular de proteínas a partir de la secuencia de aa, UPK3c tendría un peso molecular total de 28.5 KDa y dado que el péptido señal en UPK3c tendría un peso molecular de 3,5 KDa, la proteína madura de UPK3c sin posibles modificaciones post-traduccionales tendría en principio un peso molecular de 25 KDa.

b. Análisis de la divergencia de UPK3c respecto a UPK3a/3b

La secuencia de la proteína humana de UPK3c tiene una longitud de 263 aa siendo de este modo la UPK3 de secuencia más corta debido a que las secuencias de proteína de UPK3a y UPK3b tienen una longitud de 287 y 276 aa respectivamente. UPK3a y UPK3b comparten una identidad del 28% y la identidad que comparte UPK3c con estas dos es del 20% y 34% respectivamente. UPK3c, al igual que UPK3a y UPK3b, está codificada por seis exones en fases 1:1:2:1:2 con DTM tipo I en el exón seis y un péptido señal en el primer exón. Su extremo carboxilo con una

longitud de 41 aa es más corto que el del resto de las UPK3s que es de 58 y 59 aa (Figura 26).

a.

```
MDNEWRLGPAIGLSAGOSQLLVSLLLLLTRVQPCTDVAAPEHISYVPQLSNDTLAGRLTLSTFTLEQPLGQFSSHNISDLDTIWLVVALS 90  
NATQSFAPRTNQDIPAPANFSPQRGYLTLRANRVLYQTRGQLHVLRVGNDTHCQPTKIGCNHPLGPGPYRVKPLVMNDEGPFVAETKWS 180  
SDTRLQQAQALRAVFGPQSPGTVIIAILSILLAVLLTLVLLAVLIYTCFNSCRSTSLSGPEEAGSVRRYTTHLAFSTPAEGAS 263
```

b.

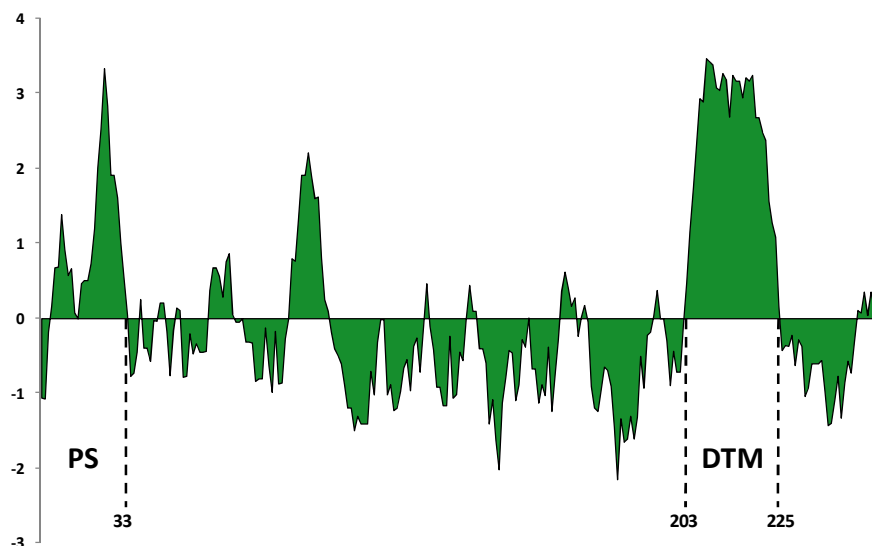


Figura 25: Características asociadas a la secuencia proteica de UPK3c. (a) Secuencia proteica de UPK3c. Las letras negras y azules representan diferentes exones. Las letras rojas hacen referencia a las fases 1 o 2 de los intrones; (b) Perfil hidropático de las copias de UPK3c en 7q22.1. Péptido señal (PS); Dominio Transmembrana (DTM). Los números representan las coordenadas en la secuencia de la proteína de UPK3c.

UPK3a y UPK3b poseen un fragmento de 12 aa muy conservado amino terminalmente yuxtapuesto al DTM. UPK3c humana lo ha perdido por completo, al igual que otros mamíferos como la vaca o marsupiales como la zarigüeya. Sin embargo, en el caso del gallo, se pueden apreciar todavía varios aa conservados debido a que el origen de UPK3c se sitúa en el ancestro común de aves y reptiles (De Salle *et al.* 2014) (Figura 8 y Figura 26).

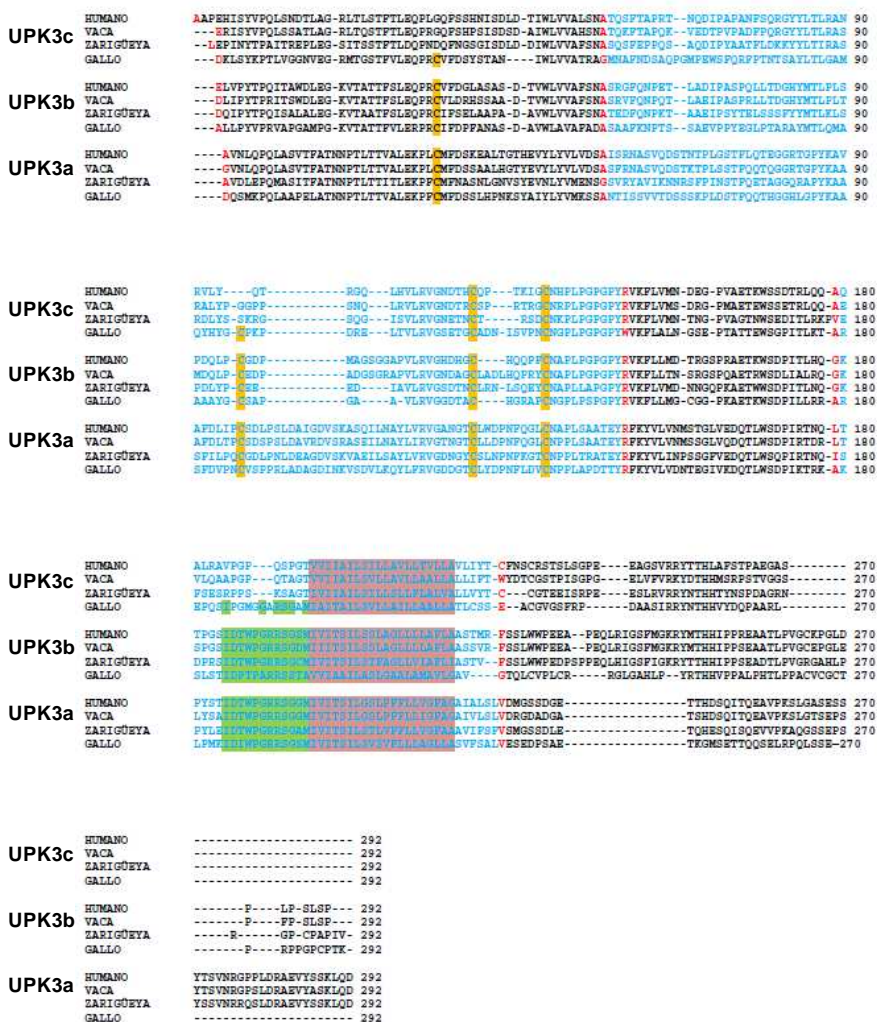


Figura 26: Divergencia entre UPK3c, UPK3b y UPK3a. Alineamiento de los exones 2-6 en UPK3a, UPK3b y UPK3c en mamíferos (humano y vaca), marsupiales (zarigüella) y aves (gallo) realizado en el servidor MAFFT con los parámetros de serie. Letras azules y negras representan exones alternados. Letras de color rojo representan la unión de los exones por fases 1 o 2. C resaltadas en color amarillo son residuos de cisteína. Resaltar la pérdida de las cisteínas C₁-C₂ en UPK3c de mamíferos y de la similitud en la región de máxima conservación entre UPK3a y UPK3b. Letras resaltadas en verde corresponden al fragmento de 12 aa muy conservado yuxtapuesto al DTM. Letras resaltadas de rosa corresponden al DTM.

Esta misma situación se puede observar en el caso de las C₁ y C₂ en UPK3a/UPK3b las cuales están muy conservadas y en el caso de UPK3c aún existen en aves y reptiles pero sin embargo en marsupiales y mamíferos han desaparecido (Figura 26).

5.Glicosilación en UPK3c

a.Glicosilación de UPK3c transfectada

Tras su clonación en el vector de expresión pCS2+, UPK3c fue transfectada de forma transitoria en la línea celular de CU T24 dado que no la expresa de forma endógena. A través de la detección de UPK3c transfectada por WB se observan dos bandas de peso molecular aproximado de 43 y 28 KDa que podrían corresponder a la UPK3c modificada post-traduccionalmente a través de glicosilaciones y a la secuencia completa de UPK3c sin modificaciones post-traduccionales según al peso molecular establecido por el servidor ExpPASy (Figura 27).

Dado que el resto de UPK2/3s humanas, excepto UPK2a, están N-glicosiladas (Deng *et al.* 2002; Wu *et al.* 1994), para comprobar el estado glicosilativo de UPK3c se utilizaron dos tipos de glicosidasas (Endoglicosidasa H o Endo H y Péptido N-glicosidasa F o PNGase F) capaces de romper este tipo de enlace glicosídico. El enlace N-glicosídico se caracteriza por unir un glicano a un átomo de nitrógeno, normalmente de una asparragina (Asn) acorde a una secuencia consenso. Aunque hay diversos N-glicanos atendiendo al número y al tipo de residuos que contengan, todos comparten una estructura común, tres manosas seguidas de dos N-acetil glucosaminas (GlcNAc), unida al nitrógeno de la Asn. Endo H es capaz de cortar el enlace entre las dos GlcNAc siempre que el glicano

esté compuesto solamente de manosas y deja libre por un lado el glicano y por otro el esqueleto proteico con un GlcNAc unida. PNGase F corta el enlace entre el nitrógeno de la Asn y el GlcNAc incluso cuando los glicanos son complejos dejando libre el esqueleto proteico por un lado y el glicano por el otro (Maley *et al.* 1989).

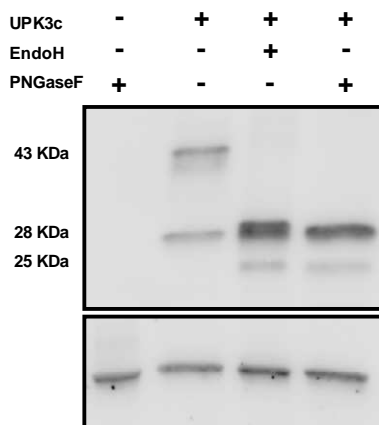


Figura 27: La deglicosilación de UPK3c transfectada muestra diferentes grados de maduración. Immunoblot de un lisado celular de la línea de CU T24 transfectada con UPK3c tratado con con dos tipos de N-glicosidasas. La inmunodetección de UPK3c muestra diferentes grados de maduración, sensibilidad ante ambas N-glicosidasas y la posibilidad de que esté N-glicosilada en varios residuos. **(Panel superior)** UPK3c; **(Panel inferior)** β -actina.

Cuando UPK3c transfectada es tratada con ambas glicosidasas desaparece la banda de aproximadamente 43 KDa, se incrementa la intensidad de la señal de la banda de aproximadamente 28 KDa y aparece una nueva banda de aproximadamente 25 KDa que podría corresponder a la proteína sin péptido señal (Figura 27). Las dos glicosidasas son capaces de deglicosilar la proteína transfectada indicando que su composición en glicanos es exclusivamente de manosas. Las bandas que se observan tras el tratamiento con Endo H tienen un peso molecular ligeramente más elevado que las tratadas con PNGase F (Figura 27). Esto puede ser debido a las GlcNAc que quedan unidas al esqueleto de la proteína de UPK3c que al cortar con EndoH dan un mayor peso molecular y por tanto una menor movilidad electroforética. El peso molecular de una GlcNAc es de 0.221

KDa, aunque difícil de aproximar la diferencia de movilidad electroforética entre la UPK3c tratada con EndoH o PNGaseF, es probable que UPK3c transfectada esté N-glicosilada en varios residuos de Asn.

El servidor bioinformático NetNGlyc v1.0 establece potenciales sitios de N-glicosilación y en el caso de la secuencia de UPK3c, establece cuatro potenciales sitios con elevada probabilidad de estar glicosilados localizados en la parte extracelular de la proteína en las posiciones 51, 76, 91 y 110 y otro posible sitio de glicosilación con menor probabilidad localizado también en la parte extracelular en el aminoácido 140. De esta manera, dada la diferencia de peso molecular de UPK3c tras los tratamientos de glicosidasas, varios de estos potenciales sitios de glicosilación podrían realmente estar glicosilados. El peso molecular del péptido señal como el de las últimas glicosilaciones que dan lugar a la proteína madura en la UPK3c transfectada coinciden y son de 3.5 KDa.

b.Grado y tipo de glicosilación de UPK3c endógena

La UPK3c endógena en cornea, pterigion y en tumor de vejiga también están glicosiladas. UPK3c tiene un peso molecular aproximado de 45 KDa en cornea (superior al peso molecular de 43 KDa de UPK3c transfectada) y tanto en pterigion como en tumor de vejiga además de esta banda de 45 KDa, presenta bandas adicionales de aproximadamente entre 40 y 45 KDa (Figura 28).

Al contrario que la UPK3c transfectada, la proteína endógena de cornea fue completamente resistente a EndoH. Las UPK3c endógenas de pterigion y de tumor de vejiga fueron parcialmente sensibles a EndoH (Figura 28) conservando bandas resistentes a Endo H de aproximadamente 43 y 45 KDa y observándose bandas que coinciden con

UPK3c sin modificaciones post-traduccionales con y sin péptido señal. En cambio cuando las muestras son tratadas con PNGaseF, todas las muestras son sensibles a ella, observándose dos bandas que coinciden con UPK3c con y sin péptido señal. En el caso de las muestras que fueron parcialmente sensibles a EndoH, al hacerlo totalmente con PNGaseF, solamente la intensidad de señal se ve aumentada en la UPK3c sin péptido señal, indicando que al contrario de la UPK3c transfectada, en este caso la proteína madura sin péptido señal es la mayoritaria.

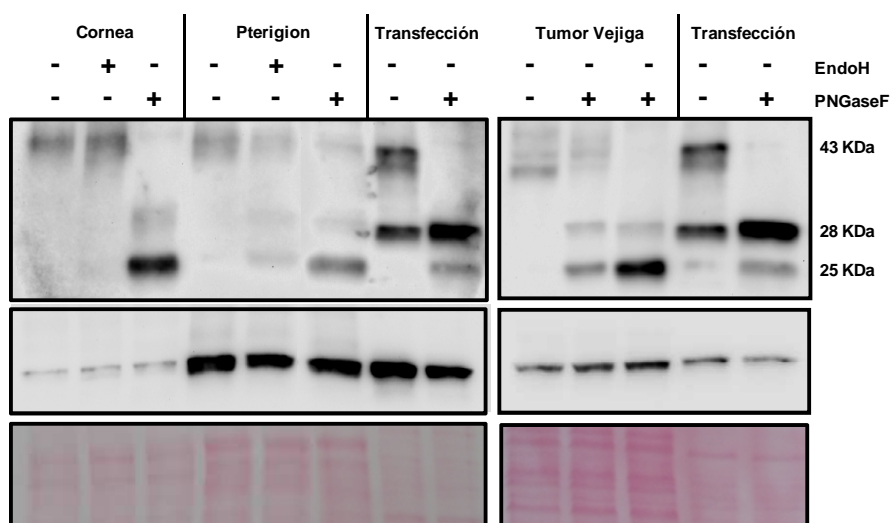


Figura 28 Diferentes grados de glicosilación de UPK3c en cornea, pterigion y un tumor de vejiga. Extractos proteicos de lisados celulares de cornea, pterigion y tumor de vejiga humanos se trataron con los dos tipos diferentes de N-glicosidasas. La inmunodetección de UPK3c en cornea sana, pterigion y tumor de vejiga muestran diferentes grados de glicosilación y sensibilidad a N-glicosidasas. Una vez deglicosiladas se observan distintos grados de maduración de UPK3c. **(Panel superior)** UPK3c, **(Panel medio)** β -actina y **(Panel inferior)** membrana de nitrocelulosa teñida con Ponceau S solution como control de carga.

A diferencia de la muestra de tejido sano de cornea en la que solo aparece una banda glicosilada, los tejidos de pterigion y de tumor de vejiga, incluyendo cuatro tumores de vejiga adicionales (Figura 29), presentan bandas de diferente tamaño entre ellos indicando una posible alteración de la glicosilación en estos tumores.

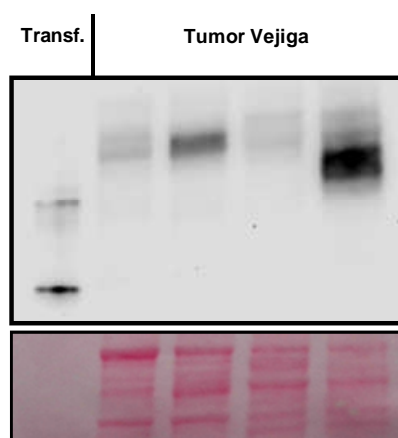


Figura 29: Diferentes grados de glicosilación de UPK3c en tumores de vejiga. Extracto proteico de lisado celular de cuatro tumores de vejiga, tratados con solución tampón desnaturizante de glicoproteínas. La inmunodetección de UPK3c muestra que en todos los tumores de vejiga, UPK3c está solamente glicosilada y el grado de glicosilación es diferente indistintamente del estadio o del grado del tumor de vejiga. **(Panel superior)** UPK3c y **(Panel inferior)** membrana de nitrocelulosa teñida con Ponceau S solution como control de carga.

6. Inmunolocalización de UPK3c en el urotelio de ratón

Para determinar la localización de UPK3c en el urotelio, se estudió su inmunolocalización en secciones de vejiga sana de ratón. Su expresión se comparó con la de las uroplaquinas que forman heterodímeros. Los heterodímeros de UPK1a/UPK2a y UPK1b/UPK3a se expresan en la parte más apical de las células “umbrella” y en su parte citosólica colocalizando Sin embargo, no se localizó su expresión en las capas intermedia y basal (Figura 30a y 30b).

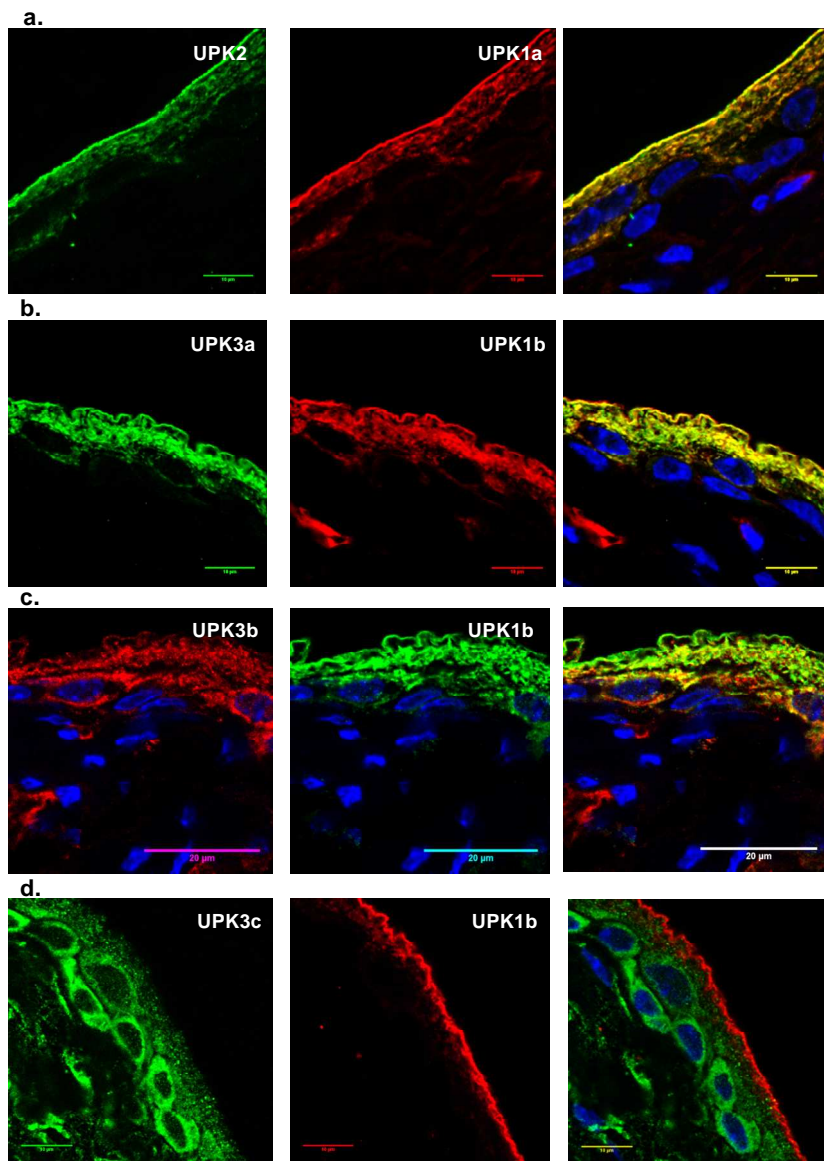


Figura 30: La inmunolocalización de UPK3c en urotelio de ratón es diferente a la del resto de UPK3s. Inmunolocalización de la familia de uroplaquinas en el urotelio de ratón sano. Secciones de parafina de urotelio de vejiga de ratón sano se incubaron con anticuerpos para las diferentes uroplaquinas y se detectaron por microscopía confocal. Emisión de fluoróforo para UPK1a y UPK1b 617nm (rojo), para UPK2a, UPK3a, UPK3b y UPK3c emisión de 520nm (verde) y para la detección nuclear emisión 430nm (azul). **(a)** Colocalización del heterodímero UPK1a/UPK2; **(b)** Colocalización del heterodímero UPK1b/UPK3a; **(c)** Colocalización del heterodímero UPK1b/UPK3b; **(d)** Localización de UPK1b y UPK3c.

En el caso del heterodímero que forma UPK1b con el componente minoritario de las partículas de 16 nm, UPK3b, se observa que su colocalización no es tal como la de UPK1b/UPK3a. Se observa que se localiza en la parte más apical de las células “umbrella” pero menos que en la capa intermedia que es donde parece mayoritaria su localización (Figura 30c).

UPK3c muestra una pauta de expresión diferente a la del resto de uroplaquinas. Por un lado, aparentemente no colocaliza con UPK1b e incluso aunque se exprese en las células “umbrella”, UPK3c podría no formar parte de la parte apical de estas y por tanto de las placas. La expresión de UPK3c en la parte citosólica de las células “umbrella” tiene un aspecto vesicular y difuminado si se compara con la expresión en las capas intermedias y basales donde su expresión es de mayor intensidad (Figura 30d). Por lo tanto, al contrario que el resto de uroplaquinas, las cuales son consideradas productos de diferenciación, UPK3c se expresa en las capas menos diferenciadas o sin diferenciar mientras que a medida que las células se diferencian, su expresión se hace más débil (Figura 30d y Figura 31).

7.UPK3c y cáncer de vejiga

a.Immunolocalización de UPK3c en modelos transgénicos de ratón como sistemas modelos de tumorigénesis de vejiga

La existencia de modelos transgénicos de ratón simulando las dos variantes fenotípicas de CU (superficial e invasivo) ha permitido el estudio de la immunolocalización de UPK3c y del resto de uroplaquinas como una aproximación de lo que podría luego ocurrir en CU humanos.

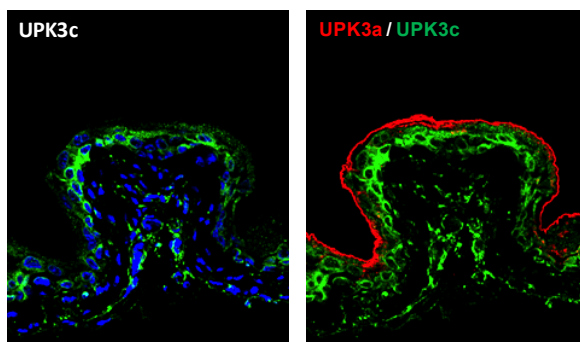


Figura 31: UPK3c se expresa mayoritariamente en las capas del urotelio menos diferenciadas. Immunolocalización de UPK3a en la capa más apical del urotelio sano de ratón mientras que UPK3c es localizada principalmente en las capas inferiores y por tanto menos diferenciadas. Sección de parafina de urotelio de vejiga de ratón sano se incubó con anticuerpos para UPK3c y UPK3a y se detectaron por microscopía confocal. Emisión de fluoróforo para UPK3a fue 617nm (rojo), para UPK3c de 520nm (verde) y para la detección nuclear de 430nm (azul).

El modelo transgénico de ratón que simula la variante papilar no invasiva se produce a través de la expresión específica del oncogén HRAS constitutivamente activado que produce hiperplasia del urotelio y tumores papilares superficiales de bajo grado (Zhang *et al.* 2001).

La variante invasiva se ha obtenido mediante la expresión del oncogén SV40T el cual inactiva tanto p53 como retinoblastoma dando lugar a un carcinoma *in situ* (CIS) el cual tiene una elevada probabilidad de derivar en un carcinoma invasivo y por tanto de alto grado (Zhang *et al.* 1999).

En el caso del modelo transgénico de ratón que simula un CU de bajo grado, se observa una hiperplasia con una parte apical continua donde las uroplaquinas mayoritarias de las placas colocalizan. Estas se expresan también en el citoplasma y en las membranas basolaterales y lo hacen de una forma gradual de tal manera que las capas más bajas no

expresan las uroplaquinas mayoritarias (UPK1a, UPK2a, UPK1b, UPK3a) y minoritaria (UPK3b) de las placas mientras que a medida que las capas celulares son más apicales la expresión de estas es más intensa (Figura 32).

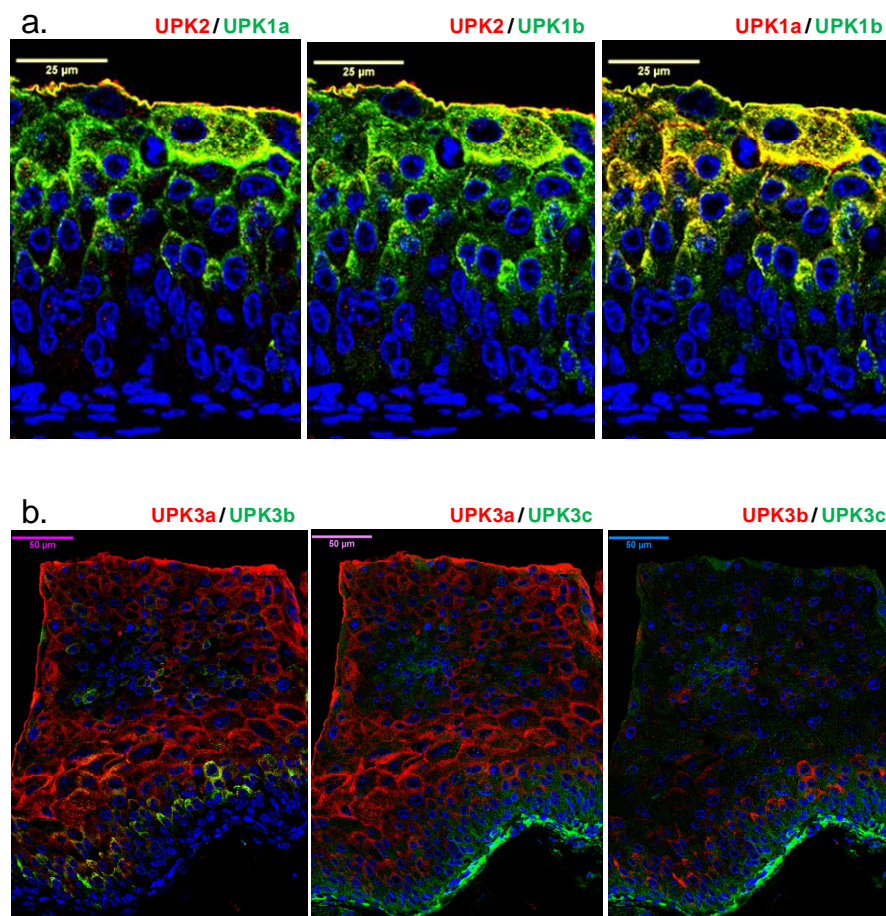


Figura 32: Immunolocalización de las uroplaquinas en urotelio hiperplásico del modelo transgénico de ratón HRAS constitutivamente activado que simula CU de bajo grado. Secciones de parafina de vejiga de ratones transgénicos con la vía de HRAS constitutivamente activada fueron incubados con anticuerpos para todas las uroplaquinas y detectadas por microscopía confocal. La emisión de fluoróforos fue de 430 nm (azul) para la localización nuclear, 520 nm para UPK1b y UPK3c, 617 nm para UPK1a y UPK3b y de 668 nm para UPK2 y UPK3a. **(a)** UPK1a, UPK1b y UPK2 **(b)** UPK3a, UPK3b y UPK3c

Al comparar la expresión de las tres UPK3s se puede observar que UPK3a se expresa solamente en las capas hiperplásicas localizadas en la parte más intermedia y apical. La expresión de UPK3b es aislada, expresándose sobre todo en la parte más baja de la zona hiperplásica intermedia colocalizando con UPK3a, aunque es posible localizarla en la parte más apical donde UPK3a no se encuentra. UPK3c se expresa principalmente en la parte más basal de la hiperplasia mientras que de forma gradual en las capas superiores disminuye la intensidad de su expresión hasta prácticamente desaparecer (Figura 32b).

Aparentemente las expresiones de UPK3a y UPK3c, al igual que como se puede observar en el urotelio sano, son excluyentes. Utilizando un aumento menor y abarcando de esta manera un mayor campo se observa que las partes basales y apicales en la sección están intercaladas permitiendo observar de una manera más clara la aparente expresión excluyente de UPK3a y UPK3c (Figura 33).

El modelo transgénico de ratón que simula un CU de alto grado a través de la expresión del oncogén SV40T, también expresa todas las uroplaquinas aunque su pauta de expresión es diferente a la del modelo de ratón transgénico con la ruta HRAS constitutivamente activada (Figura 34).

En este caso, no existe hiperplasia y los componentes mayoritarios de las uroplaquinas son expresados exclusivamente en la parte apical del urotelio de manera discontinua, habiendo zonas donde la expresión es menor o ausente, lo cual podría poner de manifiesto que la formación de las placas está comprometida. UPK1a como UPK1b y UPK2 poseen una pauta de expresión similar exceptuando que UPK1b podría también encontrarse en las membranas basolaterales de las células (Figura 34a).

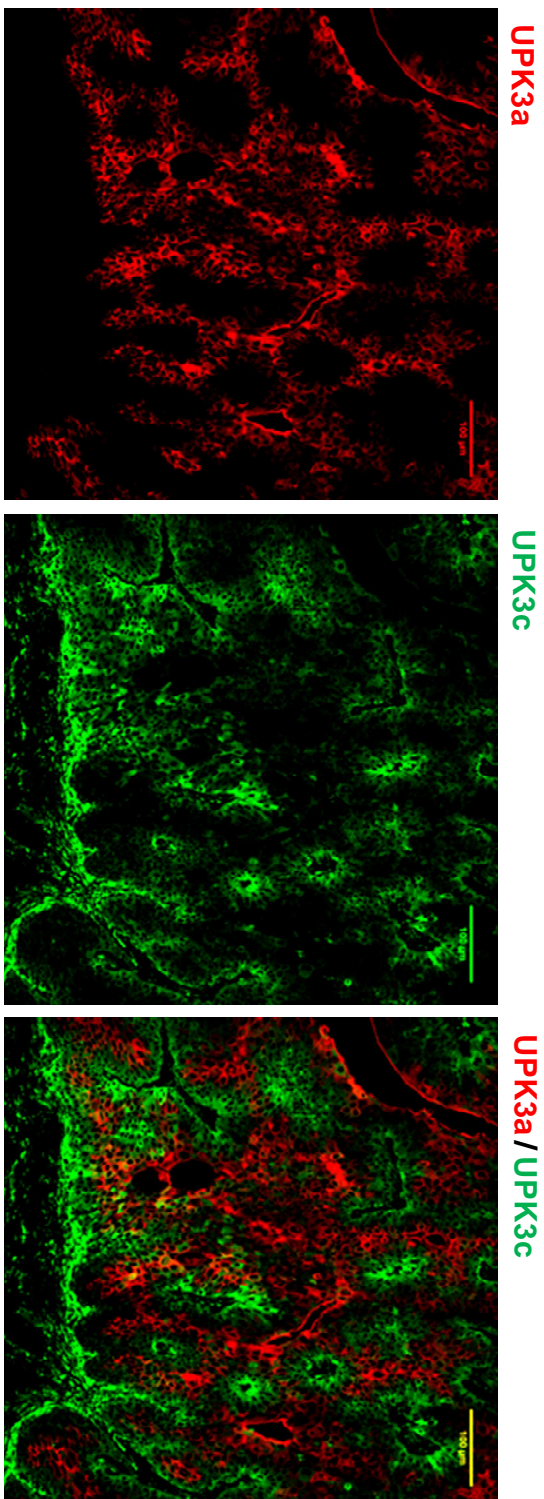


Figura 33: Immunolocalización excluyente de UPK3a y UPK3c en el urotelio hiperplásico de ratones con la vía HRAS constitutivamente activada. Secciones de parafina de vejiga de ratón del modelo transgénico HRAS activado constitutivamente incubado con anticuerpos para UPK3a y UPK3c y detectados mediante microscopia confocal. Emisión de fluoróforo para UPK3c es de 520 nm (verde) y para UPK3a es de 617 nm (rojo).

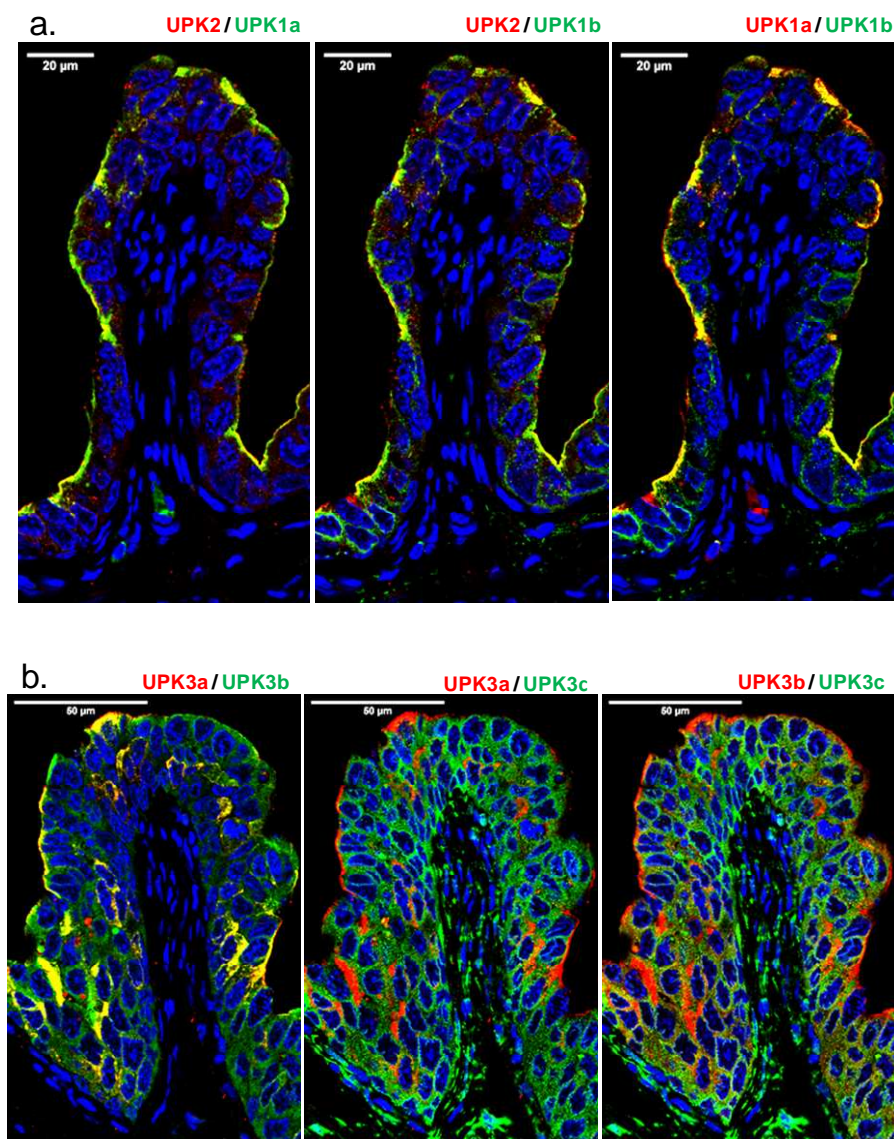


Figura 34: Immunolocalización de la familia de uroplaquinas en el modelo transgénico de ratón SV40T que simula CU de alto grado. Secciones de parafina de vejiga de ratones transgénicos con el oncogén SV40T se incubaron con anticuerpos para todas las uroplaquinas y detectadas por microscopía confocal. La emisión de fluoróforos fue de 430 nm (azul) para la localización nuclear, 520 nm para UPK1b y UPK3c, 617 nm para UPK1a y UPK3b y de 668 nm para UPK2 y UPK3a. **(a)** UPK1a, UPK1b y UPK2; **(b)** UPK3a, UPK3b y UPK3c.

UPK3a podría ser la uroplaquina de los componentes mayoritarios de las placas cuya expresión está más disminuida. Al igual que el resto de uroplaquinas mayoritarias de las placas, se expresa en la parte apical pero de manera muy aislada. Al contrario que UPK3a, UPK3b presenta una expresión más generalizada en todas las capas del urotelio y allá donde se encuentra UPK3a, UPK3b colocaliza con ella al igual que en el modelo transgénico de ratón que simula un CU de bajo grado. Sin embargo, entre ambos modelos transgénicos de ratón, estas dos uroplaquinas intercambian su pauta de distribución. En ambos modelos colocalizan pero en el modelo de bajo grado UPK3a presenta una pauta de expresión más amplia que UPK3b, la cual está bastante restringida y en el modelo de alto grado, la situación es a la inversa, UPK3b presenta una expresión más generalizada que UPK3a.

UPK3c también se encuentra en todas las capas y aunque la expresión es algo mayor en la parte más basal, no existe ese contraste con el resto de las capas como en el urotelio sano o en el modelo transgénico de ratón que simula un bajo grado (Figura 34b).

b.Expresión de UPK3c en tumores de vejiga humano

La expresión de las UPK3s se analizó en un total de 31 tumores de vejiga por WB. Su expresión, además se comparó con la expresión de dos proteínas (RB y p130) implicadas en la regulación del ciclo celular (Grana *et al.* 1998) y de un receptor de membrana (FGFR3) implicado en las vías de señalización que regulan la proliferación celular, diferenciación, migración y apoptosis (Liu *et al.* 2014) (Figura 5). De los 31 tumores analizados, acorde a su estadio de invasión, 11 de ellos era superficiales de bajo grado, 9 superficiales de alto grado y 11 eran invasivos. La

determinación del estadio y del grado de los tumores de vejiga se llevó a cabo por el servicio de Anatomía Patológica del HJXXIII.

En los tumores de vejiga analizados (Figura 35 y Figura 36), la expresión de UPK3a fue la más ubicua de todas las UPK3s siendo detectada en 25/31 tumores, seguida de UPK3b 24/31 y UPK3c en 19/31. De los 20 tumores de alto grado se observó que 6, 6 y 11 no expresaban UPK3a, UPK3b y UPK3c respectivamente,

En ninguno de los 31 tumores de vejiga analizados se pudo observar la ausencia de todas las UPK3s a la vez. Sin embargo, aunque se detectan todas las combinaciones de parejas entre ellas, al contrario que UPK3a y UPK3b, UPK3c nunca aparece como la única uroplaquina detectada.

Parece que exista una relación entre la expresión de FGFR3 y UPK3c. De los 15 tumores de vejiga donde se detectó expresión de FGFR3, en 14 también se detectó UPK3c y de los 16 tumores donde no se detectó expresión de FGFR3, en 12 tampoco se detectó de UPK3c. De los 4 en los que si se detectó expresión de UPK3c y no de FGFR3, 3 eran de alto grado, 2 eran tumores invasivos y únicamente en un tumor, superficial de bajo grado, se detectó expresión de FGFR3 y no de UPK3c.

Para facilitar la detección de UPK3c, todos los tumores fueron deglicosilados lo que permitió observar que solamente 1 presentó la banda de UPK3c sin péptido señal, otro solamente la banda de UPK3c con péptido señal y otro mostró una deglicosilación aberrante dada a la presencia de varias bandas. En los 13 tumores de vejiga restantes se observaron ambas bandas de UPK3c (Figura 35).

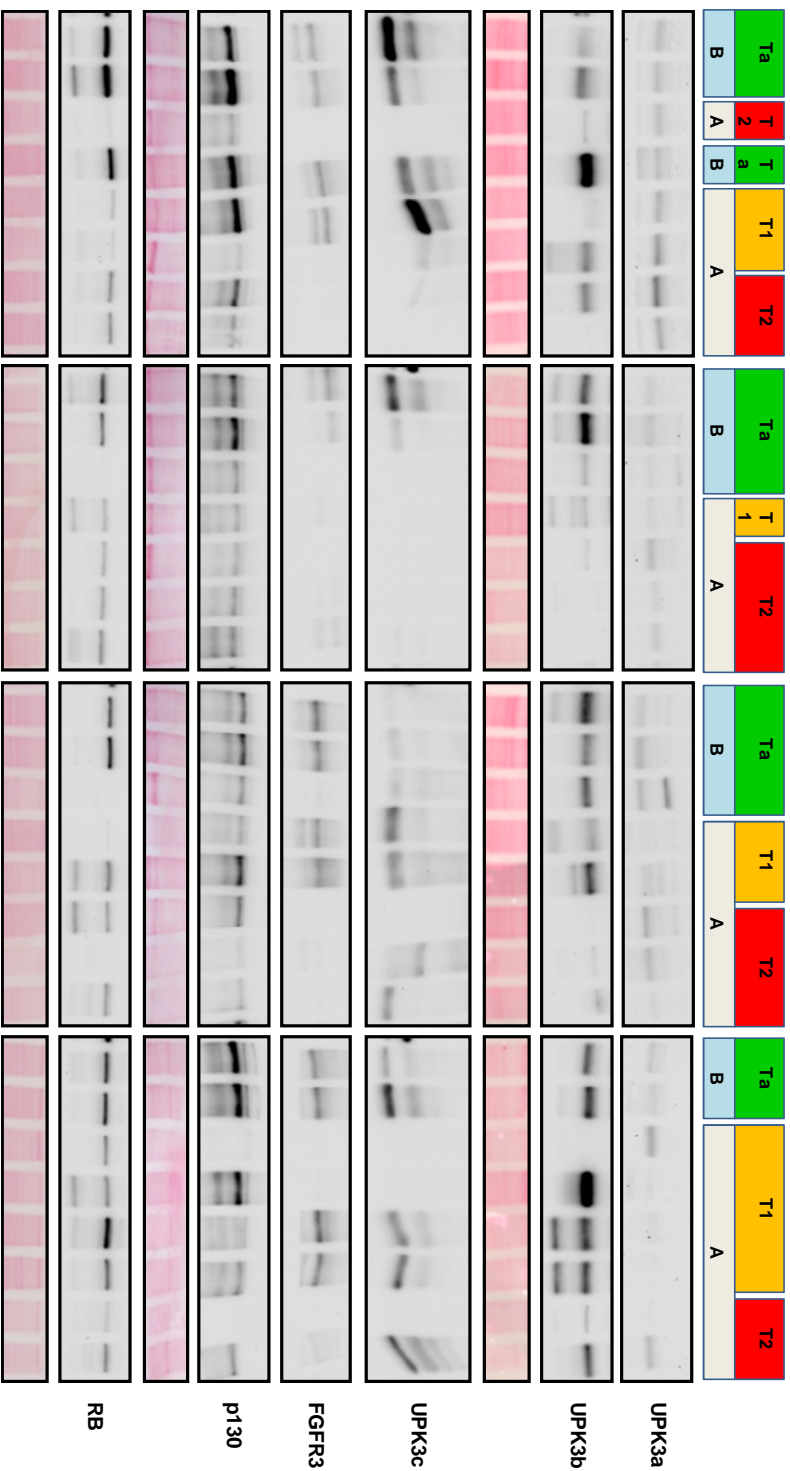


Figura 36: Expresión de UPK3s en tumores de vejiga. Extracto proteico de lisado celular de tumores de vejiga de distintos estadios y grados fueron inmunodetectados con anticuerpos para UPK3a, UPK3b, UPK3c, Rb, p130 y FGFR3. Como control de carga se ha utilizado la membrana de nitrocelulosa teñida con Ponceau S solution. Ta (superficial mucosa); T1 (superficial submucosa); T2 (invasivo); A (alto grado) y B (bajo grado).

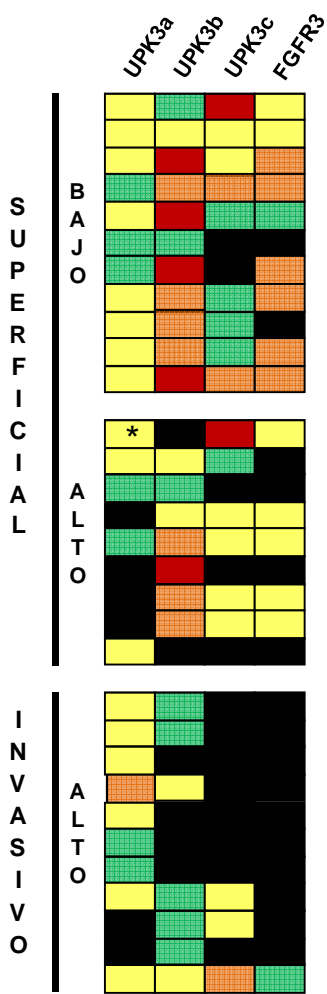


Figura 36: Resumen del análisis de WB.

Los resultados obtenidos para las UPK3s y para FGFR3 por WB en los 31 tumores fueron agrupados por superficial/invasivo y grado y la intensidad de la señal codificada atendiendo a un patrón de colores. Negro: no detección; verde: +/-; amarillo: +; naranja: ++ y rojo: +++. El * representa que el tumor de vejiga es un Ta de alto grado.

Ambas formas de UPK3c no se encontraron en la misma proporción. El $72 \pm 2\%$ pertenece a la forma madura de UPK3c sin péptido señal, mientras que el $28 \pm 2\%$ restante corresponde a la forma inmadura de UPK3c con péptido señal, al contrario de lo que ocurre en la transfección donde el $79 \pm 4\%$ pertenece a la forma inmadura con el péptido señal mientras que el $21 \pm 4\%$ restante corresponde a la forma madura sin péptido señal.

DISCUSIÓN

UNIVERSITAT ROVIRA I VIRGILI

EVOLUTION AND MOLECULAR CHARACTERIZATION OF UROPLAKIN 3C, A NOVEL HUMAN TETRASPANIN ASSOCIATED UROPLAKIN

Javier Ugarte Chicote

LAS UROPLAQUINAS SE ENCUENTRAN DISTRIBUIDAS DE FORMA DIFERENTE EN LOS DIVERSOS GRUPOS DE VERTEBRADOS

Este estudio presenta una visión de conjunto de la evolución y diversificación de probablemente todos los grupos de uroplaquinas que existen en la naturaleza dado que en los análisis se han utilizado secuencias de especies representativas de los principales grupos de vertebrados (peces, anfibios, aves/reptiles y mamíferos) y se han identificado tres nuevos grupos de uroplaquinas (UPK3c, UPK3d y UPK2b).

Nuestros análisis indican que las uroplaquinas durante la diversificación de los vertebrados no se distribuyeron uniformemente en los distintos grupos de estos organismos (Figura 7). Las primeras secuencias de uroplaquinas aparecen en un pez, la lamprea, considerado un fósil viviente perteneciente a la rama más basal de los vertebrados cuando estos carecían de maxilares (Smith *et al.* 2013). Estas primeras uroplaquinas son las que más adelante vamos a encontrar formando los componentes mayoritarios de las placas del urotelio de los mamíferos: UPK1a, 1b, 3a y 2a (Wu *et al.* 1994).

La siguiente rama del árbol evolutivo de los vertebrados extinta desde hace 250 MYA comprendía los primeros peces con mandíbula (placodermos y acantoideos).

Después tuvo lugar la escisión del tronco común de los peces cartilagosos (rayas y tiburones) donde ya encontramos todas las uroplaquinas (UPK1a, 1b, 2a, 2b, 3a, 3b, 3d) a excepción de la UPK3c que aparece en el ancestro común de reptiles/aves y mamíferos (Figura 7 y Figura 8). La aparición de las UPK2b, 3b y 3d en peces cartilagosos podría explicarse de varias formas, bien se generaron en el ancestro común de los peces cartilagosos y óseos o en el de la lamprea y los peces cartilagosos y han desaparecido en lamprea, dado que el genoma de la lamprea a pesar de haber sido secuenciado con un 5 por ciento de

cobertura total aun presenta lagunas. De hecho, en lamprea existen tres secuencias relacionadas con UPK3a y parece que una de ellas (lamprea2) estaría más cerca de UPK3b y por lo tanto UPK3a y 3b se habrían generado en el ancestro común de vertebrados. Algo semejante ocurrió probablemente también con UPK2a y 2b ya que UPK2b posee las cuatro cisteínas del dominio urofibrina y UPK2a solo dos por lo que es poco probable que primero se hubieran perdido en UPK2a y luego recuperado en UPK2b. No obstante, para responder a estas cuestiones, se necesitan más genomas de organismos pertenecientes a esta etapa basal de la evolución de los vertebrados. A este respecto está pendiente la secuenciación del genoma de otro pez sin mandíbula de una clase diferente a la lamprea, el mixino (*Myxine glutinosa*) del que se ha secuenciado el genoma mitocondrial y que sería probablemente de gran ayuda para clarificar la evolución de las uroplaquinas y otras proteínas en esta fase crucial del desarrollo de los vertebrados donde se produjeron numerosas innovaciones específicas de este grupo de organismos (Boehm 2012).

En la siguiente radiación, los peces óseos (osteichthyes) no utilizaron las UPK3b y 2b que ya existían en el ancestro común de peces óseos y cartilaginosos. Los peces óseos se dividen a su vez en dos ramas, la de los peces óseos con aletas lobuladas (sarcopterigios) que incluye celacantos y peces pulmonados que están relacionados con los tetrápodos (anfibios, aves/reptiles y mamíferos) y la rama de los peces óseos con aletas rayadas (actinopterygii). Un hecho sorprendente que ocurrió en esta última rama es la desaparición de todas las uroplaquinas en los teleósteos modernos (acanthomorpha) que abarcan un tercio de las especies actuales de vertebrados ~16000 especies y representan más del 50% de los peces actuales entre los que se encuentran el bacalao (*Gadus morhua*), el pez globo (*Fugu rubripes*) o el atún (*Thunnus thynnus*) (Nelson JS 2006) Teniendo en cuenta que se han secuenciado parcial o totalmente el genoma de 17 especies de peces de este grupo, al no haberse identificado ni una sola uroplaquina en estos 17 genomas ni en ninguna base de ESTs o proteína indica que las uroplaquinas ciertamente desaparecieron durante

el origen evolutivo de este grupo de peces (acanthomorpha) (Figura 7) que incluye el pez modelo medaka (*Oryzias latipes*) cuyo genoma ha sido totalmente secuenciado y en el que anteriormente hemos encontrado diversos ortólogos de proteínas humanas como los de Nanog, Oct-4 y Usherine (Aller *et al.* 2013; Sánchez-Sánchez *et al.* 2010; Camp *et al.* 2009). Esta desaparición conjunta de las uroplaquinas merecería ser estudiada en profundidad dado que la gran expansión de este grupo de peces se debió a adaptaciones específicas (Finn y Kristoffersen 2007) en las que la presencia de uroplaquinas fue prescindible o nociva, lo que podría proporcionar información sobre su posible función general.

Finalmente en tetrápodos (anfibios, aves/reptiles y mamíferos) desaparece UPK3d y nos encontramos de nuevo con UPK3b y 2b, esta última únicamente en anfibios y reptiles. También en aves desaparece la pareja UPK1a y 2a, y en reptiles UPK2b.

La existencia de uroplaquinas en organismos que carecen de urotelio sugiere que además de su participación en la formación de las placas del urotelio y en funciones relacionadas con su funcionamiento, las uroplaquinas deben desarrollar actividades urotelio inespecíficas. A este respecto se ha descrito la participación de UPK3a y UPK1b en la fertilización de los ovocitos de rana (*Xenopus laevis*) (Sakakibara *et al.* 2005; Mahbub Hasan *et al.* 2007) y se ha vinculado una secuencia que se denomina UPK3a-like, y que en nuestro estudio hemos identificado como UPK3d, con la morfogénesis del conducto pronefrico durante el desarrollo del aparato urogenital en el pez cebra (*Danio rerio*) (Mitra *et al.* 2012).

Por otra parte, en el urotelio, las uroplaquinas forman heterodímeros específicos (UPK1a con UPK2a y UPK1b con UPK3a o la minoritaria 3b) imprescindibles para que las uroplaquinas alcancen la superficie apical de las células “umbrella” (Tu *et al.* 2002). Excepto UPK1b que puede alcanzar la membrana apical por si sola (Tu *et al.* 2002). La distribución de las uroplaquinas en vertebrados y su expresión en tejidos

humanos no se opone a la formación de estos dímeros (UPK1a/2a y UPK1b/3a) ya que UPK1a y UPK1b se originaron en el mismo periodo de tiempo que UPK2a y UPK3a. En los diferentes grupos de vertebrados siempre se mantiene la existencia de las uroplaquinas susceptibles de formar heterodímeros, con la salvedad de que quizás en reptiles no existe UPK2a que podría ser reemplazada por UPK2b. Esta afirmación es meramente hipotética y necesitaría confirmación experimental. La expresión observada en tejidos humanos avalaría la formación de heterodímeros excepto por la falta de expresión de UPK1b en próstata y de UPK2 en tráquea pero el análisis de expresión es únicamente a nivel de RNA (Figura 22).

La distribución de las uroplaquinas y su expresión no se opone a la formación de dímeros, pero es posible que no sea una característica compartida por todas las uroplaquinas. La expresión de UPK1b en diversos tejidos sin UPK3a de heterodímero podría explicarse como hemos comentado por su capacidad de alcanzar la membrana apical sin necesidad de formar un heterodímero con UPK3a o 3b (Tu *et al.* 2002). Esta expresión fuera del urotelio ya en algunos casos había sido descrita con anterioridad (Sakakibara *et al.* 2005; Olsburgh *et al.* 2003; Adachi *et al.* 2001) El análisis de la colocación en el urotelio de las uroplaquinas que forman los heterodímeros también avala su formación pero descarta a UPK3c como veremos más adelante.

En conjunto parecería que las placas del urotelio surgen por la asociación de los dos tipos de heterodímeros y que algunas uroplaquinas podrían existir sin formar heterodímeros y estar implicadas en funciones que desconocemos.

ORIGEN DE LAS UROPLAQUINAS UPK2/3 A PARTIR DE UNA SECUENCIA RELACIONADA CON PTPRQ

Mientras que el origen de las uroplaquinas UPK1a/1b, que pertenecen a la familia de las tetraspaninas, podía explicarse por la duplicación de un gen de esta familia en el origen de los vertebrados, se desconocía cómo se originó el gen ancestral que dio lugar a las uroplaquinas UPK2/3, ya que no se habían encontrado proteínas UPK2/3 en organismos más antiguos que los vertebrados, ni se sabía que pudieran estar relacionadas con otras familias de proteínas. Fue la observación de que un fragmento del receptor de fosfatasa PTPRQ presentaba cierta similitud con casi toda la secuencia de UPK3a, lo que planteó la posibilidad de que las UPK2/3s pudieran haberse generado a partir de un gen relacionado con las R3 PTPR por duplicación génica y pérdida de dominios FN3 y del dominio catalítico.

Esta sugerencia está basada en la conservación en ambas familias de la misma pauta de intrones/exones, de cisteínas (C₁-C₄), de los sitios de corte de furina en posiciones similares en las secuencias de UPK2s y R3 RPTPs de invertebrados y, por otra parte, en los análisis filogenéticos, que agrupan en una misma rama del árbol las secuencias de UPK3s, PTPRQs y las dos secuencias de *Ciona*: la del precursor de PTPRQ y la secuencia de R3 que carece de dominio catalítico.

Esta última secuencia sin dominio catalítico y con solo cuatro dominios FN3 presente en las dos especies de *Ciona* (CionaNocat1 en *C. intestinalis* y CionaNocat2 en *C. savignyi*) curiosamente, mantiene cierta similitud en su fragmento citoplásmico con UPK3as, incluyendo dos sitios de fosforilación: una tirosina (Tyr249) que es fosforilada en *Xenopus* durante la fertilización de los ovocitos y una treonina (Thr244) en humanos en respuesta a la adhesión de la uropatogénica *E. coli* al urotelio (Mahbub Hasan *et al.* 2011; Thumbikat *et al.* 2009). Estas similitudes no implican que esta secuencia de *Ciona* sea el precursor de las UPK2/3 (falta el análisis de

sintenia de genomas basales mejor anotados) pero probablemente indica como debió producirse la transición entre alguna secuencia relacionada con el precursor de PTPRQ y la formación del primer gen de las UPK2/3. Los principales mecanismos que determinan la evolución molecular de las proteínas son la duplicación de genes, la fusión, la recombinación y la pérdida de fragmentos (Weiner *et al* 2006). Aunque en comparación con la ganancia de dominios, se sabe relativamente poco acerca de las pérdidas, se ha descrito que se producen mayoritariamente en los extremos de las secuencias como debió producirse para generar la secuencias UPK2/3 ancestral a partir de una secuencia R3 PTPR de invertebrados (Weiner *et al.* 2006).

1.El dominio urofibrina es un dominio FN3 divergente

Aunque los dominios FN3 de por si son muy divergentes y tienen una identidad de secuencia inferior al 25% (Halaby *et al.* 1999), el dominio urofibrina tiene características específicas que se han conservado desde su aparición en la secuencia de esponja. Por un lado, los dominios canónicos de FN3 tienen alrededor de 100 aa de longitud, el tamaño del dominio urofibrina tiene alrededor de 160 aa y el motivo $C_3X_{(5-7)}C_4$ NGPL se ha mantenido en las R3 RPTPs y UPK2/3 desde las esponjas hasta los mamíferos no encontrándose en otro tipo de proteínas. Por el otro lado, presenta un patrón exón/intron único entre los dominios de FN3 de las R3 PTPRs (García-España *et al* en preparación) y carece de un residuo clave en la definición de los dominios FN3 canónicos, un triptófano en la hebra beta B (Halaby *et al.* 1999). Todo ello, sugiere que el dominio urofibrina tiene características diferenciales con respecto a los dominios FN3 canónicos como el dominio 10 de la fibronectina. Aunque no se conoce su estructura terciaria, por su similitud en estructura secundaria con los dominios FN3, probablemente este basada también en láminas beta antiparalelas pero con diferencias en la longitud de la secuencia que

conecta las hebra beta C y C' (Figura 14) (Halaby *et al.* 1999; Koide *et al.* 1998).

2.Posibles implicaciones funcionales resultantes de la caracterización del dominio urofibrina

Los principales mecanismos que determinan la evolución molecular de las proteínas son la duplicación de genes, la fusión, la recombinación y la pérdida de fragmentos (Weiner *et al.* 2006). La descripción de los componentes dominios y motivos de una proteína facilitan los análisis filogenéticos, la anotación funcional, y la exploración de la diversidad de sus funciones. El hallazgo de la relación entre las proteínas R3 PTPRs y UPK2/3 y la caracterización de un dominio FN3 divergente en ambas familias puede ser de utilidad para la caracterización estructural y funcional de estos grupos de proteínas integrales.

Las cisteínas que forman puentes disulfuro se ganan y se pierden en parejas durante la evolución de las proteínas y por tanto es posible deducir la formación de puentes disulfuro en su secuencia (DeSalle *et al.* 2010; Kreisberg *et al.* 1995). De las cuatro cisteínas (C₁-C₄) presentes en el dominio urofibrina, dos de ellas C₃-C₄ están conservadas desde la secuencia de esponja pero se pierden en las proteínas UPK2a, por otro lado C₁-C₂ aparecen en las de tunicados y aunque todavía están presentes en la secuencia de UPK3c de aves/reptiles, se pierden en las secuencias de UPK3c de mamíferos (DeSalle *et al.* 2014). Estas observaciones sugieren que C₁ está formando un puente disulfuro intraproteico con C₂ y C₃ con C₄ en la estructura de la urofibrina lo que puede ayudar en la resolución de su estructura

LOCALIZACIÓN DE UPK3c EN DS Y SU POSIBLE TRASLOCACIÓN AL LOCUS GENÉTICO 7p13 A TRAVÉS DE UN INTERMEDIARIO CIRCULAR

Las duplicaciones segmentales (DS) constituyen una proporción relativamente grande (~5%) del genoma humano (Bailey *et al.* 2006; Bailey *et al.* 2002) y son definidas como dos o más segmentos cromosómicos mayores de 1 Kb, con una identidad superior al 90% (Bailey *et al.* 2006). Además, son regiones dinámicas de los cromosomas que en determinados casos se han asociado con enfermedades genéticas producidas por inestabilidad genómica (Marques-Bonet *et al.* 2009).

Los mecanismos moleculares por los cuales se producen y duplican las DS, su importancia en la susceptibilidad a padecer enfermedades y como influyen en el desarrollo de estas patologías, esta empezando a ser desvelado (Marques-Bonet *et al.* 2009, Jiang *et al.* 2007, Bailey *et al.* 2006). Especialmente, se conoce poco el mecanismo por el cual las DS se dispersan a las regiones intersticiales de los cromosomas cuando esta propagación no es el resultado de una duplicación en tándem, es decir cuando las DS no se localizan de forma consecutiva en los cromosomas (Bailey *et al.* 2006).

Recientemente se ha descrito un aumento específico en el número de DS en la rama evolutiva humana después de la separación de la del chimpancé (Marques-Bonet *et al.* 2009). Esta observación estaría de acuerdo, en nuestro caso, con la aparición en humanos de la DS3 en el locus genético 7q22.1 (Figura 17). Esta copia adicional de UPK3c podría tener interés en su estudio puesto que, al contrario que la copia en la DS2 de 7p13, se transcribe al igual que la copia en DS1, aunque al parecer la transcripción de la copia en DS1 es mayor (Figura 21). Una de las consecuencias de las duplicaciones de DS es que pueden aumentar o disminuir el número de copias de un gen (Lupski *et al.* 2005). En la DS2

humana, RASA4CP, POLR2J4 y las copias de UPK3c y HSPC047 son pseudogenes mientras que SPDYB2-L2 es un gen diferente a SPDYB2-L1 que se encuentra en las DS1 y DS3 de 7q22.1 (Chauhan *et al.* 2012). En el caso de gorila y chimpancé las regiones correspondientes no están tan bien anotadas como para poder saber si esto también ocurre. SPDYB2-L1 y SPDYB2-L2, junto a SPDYB2, pertenecen a un grupo de la amplia familia de genes *Speedy/RINGO cell cycle regulator* (Chauhan *et al.* 2012) y múltiples copias intercaladas de miembros de esta grupo (especialmente SPDYB2-L1) pueden encontrarse en 7q11.23 de humanos. Siempre asociadas a un fragmento de la secuencia del gen PMS2

En orangután, ningún miembro de este grupo de la familia Speedy se encuentra en la localización en que aparecerá en las DS1-3, lo que sugiere que tal vez su desplazamiento a la posición entre los genes *UPK3c* y uno de los nuevos miembros de POLR2J pudo estar relacionado con el origen de la DS1-3 ancestral. En este sentido, la reconstrucción ancestral de DS en humanos ha revelado que la mayoría de ellas están localizadas en regiones específicas del genoma denominadas bloques de duplicación (Jiang *et al.* 2007). Estos bloques contienen regiones denominadas “core duplicons” a las cuales se les ha atribuido un papel primordial en la expansión de las duplicaciones dando lugar a la hipótesis del “core duplicon” (Jiang *et al.* 2007). De esta manera el fragmento de DNA que contiene SPDYB2-PMS2 podría representar un “core duplicon” que pudo haber desarrollado un papel importante en el origen y/o dispersión de las DS1-3

Los cambios de localización de UPK3c desde orangután en el locus genético 7p22.1 a las localizaciones que luego encontramos en humanos en las DS1 y DS3 de 7q22.1 se pueden explicar por las dos inversiones del cromosoma 7 que ocurrieron durante la especiación de los grandes primates (Figura 17) (Müller *et al.* 2004), más una duplicación específica en tándem de DS1 que generó DS3 en humanos después de la separación con chimpances probablemente por un proceso de errores en la replicación

y recombinación homóloga no alélica (Schacherer *et al.* 2005, Koszul *et al.* 2004). En cambio, el desplazamiento de DS2 a 7p13 en el otro brazo del cromosoma que debió producirse en el ancestro común de gorilas chimpancés y humanos no puede explicarse por los mecanismos arriba mencionados de inversiones y duplicación en tándem por recombinación homóloga.

Curiosamente, el desplazamiento de DS2 a 7p13 pudo haberse producido por un mecanismo consistente con la participación de un intermediario circular, un fenómeno de traslocación de una DS que sepamos solo se ha descrito una vez en mamíferos y nunca en primates (Durkin *et al.* 2012). Esta sugerencia se basa en la presencia en el sitio donde se insertó DS2 de secuencias conservadas en gorila, chimpancé y humano compatibles con un intermediario circular. Primero, el gen *RASA4* en DS2 está dividido en dos fragmentos situados en los extremos de la inserción, si estos fragmentos se unen se reconstruye la secuencia de *RASA4* con únicamente la pérdida de 4 pb en el punto de unión, y segundo, esta pérdida de 4 pb así como la de 2 pb en la secuencia en la cual se inserta pueden ser explicadas por un mecanismo de rotura y reparación de DNA que ha dejado su impronta en el punto de inserción (Lieber *et al.* 2008). Esto indicaría que el gen *RASA4* se ha roto durante el proceso de inserción y sugiere la intervención de un intermediario circular.

La recombinación homóloga entre elementos Alu intracromosomales que comparten orientación transcripcional (ARMD) es un mecanismo por el cual se ha explicado la formación de intermediarios circulares (Sen *et al.* 2006). Como resultado de la ARMD quedarían dos elementos Alu quiméricos, uno en la zona de escisión del círculo y otro en el círculo. Basándonos en este mecanismo como posible explicación para la formación del intermediario circular, de los 122 elementos Alu localizados en la DS2, dos de ellos han podido ser identificados como elementos Alu quiméricos. En la sección de resultados hemos descrito uno de estos dos posibles escenarios (Figura 18 y Figura Suplementaria 5). En el caso de

que la ARMD hubiese ocurrido entre elementos Alu iguales, su identificación sería complicada. No han quedado señales del sitio de salida del círculo en humanos posiblemente porque los cambios que se han producido desde el ancestro común de gorila, chimpancé y humano donde se produjo la translocación de DS2 a 7p13 han eliminado estas señales. Los loci genéticos 7q11.23 en gorila y chimpancé están mal anotados y sería interesante poder estudiarlos debido a que representan los loci previo y posterior más cercanos a la inversión paracéntrica la cual en comparación con lo encontrado en orangután tuvo que ocurrir entre los genes *UPK3b* y *UPK3c* y que podrían conservar alguna impronta sobre la salida del IC o la disposición de las DS en el momento de la salida debido a que DS2 se encuentra ya en gorila, es decir, antes de la inversión paracéntrica.

En una búsqueda preliminar en la base de DS de la UCSC no hemos podido identificar ninguna DS que contenga un fragmento invertido con respecto a otra DS, es decir, la presencia de un bloque de duplicación que estuviera reflejado en dos bloques invertidos en otra localización. Sin embargo, además de en el genoma de la vaca (Durkin *et al.* 2012), este tipo de fenómeno ha sido registrado con anterioridad en peces (Fujimura *et al.* 2011) y levaduras (Borneman *et al.* 2011). De esta manera, aunque este mecanismo de translocación no parece ser muy frecuente, ha podido ser registrado en organismos de diferente complejidad poniendo de manifiesto que podría ser un mecanismo común a todos ellos. Sin embargo, un análisis más detallado sería necesario para poder aproximar la frecuencia de este fenómeno.

La zona de 7p13 donde se produce la inserción de la DS2 es una zona rica en elementos transponibles (93% de las 1000 pb de bases adyacentes a ambos extremos del punto de inserción) tal como ha sido observado con anterioridad (Kapitonov *et al.* 2005, Babcock *et al.* 2003, Bailey *et al.* 2003). Su inserción recae exactamente sobre un solo-Long Terminal Repeat de un retrovirus endógeno humano. Se ha podido comprobar que solo-LTRs pueden actuar sobre la regulación génica

(Gogvadze y Buzdin 2009; Dunn *et al.* 2005; van de Lagemaat *et al.* 2003; Long *et al.* 1998), y en el caso de que este hubiese sido activo en el momento de la inserción de la DS2, esta lo hubiese dividido pudiendo haberse visto afectada, de esta manera, la regulación de algún gen.

La microhomología de 5pb del extremo DS2-7p13 posee a ambos lados un elemento Alu, los cuales están inversamente orientados, próximos (<200 pb) y prácticamente equidistantes de la microhomología. En este caso, dadas sus características, podría ser que estos dos elementos Alu hubiesen interaccionado de manera no recombinogénica facilitando la inserción del intermediario circular ya que antes de producirse la inserción estarían orientados en la misma dirección

EXPRESIÓN DE LAS UROPLAQUINAS EN TEJIDOS HUMANOS Y DIVERGENCIA DE UPK3c RESPECTO A UPK3a

En el estudio que hemos llevado a cabo, se ha estudiado por primera vez la expresión de todas las uroplaquinas humanas en un amplio número de tejidos utilizando la misma técnica, obteniendo de esta manera, además de información específica de cada una de ellas una visión más completa de su expresión en los distintos tejidos.

La expresión simultánea de las uroplaquinas mayoritarias de las placas (UPK1a, 1b, 2a y 3a) se ha observado que está restringida a los órganos con urotelio, respetando de esta manera la formación de heterodímeros como ha sido observado anteriormente (Wu *et al.* 1994, Deng *et al.* 2002) y como se ha podido intuir también a través de su distribución en los diferentes grupos de vertebrados (Figura 7). La única de estas uroplaquinas que se expresa fuera del urotelio es UPK1b que se expresa en cornea como ya había sido descrito (Olsburgh *et al.* 2003, Adachi *et al.* 2001): Probablemente debido a su capacidad de alcanzar la membrana apical por si sola (Tu *et al.* 2002) en cornea no se detecte expresión de UPK3a o UPK3b con las que forma heterodímeros en el urotelio. Por otro lado, la elevada expresión de UPK3a en próstata sobresale frente a la del resto de uroplaquinas lo que indica que podría ser de interés un estudio más detallado de esta proteína tanto en tejido sano como en cáncer de próstata.

UPK3c y UPK3b muestran un patrón de expresión divergente frente al resto de las uroplaquinas ya que se expresan prácticamente en todos los tejidos analizados. Este patrón más amplio había sido ya descrito en el caso de UPK3b, que por ejemplo, debido a su elevada actividad epicardial durante el desarrollo embrionario del ratón ha permitido utilizar su promotor para generar ratones transgénicos con expresión en el epicardio (Huang *et al.* 2012).

UPK3c y UPK1b en cornea mostraron el nivel de expresión más elevado de todas las uroplaquinas en todos los tejidos. De hecho son las únicas uroplaquinas que han sido identificadas en los proteomas de cornea de ratón y de humano, (Dyrlund *et al.* 2012; Shao *et al.* 2011), Aunque en el proteoma humano, que está dividido en las diferentes capas que forman la cornea, UPK1b solo se detectó en el epitelio mientras que UPK3c fue detectada en todas las capas (epitelio, estroma y endotelio) (Dyrlund *et al.* 2012). Dado que el origen de UPK3c coincide con la colonización del medio terrestre por los vertebrados, su elevada expresión en cornea podría estar relacionada con la adaptación que sufrió la cornea en este periodo que le permitió un mayor poder de refracción para acomodar el paso de la luz del aire al medio acuoso intraocular. (Jonasova *et al.* 2008).

UPK3c está N-glicosilada, y al menos en cornea sana, por azúcares complejos como UPK3a/3b. No obstante, UPK3c posee una serie de características que hacen considerarla una uroplaquina divergente frente al resto de UPK3s. Además de su expresión en un elevado rango de tejidos, por un lado, ha perdido las C₁ y C₂, presentes en UPK3a/3b, y que como ha sido sugerido previamente podrían formar un puente disulfuro intraproteico. Por otro lado, ha perdido también el fragmento de ~12 amino ácidos muy conservado en UPK3a/3b yuxtapuesto al DTM y que, aunque no se haya visto que sea necesario para la formación de los heterodímeros con UPK1b, no se ha descartado que pudiese estar vinculado con la oligomerización de las partículas de 16 nm (Deng *et al.* 2002). Finalmente, en el urotelio de ratón posee una expresión excluyente respecto a la del producto de diferenciación UPK3a (Wu *et al.* 1994). Mientras que UPK3a se localiza en la parte más apical de las células “umbrella”, UPK3c es expresada principalmente en las capas menos diferenciadas, y aunque se exprese en las células “umbrella”, aparentemente no llega a la membrana apical y no colocaliza con UPK1b como lo hacen UPK3a y UPK3b.

EXPRESIÓN DE UPK3c EN TUMORES DE VEJIGA Y SU POSIBLE RELACIÓN CON FGFR3

En la actualidad el trabajo con el tamaño muestral más grande de tumores de vejiga e información clínica que ha estudiado la expresión de uroplaquinas es el de Huang *et al.* 2007. En este trabajo se analizó la expresión de uroplaquinas por inmunohistoquímica con un anticuerpo generado contra las placas y que por inmunoblot es muy sensible a UPK3a, moderadamente a UPK1a y muy poco sensible a UPK2. Las principales conclusiones del estudio de Huang *et al.* 2007 fueron que: (1) aunque aberrante, la expresión de uroplaquinas se mantiene en los CU; (2) los tumores invasivos sin expresión de uroplaquinas representaban un mal pronóstico; y (3) no existía relación alguna con marcadores del ciclo celular (p53 y pRB entre otros) y por tanto podría utilizarse de manera conjunta con otros marcadores moleculares.

En nuestro estudio hemos comparado la expresión proteica de las UPK3s con la de dos marcadores de desregulación del ciclo celular (p53 y RB) implicados en la ruta invasiva del CU (Figura 5) y con una glicoproteína de membrana perteneciente a la familia de las tirosinas quinasas (Powers *et al.* 2000; Böttcher y Niehrs 2005) y considerada un marcador de buen pronóstico en los tumores superficiales (Figura 5).

Al igual que en Huang *et al.* 2007, se encontró expresión de unas u otras uroplaquinas en todos los tumores, siendo la expresión más ubicua la de UPK3a. Por lo tanto, de acuerdo con estudios anteriores (Huang *et al.* 2007; Olsburgh *et al.* 2003; Yuasa *et al.* 2003; Moll *et al.* 1995), aunque a las uroplaquinas de las placas UPK3a/b se las considere productos de diferenciación, no se observa la esperada disminución de su expresión durante la progresión de los CUs humanos. En cambio, UPK3c no siendo aparentemente un producto de diferenciación, desaparece en algunos tumores de alto grado superficiales y en casi todos los invasivos mientras

que la expresión de UPK3a se mantiene. Esto podría estar relacionado con la excluyente expresión de UPK3a y UPK3c en el urotelio de ratón que daría lugar a una distribución también diferente en tumores invasivos.

En ninguno de los tumores analizados se detectó como única uroplaquina a UPK3c, mientras que UPK3a o UPK3b sí que pudieron ser detectadas aisladamente. Esto podría ser debido a que aunque en cantidades distintas, UPK3a y UPK3b se encuentran en las placas y en los modelos de ratones “knock-out” de UPK3a la expresión transcripcional de UPK3b aumenta (Deng *et al.* 2002) dando a entender que UPK3b podría suplantar la función de UPK3a en su ausencia y por tanto podría ser normal encontrar UPK3a o UPK3b por sí solas. Sin embargo, dado que UPK3c aparentemente no forma parte de las placas no tendría por qué encontrarse sola.

La falta de relación en la expresión de uroplaquinas y marcadores de la desregulación del ciclo celular (p53 y RB) observada previamente se detectó en este caso también para UPK3a (Huang *et al.* 2007) y para UPK3b y UPK3c.

En UPK3c y FGFR3 parece apreciarse una relación directa en la detección de su expresión. De los 31 tumores analizados en 26 de ellos la detección para ambas proteínas coincidió (14 ambas detectadas y 12 no detectadas ambas), 1 se detectó FGFR3 y no UPK3c y 4 se detectó UPK3c y no FGFR3.

FGFR3 tiene atribuido un efecto de buen pronóstico debido a que en las lesiones benignas y los tumores de bajo grado el 80% están mutados para FGFR3 (Hernández *et al.* 2006; van Rhijn *et al.* 2002, Kimura *et al.* 2001) y de estos el 85% están sobreexpresados (Tomlinson *et al.* 2007; Bodoor *et al.* 2010) mientras que en el caso de tumores invasivos, el 54% presenta una desregulación de FGFR3 bien a través de mutaciones o de sobreexpresión, siendo esta última mayoritaria del FGFR3 sin mutar (Tomlinson *et al.* 2007).

Aunque no hemos analizado en este caso concreto las mutaciones de FGFR3, indistíntamente se puede intuir una relación directa en la detección de UPK3c y FGFR3 y sería necesario analizar un tamaño muestral mayor para poder comprobar si esta tendencia se mantiene. Sin embargo, en principio, a la vista de estos resultados preliminares se podría sugerir que la detección de UPK3c podría ser un marcador de buen pronóstico.

.

.

CONCLUSIONES

1. Uroplakins originated in the common ancestor of vertebrates and diverged by gene duplication during vertebrate evolution.
2. Identification of three novel groups of uroplakin orthologs (UPK2b, UPK3c and UPK3d), all belonging to the UPK2/3 tetraspanin associated family.
3. UPK2/3 tetraspanins are evolutionary and structurally related to the more ancient family of subtype R3 tyrosine phosphatase receptors (R3 PTPRs); probably origin of UPK2/3 genes from a R3 PTPR related gene (PTPRQ) by gene duplication and domain loss in the common ancestor of vertebrates.
4. The origin of the UPK3c is linked to the duplication of the *UPK3b* gene in the common ancestor of reptiles and mammals.
5. Expansion of the *UPK3c* genetic loci in gorillas, chimpanzees and humans by the duplications of a large segmental duplication generated in the common ancestor of the gorilla, chimpanzee and human branch.
6. Identification of a 80.6 Kb intrachromosomal transposition of a *UPK3c* segmental duplications by a circular intermediate.
7. The expression of UPK3c, which is widespread in human tissues, is particularly strong in corneal tissue.
8. UPK3c contains up to 20 KDa of complex N-linked oligosaccharides in human normal corneal tissue, pterygium and bladder tumors.
9. In contrast with the localization of the uroplakins that form the urothelial plaques (UPK1a, UPK1b, UPK2a, UPK3a and UPK3b), UPK3c is not present in the apical surface of the umbrella cells and localizes mostly to the basal and intermediate cell layers of the mouse urothelium.
10. UPK3c maintains its predominant expression in less differentiated basal and intermediate cells and absence of colocalization with other human uroplakins in urothelial tumors of mouse transgenic models of superficial and invasive bladder cancer.

11. In a sample of human bladder tumors, the number of tumors in which UPK3c expression could be detected by immunoblot was higher in superficial low grade tumors, decreases in superficial high grade and invasive tumors.

BIBLIOGRAFÍA

UNIVERSITAT ROVIRA I VIRGILI

EVOLUTION AND MOLECULAR CHARACTERIZATION OF UROPLAKIN 3C, A NOVEL HUMAN TETRASPANIN ASSOCIATED UROPLAKIN

Javier Ugarte Chicote

Abascal F, R. Zardoya et al. (2010). "TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations." *Nucleic Acids Res.*;38(Web Server issue):W7-13

Acharya, P., J. Beckel, et al. (2004). "Distribution of the tight junction proteins ZO-1, occludin, and claudin-4, -8, and -12 in bladder epithelium." *Am J Physiol Renal Physiol* 287(2): F305-18.

Adachi, W., K. Okubo, et al. (2001). "Human uroplakin Ib in ocular surface epithelium." *Invest Ophthalmol Vis Sci* 41(10): 2900-5.

Aller, E, AV. Sánchez-Sánchez , et al. (2013) "Analysis of the Ush2a gene in medaka fish (*Oryzias latipes*)." *PLoS One*. 8(9):e74995

Alroy, J., F. B. Merk, et al. (1982). "Membrane differentiation in the Golgi apparatus of mammalian urinary bladder epithelium." *Anat Rec* 203(4): 429-40.

Amores, A., J. Catchen J et al. (2011) "Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication." *Genetics*. 188(4):799-808.

Andersen, JN., OH. Mortensen, et al (2001)." Structural and Evolutionary Relationships among Protein Tyrosine Phosphatase Domains." *Molecular and Cellular Biology*.;21(21):7117-7136.

Apodaca, G. (2004). "The uroepithelium: not just a passive barrier." *Traffic* 5(3): 117-28.

Apodaca, G., S. Kiss, et al. (2003). "Disruption of bladder epithelium barrier function after spinal cord injury." *Am J Physiol Renal Physiol* 284(5): F966-76.

Babcock, M, A. Pavlicek, et al. (2003) "Shuffling of Genes Within Low-Copy Repeats on 22q11 (LCR22) by *Alu*-Mediated Recombination Events During Evolution." *Genome Research*. 13(12):2519-2532.

Bailey, J. A. and E. E. Eichler (2006). "Primate segmental duplications: crucibles of evolution, diversity and disease." *Nat Rev Genet* 7(7): 552-64.

Bailey, JA, G. Liu , et al. (2003). "An *Alu* Transposition Model for the Origin and Expansion of Human Segmental Duplications". *American Journal of Human Genetics*. 73(4):823-834.

Bailey, JA.,Z. Gu Z et al. (2002) "Recent segmental duplications in the human genome." *Science* 9;297(5583):1003-7.

Blom, N., S. Gammeltoft, et al. (1999). "Sequence and structure-based prediction of eukaryotic protein phosphorylation sites." *J Mol Biol* 294(5): 1351-62.

Bodoor, K., A. Ghabkari, et al. "FGFR3 mutational status and protein expression in patients with bladder cancer in a Jordanian population." *Cancer Epidemiol* 34(6): 724-32.

Boehm T, (2012) "Evolution of vertebrate immunity." *Curr Biol*. 11;22(17):R722-32.

Borneman, A. R., B. A. Desany, et al. "Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*." *PLoS Genet* 7(2): e1001287.

Bos, J. L. (1989). "ras oncogenes in human cancer: a review." *Cancer Res* 49(17): 4682-9.

Bottcher, R. T. and C. Niehrs (2005). "Fibroblast growth factor signaling during early vertebrate development." *Endocr Rev* 26(1): 63-77.

Boucheix, C. and E. Rubinstein (2001). "Tetraspanins." *Cell Mol Life Sci* 58(9): 1189-205.

Byrum, C. A., K. D. Walton, et al. (2006). "Protein tyrosine and serine-threonine phosphatases in the sea urchin, *Strongylocentrotus purpuratus*: identification and potential functions." *Dev Biol* 300(1): 194-218.

Camp, E., AV. Sánchez-Sánchez AV, et al. (2009) "Nanog regulates proliferation during early fish development." *Stem Cells*. (9):2081-91

Chaerkady, R., H. Shao, et al. (2013) "The keratoconus corneal proteome: loss of epithelial integrity and stromal degeneration." *Journal of proteomics*.;87:122-131

Chauhan, S., X. Zheng, et al. (2012) "Evolution of the Cdk-activator Speedy/RINGO in vertebrates." *Cell Mol Life Sci* 69(22): 3835-50.

Conrad, B. and S. E. Antonarakis (2007). "Gene duplication: a drive for phenotypic diversity and cause of human disease." *Annu Rev Genomics Hum Genet* 8: 17-35.

Corpet, F. (1988). "Multiple sequence alignment with hierarchical clustering." *Nucleic Acids Res* 16(22): 10881-90.

Dehal, P. and J. L. Boore (2005). "Two rounds of whole genome duplication in the ancestral vertebrate." *PLoS Biol* 3(10): e314.

Delsuc, F., H. Brinkmann, et al. (2006). "Tunicates and not cephalochordates are the closest living relatives of vertebrates." *Nature* 439(7079): 965-8.

Deng, F. M., F. X. Liang, et al. (2002). "Uroplakin IIIb, a urothelial differentiation marker, dimerizes with uroplakin Ib as an early step of urothelial plaque assembly." *J Cell Biol* 159(4): 685-94.

DeSalle, R., TT. Sun et al. (2013) "The Evolution of Tetraspanins Through a Phylogenetic Lens" Book Section Tetraspanins Springer Link

Desalle, R., J. U. Chicote, et al. "Generation of divergent uroplakin tetraspanins and their partners during vertebrate evolution: identification of novel uroplakins." *BMC Evol Biol* 14: 13.

DeSalle, R., R. Mares, et al. "Evolution of cysteine patterns in the large extracellular loop of tetraspanins from animals, fungi, plants and single-celled eukaryotes." *Mol Phylogenet Evol* 56(1): 486-91.

Dillon, M. J. and C. D. Goonasekera (1998). "Reflux nephropathy." *J Am Soc Nephrol* 9(12): 2377-83.

Duckert, P., S. Brunak, et al. (2004). "Prediction of proprotein convertase cleavage sites." *Protein Eng Des Sel* 17(1): 107-12.

Dunn, C. A., L. N. van de Lagemaat, et al. (2005). "Endogenous retrovirus long terminal repeats as ready-to-use mobile promoters: the case of primate beta3GAL-T5." *Gene* 364: 2-12.

Durkin, K., W. Coppieters, et al. "Serial translocation by means of circular intermediates underlies colour sidedness in cattle." *Nature* 482(7383): 81-4.

Dyrlund, T. F., E. T. Poulsen, et al. "Human cornea proteome: identification and quantitation of the proteins of the three main layers including epithelium, stroma, and endothelium." *J Proteome Res* 11(8): 4231-9.

Falke, J. and JA Witjes . (2011) "Contemporary management of low-risk bladder cancer." *Nature Review Urology*.;8(1):42-9.

Felsenstein J. *Inferring Phylogenies*. Sunderland: Mass: Sinauer Associates; 2004

Finn, RN.y BA. Kristoffersen (2007) "Vertebrate vitellogenin gene duplication in relation to the "3R hypothesis": correlation to the pelagic egg and the oceanic radiation of teleosts." PLoS One. 24;2(1):e169.

Fujimura, K., M. A. Conte, et al. "Circular DNA intermediate in the duplication of Nile tilapia vasa genes." PLoS One 6(12): e29477.

Garcia-Espana, A., P. J. Chung, et al. (2008). "Appearance of new tetraspanin genes during vertebrate evolution." Genomics 91(4): 326-34.

Garcia-Espana, A., P. J. Chung, et al. (2006). "Origin of the tetraspanin uroplakins and their co-evolution with associated proteins: implications for uroplakin structure and function." Mol Phylogenet Evol 41(2): 355-67.

Giltay, J. C., J. van de Meerakker, et al. (2004). "No pathogenic mutations in the uroplakin III gene of 25 patients with primary vesicoureteral reflux." J Urol 171(2 Pt 1): 931-2.

Gogvadze, E. and A. Buzdin (2009). "Retroelements and their impact on genome evolution and functioning." Cell Mol Life Sci 66(23): 3727-42.

Grana, X., J. Garriga, et al. (1998). "Role of the retinoblastoma protein family, pRB, p107 and p130 in the negative control of cell growth." Oncogene 17(25): 3365-83.

Gupta, R. and S. Brunak (2002). "Prediction of glycosylation across the human proteome and the correlation to protein function." Pac Symp Biocomput: 310-22.

Halaby, D. M., A. Poupon, et al. (1999). "The immunoglobulin fold family: sequence analysis and 3D structure comparisons." Protein Eng 12(7): 563-71.

Hasan AK., Y. Fukami, et al. (2011) "Gamete membrane microdomains and their associated molecules in fertilization signaling." Mol Reprod Dev.;78(10-11):814-30.

Hatina, J. and W. A. Schulz (2012) "Stem cells in the biology of normal urothelium and urothelial carcinoma." *Neoplasma* 59(6): 728-36.

Hemler, M. E. (2003). "Tetraspanin proteins mediate cellular penetration, invasion, and fusion events and define a novel type of membrane microdomain." *Annu Rev Cell Dev Biol* 19: 397-422.

Hendriks WJ A. Elson, et al. (2013) "Protein tyrosine phosphatases in health and disease." *FEBS J.* 280(2):708-30.

Hernandez, S., E. Lopez-Knowles, et al. (2006). "Prospective study of FGFR3 mutations as a prognostic factor in nonmuscle invasive urothelial bladder carcinomas." *J Clin Oncol* 24(22): 3664-71.

Hicks, R. M. (1965). "The fine structure of the transitional epithelium of rat ureter." *J Cell Biol* 26(1): 25-48.

Hicks, R. M. (1966). "The function of the golgi complex in transitional epithelium. Synthesis of the thick cell membrane." *J Cell Biol* 30(3): 623-43.

Hicks, R. M. (1975). "The mammalian urinary bladder: an accommodating organ." *Biol Rev Camb Philos Soc* 50(2): 215-46.

Holland, N. D., L. Z. Holland, et al. (2015) "Scenarios for the making of vertebrates." *Nature* 520(7548): 450-5.

Hu, A., TB., Ge Zhou, et al. (2008) "Assembly of a membrane receptor complex: roles of the uroplakin II prosequence in regulating uroplakin bacterial receptor oligomerization." 414 (2) 195-203

Hu, C. C., F. X. Liang, et al. (2005). "Assembly of urothelial plaques: tetraspanin function in membrane protein trafficking." *Mol Biol Cell* 16(9): 3937-50.

Hu, P., F. M. Deng, et al. (2000). "Ablation of uroplakin III gene results in small urothelial plaques, urothelial leakage, and vesicoureteral reflux." *J Cell Biol* 151(5): 961-72.

Huang, G. N., J. E. Thatcher, et al. (2012) "C/EBP transcription factors mediate epicardial activation during heart development and injury." *Science* 338(6114): 1599-603.

Huang, H. Y., S. F. Shariat, et al. (2007). "Persistent uroplakin expression in advanced urothelial carcinomas: implications in urothelial tumor progression and clinical outcome." *Hum Pathol* 38(11): 1703-13.

Huang, S., S. Yuan, et al. (2005). "The phylogenetic analysis of tetraspanins projects the evolution of cell-cell interactions from unicellular to multicellular organisms." *Genomics* 86(6): 674-84.

Hudoklin, S., D. Zupancic, et al. (2009). "Maturation of the Golgi apparatus in urothelial cells." *Cell Tissue Res* 336(3): 453-63.

Hughes, J. F. and J. M. Coffin (2004). "Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: implications for human and viral evolution." *Proc Natl Acad Sci U S A* 101(6): 1668-72.

Hurst, R. E., S. W. Rhodes, et al. (1987). "Functional and structural characteristics of the glycosaminoglycans of the bladder luminal surface." *J Urol* 138(2): 433-7.

Jaworski, C. J., M. Aryankalayil-John, et al. (2009). "Expression analysis of human pterygium shows a predominance of conjunctival and limbal markers and genes associated with cell migration." *Mol Vis* 15: 2421-34.

Jenkins, D., M. Bitner-Glindzicz, et al. (2005). "De novo Uroplakin IIIa heterozygous mutations cause human renal adysplasia leading to severe kidney failure." *J Am Soc Nephrol* 16(7): 2141-9.

Jiang, Z., H. Tang, et al. (2007). "Ancestral reconstruction of segmental duplications reveals punctuated cores of human genome evolution." *Nat Genet* 39(11): 1361-8.

Jost, S. P., J. A. Gosling, et al. (1989). "The morphology of normal human bladder urothelium." *J Anat* 167: 103-15.

Jonasova, K., Z. Kozmik et al. (2008) "Eye evolution: lens and cornea as an upgrade of animal visual system." *Semin Cell Dev Biol.* 19(2):71-81

Jones, T. and L. Cheng (2005) "Bladder biopsy interpretation: Biopsy interpretation series." *American Journal of Surgical Pathology*

Jurka, J., V. V. Kapitonov, et al. (1996). "Identification of new medium reiteration frequency repeats in the genomes of Primates, Rodentia and Lagomorpha." *Genetica* 98(3): 235-47.

Kanai, A., M. Epperly, et al. (2004). "Differing roles of mitochondrial nitric oxide synthase in cardiomyocytes and urothelial cells." *Am J Physiol Heart Circ Physiol* 286(1): H13-21.

Kapitonov, V., J. Jurka, et al. (2005) "RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from *Transib* Transposons" *PLOS Biology*

Katoh, K. and D. M. Standley "MAFFT multiple sequence alignment software version 7: improvements in performance and usability." *Mol Biol Evol* 30(4): 772-80.

Kelly, H., S. Ennis, et al. (2005). "Uroplakin III is not a major candidate gene for primary vesicoureteral reflux." *Eur J Hum Genet* 13(4): 500-2.

Khandelwal, P., S. N. Abraham, et al. (2009). "Cell biology and physiology of the uroepithelium." *Am J Physiol Renal Physiol* 297(6): F1477-501.

Kimura, T., H. Suzuki, et al. (2001). "The incidence of thanatophoric dysplasia mutations in FGFR3 gene is higher in low-grade or superficial bladder carcinomas." *Cancer* 92(10): 2555-61.

King, N., M. J. Westbrook, et al. (2008). "The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans." *Nature* 451(7180): 783-8.

Kleinjan, D. A., R. M. Bancewicz, et al. (2008). "Subfunctionalization of duplicated zebrafish *pax6* genes by cis-regulatory divergence." *PLoS Genet* 4(2): e29.

Koide, A., C. W. Bailey, et al. (1998). "The fibronectin type III domain as a scaffold for novel binding proteins." *J Mol Biol* 284(4): 1141-51.

Kong, X. T., F. M. Deng, et al. (2004). "Roles of uroplakins in plaque formation, umbrella cell enlargement, and urinary tract diseases." *J Cell Biol* 167(6): 1195-204.

Kozul, R., S. Caburet, et al. (2004). "Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments." *Embo J* 23(1): 234-43.

Kreisberg, R., V. Buchner, et al. (1995). "Paired natural cysteine mutation mapping: aid to constraining models of protein tertiary structure." *Protein Sci* 4(11): 2405-10.

Kuratani, S. "Evolution of the vertebrate jaw from developmental perspectives." *Evol Dev* 14(1): 76-92.

Kurzrock, E. A., D. K. Lieu, et al. (2008). "Label-retaining cells of the bladder: candidate urothelial stem cells." *Am J Physiol Renal Physiol* 294(6): F1415-21.

Kwok, P. Y., C. Carlson, et al. (1994). "Comparative analysis of human DNA variations by fluorescence-based sequencing of PCR products." *Genomics* 23(1): 138-44.

Lavelle, J., S. Meyers, et al. (2002). "Bladder permeability barrier: recovery from selective injury of surface epithelial cells." *Am J Physiol Renal Physiol* 283(2): F242-53.

Letunic, I., T. Doerks, et al. "SMART: recent updates, new developments and status in 2015." *Nucleic Acids Res* 43(Database issue): D257-60.

Lewis, S. A., D. C. Eaton, et al. (1976). "The mechanism of Na⁺ transport by rabbit urinary bladder." *J Membr Biol* 28(1): 41-70.

Liang, F., B. Kachar, et al. (1999). "Urothelial hinge as a highly specialized membrane: detergent-insolubility, urohingin association, and in vitro formation." *Differentiation* 65(1): 59-69.

Lieber, M. R. (2008). "The mechanism of human nonhomologous DNA end joining." *J Biol Chem* 283(1): 1-5.

Lin, J. H., X. R. Wu, et al. (1994). "Precursor sequence, processing, and urothelium-specific expression of a major 15-kDa protein subunit of asymmetric unit membrane." *J Biol Chem* 269(3): 1775-84.

Lindsey A. Torre, LA.; F. Bray (2015) "Global cancer statistics, 2012" *CA: A Cancer Journal for Clinicians* Volume 65, Issue 2, pages 87–108

Liu, X., W. Zhang, et al. (2014) "Clinical significance of fibroblast growth factor receptor-3 mutations in bladder cancer: a systematic review and meta-analysis." *Genet Mol Res* 13(1): 1109-20.

Long, Q., C. Bengra, et al. (1998). "A long terminal repeat of the human endogenous retrovirus ERV-9 is located in the 5' boundary area of the human beta-globin locus control region." *Genomics* 54(3): 542-55.

Lundin, L. G., D. Larhammar, et al. (2003). "Numerous groups of chromosomal regional paralogies strongly indicate two genome doublings at the root of the vertebrates." *J Struct Funct Genomics* 3(1-4): 53-63.

Lupski JR, Stankiewicz P. Genomic Disorders: Molecular Mechanisms for Rearrangements and Conveyed Phenotypes. *PLoS Genetics*. 2005;1(6):e49. doi:10.1371/journal.pgen.0010049.

Mahbub Hasan, A. K., Z. Ou, et al. (2007). "Characterization of *Xenopus* egg membrane microdomains containing uroplakin Ib/III complex: roles of their molecular interactions for subcellular localization and signal transduction." *Genes Cells* 12(2): 251-67.

Mahbub Hasan, A. K., K. Sato, et al. (2005). "Uroplakin III, a novel Src substrate in *Xenopus* egg rafts, is a target for sperm protease essential for fertilization." *Dev Biol* 286(2): 483-92.

Maley, F., R. B. Trimble, et al. (1989). "Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases." *Anal Biochem* 180(2): 195-204.

Marques-Bonet, T., S. Girirajan, et al. (2009) "The origins and impact of primate segmental duplications." *Trends in genetics* (10):443-454. doi:10.1016/j.tig.2009.08.002.

McLysaght, A., K. Hokamp, et al. (2002). "Extensive genomic duplication during early chordate evolution." *Nat Genet* 31(2): 200-4.

Min, G., H. Wang, et al. (2006). "Structural basis for tetraspanin functions as revealed by the cryo-EM structure of uroplakin complexes at 6-A resolution." *J Cell Biol* 173(6): 975-83.

Min, G., G. Zhou, et al. (2003). "Structural basis of urothelial permeability barrier function as revealed by Cryo-EM studies of the 16 nm uroplakin particle." *J Cell Sci* 116(Pt 20): 4087-94.

Mitra, S., S. Lukianov, et al. (2011) "Requirement for a uroplakin 3a-like protein in the development of zebrafish pronephric tubule epithelial cell function, morphogenesis, and polarity." *PLoS One* 7(7): e41816.

Moll, R., X. R. Wu, et al. (1995). "Uroplakins, specific membrane proteins of urothelial umbrella cells, as histological markers of metastatic transitional cell carcinomas." *Am J Pathol* 147(5): 1383-97.

Morgenstern, B. (1999). "DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment." *Bioinformatics* 15(3): 211-8.

Muller, S., P. Finelli, et al. (2004). "The evolutionary history of human chromosome 7." *Genomics* 84(3): 458-67.

Mysorekar, I. U., M. A. Mulvey, et al. (2002). "Molecular regulation of urothelial renewal and host defenses during infection with uropathogenic *Escherichia coli*." *J Biol Chem* 277(9): 7412-9.

Nelson JS., "Fishes of the world: 4th edition" Wiley

Ohno, S. (1999). "Gene duplication and the uniqueness of vertebrate genomes circa 1970-1999." *Semin Cell Dev Biol* 10(5): 517-22.

Olsburgh, J., P. Harnden, et al. (2003). "Uroplakin gene expression in normal human tissues and locally advanced bladder cancer." *J Pathol* 199(1): 41-9.

Ono, K., H. Suga, et al. (1999). "Multiple protein tyrosine phosphatases in sponges and explosive gene duplication in the early evolution of animals before the parazoan-eumetazoan split." *J Mol Evol* 48(6): 654-62.

Ono-Koyanagi, K., H. Suga, et al. (2000). "Protein tyrosine phosphatases from amphioxus, hagfish, and ray: divergence of tissue-specific isoform genes in the early evolution of vertebrates." *J Mol Evol* 50(3): 302-11.

Parsons, C. L., D. Boychuk, et al. (1990). "Bladder surface glycosaminoglycans: an epithelial permeability barrier." *J Urol* 143(1): 139-42.

Parsons, C. L., C. Greenspan, et al. (1977). "Role of surface mucin in primary antibacterial defense of bladder." *Urology* 9(1): 48-52.

Petersen, T. N., S. Brunak, et al. "SignalP 4.0: discriminating signal peptides from transmembrane regions." *Nat Methods* 8(10): 785-6.

Porter, K. R., K. Kenyon, et al. (1967). "Specializations of the unit membrane." *Protoplasma* 63(1): 262-74.

Powers, C. J., S. W. McLeskey, et al. (2000). "Fibroblast growth factors, their receptors and signaling." *Endocr Relat Cancer* 7(3): 165-97.

Putnam, NH., T. Butts et al. (2008) "The amphioxus genome and the evolution of the chordate karyotype" *Nature* 453, 1064-1071

Rastogi, S. and D. A. Liberles (2005). "Subfunctionalization of duplicated genes as a transition state to neofunctionalization." *BMC Evol Biol* 5: 28.

Ravi, V. and B. Venkatesh (2008). "Rapidly evolving fish genomes and teleost diversity." *Curr Opin Genet Dev* 18(6): 544-50.

Ronquist F, M. Teslenko et al. (2012) . "MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space." *Syst Biol.* ;61(3):539-42

Sakakibara, K., K. Sato, et al. (2005). "Molecular identification and characterization of *Xenopus* egg uroplakin III, an egg raft-associated transmembrane protein that is tyrosine-phosphorylated upon fertilization." *J Biol Chem* 280(15): 15029-37.

Sánchez-Sánchez, AV., E. Camp et al. (2010) "Medaka Oct4 is expressed during early embryo development, and in primordial germ cells and adult gonads." *Dev Dyn.* 239(2):672-9. doi: 10.1002/dvdy.22198.

Schacherer, J., J. de Montigny, et al. (2005). "Duplication processes in *Saccharomyces cerevisiae* haploid strains." *Nucleic Acids Res* 33(19): 6319-26.

Schonfelder, E. M., T. Knuppel, et al. (2006). "Mutations in Uroplakin IIIA are a rare cause of renal hypodysplasia in humans." *Am J Kidney Dis* 47(6): 1004-12.

Schroder, S., F. Fraternali, et al. (2011) "When a module is not a domain: the case of the REJ module and the redefinition of the architecture of polycystin-1." *Biochem J* 435(3): 651-60.

Schulz, W. A. (2006). "Understanding urothelial carcinoma through cancer pathways." *Int J Cancer* 119(7): 1513-8.

Sen, S. K., K. Han, et al. (2006). "Human genomic deletions mediated by recombination between Alu elements." *Am J Hum Genet* 79(1): 41-53.

Shao, H., R. Chaerkady, et al. "Proteome profiling of wild type and lumican-deficient mouse corneas." *J Proteomics* 74(10): 1895-905.

She, X, Z.,Jiang, et al. (2004) "Shotgun sequence assembly and recent segmental duplications within the human genome." *Nature*. 21;431(7011):927-30.

Shpakovskii, D. G., E. K. Shematorova, et al. (2004). "[New genes on human chromosome 7: bioinformatic analysis of a gene cluster from the POLR2J family]." *Bioorg Khim* 30(6): 621-5.

Smith, J. J., S. Kuraku, et al. (2012) "Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution." *Nat Genet* 45(4): 415-21, 421e1-2.

Staehelin, L. A., F. J. Chlapowski, et al. (1972). "Luminal plasma membrane of the urinary bladder. I. Three-dimensional reconstruction from freeze-etch images." *J Cell Biol* 53(1): 73-91.

Suga, H., Z. Chen, et al. (2013) . "The Capsaspora genome reveals a complex unicellular prehistory of animals." *Nat Commun.*;4:2325.

Sun, T. T. (2006). "Altered phenotype of cultured urothelial and other stratified epithelial cells: implications for wound healing." *Am J Physiol Renal Physiol* 291(1): F9-21.

Thompson, J. D., D. G. Higgins, et al. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." *Nucleic Acids Res* 22(22): 4673-80.

Thumbikat, P., R. E. Berry, et al. (2009). "Bacteria-induced uroplakin signaling mediates bladder response to infection." *PLoS Pathog* 5(5): e1000415.

Tomlinson, D. C., O. Baldo, et al. (2007). "FGFR3 protein expression and its relationship to mutation status and prognostic variables in bladder cancer." *J Pathol* 213(1): 91-8.

Truschel, S. T., E. Wang, et al. (2002). "Stretch-regulated exocytosis/endocytosis in bladder umbrella cells." *Mol Biol Cell* 13(3): 830-46.

Tu, L., T. T. Sun, et al. (2002). "Specific heterodimer formation is a prerequisite for uroplakins to exit from the endoplasmic reticulum." *Mol Biol Cell* 13(12): 4221-30.

van de Lagemaat, L. N., J. R. Landry, et al. (2003). "Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions." *Trends Genet* 19(10): 530-6.

van Eekelen, M., J. Overvoorde, et al. (2010) "Identification and expression of the family of classical protein-tyrosine phosphatases in zebrafish." *PLoS One* 5(9): e12573.

van Rhijn, B. W., R. Montironi, et al. (2002). "Frequent FGFR3 mutations in urothelial papilloma." *J Pathol* 198(2): 245-51.

Varley, C. L., M. A. Garthwaite, et al. (2006). "PPARgamma-regulated tight junction development during human urothelial cytodifferentiation." *J Cell Physiol* 208(2): 407-17.

Walz, T., M. Haner, et al. (1995). "Towards the molecular architecture of the asymmetric unit membrane of the mammalian urinary bladder epithelium: a closed "twisted ribbon" structure." *J Mol Biol* 248(5): 887-900.

Weiner, J 3rd, F. Beaussart et al. (2006) "Domain deletions and substitutions in the modular protein evolution." *FEBS J.* 273(9):2037-47.

Wu, X. R., X. P. Kong, et al. (2009). "Uroplakins in urothelial biology, function, and disease." *Kidney Int* 75(11): 1153-65.

Wu, X. R., J. H. Lin, et al. (1994). "Mammalian uroplakins. A group of highly conserved urothelial differentiation-related membrane proteins." *J Biol Chem* 269(18): 13716-24.

Wu, X. R., M. Manabe, et al. (1990). "Large scale purification and immunolocalization of bovine uroplakins I, II, and III. Molecular markers of urothelial differentiation." *J Biol Chem* 265(31): 19170-9.

Wu, X. R., J. J. Medina, et al. (1995). "Selective interactions of UPIa and UPIb, two members of the transmembrane 4 superfamily, with distinct single transmembrane-domained proteins in differentiated urothelial cells." *J Biol Chem* 270(50): 29752-9.

Wu, X. R. and T. T. Sun (1993). "Molecular cloning of a 47 kDa tissue-specific and differentiation-dependent urothelial cell surface glycoprotein." *J Cell Sci* 106 (Pt 1): 31-43.

Wu, X. R., T. T. Sun, et al. (1996). "In vitro binding of type 1-fimbriated Escherichia coli to uroplakins Ia and Ib: relation to urinary tract infections." Proc Natl Acad Sci U S A 93(18): 9630-5.

Yuasa, T., T. Yoshiki, et al. (2003). "Molecular cloning and expression of uroplakins in transitional cell carcinoma." Adv Exp Med Biol 539(Pt A): 33-46.

Zhang, ZT., J. Pak, et al. (2001) "Role of Ha-ras activation in superficial papillary pathway of urothelial tumor formation." Oncogen 12;20(16):1973-80.

Zhang, ZT., J. Pak, et al. (1999) "Urothelium-specific expression of an oncogene in transgenic mice induced the formation of carcinoma in situ and invasive transitional cell carcinoma." Cancer Research 15;59(14):3512-7

Zhou, G., W. J. Mo, et al. (2001). "Uroplakin Ia is the urothelial receptor for uropathogenic Escherichia coli: evidence from in vitro FimH binding." J Cell Sci 114(Pt 22): 4095-103.

ANEXOS

UNIVERSITAT ROVIRA I VIRGILI

EVOLUTION AND MOLECULAR CHARACTERIZATION OF UROPLAKIN 3C, A NOVEL HUMAN TETRASPANIN ASSOCIATED UROPLAKIN

Javier Ugarte Chicote

Figura Suplementaria 1: Alineamiento de todas las secuencias de proteína de UPK2/3. El alineamiento se generó con el servidor MAFFT y comprende desde el comienzo del segundo exón hasta el final del quinto exón. Las C₁-C₄ están resaltadas en color amarillo y las fases intrónicas marcadas de color rojo.

UPK3a.human	---AVNLQP-QLA--SVTFATNNPTLTTVALEKPL--C	---MFDSK-----EAL
UPK3a.mouse	---TVNLQP-QLA--SVTFATNNPTLTTVALEKPL--C	---MFDS-----EPL
UPK3a.cow	---GVNLQP-QLA--SVTFATNNPTLTTVALEKPL--C	---MFDS-----AAL
UPK3a.elephant	---AVNLQP-QLA--SVTFATNNPTLTTVALEKPL--C	---MFES-----TAL
UPK3a.dog	---GVNLQP-QLA--SVTFATNNPTLTTVALEKPL--C	---MFDSA-----ATL
UPK3a.opossum	---AVDLEP-QMA--SITFATNNPTLTTITLEKPF--C	---MFNAS-----NLG
UPK3a.platypus	---AQIKP-QIA--GLAFATNNPTLTTIALEKPF--C	---MFDDGA-----GPA
UPK3a.chicken	---DQSMKP-QLA--APELATNNPTLTTVALEKPF--C	---MFDS-----LHP
UPK3a.aligatore	---AQSLRP-QIA--SPQLATNNPTLTTIALEKPF--C	---VFDGS-----LSP
UPK3a.turtle	---ALNLRP-QIA--NPKLATSNPTLTTIALEKPF--C	---MFDDS-----LSP
UPK3a.Xtropical	---AGADMP-LLA--NSDFFSLNPTQTITIALEKPF--C	---MFKSA-----V--
UPK3a.Xlaevis	---ARSAVP-LLA--NSDFFSLNPTQTITILERPF--C	---MKDA-----I--
UPK3a.salamande	---TFAVTP-QVA--GPALVPNTPTLNTVALQKPV--C	---LFDSS-----TGG
UPK3a.coelacant	---AQFPKP-EIA--NPQFVAGNPTLTTISLEKPF--C	---VFDRT-----ISA
UPK3a.spottedga	---EALKVKP-EAV--SPRLLRFNPTQSTVSLAKPL--C	---VFDVS-----KPT
UPK3a.skate	---PLSFKP-EVS--SIALTGLRSTTTVTLAKPI--C	---VFSSG-----D--
UPK3a.shark	---ILQFKP-EMP--SVSTIDGLRARTVTLSKPL--C	---VLPTG-----H--
UPK3a.3.lamprey	-----GERYGNLFPSTSSSTRCANQF--Q	---SFVSS-----AHS
UPK3a.1.lamprey	---GSGTKP-QLV--SSTAVPYNPTETTTIVWSKPF--C	---VFQKP-----VPT
UPK3a.2.lamprey	---VSQVVP-TVV--NPNL-LGAVTQTTVALQAPF--C	---SALDAEVVVAL-----SSV
UPK3b.human	---ELVPYTP-QI--TAWDLEKVTATTFVLEQPR--C	---VFDGL-----ASA
UPK3b.cow	---DLIPYTP-RI--TSWDLEKVTATTFVLEQPR--C	---VLDRH-----SSA
UPK3b.mouse	---DLIAYVP-QI--TAWDLEKVTATTFVLEQPR--C	---VFDEH-----VST
UPK3b.elephant	-----DPLHA--TDNGLGGKVTATTFVLEQPR--C	---VLDGH-----ARA
UPK3b.dog	---ELIPYTP-QI--TAWDLEKVTATTFVLEQPR--C	---VLDGQ-----ASV
UPK3b.platypus	---LMSYKP-QV--SAHPLEKVTASTFVLEQPR--C	---VFDGQ-----VVS
UPK3b.opossum	---DQIPYTP-QI--SALALEKVTATTFVLEQPR--C	---IFSEL-----AAP
UPK3b.Xtropical	---DITSYVP-QL--TLSPVIGTVTSTTFVLDKPK--C	---VFGNT-----
UPK3b.Xlaevis	---DITTYVP-QL--TLMPIQGSVSTTFVLDKPK--C	---IFGSR-----
UPK3b1.Xtropical	---IPYVVP-QI--TTKPIILGKLTSSSFVLEQPK--C	---IFQQY-----K
UPK3b2.Xtropical	---DVGSYVP-KI--TTSILGNLTFSTFVLEQPK--C	---IFSSNY-----P
UPK3b.chicken	---ALLPYVP-RV--APGAMPKVTATTFVLEPR--C	---IFDPF-----ANA
UPK3b.aligatore	---ELVPYTP-QI--TGNLSLEKLTASTFVLDQPI--C	---IFDQY-----VNA
UPK3d.coelacant	---DVPYLP-EV--AFNNVAGSVTATTFVLEQPR--C	---IFKDV-----FSS
UPK3d.1.spotted	---VTSVEYTP-EI--TAYNMAGRVTGTTVMVKQPR--C	---YFDNQLLLPC-----TPD
UPK3d.salmo	---IVSLDHP-EI--LPYEASGRVSTTTIVLQPL--C	---YFNTLTQLKC-----SQS
UPK3d.trout	---IVSLDHP-EI--LPYEASGRVSTTTIVLQPL--C	---YFNTLTQLKC-----SQS
UPK3d.zebrafish	---GQIFQP-QL--APANFLGRITSNVILQQPV--C	---VFQTK-----TCP
UPK3b.skate	---ADEADFVP-EI--FKDNVIGGVSTTFVLEQPK--C	---VFDPR-----PACT
UPK3b.shark	---VKNYTP-EV--TNANLAGRITATITFQQL--C	---QFDS-----INA
UPK3d1.salmo	---IPVNNTP-EI--TPSSLAAKLTNSVILTSPS--C	---YFNLGNLSC-----NSTT
UPK3d1.trout	---FVNNTPI-AI--NPSSLAAKLTINSVILTSPS--C	---YFNSLANLPC-----NSTT
UPK3d.spottedga	T-VPSVDYVP-EI--IPMKLAGRLTGTFFLRQPL--C	---YFNSQQGLNC-----SLS
UPK3d2.spottedga	---VPYAP-EV--TPHNLLGRVSTTTLLTQQPV--C	---FPNNLSGLPC-----APD
UPK3c.human	AAPEHISYVP-QL--SNDTLAGRLTSTFTLEQPL--G	---QFSSHN-----ISD
UPK3c.mouse	---ESINYAP-QL--LGATLEGRLTQSTFTLEQPL--G	---QFKNVN-----LSD
UPK3c.dog	---EHISYVP-QL--SNRSLAGTLTQSTFTLEQPR--G	---QFSHRS-----ISD
UPK3c.cow	---ERISYVP-QL--SSATLAGRLTQSTFTLEQPR--G	---QFSHPS-----ISD
UPK3c.elephant	---ELISYVP-RL--SNATLAGRLTQSTFTLEQPR--G	---RFSHK-----ISD
UPK3c.opossum	---EPINYTP-AI--TREPLEGISTSTFTLDQPN--D	---QFNGSG-----ISD
UPK3c.platypus	---KIAATP-VI--TKLPMEGKITLSTFTLDQPL--G	---QFNSSA-----VDD
UPK3c.lyzard	---VNYTP-RL--ASENLGGKVTASTFTLDQPR--C	---VFNDV-----VNA
UPK3c.chicken	---DKLSYKP-TL--VGGNVEGRMTGSTFVLEQPR--C	---VFDSS-----S
UPK3c.turtle	---SIDYTP-VI--TARELEKITSTFVLEQPR--C	---VFNDY-----VSN
UPK2a.human	---DFNI-----SLSGLLSPALTESLLVALPP--C	---HLTGG-----
UPK2a.mouse	---DFNI-----SLSGLLSPALTESLLIALPP--C	---HLTGG-----
UPK2a.dog	---DFNI-----SSVSGLLSPALTESLLVALPP--C	---HLTGG-----
UPK2a.cow	---ADFNI-----SLSGLLSPVMTESLLVALPP--C	---HLTGG-----
UPK2a.elephant	---DFHI-----SLSGVLSPALTESLLVALAP--C	---RLTGG-----
UPK2a.opossum	---EFNI-----SLSGLLSPALAESLLVALPP--C	---HLTGG-----
UPK2a.platypus	---FNI-----SLSGLLTPTLAESLLVALPP--C	---HLTGG-----
UPK2a.salamande	---QTAQNFT-----SLADLPINPLQTLAIVAFPP--C	---CWVGVQSNA-----
UPK2a.Xlaevis	-----QNT--SLADGVLTPLSTSVIAPFG--C	---KDSGK-----
UPK2a.Xtropical	-----QNV--SQATGVLTPLATSAIFAPD--C	---TYSQG-----
UPK2a.coelacant	---NFNT-SLS--DNAALISNVYSSFVLSLPP--C	---TYAGK-----
UPK2a.skate	---GFTI--SLANDNMGEVVASRRSMSAIITMDPNSC--	---NLAGE-----
UPK2a.zebrafish	---DIPI--SLLNPNTDGVLASTFPNSFLLQMPD--C	---SIYGNQ-----
UPK2a.salmo	---EFQV-SLLK--ESDGVVTRGFADSLLSLPP--C	---ALATQ-----
UPK2a.lamprey	---DYQV-KLL--NTSVVRAAQTPQSVAFVAPV--C	---CEI-AMTAGT-----
UPK2b.snake	---MDFFR-----NNSMLLTFTTFTTIVNIK--C	---VSSSK-----FSP
UPK2b.turtle	---TLDFFP-----NASLILAAARMSTSFIVNIK--C	---ISSSQ-----FTF
UPK2b.lyzard	---ELDFFP-----NNSMLLTSMTYFIVNVK--C	---VSPKQ-----FSP

UPK2b.aligator ---TLNFFP----NMSNILAARLSSSFIVNIPK--C---ISSH-----YTP
 UPK2b.salamande ---PFSFFG----NQDNIIGNRMGYSFITNIPS--C---ISGAG-----YTP
 UPK2b.Xlaevis ---SNFQFFT----NTDDVLGAVLSQSPFVNVPSC---INAVG-----YVP
 UPK2b.shark -----DYDI-----ALFNGEVSRLATLVLVSPAY--C---LPEKWVEQRGGELTVRS
 UPK3a.human TG-THEYVLYVLVD-SAISRNASVQDSTNTPLGS---TFLQTE-----GGRTPGY---
 UPK3a.mouse SG-SYEVLYLAMVD-SAMSRNVSVQDSAGVPLST---TFRQQTQ-----GGRSGPY---
 UPK3a.cow HG-TYEVLYVLVD-SASFRNASVQDSTKTPLSS---TFQQTQ-----GGRTPGY---
 UPK3a.elephant VG-TYEVLYVLAE-SASLRNASIQDHTSAPLSS---TFQQTQ-----GGRTPGY---
 UPK3a.dog NG-TYEVLYVLVN-LASSRNASVQDGARAPLSS---TQQTE-----GGRTPGY---
 UPK3a.opossum NV-SYEVNLYVMEN-SGSVRYAVIKNNRSFPINS---TFQETA-----GGQRAPY---
 UPK3a.platypus QL-PSEIYLYVMM-DADAH--PVLNNGSRPLRT---TFQQA-----GGRTPGY---
 UPK3a.chicken NK-SYAIYLYVMKS-SANTISSVVTDSSSKPLDS---TFQQTG-----GHLGPY---
 UPK3a.aligator GK-SYEVLYLAMMD-SMSSIAVTDNNSKPLDS---TFQEVN-----GGQLGPY---
 UPK3a.turtle GS-SYEVLYAMAD-SESTVSAVTDNNSKPLNT---TFQDTN-----GGQLGPY---
 UPK3a.Xtropical ----NVYLIGIVA-GAPNT--PLYDGNKKVNAS---TYSGTQ-----GKTPGY---
 UPK3a.Xlaevis ----NVYLFAIVK-GATNI--QVADAARKVIAS---NYTGTQ-----GGLLGPY---
 UPK3a.salamande DPANYQVELFAMAA-SAPSS-----TPLASGN---TFRNTS-----GGTTPGY---
 UPK3a.coelacant SSGDFTYAVFAVKS-TVAVA-----ENINDFTQ---TYQSSR-----EGTTAPY---
 UPK3a.spottedga NE--MMVDVYVHVS-LSATL-----TFETGK---TYKTN-----GGTETPY---
 UPK3a.skate ----VVEVFGVQT-TADSI-----PIEIGNRTILTYQQTG---GGARGPY---
 UPK3a.shark ----MIEELVVQG-NVTP-I-----ATNLDDG---TYQATK-----GGATGPY---
 UPK3a.3.lamprey S-----SLLRLYD-TTTERSAPGYSF-SRVLGA--VGSRVATG-----AGDRGPY---
 UPK3a.1.lamprey TQ---YVVDVYASIT-NNSYA----FDNSIGAVLS---SY-WTN-----AVSPSPY---
 UPK3a.2.lamprey SA---DLRLFVMATAQRNVSNEMITSSSTIGLDK---GY-----
 UPK3b.human SD---TWLVVAFS-NAS-----R--GFQNPPE---TLADIP---
 UPK3b.cow AD---TWLVVAFS-NAS-----R--VFQNPQ---TLAEP---
 UPK3b.mouse KD---TIWLVVAFS-NAS-----R--DFQNPQ---TAAKIP---
 UPK3b.elephant TD---TWLVVAFS-NAS-----K--DFQPKK---TQAEIP---
 UPK3b.dog AS---TWLVVTFS-NAS-----K--DPHNQ---TLAEP---
 UPK3b.platypus TD---TIWLVVAFS-NAS-----R--DFQNPQ---TAAIP---
 UPK3b.opossum AD---AVWLVAFA-NAT-----E--DFQNPK---TAAEP---
 UPK3b.Xtropical GN---QVWLLVARS-NVS-----A--NV-----VLTTPS---
 UPK3b.Xlaevis TN---QVWLLVARS-NVS-----V--SITN-----AMLKPPS---
 UPK3b1.Xtropical TS---LWLVVALN-RVI-----P--QLSY---TQLSNPA---
 UPK3b2.Xtropical TQ---DVWLVALD-TVE-----P--PFLTD-----TNLSTFV---
 UPK3b.chicken SD---AVWLVAFA-DAS-----A--AFKNPT---SSAEPV---
 UPK3b.aligator TD---D1WLVAFA-NAT-----S--SLKNPT---SRDIP---
 UPK3d.coelacant CPL-CQLWVAVATE-KGA-----N--NLITTVG---RPVTPA---
 UPK3d.1.spotted KC---E1WLVAVGSAGI-----Q--NFDADK GKST--ILSESYPY---
 UPK3d.salmo TC---QVMAAIASG-PGI-----N--NFDIDKLVAV-QIVSASYPY---
 UPK3d.trout TC---QVMAAIASG-PGI-----N--NFDIDKLVAV-QIVSASYPY---
 UPK3d.zebrafish GC---E1WLVAALS-TGT-----G--NPNALNVISSPISLSVSPYP---
 UPK3b.skate LC---E1WLVDNP-ANV-----S--TFDTN-----MNLTPSA---
 UPK3b.shark TP-GLIFWLVVANSQRGQ-----N--TFDA---ITFPQSL---
 UPK3d1.salmo TC---E1WLVSAKD-TGV-----S--NYDADNKMPY--IDTRSPYP---
 UPK3d1.trout TC---E1WLVSAID-TGV-----S--NYDADNKMPY--IDTRSPYP---
 UPK3d.spottedga TC---E1WLVAVARE-AGV-----N--NFDTDKVPQSFDIVSASYPY---
 UPK3d2.spottedga NC---E1WLVIA--RGV-----P--KFEAIKNTS--VLMSSYPY---
 UPK3c.human LD---TIWLVVALS-NAT-----Q--SFTAPR---TNQDIP---
 UPK3c.mouse PD---P1WLVVAHS-NAA-----Q--NFTAPR---KVEDRH---
 UPK3c.dog SD---A1WLVVAHS-NAT-----Q--NFSAPQ---RVEDIP---
 UPK3c.cow SD---A1WLVVAHS-NAT-----Q--KFTAPQ---KVEDTP---
 UPK3c.elephant FD---A1WLVVAHS-NAT-----Q--SF1APQ---RVKDSF---
 UPK3c.opossum LD---D1WLVVAFS-NAS-----Q--SFEPPQ---SAQDIP---
 UPK3c.platypus LD---D1WLVVAYS-NAT-----D--NFSNPV---KPDEY---
 UPK3c.lyzard TD---GIWLLVARS-DAA-----R--NFTRPQ---SPSELP---
 UPK3c.chicken TA---N1WLVVATR-AGM-----N--AFNDS---AQPGMPEW---
 UPK3c.turtle TD---E1WLVVALS-NGTS-----IS--TFTNPT---SLQSLP---
 UPK2a.human ----NATLMVRRANDSK-----
 UPK2a.mouse ----NATLMVRRANDSK-----
 UPK2a.dog ----NATLMVRRANDSK-----
 UPK2a.cow ----NATLTVRRANDSK-----
 UPK2a.elephant ----NATLMVRRANDSK-----
 UPK2a.opossum ----KASLTVRRVNESAS-----
 UPK2a.platypus ----NATLKVQGVNGSS-----
 UPK2a.salamande ----NLLIVSKNSSGSS-----I
 UPK2a.Xlaevis ----TVNLIIVANGTTT-----
 UPK2a.Xtropical ----TVSLVINTSTSTATI-----
 UPK2a.coelacant ----NASVITYSKNSTSEDS-----
 UPK2a.skate ----TVIVTVNNTSNGQVI-----
 UPK2a.zebrafish ----SVLILLYTEAPTNL-----
 UPK2a.salmo ----SVTLEYNNTDTNESK-----
 UPK2a.lamprey ----MLNIFVTPNNTGVEV-----
 UPK2b.snake A----TVRIAIAQLPDASSLP-----GITDDQIE---NLRRTPQALV
 UPK2b.turtle T----TIRPAVAIVGDKVAMP-----AVTDNQIQ---SL
 UPK2b.lyzard V----KIRLAVAQVPDTSTLP-----GVIDTDDIA---NLRRTPQATI
 UPK2b.aligator T----TIRPAIVLGETVGLP-----AVTDNQIR---SLRNPQDAPV
 UPK2b.salamande A----S1LLAVS--TSAAVP-----GVSNTDAIK---SLLNDSRAQI
 UPK2b.Xlaevis S----TLKVAVANRNPTC-----MVDTDSIK---SLKTDPTAPV
 UPK2b.shark SA----SQVEELLREDNATFV-----
 UPK3a.human -KAVAFDLI-----PCSDPLSLD---AIG---DVSQASQILNAYLVRVGA
 UPK3a.mouse -KAAAFDLT-----PCGDPLSLD---AVG---DVTQASEILNAYLVRVGN
 UPK3a.cow -KAAAFDLT-----PCSDSPSLD---AVR---DVSRASEILNAYLIRVGT

UPK3a.elephant	-KAAAFDLQ-----P	SDLP	SLD---	AVE---	DVSRASEILSTYLVRVGA	
UPK3a.dog	-KAVAFGLI-----P	SDLP	SLD---	AVG---	DVARASEILNAYLVRVGA	
UPK3a.opossum	-KAASFILP-----Q	GDLP	PNLD---	EAG---	DVSKVAEILSAYLVRVGD	
UPK3a.platypus	-WAATFSVP-----R	EDLP	QLW---	DAG---	DPARAPQILDAYLFRVGG	
UPK3a.chicken	-KAASFDPV-----N	CVSPP	RLA---	DAG---	DINKVSDVLKQYLFVRVGD	
UPK3a.aligator	-KAAVLNVP-----D	CASPP	KLA---	DIR---	NVKKASDVLKQYLFVRVGD	
UPK3a.turtle	-RAALFNVP-----N	CASPP	MPLA---	DVV---	NVKKASDVLKQYLFVRVGD	
UPK3a.Xtropical	-IVAKLPNQ-----Q	CINI	QALS---	NMA---	DPTQVQSILSKYVVRVGA	
UPK3a.Xlaevis	-QVAKLDNP-----K	CENIQ	ASN---	IMA---	DP---NKYIVRVGG	
UPK3a.salamande	-VAGKFGVP-----N	CTLVG	SLT---	PSQ---	D---DFLQRQYIFRVGD	
UPK3a.coelacant	-KAASFVAV-----N	DSPL	VLS---	NPV---	NVNVIRSLNQLYLRIGD	
UPK3a.spottedgeda	-KATSFGIP-----N	CTSP	PNPA---	DLS---	VPQRIDKTLDEYLRVIGS	
UPK3a.skate	-RAARFMTP-----I	CTSLP	-FV---	PSR---	DPAVIRVQIEQYLFVRVGD	
UPK3a.shark	-RAARFANP-----E	CTSS	-FT---	PST---	DPTKIQTLLIDRYFIRVGS	
UPK3a.3.lamprey	-LAGYFVPV-----P	CAQPP	PSVV	VTGT---	GSSVALALLGMVIFRLLG	
UPK3a.1.lamprey	-LAATFKVP-----D	CASQ	PSIY---	DAM---	AVKTNATFRLLG	
UPK3a.2.lamprey	-AGSGAGTS-SWY	LALG	RRP-LQNC	TI	STPSL-----STPTSYRVGA	
UPK3b.human	-ASPQLLTD-GHY	MTLPL	SDPQL	CDP	MA-----GSGGAPVLRVGH	
UPK3b.cow	-ASPRLLTD-GHY	MTLPL	TMDQL	CE	DPAD-----GSGRAPVLRVGN	
UPK3b.mouse	-TFPQLLTD-GHY	MTLPL	SLDQL	CE	DLTG-----GSGGVPVLRVGN	
UPK3b.elephant	-TFAQLLTD-GHC	MTLPP	SPVQL	CT	DSVG-----GSGSALLRVGN	
UPK3b.dog	-AFPRLLTD-GYY	MTLPL	SLDQL	CE	DEPG-----GGRSIPVLRVGN	
UPK3b.platypus	-AYPRLLTD-YYM	TLKVS	PDLYP	CA	DA-----GGLSVLRVGT	
UPK3b.opossum	-SYTELSSS-FYY	MTLKL	SPDLYP	CE	EE-----DIAVLRVGS	
UPK3b.Xtropical	-MYSSFATQ-GYY	HVPFG	TESLYH	C	SNTA-----EYIRVGD	
UPK3b.Xlaevis	-MYSSFATQ-GYY	HVPLG	TEASYP	C	SNTA-----DYIRVGD	
UPK3b1.Xtropical	-NISSFETN-GFY	HLPV	GGDYP	C	ADTS-----GQLSAMLYVGS	
UPK3b2.Xtropical	-TYSSFTTN-KFY	HTRV	RGADYP	C	FNES-----AMSLALLQVGA	
UPK3b.chicken	-PYEGLPTA-RAY	MTLQ	MAAAAYG	C	SAP-----GAAVLRVGG	
UPK3b.aligator	-PYQRLSTA-LHY	XLRT	TI	XFFP	CP	RST-----NTSVLRVGS
UPK3d.coelacant	-GYQKLSQR-GYY	FTR	SSKDL	YK	CVEN-----DDKVRVLRIGA	
UPK3d.1.spotted	TAFTGSPPK-NYY	LTKV	GRQV	QV	PP---CQ	QSH-----GIVYFRVGD
UPK3d.salmo	IAFSSQTNR-MYF	VTKL	GRPK	DF	PP---CQ	QLP-----GKIFYFRVGA
UPK3d.trout	IAFSSQTNR-MYF	VTKL	GRPK	DF	PP---CQ	QLP-----GKIFYFRVGA
UPK3d.zebrafish	TAFLP-SSA-QFF	LTRV	GLAN	F	NTAP-----AFFYFTVGA	
UPK3b.skate	ATYKDFLLN-GFY	LTVK	TART	DY	APQT-----MGQVYTLRVGD	
UPK3b.shark	MTHQASTA-DY	YTLR	TLAERY	L	CP	TSS-----SNGEMYFVRVGG
UPK3d1.salmo	TAFSSNTSK-KY	FLTK	LGLQ	KAYP	CP	IVA-----GTGYFRVGS
UPK3d1.trout	TAFSSNTSK-KY	FLTK	LGLQ	KAYP	CP	IVA-----GTDYFRVGS
UPK3d.spottedgeda	EAFQ---NK-NYY	VTR	LGVQ	NNFL	CA	ELP-----GIRYFRVGA
UPK3d2.spottedgeda	DAFRMNSSP-NY	FLTK	LGVQ	N	AF	CAEST-----GFRFRVGD
UPK3c.human	-APANFSQR-GYY	LTLR	ANRVLY	--	QT-----RGQLHLVRVGN	
UPK3c.mouse	-APANFDRN-GYY	LTLR	ANRVHY	KG	QP-----DSQLRVRVGN	
UPK3c.dog	-VPEDFTRR-GYY	LTLR	ANRLLY	PG	NP-----GNQLRVRVGN	
UPK3c.cow	-VPADFPQR-GYY	LTLR	ASRALY	PG	GP-----SNQLRVRVGN	
UPK3c.elephant	-APADLPLK-GYY	LTLR	ASRALY	PG	DA-----GNQLQVLRVGN	
UPK3c.opossum	-YAATFLDK-KY	YLTR	ASRD	LY	SSKRG-----SQGISVLRVGN	
UPK3c.platypus	-NLSELSKK-QY	YMT	RATRD	LY	PG	GNN-----SSFLHLVRVGG
UPK3c.lyzard	--FDQLEKN-GLY	LTLN	TAP	ASYP	CP	EPGA-----AGGPLTVLRVGN
UPK3c.chicken	-SFQRFPTNTS	AYLTL	GMQY	HYG	CP	PKP-----DRELTVLRVGS
UPK3c.turtle	-AFQKPPGS-PHY	MTGT	SSLN	Y	PEKS-----SGQITVLRVGN	
UPK2a.human	VVTSSFVVP-----P	CRGR	REL	V---SVV---	DSGAGFTVTRLSAYQV--	
UPK2a.mouse	VVKSDFVVP-----P	CRGR	REL	V---SVV---	DSGSGYTVTRLSAYQV--	
UPK2a.dog	VVKSSFVVP-----P	CRGR	REL	V---SVV---	DSGAGFTVTRLSAYQV--	
UPK2a.cow	VVRSSFVVP-----P	CRGR	REL	V---SVV---	DSGSGFTVTRLSAYQV--	
UPK2a.elephant	VVKSSFVVP-----P	CRGH	REL	V---SVV---	DSGAGFTVTRLSAYQV--	
UPK2a.opossum	GMTHNFVVP-----P	CRARR	DL	V---SVV---	YNSGSPSITRLSAYQV--	
UPK2a.platypus	VLQQRVFPVVP-----P	CRGR	REL	V---SVV---	DSGAGFATTRLDAYQI--	
UPK2a.salamande	TNNQTVFPVVP-----P	CLRR	DAV	V---FSS---	DSSSGTIVITNIGFRV--	
UPK2a.Xlaevis	VQNISLQVP-----Q	CLRK	RDV	V---VIN---	NSQSGMVTMNVGYQI--	
UPK2a.Xtropical	IQNATFQVP-----Q	CLRK	RDIV	V---VIN---	NGQSGMVTMNVGYQI--	
UPK2a.coelacant	-KTESFVVP-----P	CRFR	REV	V---EVA---	RQMEGFTVTDLLGFRV--	
UPK2a.skate	-AQPNFVRP-----V	CRNR	DLI	V---SLV---	SNADGTPQTLMLGYML--	
UPK2a.zebrafish	NNTVNFVVP-----P	CPV	SQ	S---	WYLL-----	
UPK2a.salmo	TLVNI	FKVL-----P	CRFR	RDII	V---STI---	ENNAQFTSRNLGYQV--
UPK2a.lamprey	-IVGQVAVP-----G	CRVAR	AVSG	-QVVS	GN	DTGPGIPVPSGVAYRV--
UPK2b.snake	YFADEFKDI-----S	CRVAR	DL	V---VLD---	LDDSQYELITVVGQVGG	
UPK2b.turtle	-HSSKFKSV-----S	CR	IARD	V---SMD---	IDDGNVKTITVVGQVGV	
UPK2b.lyzard	YFAGEFDSV-----S	CRVTR	DL	V---VMD---	MDDSQFELITVVGQVGA	
UPK2b.aligator	YFAGEFASV-----S	CRMAR	AL	V---KMD---	IDDENYKLTITVVGQVGA	
UPK2b.salamande	YYAGQFNVT-----P	CRVSR	DV	V---QVSSK---	AADNSFTLITVVGQVGS	
UPK2b.Xlaevis	YTGQMKVP-----Q	CLRR	RDLE	V---PVK---	MNS-----MRDLGYQVGT	
UPK2b.shark	-LPQRYSPV-----F	Q	NDFS	PKPA-----AVSTPLAYQMGP		
UPK3a.human	NGTCLW---DPNF-QGL	CNAP	LSAATEY	RFKY	---VLVNM-STGL---VQDQTLWSDP	
UPK3a.mouse	NGTCFW---DPNF-QGL	CNAP	PLTAATEY	RFKY	---VLVNM-STGL---VQDQTLWSDP	
UPK3a.cow	NGTCLL---DPNF-QGL	CNAP	PLSAATEY	RFKY	---VLVNM-SSGL---VQDQTLWSDP	
UPK3a.elephant	NGTCLS---DPNF-QGL	CNAP	PLSVAATEY	RFKY	---VLVNM-STGL---VQDQTLWSDP	
UPK3a.dog	NGTCLS---DPNF-GGL	CNAP	LSAATEY	RFKY	---VLVNM-STGL---VQDQTLWSDP	
UPK3a.opossum	NGYCSL---DPNF-KGT	CNAP	PLTRATEY	RFKY	---VLINP-SSGF---VQDQTLWSQP	
UPK3a.platypus	DTACMW---DPDF-SG	CNAP	PLAGET	TY	RFKY---VLVNT-TSGS---VEDQSLWSDP	
UPK3a.chicken	DGTCCLY---DPNF-LD	CNAP	PLADTT	TY	RFKY---VLVNT-TEGI---VKDQTLWSDP	
UPK3a.aligator	DVSCCLY---DPNF-LG	CNAP	PLADTT	TY	RFKY---ILVDE-TLGI---MKDET	
UPK3a.turtle	DVTCCLY---DPNF-PG	CNAP	PLAQD	TT	TY	RFKY---LLVDV-NAGV---VKDQTLWSDP

UPK3a.Xtropical DVTCLT---NPNF-VGYCNAPLQGNQY^SFKY---LFTD--SGDI---VQSETSWSLG
 UPK3a.Xlaevis DVNCLT---DPNF-KGICNPPLQNNLQY^RFY---VFT---IGDV---VOYQTDWSP
 UPK3a.salamande NPTCLT---DPNF-SGICNPPLANSTAY^RFLY---SLVHN-NGS---RVANTHWSDR
 UPK3a.coelacant DTMCNMM---PGICNGPLTPNTAY^RFKY---VLLNG-NR---PAAQTWPSSM
 UPK3a.spottedgeda NPTCVG---EPEA-EAFCNAPLSDGTSY^RFKY---LLVNG-TT---TIARETWSSES
 UPK3a.skate DQOCLN---QAPFVSGNENAPLKENVAY^RFKY---AVLNS-STNI---ILNETSWSDP
 UPK3a.shark DTQCLS---GGTAQSVPCNAPLNNVNY^RFKY---VARDP-ITFI---LRDETVMWSKP
 UPK3a.3.lamprey DPACATGIADDA-ALVNGPLWNGTAY^RVKF---VLLAR-SLLS-P-PIVAMPTWSDP
 UPK3a.1.lamprey DTACVN---SIGPSTSVNGPLVPMGY^RVKY---TLSEE-SPOFP-R-PMDETWSWSDP
 UPK3a.2.lamprey DSKCSV---AVTNGNPLNAGTIY^WFKY---IIGTV-ATHGAIDRSYLESSWSKP
 UPK3b.human DHGC---HQ-QPFCNAPLPGPGPY^RVKF---LLMDT-RGS---PRAETKWSDP
 UPK3b.cow DAGCLA---DLHQ-PRYCNAPLPGPGPY^RVKF---LLTNS-RGS---PQAETRWSDL
 UPK3b.mouse DFGC---YQ-RPYCNAPLPSGPGY^SVKF---LVMDA-AGP---PKAETKWSNP
 UPK3b.elephant DARCLA---DLQO-SPYCNAPLPSGPGY^SVKF---LLMDT-EGS---PKAETRWSDS
 UPK3b.dog DPGCLA---DFYE-PPYCNPLPSGPGY^RVKF---LLMDA-RGS---PQAETRWSDP
 UPK3b.platypus DPGCLR---DPGR-EYCNAPLPAAGPY^RTKF---LVMDE-ASH---PKAETQWSDP
 UPK3b.opossum DTNCLR---NLSQ-EYCNAPLLAPGY^RVKF---LVMDN-NGO---PKAETWSDP
 UPK3b.Xtropical TAQOND---N---TNCNGPLPDPGYPY^RVKY---LVMNN-NA---LVSQSLWSQQ
 UPK3b.Xlaevis TVYCTD---N---TYCNAPLSDSGPY^RVKF---VVMNN-NA---LVSQSLWSGL
 UPK3b1.Xtropical DVNCSN---PLFCNGPVPSPRGTY^RVRF---VVLNG-TV---METGRWSEV
 UPK3b2.Xtropical DEKONE---SFCNGPLTSPGTY^RVRF---VVLNN-TG---MVAKTNSDL
 UPK3b.chicken DTACHG---RABCNGLPSPGTY^RVKF---LLMCC-GG---PKAETKWSDP
 UPK3b.aligator DAFCRN---DNSQ-QHCNGPLNPGPY^RVKF---LILDS-NG---AKAETRWSER
 UPK3d.coelacant DTSCPQ---SIGE-SDCNGPLPDSGSY^RARF---LIVDS-QSGN---MLKAQSQWSSK
 UPK3d.1.spotted EGNCTS---ANCNGILPAGSTV^RVKY---VLVDP-ASRN---VTTETYWSQN
 UPK3d.salmo EGNCTN---TNCNGILPAGSTV^RVKY---LLIDP-VSRG---VVSSEKWSYP
 UPK3d.trout EGNCTN---TNCNGILPAGSTV^RVKY---LLIDP-VSRG---VVSSEKWSYP
 UPK3d.zebrafish DGICTG---INCNGVLEPVGSIV^SFRY---LLIDP-SNTY---VNMNTNWGGP
 UPK3b.skate EDPCTT---PNCNAPLNAGSLV^RVRY---VMINP-LATT---NMVIAVTKWSNP
 UPK3b.shark ETTCRT---ETCNVPLNNSKSY^SVKF---LILNP-SLQNS-PNVIAQTNWSSE
 UPK3g1.salmo DGNCSST---PNCNGILPVGSTAR^RFKY---VLINP-ANKT---VVAESLWSNN
 UPK3d1.trout DGSCST---PNCNGILPVGSTAR^RFKY---VLINP-ENKT---VVAESLWSNN
 UPK3d.spottedgeda EGNCSST---PTCNGILPAGSTASCAVLGLAVVSQ^LQDL---VKGALLWAQQ
 UPK3g2.spottedgeda EGLCST---SNCGNGLPAGSTV^RAKY---VLLDP-GSQJ---VVSSEQWSSP
 UPK3c.human DTHCOPT---KIGCNHPLPGPGPY^RVKF---LVMND-EG---PVAETKWSSD
 UPK3c.mouse DNNCSLE---SQGNSPLPGAGPY^RVKF---LAMSA-EG---PVAETLWSEE
 UPK3c.dog DTSCSPT---KRCGNHPLPGPGPY^RVKF---LVMSD-KG---PVAETWESNE
 UPK3c.cow DTRCSPR---TRCGNRLPGLPGPY^RVKF---LVMSD-RG---PMAETWESSE
 UPK3c.elephant DTRCSLT---TKCGNHPLPGPGPY^RVKF---LVLSDDRG---LVAETWESGE
 UPK3c.opossum ETNCTR---SDCNKPLPGPGPY^RVKF---LVMNT-NG---PVAGTNWSED
 UPK3c.platypus EANCTS---KSCNGYLPGLPGPY^RVKF---LVMNE-RG---PVAETDWSDN
 UPK3c.lyzard EVQCAS---NRAR-PDCNGLPRPGPY^RVKF---LAINP-DG---VTAESSEWSEE
 UPK3c.chicken ETGCAD---NISV-PNCNGPLPGPGPY^WVKF---LALNG-SE---PTATTEWSP
 UPK3c.turtle ETGCVS---DTTR-PDCNGLPLGLPGY^RVKF---LAMSP-VTG---PTAETRWSDP
 UPK2a.human -----TNLVPGTKFY^ISY---LVKKG-TAT---ESSREIP
 UPK2a.mouse -----TNLTPGTKY^ISY---RVQKG-TST---ESSPETP
 UPK2a.dog -----TNLVPGTKY^ISY---LVRKG-SST---ESSREIP
 UPK2a.cow -----TNLAPGTKY^ISY---LVTRK-AST---ESSREIP
 UPK2a.elephant -----TNLVPGTKY^ISY---LVKKG-TST---ESSKESA
 UPK2a.opossum -----TNLIPGTKY^FVY---SVEKG-TAV---ESSNKVQ
 UPK2a.platypus -----TGLRPATTY^VXXX---XXXX-XXX---XXXXXXXX
 UPK2a.salamande -----TNLTANTY^TASY---QSNGV-TIGLPTN-----
 UPK2a.Xlaevis -----QNLQPGAIY^TTY---AVDGS-NIP-----SIT
 UPK2a.Xtropical -----QNLQPGTNY^MATY---SSGGI-SGP-----SFQ
 UPK2a.coelacant -----GNLQACTQY^DFRY---TIDNT-TNV---LKSNI---IQ
 UPK2a.skate -----EMLQPSTTY^NVYL---RAGTI-RSNMLG-----
 UPK2a.zebrafish -----GNLNGTTY^SMSY---KIGND-TSSVL-----
 UPK2a.salmo -----TNLTTGSTY^RFQY---VVGAE-KSNILE-----
 UPK2a.lamprey -----TGLTPSTYS^VVL-RHSTLGLQ-SVPGI-----
 UPK2b.snake EF-CRQTK---GPFQCNQALKPSTFY^RVNF---PFLDD-KSV---IRAHTDWSSTA
 EV-CEKTK---GPFQCNQALQPSVY^RVNF---FILDE-KAV---IRAHTGWSDV
 UPK2b.turtle EF-CRQTK---GPFQCNQALKPSTFY^RVNF---PFLDN-KSV---IRAHTDWSSTA
 UPK2b.lyzard EF-CRQTK---GPFQCNQALKPSTFY^RVNF---PFLDN-KSV---IRAHTDWSSTA
 UPK2b.aligator EV-CENTK---GPFQCNRLKPSSTFY^RVNF---FVLDE-NAV---VRAHTDWSDP
 UPK2b.salamande EV-CTSVK---GLYCNQVLEPGTPY^RVNF---FILDA-SNV---IRAYTDWSDI
 UPK2b.Xlaevis EN-CTEVS---GPFQCNQALQPSVY^RVNF---IILDE-TDT---PRAYTGWSEP
 UPK2b.shark NVACID---GSCIEHVLPGQR^FVRY---IISV-SEE---ALVTTKWSAP
 UPK3a.human -IRTNR-LTPYSTIDTPGRRSGGMIVITSILGSLPFFLLVGFAGAI^LSLV-----
 UPK3a.mouse -IWTNR-PIPYSAIDTPGRRSGGMIVITSILGSLPFFLLVGFAGAI^LLSFV-----
 UPK3a.cow -IRTRD-LTLYSAIDTPGRRSGGMIVITSILGSLPFFLLVGFAGAI^LSLV-----
 UPK3a.elephant -IRTRN-LTPYSAIDTPGRRSGGMIVITA^LILGSLPFFLLVGFAGAV^LLSFV-----
 UPK3a.dog -IRTRN-PTPYAAIDTPGRRSGGMIVITSILGSLPFFLLVAFAGAV^LSLL-----
 UPK3a.opossum -IRTNQ-ISPYLEIDTPGRRSGGMIVITSILSTLVPFFLLVGFAGAAV^LFSFV-----
 UPK3a.platypus -IRTRK-TLSWSRVDTPGRRSGGMIVITSILSSLMFVLLVGLAAAV^TCRV-----
 UPK3a.chicken -IKTRK-AKLPKMDIWPGRSSGGMIVITSILSVSVFLLLAGLASV^FSAV-----
 UPK3a.aligator -IKTNS-VKKSSTIDTPGRRSGGMIVITSILSVLIFLLLAGLFAV^FSAV-----
 UPK3a.turtle -MKTRR-VKQSTIDTPGRRSGGMIVITSILSTLMFILLVAGFLASV^FIVM-----
 UPK3a.Xtropical -ITTVN-GKASSTIDTPGRRSGGMIVLTSILSTLMFFVPIAYVIGFAYSIA-----
 UPK3a.Xlaevis -ISTVN-VKSSGTIDTPGRRSGGMIVLTSILSTLMFFVFPYAVIGFAYSIA-----
 UPK3a.salamande -ISTKN-VKTPDLDTPGRRSGGMIVITSILSTLFFLLSGFVAAAV-----
 UPK3a.coelacant -ISTRK-SKPFEDIDTPGKRTGGQVVVTTILVILLFLLCGFV^TLVASI-----
 UPK3a.spottedgeda -ILTRK-ALSPDEIDTWIGKRS^GGGIVVTVILSLLFLLGAAIFPMGV^LDI-----
 UPK3a.skate -IPLLR-VADFALIDTPGARTGGMVIT^TLLVILLSSLLLCGVALLYAC-----
 UPK3a.shark -ITLLQ-VQDPARIITWPGARTGGMV^LTTLLVILLFLLCAFV^AL-----

UPK3a.3.lamprey -ISTKP-VVSVEQIHVWPGKRTGGMVVVTSVLCSILFVLLFALTAALLAPS-----
 UPK3a.1.lamprey -VSTKK-SPAASTINTWPGKRTGGMVVTAVLSTLLFLLAALLLVVIFKACS-----
 UPK3a.2.lamprey -IRLNK-AGELNAIGVTPGPXSGGMVVVTVILVLLFIAVDRARPALGQ-----
 UPK3b.human -ITLHQ-GKTPGSIDTWGRRRSGSMIVITSLSSLAGLLLLAFLAASTMRF-----
 UPK3b.cow -TALRQ-GKSPGSIDTWGRRRSGDMIITISILSSLAGLLLLAFLAASSVRF-----
 UPK3b.mouse -YLLHQ-GKNPNSIDTWGRRRSGCMIVITSLSALAGLLLLAFLAASTTRF-----
 UPK3b.elephant -TALHQ-GRAPGSIDTWGRRRSGGMIVITAVLSSLAGLLLLAFLAASTVHF-----
 UPK3b.dog -ITLHQ-GKAPGSIDTWGRRRSGDMIITISILSSLAGLLLLAFLAASTVHF-----
 UPK3b.platypus -ITLQK-GRDPSSVDTWPGRRRSGMVVIAISILSVLAGLLALLLTAAGACTV-----
 UPK3b.opossum -ITLQK-GKDPRESIDTWGRRRSGCMIVITSLSTFAGLLVIAFLIASTVQF-----
 UPK3b.Xtropical -ITLLT-GKSSSQLDTPGRRRSGGMIVLTSILSVLMGILTLCLFAAFVFG-----
 UPK3b.Xlaevis -ITLRT-GKNPSTIDTWGRRRSGGMIVLTSILSLLMGLITLCLIAAFVFG-----
 UPK3b1.Xtropical -ITMHI-AINNSTIKTEPKRPPSGMIVITILASLLFLLVCLIAAVSLGS-----
 UPK3b2.Xtropical -IRLPI-GINYYTIDTWPLSRSGSMIVITLISLLAVLLACLALLALCE-----
 UPK3b.chicken -ILLRR-ARSLSTIDTTPARRSSTAVVIAAILASLGAALAMVAVGAV-----
 UPK3b.aligator -ITLQK-GHRSSTIDTWGRRRSGTMVVITVILSSVLGILTI AFLCTAYS-----
 UPK3b.coelacant -IQLKR-PKDPASIDTWPGKRTGGMVVVTILVLIALLLFLIVALITAC-----
 UPK3d.1.spotted -ITLYS-STDPPVVFDPGIRQSRAGMIVITSLSVLLFLLLLL IAGLIYAK-----
 UPK3d.salmo -ISLTS-TRSWSSIDEWLGKRRSGGMVITVSSCLLAVLLLLLGAVLLLDG-----
 UPK3d.trout -ISLTS-TRSWSSIDEWLGKRRSGGMVITVSSCLLAVLLLLLGAVLLLDG-----
 UPK3d.zebrafish -FNLT-LLSYQIINDGLSARSGMVVITLCLVAVALLLVFFMLCVSC-----
 UPK3b.skate -IQLEN-AVDPNIIDT-STRRSGAMVITLISILLFLLLVFLVMLAA-----
 UPK3b.shark -ISLKT-VVEPRNIDPSRGRSAGMIVITAILSVLLFLLLAFFVAMLLMVC-----
 UPK3d1.salmo -ITLYS-LKDPEKIDNGFAGRSAAMIVITAILCSFLALLLLLMLLIYVL-----
 UPK3d1.trout -ITLYP-LKDLESIDHGFAGRSASMIVITAILCSFLALLLLLMLLIYVLC-----
 UPK3d1.spottedga GLVVVD-VMPSSIDDWTKRRSGGMVITVITSCLLAALLLLLVAALLLGG-----
 UPK3d2.spottedg -ISLIA-LSDSAGIDEWIGKRRSGGMVITVSSLLAALLLLLTAFLVLR-----
 UPK3c.human -TRLQQ-AQ---ALRAVPGPQSGTGVVIAIILSILLAVLLTLLAVLIYTC-----
 UPK3c.mouse -IYLQQ-AQ---TFREAPGQSGKGTVVIIAFLSILLAILLVVFLVVISAC-----
 UPK3c.dog -THLQR-AE---RLQAAPGPQSTGTVVIIAIVLAVLLVLLTALLALIYTC-----
 UPK3c.cow -TRLQQ-AE---VLQAAPGPQTAGTVVIAIILSVLLAVLLAALLLIPTW-----
 UPK3c.elephant -THLQR-AE---VLQASPGPQTAGTVVITFLSVLLVLLTALLALIYTC-----
 UPK3c.opossum -ITLRK-PVEFS--ESRPPSKSAGTIVIIAIVLILLLLPLALVALLIYTC-----
 UPK3c.platypus -IALRE-AKDPG--DPNPTRRSTFMIVITLISVLFALLGALTVGLIQA-----
 UPK3c.lyzard -IALVQ-AQSPETIDVSPGRRSASAIASLSILCAVLLAALLIALVYKYT-----
 UPK3c.chicken -ITLKT-AREPQSIPIGMMGARSAMIAITAILSVLLAALLAALLATLCS-----
 UPK3c.turtle -ILLKA-GKDPATIDTWGKRSAGMIVITLISILLAILLACPIAALTYRC-----
 UPK2a.human -MSTLP-RRNMESIGLGMA-RTGGMVVITVLLSVAMFLLVGLFIALALGSRK-----
 UPK2a.mouse -MSTLP-RRNMESIGLGMA-RTGGMVVITVLLSVAMFLLVGLFIALALGSRK-----
 UPK2a.dog -MSTLP-RRQAESIGLGMA-RTGGMVVITVLLSVAMFLLVGLFIALALGARK-----
 UPK2a.cow -MSTFP-RRKAESIGLAMA-RTGGMVVITVLLSVAMFLLVGLFIALALGARK-----
 UPK2a.elephant -MSTLP-RRKMESTGLGMA-RTGGMVVITVLLSVAMFLLVGLFIALALGAQK-----
 UPK2a.opossum -MATLP-RRKVETLGLGMA-RTGGMVVITVLLSVAMFLLVGLFIALALGVHK-----
 UPK2a.platypus -XXXXX-XQKAEALLVVGMA-RTGGMVVITVLLSVAMFLLVGLFIALVLSGSD-----
 UPK2A.salamande -FTTVQ-PTNYTAM-PEVFARSGGMVITVLLSIAMAILVIALIILTFVMGRKK-----
 UPK2a.Xlaevis -FST---RSVSQTV-PDIMARSGGMVITVLLSIAMFVLLVGLIIVLVI GRK-----
 UPK2a.Xtropical -FST---RTVYPVAV-ANIMARSGGMVITVLLSIAMFVLLVGLIIVLVI GRK-----
 UPK2a.coelacant -ITTSI-VTSNVLIDEGFKLHSGGMIVITLISFAMFLLIIGVIVVVLGNKS-----
 UPK2a.skate -VTTIS-PVDYRTIDLGFQ-RSGAMVITVILSIAMLALI IAFIVVVLVLSK-----
 UPK2a.zebrafish -TNMTT-NNDYQQIDTGLRARSAGMVVITVILSLAMVFLVGLIIVVFFPSG-----
 UPK2a.salmo -VSTRQ-VKDHNQIDSGLPACSGAMVITVILSVSMFILLVALIFTVAHSLGGD-----
 UPK2a.lamprey -LSTANARTPRSAVDAAEFRRSGGMVITVILSVLIFLLAILVAALLGNKN-----
 UPK2b.snake -IQTRN-VTNYESADVMFEGRAGGMIVITILVSVGGVLLVALIIVAAVLSKK-----
 UPK2b.turtle -IQTRN-VTTFMAYDGSFIGRAGGMIVITVLLSVAMFLLVGLIIVAAVLSKK-----
 UPK2b.lyzard -IQTRN-VSDHESADVMFEGRAGGMIVITILVSVGGVLLVALIIVAAVLSKNK-----
 UPK2b.aligator -IQTRN-VTSFAYSADGSEGRAGGMIVITVLLSVGMFLVVVGLIIVAAVLSGRKS-----
 UPK2b.salamande -VTTLN-VTNNALDLSGLRRRSGGMVITVLLSIALFMLVPALIAITLVVGREKSP-----
 UPK2b.Xlaevis -RTTRQ-VRRLDAVDLGLSGHSGGMVITVLLSVSVFLLVGLFMVAVVVRAPNFTSES-----
 UPK2b.shark -IATRDDPPSYFSIDATPTPRSGAMVITVLLVVALFLLVGLFVSMLAARSKFAR-----

Material Suplementario 1: Secuencias y código de todas las secuencias de PTPR subtipo R3. El dominio urofibrina y el dominio transmembrana están resaltados de color verde y rojo respectivamente. Los exones están marcados en color azul o negro de manera alternada. Las letras rojas indican la zona de unión entre dos exones adyacentes por fase 1 o 2. Los residuos de cisteínas y el motivo CNGPL están resaltados de amarillo.

PORIFERA

Spongue XP_003383454.1 curated by pteoin profofiling [Amphimedon queenslandica]

MFFIIDILIKNFYLGFPADVINFVNSGNSTSLALSQRLRYPLVPPDMYYIAYNYTELSGANPRSDNSTIMIDEEDDVTITTT
DYILYTLDDLLPFTSYVTLFPQYGEEEGEGLSDSGNTGEGMMSSQVNDTNVDIDRDNMRLNISWAPPSPRPNAGNLRYEVT
LSIRGKICLITINTELSVIAEFKFNVTYITIRITPFNSFGSGTTRELDMFATPEGAPTPPLNFTLTELSSLKASWEEPST
LNGVLANYTVYCNLSSLQFYSQQLMLSFNAPPTVDPDSTANITGLYPTNYCYATASTGGESSRNSNDARSARTSEEEP
AGPPEGFNVTDVATSVSLEWRRPVPNGVILHYFLQYSAYNVTIPVTFNTSSDEYSVTSYTVESLNEYNTNYTFTISAVTS
GGAGPLATTSTRTSEA VPGTNLDGFNTSSMSLRITWTELLEDQNGVIIGYNIYSYFSLPAVGQPINNFTSDTSYNISGLDV
YTDYNVSVAAAYSAGTGPFDVSVIRRTDSTVPTSPSDVSYINISSTSEI EVSWNPPTDFNGPNEGVIYITIRLESDETESMSTN
RLTGTSFVINELEKYEQYSVTVAFTDKGPGASSDVLVLTDEDLPGPPSNVSTMSNTSISITWSPPLDPNGLLLSYSIN
VTLNSTYAQYLSFDVMTMSVSNIFSYTLFDLLPFAGYDISLQASTSVGLGTPAIEETETTLQAAPAAPVANPMASPVSSA
VNVSWLPPNLSNWNGLITNYTIEYRTNDEYIRPSIEVTATLSSFANSDDPRAATTPQLQESIIIPSLHEFVNYSFIIITLSN
SIGQISSLPVHIWTLSTPGSPPLDVSVIVDSPSTARISWSPPMYIDRNGIIVNYTVRIITTVRGTIRETNI SNVSGNYD
ASDLPOFASFNVTVAAATSVGLGYPSPVSNMTFEGVPSADPVPVIVKRRNDTAIQVNWTRPAEPNGIILGYLIYYIGTKNN
TGTEYSNINVLIIINVTDPNTLSYLITNLLADTQYFINVTAITSAGLGAVGDGDSILLPTGR PTPVPPNVTIGTPPTGGTS
DPTTATITRIEVTIPEELNANGPLTRIRILIRIFLSRNDTISTWYESQKFPNSVAPPWQATQLPLNQGRRRKRQAGGETVA
ETIGTNNSCGPNDIVCNGPLKPGTQYQFKYRVYNSDDDDSYVESQYSGPIRTGPIAEENNTGTT VYAGVYVIVLVAIHLG
ILYIVYVIVYKRRRKRKAYSFAASFNKGKEELITAGEGKATLSKAIQLTPTDPTIIVPLTENTIVNPTAVRTASVKATPNR
KVYLRDFPSYVEDMLNDAGFKFSEEEYKVS AVGLDHSKDALLPENRAKNRYTNILAYDHSRVKLESIDDEPGSDYINANY
IPGYRMRRAYIATQGPLPSTFDDFWRMTWEQNSHVIVMLTQLVERGR TKCHRYWPGAQPEVYGEINVDMLSETEKSDWIVR
KFKITKEKRSRTIITHYQFVSWPDHGVPEAGPALDFREVHEVASSAFGPVIVHCSAGVGRGTGFIALGTLQLQHKDHDV
DLFGLASEMRQHRNHMIQTEPQYVFIHKAMVDACKSSMKFGGKIKVLPSPFIYSNIGPASQLSDFYDMLSEEAASYQYK
TSGNGKTYDESTF

PROTOSTOMAS

Nematode.2 CeNTP10D C. elegans. F44G4.8a DEP-1 NP_001254192.1 [Caenorhabditis elegans]

MIRWKYELHSLIWLFLVLHLKSKCQSDSLTSAEQHELFAIKKDSLSPWSQILVSLPRRHPLYQSFAAKIQDVTEINISDEV
DSNKTFSVSSDDAPYSIRIHALRAGHRYSAI IHGQKDGSTSLIKEESVVMPPRAPDFRSMDSDIQVAEHNITMRTIKNDSYL
QDSFSIEYRQINPDKKFPVLQILDIPEQKNEFYLGNLNSGFDYSVRVIAHKDGMSSRPWISTLTITKPSPLKEVININQAG
SCVEVSWQNDDEFSGADFYTIQYSLQSTPNNSTNMTIPSTESSICDSMLQGEAYQI IATVQKGGQVSEPLITKFLRPLP
PIDFRVRADLKRKGYLLAELPTSSKIDKQITVAGDEAERSVNYANIEQTKSGHKICWFNFALSPGERFDFSISSMANES
ASQKLQKSIVLTPAFDFNAFGLTLQESNGGIELIWPKSEVFMTRVKDIWNKVVGAESSLNMRITP IGNNDTDTLTKFETS
PKNIDPVFAKNLVKGCYRVQLFTVTKTGI ISETRHNETIRMSPAVNVLSVTRSSATLRIVFSTHHDSTISNCQMHI
VVRDMNGKSVFDRKMLTATFAPLLNLDGLSPFKYTVNTQIICGSGSSETPQCPAATRTRMQLSFSTRQDKPAAVQDLKV
EPLNSYSVMLTWLPPALPNGILHYAVNVTKIGSDETRTIDVGVSSNRSDHTVQVVIDELFGGHTYSFVRAVTEAGFGEN
SPVVPVTSMPIMAPPVPTVAPMIMKESVSGSHNMIVREPTTFMDFNRNGEIKQFALIVSETTADESINRWIESDNGTYTWOQV
QRFVDVWPSYVAKLQDIQKVQDQVDSIFEEELGEDETCLEVRADRICNGPLRSASKYRVRIRLFTSPTLFTDSDPPSQVMTG
SATPAIPLLTIVVAVLIVIAFVGVIGTIFLFFWNRTKKARLAAAFAFKNGPSKEKESQWEALKMMMAERAADCLAKGLDATT
PPPSSTSSNSPTSTSTMTDCGSPHLAGPAGNAGGHRRTSLRERTGVEHRLERLSSGPVHRTPLYTVVTGANTNKRSPVR
IEDFADHVLMSADSDRFSEEDMMRNVGVGQSVAASELFINRPNRFTNIPSYDHSRVKLSNPNNIEGGDIYINANYVPG
FSSRRFIAAQGLPPTTRDHFWMTWEQQCPAIIALTKCVEKGRDKCHQYWPDENVPVLYGDLEVTIVAEKEPDEFVIRD
IRLEKSGPDGRVTRFVRWHYMAWPDFGAPSHPNGIIQFSRMFRHHLPHSPHNAPTIVHCSAGVGRSGTFISIDRLLQSS
FGDPIVDFGTVCEMRYERCQMVQNEQQYIFIHICYILQVLQSSSPSTSTSTGAHNNAGFVQDGMIVESGF

Nematode.3 HcNTP10D Barber.Pole.Worm gi|560121775|emb|CDJ93597.1| [*Haemonchus contortus*]

MVGLDVALEVVMVIRVTLTLATAIVFATKELRAASNAESEDTHEKKLDLLEFKIEKDQKLPDSQLTVKLPYNHPSYDQ
FIAKVVDISSTIDPSIRDVNRFTLSSADGSGTIRIHTLHPGHKYSVSVIVGRKGENSSVIKEELVVMDFRVEPFTTQNSDI
TMHNI TLRAVKNKFKLEDSFLIEYRQLEPEQGYVLEVLDIPEQKNLEVYLGNLNPGRDYSVQVAVKTKGKSRPWSTTL
STKPGAVSDLIVSENSCLTVEWDVPPNSGADSFLLRRLYRQLNAAVNL SVTI PGNDRFVDLCDGIVPGTAYALGVVAKKGG
SVSDETIKTYTVRASPSTDFQVSPDITKGYRLTFDLPPNSNYDGHVSVVSETLEALEDDGVEVEKEGKSCSILMPLV
PGERFEFTLSTYVENVTSTKLHRSVLTLPADFMTGFGLSLQEQYQSGVKLSWPQSDVFMRSRMDIWNKVVGPKSQQLMRLF
PADGLEKGSRLQGDPHREATPLIVGALKKGSYKQVIFTVTKSGIVSEKRYNDFRMSAPLVNITVKSITRSSAVIAHAAV
SPGEADPECLLNAILVLDHMSHVLDKTLKAQSHNFSP IELGGLRPFHKYTVNSKITCASGPSDCISTSRITRPMFTFSTMQ
DKPGPVLSSLARLPNPNYSQALMMLPAPLNGILTHYVVDVKSEDDPNDRSLNVAIGTRADHFVETVIDGLSGGERYTF
TVRAVFEAGTGEPLTALLEVRMPLIPRPTSTVPLVIAESITSHSLTVKYSPMFMNKHGRIIRSTLLVAEVTEFGQIS
ETWMNAENETYTMQVQRFDVWPLVYAAAVDEISARGSPSLSQAVGIDQTCEDLPSDVI CNGPLKAGTNYKFKRLRFTATN
LFSDDTFSEIITTEAAARGA LKIAVLALALAVLLGLGLVLFSTYWNRSKKARAYASSKESQWAALKMIMAERAADCLA
KLGLDGSMHNGEIPSPQSHLQPTFGHRRRCRSLRERTGVDRHLERLPSGP INKTPLYTITGVNTNKRSPVKIADFAEHV
RMMSADSDFRFSEYEILRNVGCQGSYTAELPVNPKPKNRFTNILPYDHSRVRMLQAGDQEGADYINANFMPGYSRREF
IAAQGLPPTTRDAFWQMAWEQYCPAIIALTKCVEKGRDKCHQYWPDSSEHSSVYSDIEVTLNNESSYDDFTIRELRLLK
HESTAPRIIWHFHYMAWPDFGVPDHPQGIIRFALMFRSRLPHSALNKPTIVHCSAGVRSRGTFFIADRLLQTIQADRPID
VFGIVHEMYRERCHMVQNEQYIFIHHCILHAIESIASERTGMTEHLHQNPVFEDDDTIAESDF

Annelid gi|443734862|gb|ELU18718.1| CAPTEDRAFT_229306 [*Capitella teleta*]

MSQPVTINQLAGFVLAALALICVRSSESSSTTTAAVPTTNFSTNIPPTTQGLEVEVTTSMVTLVWSYDGEQLTFVVEYV
EKDADNATSFVLEQQGVTSCVVEGLQAGWTYTFEVSVQLERDMLGEPAYANATTIPGKPINFRVTDNSVGATFFTLWESAA
GSQQDDFVVYRKIGSSDVVQVFTGGATSVTLHGLDAGVLYEVKVMVARGQLSTPAIITEVTRPPTPRIHNITAVDTKSV
TVQYSAGAGGSQNSFVVTVYNQEGQTNPEVCSYNTCTFDVSVFGVKPGEHFTVKIKATQDMVSSAWASMSGNTKPLPSPA
PKLFRSETTLTVIGDVPSTNSVFDGCKASVTPAIQNGPVYFKRNGGYEIRLQNPVTPGSKYAINIWTYSREESAHVSGD
FYSMPAQPGVISRHTGIATTTKQLAVSWGAPNSGQVGGYVAKLYEHGKSAIESKTLTNTMRQATFHNLTGQYVDVSVSS
YVGTGVYSAFSTIELSTKPESANLELKTTSSTLVQWDEPSTGFYDEFVLR IQSHGETEKITKSSQSTYEKEFDNLSAG
TEFTIVIVTVSHGQQTATIGIFYTKPERPSNLKKESTSDSITVKNAPVSAFDSYRITINSTDHENEVLDSSESSTF
SGLMPAQYDITLISLNLNLSLESSMLSIQVYTNPAAPTDLRILSRNTNMSFTWSPPTNPGTIDDYKYSITISGAATPEIER
SGTINEYSHQLSAGKTYIEVRAVTNNQRGLPVSLTFMTIPNDPGNVREMERGQSYLKI EFDAPWDTYDGFVTVWKEGSG
PSSVADLLVTSYNDDSITVEWEKPTGTTVNGFTISIDPPSPGQSGSKDLTENIQEYTWGLSAGREYNLIEVIFADTKSE
KVERRQTKPKRALSLSANVSTDSISLSWKVDQSSQDNFKITYEVTGSSTEWRETTFQGGQLQYSYLIEGLSAGSTYS
ISVIKAKSSVQSDPLEITQIQPKSRELKAEVIPSIGLSLWLPGFDPSTQNSYRYQYQGNVKNLNI VPVVWINSVLDLQ
LFPFERQYNYVDIAISNDQYSPNSQSTLATTYPLPPTDLTVDRSATTVSSVRVQWKDDVTRSYITSWDKIADRGTDNIRRV
GSTHERTNLNYEIPNLTAGKNYTVYYKSGNGQHSRDASTMDVTAKPVIRSTLSEDEKTTEDTIAVYTESKAGVPDHYL
FSLNNSDVTIKQRSNTDRNIRFENLVAGVKYITLARTVSGVEESMTTQKSITTKPNKPKKHCNQGTDKLTIILIKPSGFV
DEYILKCLNEDCNSTBELSNVFBQDPFNLLKPYTTYHFEAYTKRFEKLSNKFVPCKTGEGPPGPVQEFYSSENSLKPFEIR
LTWMEPRNPGKI IKYHIKYNIGIKENQVPHTGSEVDEPDMHEQLIKGLKPGFEYTFEIAAETNAIGKKTTLVKTMP IKA
IVMPNSPPPQAITTAISHDKIRIILTNPFLNTNGDVVAFSVFVTTDPNERFMANSPLRTWADVKGSPPMASYFAVYCANI
FDGNDQCSSGPARKRRAVQPRNTEFTVGGDSSCTTNADDY CNGPLDAESTYYVALVGYTENDLYSSGSPSEPIRTTAPT
NLLL IIVVVVVLVLAIVAAIGVIVYMKRSSDNNEDKLVCCVLVIGVPAPEWPPGNSNGMVIIPSKPKVKVDFVQHV
ASMAADSEFKYSEYEELSLVGKDQPKAASDLPCNRSKNRFINILPYDHSRVLKLLPIDDEDGGDYINANWIPGYNSRREYI
ASQGPLPSTRDDHWRMIWECNCQCIVMLTKCVAEARQRCDHYWPIDSEAIYYGDLQVILNETKSTEDWISSEFTIISKGDQS
RRLKHFHYHSPDMKAPNKTGSIVPFRVLRIRAKVQLDGGPLLVHCSAGVGRGTGTFIALDHLMQHIEEHDDIDYGTVYQMR
KHRCMVQTESQYIFIHKCLEFVLEGGEDDEYANTGAIGTENRAFEGLRSESDLVAEIFFNLPRLQALTNFDDDEGIDVKS
D

Fruit Fly DPTP4e L20894 FBtr0301793 *Drosophila melanogaster*

MDCATRKQQQLRAHQQQQIQIQTGRKRQQLQQRHNNHHYQNSQQQKHVWLVVGIITFLAQHANAADLVINVPNA
SSNANAFYRIDYSPPPGFPEPNTTIPASDIGKDKFRRALPGTEYNFWLYTNSTHREQLTWTVNIITAPDPPANLVSQLR
SSKSAFITWRPPGSGRYSGRFIRVGLTDLPLFERSYSLEGNETLQLSAKELTPGGSYVQVQAYSVYQKESVAYTSRNF^{TTK}
PNTPGKFIVWFRNETLLVWQPPFPAGIYTHYRVSITPDDAIQSVLYVEREGEPGPAQAFAFKGLVPGREYNISVQTVSE
DETSSVPTTARYLTVPERVLNVTFDEAYTSSSFVRVWEPRTYSEFDAYQVMLSTSRIRFNVPRANGDSVYFDYDILE
PGRTYEVVVKTIADNNSWNPASGEVTLRPRPVRSLGGFLDDRSNALHISWEPAETGRQDSYRISYHEQTNASEVPAPFPVA
AESQITNTLNEYTLDLLAGRRYLIAVQALSKGVASNASDITRYTRAPAPLIQELRSIDQGLMLSWRSDVNSRQDRYEVHY
QRNGTREERTMATNETSLTIHYLHPGSGYEVKVHAI SHGVRSEPHSYFQAVFPKPPQNLTLQTVHTNLVVLHWQAPEGSDF

SEYVVRYRTDASWPQRISGLHENEARKIDMHYGERYLQVNTVVSFGVESPHPLELNVTMPQPVSNNVPLVDSRNLTLEWP
RPDGHVDFYTLKWWPDEEDRVFVKNTQLLEDLSSPSVRIPIEDLSPGRYRFVQVASSNGIRSGTTHLSTRTMPLIQSDV
FIANAGHEQQDETTILSYTTPADSTRFDIYRFSGMDPTIKDKEKLANDTERKLSFSGLTPGKLYNVTVVWTVSGGVASLP
VQRLRLHPLPIISDLKAIQVAAREITLHWTPAPAGEYDFELQYLSADEEAPQLLNQVTKNTEITLQGLRPHYNYTFTVVVR
SGSIQGTDFADVSVSTLMRSSAPISASYQTLTAPPKGVDFYQPSDVQPGEVTFEWSLEPAEQHGPIDYFRITCQNADDAAD
VSSYEPVNAVATQKIDGLVPGNHYIFRIQAKSALGYGAEREHIQTMP **L**LAPPVPEPSVTPLEVSRSTSSTIEISPRQGYFSN
AHGMVRSYTTIIAEDVGKNASGLEMPSQDVAQYTVVWLPYQAIIEPYNPFLTSNGSRKSSLEAEHFTIGTAN**C**DKHQAGYCN
GPLRAGTTRYRIKIRAFTEDEKFTDTVYSSPITERSDITVIV **A**IVSAVILAVAVLVVAVYCYQHRCQLIRRASKLARMQDELA
ALPEGYITPNRPVHVKDFSEHYRIMSADSDRFSEEFEEELKHVGRDQACSFANLPCNRPKNRFNTILPYDHSRFLKQPVDD
DDGSDYINANYMPGHNSPREFIVTQGPLHSTREEFWRMWESNSRAIVMLTRCFEKGREKCDQYWPVDRVAMFYGDIKVQL
IIDTHYHDWISIEFWMVSRNCESRIMRHFHFTTWPDFGVPEPPQSLVRFVRAFRDVIIGTDMRPIIVHCSAGVGRSGTFIALD
RILQHIIHKSDYVDIFGIVFAMR**K**ERVFMVQTEQQYVCIHQCLLAVLEGEHELLADSLHLHANDHGYEV**V**KIYLERQPQTKMG
LPIFRASLAMAELKLDADLMTNKDEDEDQEQQQQLQLATEVFKPKGSNDDEEDEDDDDDQQLLNNETTATLSSASCS
SSTHDVHVVLQEAIEKPKQE**Q**ERICAGTQSHADTESDNTDSDDDDEDGDGKVKADGAVADEGDGWY

Fruit Fly PTP10D M80465 FBpp0303419 Drosophila melanogaster

MLYQLSKATRIRILKRQKAVPQHRWLWLSLAFLAAFTLKDVRCADLAIIPNNPGLDDGASVRLDYSPPFGYPEPNTTIIASR
EIGDEIQFSRALPGTKYNFWLYYTNFTHHDWLTWTVTITTA**P**ADPPSNLSVQVRSKNAIILWSPPTQGSYAFKIKVLGLS
EASSSYNRFTQVNDNTFQHSVKELTPGATYQVQAYTIYDGKESVAYTSRNFT**K**PNTPGKFIVWFRNETTLVLWQPPYPA
GIYTHYKVSIEPPDANDSVLYVEKEGEPGPAQAFAFKGLVPGRAYNISVQTMSEDEISLPTTAQYRVTPLRPLNVTFRDRF
ITSNFRNLWEAPKGISSEFDKYQVSVATRRQSTVPRSNEPVAFDFRDIAEPKGTFNVIIVLSGKVTSWPATGDVTLRPL
LPVRNRLSINDDKTNTMIITWEADPASTQDEYR**I**VYHELETFNGDSTLTTRDRFTLESLTLPGRNYSVLSVAQVSKMESN
ETSIFVVTSPSSPIEDLKSIRMLGNIISWSDVNSKQEQYEVLYSRNGTSDLRTQKTESRLVINKLQPGAGYELKVFVAVS
HDLRSEPHAYFQAV**Y**PNPRNMTIETVRSNSVLVHWSPPESGEFTEYSIRYRTDSEQQWRVLRPSVRSTEADITDMTKGEKY
TIQVNTVVSFGVESPVQEVNTTVPNPVSNIIQLVDSRNI**L**EWPKPEGRVESYILKWWPSDNPRGVQTKNVSENK**S**ADDL
STVRVLI**G**EGLMPGVQYKFDIQTTSYGLISGITSYLPRTMPLIQSDVVVANGEKEDERDITILSYPTLPQSSSKFDIYRFLS
GDAEIRDKEKLANDTRKVTFTGLVPGRLYNIITVTVSGGVASLP**I**QRDRLYPEPITQLHATNITDTEISLRWDLPKGEY
NDFDIAYLTADNLLAQNMTRNEITISDLRPHRNYFTTVVVRSGTESSVLRSSSPLASFTTNEAVPGRVERFHPTDVQPS
EINFEWLSLPSSEANGVIRQFSIAYTNINLLTDAGMQDFESEAFGVIKNLKPGETYVFKIAQKTAIGFGPEREYRQ**T**MP**I**
APPRPATQVVPEVYR**S**SSTIQIRFRKNYFSDQNGQVRMYTII**V**AEDDAKNASGLEMPSLDQVQSYSVWLPYQAIDPYYPF
ENRSVEDFTIGTENC**D**NHKIGY**C**NGPLKSGTTYRVKVR**A**FTGADKFTDTAYSPFIQ**T**QDNTSL**V**AA**T**PP**C**PI**L**D**V**Q**V**
LLFYKRRRNCRKTTKDSRANDNMSLPDSVIEQNRPILIKNFAEHYRLMSADSDRFSEEFEEELKHVGRDQCTFADLPCN
RPKNRFTN**L**ILPYDHSRFLKQPVDDDEGSDYINANYVPGHNSPREFIVTQGPLHSTRDDFWRMWESNSRAIVMLTRCFEK
REKCDQYWPNDTVPVYFGDIK**V**QILNDSHYADWVMT**F**M**L**CRGSEQRILRHFFHTTWPDFGVNPNPQTLVRFVRAFRDRIG
AEQRP**I**VVHCSAGVGRSGFTITLDRILQQINTSDYVDIFGIVYAMR**K**ERVMMVQTEQQYCIHQCLLAVLEGEKENIVGPAR
EMHDNEGY**E**QQVQLDENGDVVATIEGHLSHHDQQAEEAIDDENAAILHDDQQLTSSFTGHHHTMPPPTSMSSFGGGG

Fruit Fly.Ptp52F-RB FBtr0289985 Drosophila melanogaster

MEATNWQSHRRIHALILILWIAGVQQTVGSLVAVSESQELAE**K**VEVTVVSTVYTLGFSISNEADYSIDKVICRNGSGSE
TEANA**A**H**V**CENLN**P**CTFYTSVVSFRSNVSGKPPSDQTIYAYTEY**K**PKTTVTSVVATANTIKVTWQ**N**DRACVESFGITA
KATDYTKS**F**QLLN**D**KSSQTFVNL**S**ACL**T**HTITL**D**TRNNASVVVD**T**DATD**V**DTQY**A**EPGD**L**VMNV**N**TLASGITMITW**G**DP**S**E
KNCISNYVFKWR**N**DC**G**T**D**Q**N**Q**D**T**T**DD**V**ST**E**T**T**NEIDYV**T**T**P**ME**T**TP**D**TP**S**DE**V**CK**E**WT**D**V**S**SD**G**KL**R**EY**L**L**D**L**Q**GC**D**
LYTFQ**V**F**I**NE**N**STAKAS**Q**FT**S**AE**K**L**V**SAVYEP**S**PTAY**P**TL**H**W**T**W**F**SP**Q**N**H**PK**V**AN**S**V**T**L**T**GP**I**QR**S**EN**K**T**M**N**V**I**T**E
TF**A**I**F**DD**L**DP**C**GI**L**VE**I**V**N**QL**N**GS**A**G**T**K**Y**Q**E**Q**S**T**V**GED**Q**PS**V**I**Q**DP**V**VE**A**E**A**YS**M**EV**S**W**K**TP**E**Y**A**D**L**C**I**D**G**Y**R**L**S**G**W**M**E**
DD**K**L**R**E**A**L**S**I**T**T**Q**NT**V**V**F**D**K**N**L**L**A**C**Q**V**I**I**Q**I**P**Y**T**K**E**N**L**D**G**Q**L**R**Q**V**G**VE**T**KA**A**I**V**D**T**K**L**E**M**K**N**G**S**E**F**I**D**L**I**A**F**
N**A**D**Y**NN**S**C**P**T**I**F**A**L**T**C**N**A**T**T**Q**V**R**N**P**Y**A**E**R**Y**V**E**G**H**S**K**Q**G**F**N**A**S**L**S**P**L**S**P**Y**A**T**Y**V**CK**V**I**L**Y**N**V**A**GP**S**E**P**V**D**AD**M**H**T**T**T**Y**F**
EQ**P**ES**V**M**L**E**K**ST**V**SS**L**L**F**N**W**Q**P**PT**Y**T**N**GP**I**K**Y**Q**A**FL**M**R**H**E**A**S**Y**F**V**PA**D**CA**I**VE**Q**DT**K**SE**T**K**G**DP**S**V**N**FT**G**L**A**PA**V**R**Y**MM**Q**
V**A**A**Q**ND**F**GM**G**V**Y**TE**P**V**I**GI**T**LP**A**V**S**DS**V**T**Q**LT**V**LT**Q**GP**V**NN**N**AV**E**AN**V**TI**T**W**K**V**P**CK**S**NG**D**IE**Y**FL**A**NG**T**R**N**N**F**AP**V**S
F**E**RR**V**EL**D**T**G**N**K**Q**R**M**S**Y**T**E**M**Q**P**FD**Y**T**V**E**V**S**V**K**N**R**D**VE**Q**L**S**SS**V**P**G**SW**Q**SP**A**G**L**PT**I**PS**D**EL**I**Q**M**R**N**AN**V**E**T**S**N**PT**K**
TAI**V**RL**P**AD**I**MT**S**AS**G**DI**K**W**M**AL**M**I**S**Q**K****C**AG**V**PH**L**K**Y**D**V**SS**D**W**P**K**V**LS**Y**Q**E**AG**A**D**G**T**G**D**C****S**LE**Y**Q**T**TE**E**R**W**HP**E**P**V**Q**R**
RD**G**EV**S**D**E**IE**V**FT**I**GL**D**R**C**SE**V**Q**K**T**Y****C**NG**P**L**P**ED**T**D**Y**N**V**V**R**L**F**T**A**S**G**Y**S**DA**V**L**N**F**K****T**KA**E**I**K****V**IA**V**SV**N**CA**V**
VDR**S**Y**V**W**R**K**R**L**A**W**K**R**D**SG**Q**Q**I**ED**P**FG**N**VI**A**KN**F**A**I**F**Y**TE**V**AK**P**E**L**L**A**REF**K**E**I**T**V**VA**L**E**L**S**Y**S**A**S**E**L**G**CH**K**N**R**Y**A**D**I**F**P**
Y**D**KN**R**V**L**D**I**D**A**B**G**SD**Y**I**N**A**S**F**I**D**G**H**T**R**K**K**E**Y**I**A**T**Q**G**PK**P**ES**V**M**D**F**W**R**M**L**Q**Y**N**V**R**VI**Q**V**T**Q**F**REG**N**T**I**K**C**H**E**Y**P****I**V**N**R
GL**T**V**T**L**K**S**E**V**L**E**L**Y**D**R**T**E**L**T**V**H**D**K**Y**L**E**K**E**V**I**H**Y**F**F**K**W**PD**H**GV**P**ED**P**M**H**L**I**M**F**V**K**K**V**CA**E**K**R**PS**Y**SP**I**V**H**CS**A**GP**V**R
T**G**T**F**I**G**L**D**L**I**M**Q**R**L**K**S**E**S**K**I**N**I**F**E**T**V**K**K**L**R**F**Q**R**M**K**M**V**Q**T**Q**Q**Y**T**F**L**Y**A**C**T**Y**E**L**V**K**H**K**I**P**R**A**L**K**M**D**GR**K**PS**V**TV**P**A**I**PS**K**
K**V**S**F**P**D**V**D**V**G**SE**Y**SS**A**PI**T**D**L**DD**G**R**P**I**V**Q**L**PS**R**FS**G**L**R**NS**P**PG**E**ND**N**P**T**SS**S**MM

DEUTEROSTOMAS (INVERTEBRADOS)

SeaUrchin1 gi|390362950|ref|XP_790935.3| XP_011667085.1 [Strongylocentrotus purpuratus]

MEGNRPCRHLGLSILGLILICTLQDSVAIDAPDIVTVENEGNTDSLKVTTWPTTGATDQVVTWSPADGSGSDTSFGLSGTA
TEHTIPLPTGLTYNVTVAQNQGTDETSDDSEQGRKPSVNLTESGATNDSISAETWPKSGAIDQYTFSCSEGETEDPSSP
ITDVSNNMYEVSVCVGLSAAGADYITITVTSLSGNEISNESSTTTITLALPNMVHIEAGDATNSSFVASWNHPEGEMDSFQVFCY
ENDTAPAEIISDVGSNSTMYEVTCESLPTAGTYQFSVVISISGSKNVTSETVLYTSKLGASFGIRLRYFHTSSCNIWLILG
PFCNLSLVPQMHPVTHLQPLHVLHRQHFITLPPNPVELTASDSAVDSVRATWLPAGEVTSYEVQCSNGTVVPPVQTDG
PFTASCEDLPTPGDEYTMNVTSVSNKGSSDVKTIALRALPEAVTLNEERGVMVTTTTIAASWTMPSGIVEYVDVFCVLGDP
PVMGNPSGNLTASCVNLTMPGDDYINISVTAVSNQRSETDITITLALPLSANLTAGNSTINSVSAMWPPGGLVDVFPVQNC
SNGTESNVTVLVNQTNPDPSYLVSCDGVWNPGGDYTMVTVSVEQENSASIMLTALPESVFLNEERDRVNTSTITASWTM
ANGIVDYYQVCSNGTASPAQVNPSEDLASCTRLSTPGDNYTISVTTVSNQNSAVSTITITLALPSQVSDLAESGATSS
VSATWTKPDGVVSSYVTTCPDGSIQTVPEPNDGSGTHNVTCISILPDPGGGDHIMVVTLTSGTKASDPATVITITLPSQVTD
LAESGATSSVSATWTKPDGVVSSYVTTCPDGSIQTVPEPNDGSGTHNVTCISILQDLGPGGDHIMVVTLTSGTKASDPAT
ITLALPSQVTDLAESGATSSVSATWTKPDGVVSSYVTTCPDGSIQTVPEPNDGSGTHNVTCISILPDLGPGGDHIMVVTLTSG
TKASDPATVITITAFPNNSVLRREDGVTTRQITVQWDPVGEEDSFIVDCGDDGSDNITLPGDADTNTFEATCNTSAGAPY
DITVTSVSGDKESPVTGTFTAATPDGVTFEEGSASTAVTVTIALARADGFRVNCSEGTSPSHEGTPSPDSPLPGSATEVT
CNGVTPGGNHSVSVITLENNGESPAAVMYVAVPLSADLAADTSTINSVSAMWPPGGLVDIFEVNCSNGTERNATLAANE
DLNSSVYVSCDGVWNPGDNYSMVVTSLNNGEENSASIIILALPVSVNLTAGGSTSSVATWIDIPGGIVDEFEVECSNGTA
STPGNSGDQDYLASCDDVDSPGDNYTMTVKLSLNTQRNSATIVLTLALPESVSLNEAIDMVTTTTITASWTKPNGIVDYEV
SCSNGDPSVPMGNPSGNLTASCTGLSTPGDDYITISVTSVSNQRSETDITITLALPEAVLSEGESNTTIVISATWTVRNSV
VDTFNITCSDGTASPPSISVSSQAGQQLTAYCVGLPTSGEQYDLSVAISKEKLSAVSTVGVFALPASVESIAVTEPTTT
TVDVQWNLTCQCECVYNYFLTFQPDSQEAIRVDYMGAVNEYSSQVSGLIQGRSYNFTVVSQVGVGEDATLKTSEKSV
QRTGDADKRLHGSIIDIVXXXXXXXGXTRASSLINLVAPEAVGNISVTVQWNTTADVQWNLKQSDCCLYNY
FLLTFEPDSQEPVIRVDYTTGVNEYSYQVGLTEGRNYNFTVVSQVGVQDATALKTSEESVDQRTVPATPDLISVPGQR
ELNLVWAYQGDADNFTITVTPNQGIAMFNGDFNDPAEVTGLTPGTMYDIAIVTASGDERSEPITQTNRLTDPKPSVGNP
QAEAVDKNTITLTYESPLEPENGIDTYRISYIGTRDNTDHEFSDFYSPVPLVTVYNDLYPGFSYFTPIIAVNDGFPESDP
ALTEPVETPQEPSPVPEDYPYEANTVFSSETSTTSFAVVLPPDDLFSHENGELLTFAVITITIDNDPTVSSTELTYAARAEN
AYITAEIPYPPSPFGSNRRRRATDPPGTVVIGDGTGAGSQNEYCNGDLDDNTEYYPFRAYNGMGNVTSSTFGPVKTVK
DNTAGLAAGVSVSLIILIVAVVVVVVFLRRQPKEPSRPSLEGRENQYDGGPDHEMRPTRRKTGSMHRKSHSKPIVLNK
FEGHYGRMKADSDFRFTEEYDEIRLVKGDQAIIVSALEMVNRKRNFTNIIPLYEHSRVKLAALADSDTDYINANYIPGYS
PREFMACQGPLPGTVDDMRMIWEKKTSIIVMLTQLVEKGKIKCHEYWPADYNPVTYGSIQVSVQALQQYDHWVIREFISIS
QGDIEIKLKTQYHFMSPWDPHGVPDKMTMLDFVRTVREAIQKTAGSRIVAHCSAGVGRGTGYIALDRIMQAMQENDYIDIF
GIICEMRMQRNHMVQTEKQYIFIEHCVMDLLRRGEEDTGESIYVNLPSNGFKDPDDVNKNLLDQAV

SeaUrchin3 XP_792888.3 SPU_001461 LOC588094 [Strongylocentrotus purpuratus]

MEGGPNAITQLTSSLTSTNITLWNSALGVVSYEYVAYDPPNGNSPRRTDVRADPRTLTSLGLFPDSQYTVTVTWSGT
GEDATSSPFPVMDFTAPILPLQVIVRFVNETEAQLTWGLVLGEDTYIITVDDSGANP IPSSETVPEGETPLSVRKLTPG
SLYTANVQTTTATGPDQFRFTSPYPLDLVDADSTPTSIDVAWAPPRAAGVDFEYILSYSIGDRIRTEVGRFAPDVTTATI
NNLNTATTYNIYLVTAEANDGGFPQSSDVIITREGRETEIPAPGEIVIINRTPTSFRFQWGPSIVPDVYITTYSPSDGSKLDT
LSIRQSDVTAIPGRLYELTVFVLPDQDETGLVRLTLPSSPGLLAVTAVDYVSVTLTWVPGTGEYTFGFIISYKPTGITYM
DVEEVQDVRTWKEGLSEVDVSYDFQVAVSVEHNTTYSEPTRRINTTTLSIAADAFVLLDYNSTSIIVAVRDNNTNIIVN
IEPQDDEARNVNVGDRPVHFTGLNPSTTYITLSLQSNLVTQVRNTPASPSNLEVSKTSTSFISIEWAAESEVSQYKVD
VKCADMSCSFNDVIYPNERLVYELGLVYLKEYLVTVSSCLMDGQFAEQMGDEPLESSIITDVPDDEIVVQETTTDNIT
IFYATGNRRGFVEAIYLAGTAPAQVGSNSFVTGAVVKTDFGLVPGTLYTIRITELTNTIIPRTVNVTRPIAPSAIFGSP
STVPIPIWFASFSGSISFYEVTVSPDGTTPAVSRIDASGPLSLDIVTLPLKQVYTVFVAVTGVNDMPSISKAVSGMFTL
PAQPGBEILRRITTDITITVLSVSVLQYIVQLNIGTDLQSANPHVLPNVEGEFDELTPGQLYTLVLTVPGTQGLDIAIQ
VRTDPNPPGVSVELTSPVYSSLLAEWEAPPAGLSGVEVSVRPADQVLTMTMEVVDSSSTTELLLDNLLPEVNYTVEVRTI
SAGEGFSSTSDPTEALGRAPFPEAILLVGTGYSSTISVTRWKKEGVDYLAQFIEVQGEIGEDFQGLVQSSSSTTTILSFR
YTDLVPGRNYELIVRSNDQSEDTIRVSQRTLPSPPMNAAPLSAESIRVWTAATNDFEFYEVTFYPSADSTSTAMTIP
RDQFSVIDSLADTEYVFDIRSVSGSDSLQKSAITASATTEPIIMRFLSAVENLLIDFTPIVIGLFDYRYGRALPS
GTTIYEPDLSGTSQFGIQLRSGALYEITLDTRIGSIYTEVDRIQARTRPSAPRLISPVTFDFENSFTLEWLAPSTAEVDF
DGRFRVLYTPADGALESPLLDKTRTSLFLTRLEPGTEYMTLVSVSQQGSTQQVSFEEDVSVTTADGVPGGVNLIKELTETS
IRITWNSALGATYIPTITPDEGTRILDTAAIFEELTPGQLYTLIMTQIGGSGPPTQYTLNPNAPMDVAISRRTAFELD
VSWSPAEGFYNSFALSTYRVGSEVITIPVPASDAPSVTLPDLLPETSVAVSVVGTGFLSISLQFLESGAETPQREALPV
LVERTTTTTISWGASTDALATYQVIRIPEPGETSILFNLGQAITTYNFMNLESNTNYIINLQIPLSSEAPTELPVTTYPQ
TPGAIQLRRTTQTSITFAWGSAGGNFDRYSINLYPDGSQTAFLQDVPGRSDLTVSKNYLAPGTYLRIEYAVTGTTRSALR
ETTASTAMELLVFTITDAFVAVSWPNAPDFSNYLEYDPSNGVPPSPVPTQDSVFLNLSLTPQQIYNFALSSTPTDGTQ

SLVRATYVLAFLAAPTIEIMILSPPTILIKILPASSGVLDYYRVGVTRARQTGLGVGQNGVASFNI PREACPEVFTIDLVPGG
TYDVEVVAVSGETDSAVSSMDVSELMVAGQLVYVNDRTTTSLEVVWGAVDRTPTNVVNLFLNGTRTFDQSQGINERKVK
LFEGLLPGTEYSLNLLVEGDAITQSTILTATAPLTPGSLSFLLTVEGQEVTLAWPARNSQVDSYELCWTSPSQTRINTEADAT
QELLTLEQSDVEYTIISLYALVSVGVGDVKTTRSDAITRFRITLSEALEGELNVTDFFDSNSISVSEVSVSIEFENYRLQAT
PEGDGSPIFAVIPPEDPRMHTFENLIPGVRYTVNNILQGSDDPVVTRTVSQRTRPNPVSGLTVVDTTSSQSVTVNWQATGSGF
NDYVVYTLADEPTTVIYADRIPSDGTRVSVIDGLLPDTSYVVTVYTSSTGDADETTSTESVPAITFALPSCSIQVTETT
YTEITVQWGSACAVVNSYELTIEPNSANAPPGNVPIGGNFYFTFNSLVPGQRYIVGIRGTGSTPEYATDSVYTDPNTPDNL
YEVEVSTSEIALSGLDPSALYDVQIFSKVTSSELTSEKSEPMDSFTTGALQEVGIEFRDKTTTTLIEVLFGTAPLSSEVVO
NQFTTDSLDAQLSVIDFDSFSITVVWGAVEGITFDVYIIQFQPPTGNPIYNRVERTQPRESTVTSLIPGQLQTLISLILID
QVDPISVIQQRAPLPPRLDQTPTVEGSTYSILLWSMPVPSAIDQYELTYSADGLSESEFIVLEANDEFYTFRELTPST
TYEFMRKTVAGEGDTRTESSEVEMTQRTLSAPLQGQLLLRSKTTVTISVTVGWGNTANTIEFGNVVVRITPGDIADVLVNE
ERNVDFTLLNPATLYSIRLIEQNDVSSIDVLTDPVAPNSISFESIESTSATVTVNRFVGHVDNFEVWYRQTSSINDYTL
AGEVGDTESEIALSGLDPSALYDVQIFSKVTSSELTSEKSEPMDSFTTGALQEVGIEFRDKTTTTLIEVLFGTAPLSSEVVO
YRVTLTSEEGASTSFNVAPTSEPFVSVVTFDGLTPGRLYTVLSLQAFGITSTYPTKDVTRFPAPVRGLFPFIEIGTTYVLPFTWG
LPIGDYDSYELSLKGTAREEEII SASETNVRIEFLDPNTMYTFNMVTLAALRSEASDQLFTTSLAVNEIAVRDVSTTE
IEIVWGPPEGVDFWTLTTPAVSSNLIYAQSPTYRFSGLTPGQEVYITVNVNNTSQRTELEMPQNLESVGMQLFLELQWSP
PIGVYDGRITRYTLDGTRVYIDKPLPSQPPRYTLRDLDPGTDYIVSVQALSGGLESEVPVTRTLTTGVFDQLGELSVTVSNTS
IALVWGIYEDRDLNPELTNYELSIPTPDAISLIRGLDVFRAFTRERLIPGQEYSIQKLTTLTDEPVSVVQRTPNDPPAFLR
STRSNFTEIDLAFEAPLSRFDFFTLYEKSDGSDLSNMRTIAGNEDSVTVEELQPGTSYDFSLTAKSGTPEFTFASQAVE
DTFTTRAVPALTVAAAITCDTIQIFWAPDDMLYRIGIQPGPANDFRRTTEYIFSGLNPAATAYSISVTVVAFESNSQS
EYLPLIQQTRPSGAGALVADKTTTTLDVNWLPAGQNFDSYRVVYVTSSENGLPKAYPPISEATEERRLSITELDANVLYTVSI
ETISFFGTQELTSCETSTLDVTRD DALDANEIAVTAFTSTSTITVIFGKREDATGYSITLNRDGNAPRIDITVNLPPSESSRV
TFDTLNPQVNYEITLDSQEDTTIEQRTKPVAPATVLTQNPVTVSFDVIWTPQGVVYHFLLVITNPDGTRKPEI AALGPLDA
VREVTLELLPGNQYDIQVYATVGTGEDLIRSEPAQAQITNSLGALEVIYIDRTTDSLEVWANGASPRPNPQVYRVSNFTEGL
TDTPLITQDRNIRGTSLVAGQLYVNFVAENYQDQSGILNVRTIPFVVRNLTALTEPTAISFSISWNQPASGMYDGYAVY
IAGQDNVERLTSRLDDPTTELLIEDLDPNQEYTVSVYAVAGFGQDSEVRSEVTEATVSTVSLDALDLETYSVTETTIGVA
FGPALVAPSGYTLSTLSDGITIGLIIIPVAGSTKVFVPTGLTSGTLYRAQVQVTGSSVTSDDPLDIRIPRQPSFMDTLVTA
DEISLSWLRPPSGNFPDSYELSYTPENGAKVTLDPVATDETSVLLENLATATEYVILLRQAGEDTFSEPELTLVTTR SALL
ALQALPDPTSVDVSLDWREAGSIQRMYLIYDPPNGDPI SP IETGIPGPVAISGLMPSGQYTFRLFSVLTNNVQFLSSSVSH
VTTPQPPGNVAVTEEMTSTITLDSWAPSQEQFFGYLLEYEGTGFSTELGRTPATSVLGRQESVTLKGLTPEGTSYSIKISA
FVEYDDVTRSNIVTKIGTTVSTSSLNIVAKFVNQTSFTIEWQPAVGASSYVTRVSEVTSQVEISSGTLINGDYERTEHRL
VGRNQIEVSATGTGQSDTVEQRTLPSAVSDVDVPEAAALTLSDWDPVFGDVGHYASQYKLNENPYQISDLFVYVSDPSNP
TIRLTGLSPLSEYIITLTSFSGSGDQEQISSSTVEASTVNSPAEIVIETVTSSTSIASLGSVSTSPASSYVIEASDAGNV
VHSTSRALSQRFFIIISNLTPGLTYTIRITVTVGSSERDTEIALVTPAPRSLSTDSISQSVIAGTSSWQGLGNVDEYDIYLR
QAGESETPAQIVPPSQTETTFENLEPDTYYSMRIISVVRAQNLQORSIPAVTNFTTQPRIINNVETGRNLRSLWANDFGND
ILKYQLIYIEDKEGATNSNEFVDPADAPLVTLMNLMPGQTYTLILTAVDTLGNRAQIASTTYTLKPLRPTFTVNRSPSAV
SLLGNLLGGILKFFRIRVTPQVTGAGAGTPIINDVLIPEACEPIELTGLIPNDRYQVTLQAESGVASSDPQIITFTT PPL
TQQGELSFGEITPSSGFVYWTFFLGRTSYELSNINQAGLVVPSAGDAVQSVYVVDGLSSGTEYTVSLVDGGTGAI RTRPLAP
GTPLATQVLDTSIQLQWAGSSDGNVAGVEICYSPFDGSSSPVYIAGRITTYTLEGLSDTTYIISIALTGVVQVKSQKQT
IFVTTIADTTIPVASNFAVNTQAADHVIILEWDVPERASYDYFEVSYTPDVGYPSSPVRLPKAAATSLSIMNALADKVVYQF
SLVSI RGDERTDPLTTTQGIANNLIVVIGERTTTSLEILWASASGSDVEYQVSVYVGTPIVVDGSSDRRVVIGLSPG
TEYTVTVQAINAGAAGVDSMETASTLHEINGLSAMVTDQASSMIHLTITPPFEDPMRYDGNVNRVRLGGVSSPVVQPIVPI
AKSACPEIWIKKLEAATMYEIDIEWYSGLDISEAFRVQSTVAAPLIDITLDELETDSMKVSWQEGTGSFNLYLVTVSPVG
STQTPMLVAKGEPRVLDLGLVPGTAYTIMVQQGQTPAQETHVQLTLPEAINIDSLSVTRSKNALDATWASPSTGNVDGY
RVCHYPRGTQNSPFTSGTTPSINLPLGNSLTYLVTVSSSEESTYSEPTTI TQQTGLIGDPGDINIALEDVDTNSVRI TWVA
INSITRYTVGGGTGASQTVTVNSPEEAETFTNLVPGTSYTFVFTPLGTQRREREQYTRDMPNRNLLFNTITITNSISIGWE
KPLTGGVDSYRVYTPSSDGPSPASPHITIDNGQITQTLTVTNLLPGTTYTFNVYALVAELVSEPVTRGSGTIGMDIGAGY
DSATILHDTLQSLLLTNLTVAATVYDISIRTVVEGPTQNMKIKTSAALIMAVTTFKPPQNLMQVETSDSITVSWGAPL
NVVNYMHIIVKDEGDADVADATVPSSSETSYTIENLNPYTKYIITVAARFNGLSDSAIEIEQTTQFKPSVVRDPVVA INSP
NSVFLTWTPPDSPNGILLTGFVVEITGEDLTSRRKREAGFTRTETGYTVFSATIENLPTGCSYTFISKAAANSADQSDAVIA
GPLDLPHSAPPTPADRVVRLVSHGTMTMSVSDSSIFSDVYGPVTRYALLVAQSSPDTIPSTPSPAINWQVSGSQPYQT
LLTYAPFTNGNTRGFYTVGANNSTCQDQSVGYCNGPVPVPLTDYF FALRAYGSDGKFFTDSAWSPLYNSINTT WYAAA VVS I
IITLVLVLLVMLLISRQCCGEARSSALYERGNQHLPLVNAAYPSAHNQDSMHKTPLEDAFIRMGPIPVRENAYVSMRDD
RPKAPL PATLPR SAPPLTNPVQMRKFPDHVNRMSAGNNTGFVEEYNSLSNVGRDKSTASRLAVNATKNRYRNILPDETA
YHQLDLEEGQTSYINANYIQGLHSMKEYIASQGVVSDVNDWFEMVFNKSTTIAMI TGLVEGGKTKCEHYWPDDDTPVN
YGSVTVMTHTQEMEQTWVRSFLLEKGMQGFETHQYAFKGVWDHVPNNARPMIEFIRITDVAHDANLGPITVHCSAGIGR
TGVFIALHLKIQVETSKPNDYIDVCGTVARMREQRFNMVQTVKQYMF IHQCLLLMR

SeaUrchin4 SPU_016937.1 XP_783483. LOC578203, partial [Strongylocentrotus purpuratus]

MVLLPYLPLVLIAGARILSVVNTANALFLWDEQNPETFTGYLITSASINEVLPKQSEETSQKSIISNLTPGQLYEVSL
FLQPIITTFPAPLDQLQVRAAPSPPELSRVVDEITITLLRIEWDSPASGLYRGFEVNISSPTALQELPRIVGSDIRGFTTFVGL
MHNTEYTVSVTTIMGEGLERSTTAETKVMTARGDGDEVYLLSVSSTDAQIGWTIPPTTVYLDGKQITVLSQPHTIPSL
TPGRQYTFVVEDDLAFTVPERPDGLELESVGPYSLDWKVDTSLNLDLNYERLEVSYSPESSQDTSFDFISRCTDGTFTLE
GLQPDTEYSVSVTAISGDTRSDVLTRYGFTTCALEGQLHVSVLSSDEIALAWGPVNNIQGYAVRFLNAIGINMRTLSSAN
SMSATLSDLSPASNYSVQLIVMDTDGSRQTSVLQFVTDPPAPPGPITLQSSQTRLSFTWTPATSNFDSYELSSISGGDMPEP
RPVGVVYPNEGNEFEVGGLLPDTTYNVRIYSVFFESRSIASEASFNTDEQILEFNVGNSQNLTVSWQVQEGAESRYLIYTT
TNRNTNSIARPATSTSITIDNLLFGAIYNFLRATIAADGTEMEVANADFVQSPLTPSTNISKIKATTALLTFSLREGIFDD
YVVMLSSPTDCSGARQPVFTLSREACPELLLDLNYNTMYTAQVFARTGERGLSLPSETTLQNFHTTATVNVNEVEISKVGT
NSLEIIWKTDNVDYELLHYSYGPGLMSDIVPPIEINYYTFGELTPGTQYSISVISLINRERQDALEVMQYTKPDPISNL
VLTPQENTNLWYSPSTVMYALAGEERSDATEITAMTRQQNSEELRLQTRTSTISVSWMGNTTRPEVTFVLRITFQEGDA
ITLTETSPILKDIATDTSVSWVTIPGVVDYLSYSPPDGNEQATQDIDESSAIAEGLNNIYVNNFTGLTPGREYTFVTV
WPLAFNYSDAIRTNPLPIYDPSVTNTTSCNLLYQWCLPDSDFDNISITIVPDSMRTLRSRNGTGMCDYQTVYLDGLEP
ATSYNVSVVEMSGDKSSEVVMVYSTTEPLPVMNVTLDDSTIFTLTISWENPIDGSFNMFLVEYAPVRSVSPSVMALDPINMG
TNTAIIPNLQPGTLYNIMVYALAGEERSDATEITAMTRQQNSEELRLQTRTSTISVSWMGNTTRPEVTFVLRITFQEGDA
IDSIKNVSLGDLVGEFSDILPGTNYTVYLLVEVEGQLDVAVHSLDVTLLPGMPGNISCDPTKTKTQSLLEIEWGSTQEGNF
DGYAISTQTGDEEIPLESVEKDEQLQATADYLDMDNTIYTFIVRSTLGDLRGEPPTMTCTRDPQNLVWSTSTSPSGGQEVF
LDWQDYGASSYKVEYTPQEAAPLPPVQVIVLRRPSTFTIQITLPRRGRYRFLRLRVRRRSQGLDAGPDGDEFLIPTDACPE
YTFNNLLPDTDYAVDENTVAEDGRLEESSPSTMDVTTPRMEEGDISFNEVTTTISFAWGMVSAFVNGREDFYLLNLKNELS
IDVQNVQVNTSETREHTFGLMPGELYRVELIVNSNRQRSAAPMYTRPDKPDTIMLSDLTTLTSLTIDWKAPLTSYQGFRCLW
AYGETNSLELGPAMERTLSDELAAGAEYLVSLAAVVGTTGANQITSEKFTTIVTLFPFAELHVNITSFNTTAVSIVWGTDIP
LLDVSSYVVSWEATANNQSVQDISVDPPSQTDTYIDGLVPGTRYSMVDALGTVAQVTSIGATQYTNPEVVAEIRVIPSIT
CSVRLWDRPEGNLDSYIFHNVTRGTIVGAPLQLDVAINDTAETLLTYLLEGLEKDDSSDFVSLVGVVTTVAGAGEDQVES
LQQVVKLDDVNVVTITILGRDQNSIQAEVNSSVSNVSEITGLYNAEMVTSSTTRNCSTMDCFSTVDFDLNSGTLYTLY
AAVQSSGREVVPLAKAATIPESAVDLQFTSIGRDYVVLWTDNPNAGMIDSYNISSYPVNDITKLMFEVQAAAESNVLRV
DDLNEGMNYSFTVVSLLEVADLQEMGAPVEVFAVVGVLGSLNITAFDETITISIEWEQVDEEYILSYDALEGGSGNVTVQ
STDPNSYTFGLTPGTQYITGVQPGIYAALKQSTVPNPPTNLRTTMVAAQSIIVSWDPPSEGGSDTYNITYNVGTGGQMP
SPHTVSTSETSVTFSLPTLPEYFFTIIVNWKSLPSRSELTRAITSAQTLSLAVTPEDNTEVFSTWEDDLTEAFQYCVMY
EPYEAQENRQESFPVARNIRSAFYGLDPGEYTVVYVTSVSSVSDPTSHANERISFRLDPAIAEIRIMDGLSIPRAVML
EWDQPEGIRGLYEVTLDPADDEGSDVDGDELVPNITGVMTSIKLLNLAAGKGYTVGITAISGDKRSNTTFLFYTLDPDPPQ
NLVISEIGEDTATLSWADPAEMNFIDILEVVMTTMPATKTTILKLNNTVELTNLVPQGTYTVQLASVIRPDDALNQPEVIG
EPQVSSPFLMKPASPSDISVSNLNRVTVTWQFSEGVARYSYTYGLPGSTVTASETTTTDTSIQLSNLTPSTRYEFRL
WAESDQTEGSETSSDVVTNIVITVFFNVPSPVSAALPVTTLPSMPSAVGGLTITKTGIVDAEVSWSLPLPNGVIANYS
LTIHISHDNGTSLSPYDGTVTISAVAEQTEYSQDVIGLAGQYVVSVAATNQVASSNQRSTRTLFAETAPPSPRGGATPS
LVSRGSTITITFPNLPDFAFRIVRFAIVQERVDGCTVVIITAKRQTSSLTWAEARRTRFVPSYQTTTPDDYNFADGAGV
TFRVGSQSCDDPDLLTAYCNGALYPGTYSYRFAIRAYGVDGKFDTEFSSPFRTPDRGFWIIPANLAINVPIIIPAM
GFGCGKPPDDGMYSDDSNNGYDYSNGNGYNSNGNGIMKRALLINDGVNTYAYQQQIMEHSFTNIIPDYSTIPNRRQTQ
VHQPSSTHTRPTKKNMFAQHVAKMSARNAGFSDEYNLNPAMEKTRAVTAARQARNDDKNRYRNLIPDYTRVQLKGVAGT
DYINASYIKDDHGKQKIATQGPLNPIEHEFWEMVVENQSTIVMMTALVEGGKTKCEHYWPAGEEPQLHGNVTVTLVGSN
QTDNFIERLLEKEGERTERTVTHYQLAWTDHGVPPESTAPLVGLLRQVKVNTQEDGAATGPVHVCSAGVGRGTGTFIAADM
LMDAIQRSTADYIDVAGTIAKIREQRALLVQTLTYQYIYLHRTVLSLIEEQ.

AcornWorm gi|291235644|ref|XP_002737754.1| [Saccoglossus kowalevskii]

MGTQDDTLTGRVLIIEIQVFI FIGFSNAGYKCTDDCAYSHDNLCDDGGVDSYSNYCYLGSDDCGPREVPDPTTVAPPTTP
TPTPTPSSTRTAMHAMQTFVHTTTAASEAEHAHTLNTSTDAKSTDLQTEGHETS GAESTHTPEIVTDVHSSRTEY
EVMSSMTTNEENNVNMYNIRTEEQETEVEVDSSTLVGTEIVITGFTPEEQETEVEVDSSTLVGTEIVTDFPTPEEQETE
VVDSSILVTEIVITVDFPTPEEQETEVEVDSSTLVGTEIVTDFPTPEEQETEVEVDSSTLVGTEIVTDFPTPEEQETE
SSKLVGTEIVTDFPTPEEQETEVEVDSSTLVSTEIATNGFKPTEEQETEVEVDSSTLVGTEIVTDFPTPEEQETEVEVDSST
VVGTEIVISFTPEEQETEVEVVKLTLTDFPTPEEQETEVEVDSSTLVGTEIVTDFPTPEEQETEVEVDSSTLVGTEIVTDF
FPTPEEQETEVEVDSSTLVGTEIVTDFPTPEEREETEVEVDSSTLVGTEIVTDFPTPEEQETEVEVDSSTLVGTEIVTDFPT
TEEQETEVEVDSSTLVGTEIVTDFPTPEEQETEVEVDSSTLVGTEIVTDFPTPEEQETEVEVDSSTLVGTEIVTDFPTPEE
QETEVEVDSSTLVGTEIVTDFPTPEEQETEVEVDSSTLVGTEIVTDFPTPEEQETEVEVDSSTLVGTEIVTDFPTPEEQETE
EFVGSSTLAGTEVTTGFTPEEQETEVEVSSVVVGTGVTGGAQTGEPETEVEVDISTVHTEMVTTGGTPEEQETEVEVDSSTLVGTEIVTDFPTPEE
GSSTVMGTVMTGTQTEKQETEVEVSSKVDGTEVTVSGLMPTVEDGTGVIDSSRFASVTKVAKRISTLRHSTTVMRPTTEED
TESTTVGLSTLPTLAVVYVSTTEIQQMADDSYLRWSDSVALTFDLKLSTNBSLQSVSLNNNDWLIYIRLDSSTNETVNER
LRPYEVEVDNFQSSGDLVFNDIIVMNNLSGLLCEMITYICIELIADGNGTDLVFIREPVDVCSILCRGVIEIAPFDVDT
GSKLREGMDDQSPFNLLIYDPDEAGSIEGEYLWLTLSFFSQSNDGYGNTDAGTVHSLTTDQSGMSLYSGIPASWTGFRSS

AINITDKLCSVDVSKYFCVNIKSSLSNPDFTFGNPVMCTPPTCTGVDDVDTRLISIVSGIIVVKEHSSQYFNVFVTLSSP
DFGDI EGASLWNVMTYVNTQGNANVGDVSLAVADIPRDMRDLSSLAGSNAILEGVEAFMDLAEINCRQIRYLQVDSVRDQ
TSVPAFTLTKGLQTCAEITCRGVEITDTELVKSGTIVERSGNHVSFTFDVLLSSNPAASISGDRLWEMYVYLSDFVDD
VHFKGADNAEFPDFTLGLVLDENALPGCAHLSCQRFVAFNISAINVTSSFTIRWDEAVGDFDGYIIDVSPALTLGLSSSFI
GVIISTDLQLSDDFYLLLEGVDNQHISYTMFLFASSAGAGVSGMNLWDIRTQLRSIEDAIVDTSNGTIQEVNTTLNAGGVLL
MSNVKSVLSLEGLVLCQFFNICAILNKNSNANTYEYTDREYDRSTLVSCRVSCKGVEIIEVAPTIMNGDIVAEADDAHAVSI
NVGMTSSYGTNRGTAGSVALEVFVTSDDVRMPKLVNLYATLAAQQDTELGIGPVSLSNIEVELDVSAQCSEFTQLCV
VHFKGADNAEFPDFTLGLVLDENALPGCAHLSCQRFVAFNISAINVTSSFTIRWDEAVGDFDGYIIDVSPALTLGLSSSFI
TER DEPREASFFDLESQGEYVIVSAGIKNVFPKSTGVTQYQRTFPEKVDLSLRVSMVTETEHVWVDRPDGVVSTYVISIQPSE
GTATLPTLPSDSTTASFGLSPMSQVVIITVYAGSLQSKVEVPVKVITGTIPVLEVTSTMTDITVATVVDNRLSHGML
HLSYFPPTDAEPIPIELDLNKTSYVITDLVPGRLYTLVKKIVADIDGGRITNSNDRKYQQTPEPSPPGSIIVSRITFEFGAIVN
WEPSSGDNFVLVYSSASVNTALPIRVNHGQTPQLSGLSPATVYTVHVSTRGELTSQSVSTVNTLAPRGLSVNG
VPTSSSQISWLPASISGLAVEYYTYAFPVEGATPASPFTLNGTSVNIITNLYPFTIYTVVQTRIDDMYSRERGITATTYTPD
LPSSVKGFNVTVESPYQVTLTIQPPSFANGILESYIVNVVGYKEGFPFHLDFFEIPDDETDITTEYLLEDVIAGYTYTFTV
KPQNSLIGIGPLRTAGVQVIMPIYFPKPTASRITANVRFVETTAYTLRIHLSNNLFDDRFGLIIVAYLVI I AEDGEYSPLPEVY
LPSYDEVIDSSPPWQYTSQFPNPFETLSIGRRRNTLQTVNYIIGAKENC DVS L V Y C N G A L R P V T D Y R Y I R G Y N E L G N Y
T D T D W S L P Q K T E L D P F W I L Y **V V A G R L E N G I E I S C N Y C V N C** C R R R R S S T Q D S D K L S E P P S Y T P A C E N G V F P N G N E P T L E
KTEKLYTLSLRRKRKSRPVPLKGRFRDYNLEDEDDRKGFEEYDDLKLVVRQPTSVGQRMHRSKNRYTNIVPYDNNRVI
LSGSGNSYINASYISGSKGSRSYIATQGPLESTCGDFWKMIWQQRVTTVMMTQCNELGKSKCHHYWRDQTQI SHSDLT
IRLTSETRLPDWIRDFSIENGEIRALRQFHFTSWPITGTPYDADPLIRFIEAIRIQVLPNSGPILVHCSAGVGRGTGVFI
ALYHLLYFYTMIQDIFGKVIKMRKQRPFMVQTQGGYEFLYYAIQGHKIKPLGTNWSMRKTMSNWTIITYGGTVDPDSDS
GPGFFDPYDPEEV

Acorworm1. XP_006823994 curated profiling | partial [Saccoglossus kowalevskii]

MSLKHAFCFPANVNVIDYGTDIYITVNWTHSGEHVDFEVITAIATENEAANGTETVNIITRGTVDGLTPGAEYITVVAVSQ
DDVESLTPSDEVIQRTIEPPPYNVTVESYNTDSIHLRWEPPLSNVFTHYNVTVYVADSDPGMTVNVNKTSTESSLTGLISGE
TYNITIKTISGDESSGASDAVQVTEPNSVVLARDEQIITDITIAVWSSYSGDIVDTYTVQCSSNGTAEAEATVDYVNGT
DTYSAKCTGYETAGKLENITVAVSGTSGPSQSLSEPSMIEIHTKPNSSVVLARNDQSITNDTIAVWNSNAGDEVSYTYV
ECSGGTTPQAMIIYDNTVTVYSAKCTGYGTAGKLENITVAVSGTSGPSQSESEPSKLAIHTKPKSVTLARDEQIITNDT
IAVWSSYSGDIVNTYTVQCSSGGTAEAEATVDYVNDVTVYSAKCTGYGTAGKLENITVAVSGDINTTGESESEPSTMDIH
TKPNSTLARDDSQITDITIAVWSSNSGDIVDTYTVQCSSGGTAEAEATVDYVNDVTVYSAKCTGYGTAGKLENITVAVSGD
SGDLNTTGSKSEPTMEINTIPNQADPVRNDNLVTNTTIGVKWTDPPGGNINHYEIGCPDGGTPSANVHTHPDKEMTAQCT
ALATPGLGLYSITVTSVSGNSKSDPAVTSMTKPNKVRLERDDNITNDTIAVWNSNSGDEVTYTYTECSSGGTAEADSRV
DYVNGTDTYSAKCTGYETAGKLETITVAVSGDLSTTGESEPSMMEIHTESNAITLTKDSEYVTTTSIAVWTVNPPGAV
AYYTVCESEGGTAENGSIPIYDNDQHEYSKGCINYNPNGLNETINVTINADNHRGETSSMEINTEPLPVTDLTKVDATT
AFIEFVTEPESTGYSYDGYLSEDTQGFSEYLSDDVHTIITSHTFNNLTAGVLYDITVITVSGEEMSTNETEPMRTIP
NAVTDLVLTTPNQTNPEIVIMWSPVEDEVSHYRVKCNDSDDVTTEPDDQIIPHENVTSLTTECSDLFPGRLYNVSVSSIS
GYEGVDETESEPIVKAKRTHPKPVTSLSLTPDNSVDTIHVSWTAPEDSDVDGVRVCTETSDGKIDDDVEVTVYNYDFS
GLTAGRLYTDVYYSYSGSDNDEKLYSVNTTDDVKRTYKPKVTSLSLTPDNSVDTIHVSWTAPEDSDVDGVRVCTETSDG
DKIDDDVEVTVYNYDFSGLTAGRLYTDVYYSYSGSDNDEKLYSVNTTDDVKRTYKPNSTLARDDSQITDITIAVWSSNSG
GYRIEYVQDNLANAVLRREDISDPEQKIVHHTIYDLYAGTLYNVTIFTVSSGIDQLESLEPDTTETDRVPATAYNVTEVE
SSKTSYSELEFSWKINTTSLHDYEVYRGLDTSFRNTTSDSENVSLVDLRAGDTYNVTIITISGKKRSAPSEPAYGTT
PLSPGAVSVIDVTEIEIKIQAQACENCTDYVYVHHEPEDGEVDVDSSSLHAKLYDLNHLGTLNYSVVSISNDVPSEPTY
TLARTLPDTPENITIEITGMRSNLNVTWDEPEGEYAGYNISIEPSTNRAGKDIIVDVPDKGIEECIEICECDPYNTYKITIIV
HSDEPNQTFSEPAVSHAMTETA**PPPPRNINVPSSSESDIEVSGTITVILFTDDBYFNDSNGPLLNFTVIVTEDEGTEEVDK**
HLNGEIPRVYCKDCENVPGLIEEGEVKTWAEVQNTPARYQIDYPYDPLSRKKRDTTGLIEITIGSDDSCTKDEGFCNGP
LKDGTSYRYHFRAITDVGYSDFVLSGPIRTSTNMSWLW **HLNGEIPRVYCKDCENVPGLIEEGEVKTWAEVQNTPARYQIDYPYDPLSRKKRDTTGLIEITIGSDDSCTKDEGFCNGP**
YKRRYEEETTRTVEGTANPE
FIGLRDENGKPPIAQPKTKQHSKPKVLAEYHYGFQIMSADMEFRFSEYDELRPVGREQSWDAELPENRAKNRYTNIL
PYDRTRVKLSQVEDDEATDYLNANWMPGFNSPREFIAVQGTLPKTKDDFRMIWEYNVSTIVMVTQCQERGRVKCERYWP
TDDNPPVYGDVLTVTTHENELTDWVIREFTVENGKSLRRIRHFNTAWPDHGVPEETGSLIKFVRSVRAQIQNDGTPTVV
HCSAGVGRGTGFIALDRLIQHIKEBYVDIFGIACEMRMHRVMIQTESQYIFIHLCVDDLLKEREQPEDEPEDSTPDT
TEHIYGNVPSVTVMRNGPDVVTANLGDAYKRNSVIAQA

Oikopleura gi|313238213|emb|CBY13306.1| [Oikopleura dioica]

MNIIANETFLQAPVPRRLVAADPASSKVRIDLSPALGETASTRVDSYNFWYFKENESSILRENTTIPAKIPNTFTIVKP
GSLYHYVAQAVNNGVASESTTVDFRVPPLQVSDLVDAVNTSSITLWSNPLVSDSDFEFVDDIIVPGIADKSSVPAGTANI

TLSGFAGSTLANINVFTRTAGPPTAQSAPAIISVYTLPPRIDPESFNATQVHNKNAIALEWEPLSVFFTKYKIIVDYLLAF
KPTRKEIILSPTTTEYLLKSDIASIISGNVYRVSVIVALSKTNAGGDDEQVSSYPTTKMVLILPEAPTFLSVNTKKSSSKS
LTFQWSPPSSGLDYDFELKWKQPDSDCPAKFEQVRNLSIQQAQKSMDPITNPFVQYKIEELHPYCYDYDFSVKAVLNSTSPYIAG
FDEASLKTTEFVVVTGKTNAEPPLLVQRTNTPVLNMSNVVIGKGTISFDVASLFFSDAFGLLRFVAVWRKDTGDSLKKP
SIPKWTWENNENREWIAFVVKLKGDSSERTLVIGDGKESQNKDRYKNAPLESGEKYRVSIQVENEAGASALYYSQPIEMND
AKKAIAIVPIVVLIVAVIVLYYVIRRRQLSAPKDNQELGLANGTAGPGGGHVNPFVTDQRPRRIELNSVERKSKVIQ
LVDFKQVYDELQRDSSYKFAGEFFELKEVGKDTTEASLLTANRGNRYTNILAYDKTRVKLGATDDDDPGSDYINANFITG
NNSKREFIATQGPLPGTKEDFWRMCWEQNTPIIVALCQVVERGRKICDHYWPDSDPEWVGSDQLVMVTSFCETSRRRIKQ
VSVSNQRKMTNIVLFQYHFTVWPDNGVPSLEPLVKFVVRQVRFEMNRFDRGPTIIHCSAGVGRGTGFVGYDRLLQDYENK
SYIDIFGMVYEMRMSCLMVQTEAQYICLHELVDLQIGKYNDRDNGPSDNGELKTEAQTIGAPQAGITNPALIDTSSS
SSGSLTSDEEVILPKTPPIQLEPDENSNQANNQLNDAASL

Ciona Ciona Savigny ENSCSAVT0000019606 plus GENSACAN and profiling profiling

MIKQVILLFFRFLIHFRTLNKSIAESSTFSLEGQLQLDSLARTTTLLNLEPHRHVLRITISNHTTPRGVWTSSEST
LSFSTRKLPDIKRYLFPFITSPALAPPRSLHVNHTSITEESFEIAWKSAVGLVEKYRIVLTKKSADETKILVANNFATS
FVDGLSPGESYEIAMISLKGIDQSSAVFVSTTRAISDMAPPVNFMRMTRGETEITLGGWDPAPGVQVVMVRCBGFQEP
AEVDIIRPDEKRELVRGLKPGARYNFTIQTQNGQDISKAKRSLITVPRTPLSLDTVSRSDTTISLRWDQPTGILDGKI
SYTSRGGQLEILFVQPGIFEKTLNLRPGTYEISLVSTSSIESEPIILTEITELREIENLSIARLGETWAFISWVPG
GYSGFDYIVQVATVLDKDRSVARNITIQTRPLPPNDVRMDNLTSSVHVWVREPETFDQFLVYKDDFNTKMLTQQT
INDLTPGSSYVIDICVIFGVSXVNVNEITTYIIAEPGTPALVLPVSTTELLTEWEIPLNPNGLIRRYIIRFKQNYPH
PSTNYTLIETNQTRYLLGNLEPGAEGYVALFAAVNSAGVGNFTESYFFKMPESKPGVPSDLHVSNNVAISATVAVQP
GEIIRVVLNISTDHLVLDLKLNRSDKNCLDFDHKSKSDKALVSVTSLQNNVSEALNDEIYFDPLCPVFNVTSL
LPFTLYRFRVWTSSGEGDSRIREEMTLEDVDDPRRHVRLTSATHTSLNVTFPPNQPNIVITYTIIAVLQSDKMFQSNL
TYDINELRPFYKGMVRAETSIVGSPWSEILFADTQQGRPGSVQNLTAQVTSRSTLQWNPPIPNPVIITGYWIFAKV
KNVTQLTLWLEGVLSNLSGNLRLTQDVPSSPINLSYVNSSTAVNITWDEPLEANGVIKYVYVYMTKDFLVDVTERF
ALIKDLKIFSPYEVIRVPRPYTLGDPGASGTLIHTDEDFPASPVRNVTSKNITSSVVELSWTPPLIINGRARSYTVY
LLNSYNAFNSYSNKTSFAFLTNLRVFTAYKVQVACQTKKGLGNLSKTIIFIQTEEGQPATPPFNVSQNLSTKVRLLWRRP
LVPNGIIQFYEISLTSKNNKIRATTENDVAVTVVDHLTAYSEYATVRANTKFGDGSQQSHALTIHTLEDVPGSPVNT
YVNLTSSSIGLEWSLPKEPNKILKYSIRYSMLGEMNQYKAVNDTFVLFDLNKYTRYNVSVSAHTVAGEGPPAYISLHT
DEDEPESAPYDIFQQNSTIALTWRPPVKPNGIIVNYTVYSSNEDKVMKTTTKPRTVLRNLEKFTYEVYLTASTKLG
NGKISAIETFTQLEDAPADPRDVVVKALSSTSISVWSTPATPNQIQFYTVFYTDKISAVHATNVTVISMNRSTSS
KILLQITNGTNAVIKDLNKYSYDLVWVSGSTALGDGNQLSDAKKVRTMEDVPDNGVSDLSALAINATTIKVTWQPGIPLTG
PTFFHIQYVLMNTPVHDLNFIISQPHTEYVIVKYPRVVGVDAGVSSIVVTTMESAPSSPVQNLHVYVVTETIVKLKWK
PSEANGVITSYTIITVKQNYNKTGTNETETVIEDLRANSHYVSVTSSTKIGEGPSTILLFRTEKGEVQDAVTDIQLINL
TSDSALIKWLAPRQPNVITHYTVHYGRNSTIQTNVTEATLRDLHPPQLYWIYVYVWTSAGIVKPKSEKMAIKTESAP
PTSLNCSVASNFKSVKISWDLPMRTNGIIRGNLEYRSMDKNKVKFVQPSQTKIISNLKPFANYLIKLRALNLAGKQWTH
CNVTTLEGYADAPPQLQLKNSDRSITIEWSRPLQWNGRFHGYLITYKPPDSCPNPANDTQVCSFSTNLSVTLEALS
KYR AYNISVGSITGRVGTMRDIFVSTLVGVDPDSPVRDVHVTVSSSTINVTWNLPSYAGPTTYKLYSELVPLDYFDHTAQ
YLRVHNHENSPPFVGLDEDYTVGIVITMTTVGELSSIPVSDTKEDAPSDPFPEVSVGATVNLNNSSVHVLKFEPLDPNGK
LINYTIQHRRLDLSKLSIVVSDQLTAPTTEHEPYSVSVSGLLGGRYQYFRVRAATAVAGTNSWLTNKLVLPTAPPVE
DTWPEVMNKIDDSTMASVTSSTITVKKPCMFSNENGPASLSVIAEDGGNIEAEPTYWSQAPLPQSPPYKVLVSENPT
DYCNTRKRRSINEKGFVIGTSDCFSELNTHCNGPLKSSTDYRVKFRATESNGLKTDTEFSEIIRTSPSFLAHHMTLKGIG
RANGDFTLVVINAVALRVMQSIQYGFLLNNVFPIMETSIHNTMAFIEGASMDKQVEESVHGNEFSRPFVFRDEFPNHVS
SLSQNKNGKGFSELEFDDIRGIPYAGTTAIAENSCNKTKNRITKLVFPDCHRVKIEGPAIQGSNVINASQIPLDSEPEQYIA
TQTPLDHTKKDFWRMLWETGSTNIVMLCGSVNAGKRCDEFWPKQTEYFGNLAVQKKEEIRHDEWIIRQFIVTMRDKVRH
VTQHFFIKWPQLEQAENSLPLVRFIKYRLLRDRSINPTIVMCSNGSGRCGVFGLRLDNGRDRVDFGTVAALRYR
YMWQTLSEYIYLHQCVLKFIDD

Ciona.1 Ciona Intestinalis gencode and curated by profiling gi|459174018
[ref|XP_002122888.2|]

ENSCINT0000033578MPSLYVISSYEISLTLATSGTAVTKTYTVIAPALTYTVTGLTPGESYATVQAVSSSVSVS
TPSLPQRDVFENQVSSVVSTVSNLVDPAEELAVSGEADSLLMTLDTVGEKVELDENGVEYEDVTSNVGMAALLPSNKK
RSLGLVSELSINSEPGFRSDPQYALIQEIPNNVQSFITLPEEAGMLWVSNMQIINACLFSVFDGSDGNAKRIIFYAFPNK
LFLSPNNTYDVMNMMVLSATVEQTSVNNLQEPVYIQLVTNTHAVDAPSGITVDNIKSTEFVTSWTTPTDAISKMTVDIAG
TASNENDAVDVVSPHKVINTAGAAIVGNTYTYITVYAVSSTDATDFKATNQTTIFSAVPLTSVGTGNSTIAVDLSWTY
DNGGTIANVSEYLIKWDGGGSGTSPAGSGSTTATISSLSANTEYTFPSITAVSATVVRGTSAPSSATAGATTTSIDLWT

APAVGGKNNVLAYTIQWTGGAGGSKETSQTDDTISLSANTAYSFTVAAKSKAGTGEASTPLOAITLPSLPEQPTLTRST
TNPPTVIDVSWPAVTSGETVVDVVEWTPDEGPAANKAVTGTTTIDSLVPGQQYVTVRAHNSAGNSEPSPAARLRNPP
TPTGVSLSNPTNGQTSCLKVDWQITKNNFVYSSYEISLTPATSGAAVTKTYPANANSFPYIVTGLTPGESYATVQAFSSG
VSGSISASNSQTTVPPTPTGVTLSQSAGDQTTSLKVDWVMPSLYVSSYNI SLIPESSGSAIYVTFPTPSASTTDDTDT
GLTPGESYATVQAFSSGSGNNSDASASLRTVSAVPGTPNLYQPTDGSDKTTILYANWTVPTGVVDSYQLLVYLGSGVGT
LVANQTVSTNFANITSLIPGKRYQATVRAFSAVAGVSGNSNFAKTNPPPTPTGVTLSQPNVNTTSLKVDWVMIPTSFVI
SSYIISLISTSGAVVNKYTASAGSAPTYVTGLTPGESYATVQAVSSAYDPSTQSDVASSNTQRTDPPPTPTGVQLFQ
PTVNQTTSLKVDWQITETDFYSSYDITLTPSTSGAAVTKNYPANANSFPYIVTGLTPGESYATVQAVSSGSSSVSTP
SNPQRTDPPPTPTNVTLSKPTVNQTTSLKVDWHITEAYFVSSYEISLKITGTTGAAVTKTYSANDASFPYIVTGLTPGES
YTATVQAVSSGSSSVSTPSNPQRTDPGVPGIPLTSLQPTDGSDKTTILYANWTVPTGVVDSYKLLVYRQDVGVTVDANLIV
TTFNFANITSLIPGKRYQATVTVAYSVGASVNSGNSNYGKTNPPPTTNVQLSQPTVDKTTSLVNVNQITPDSVINYNIAT
TAMSTGTITNFYVTLTEKSYTMPGLIPGESYAATVQAFSAKYPPATQSFVGSLLINQRTAKQNPQLIQLYQQHNTLYVQD
GDASTQVTLENATTSYNIITGLTSGASYITSVVTSINGVRSDEAKQFSTITKGVSNLMAQSQTTNSIYVWTPPMDVL
FTNYFVEYSKMNIIFLMPEDHKENIYLRICITLLLSISGYLSYGFYSLWVNVFTQPKNGANIYIATLTGVTINEQRNGNS
SSPVQVGGSLSPGHYSAVVSASASATILVEGVASSAIHTIKPEPVTNFRSTEQTSYSIYMSWDLPTLGFQSQNVTVVPS
DSENAIVVSLGNETTSYNIITGLTSGVKYITISVITISNNIESDPATDESMTGPPVATNISITASNTTCVHGTTIYVGRGII
HPYSLTWGPHGSKLVYPYHPHITHFNLCGLQAGVEYDINNVSLAGNVSSYNIISTAYTEPNSVLSVMLSSVITEISVQWT
RNSGGVSGYEVIAATNSNCVVEANITIAAGNTTSLTSLGLTPGTVFNQVKALFSALKSPFVYQSAPTCELVNYTDEGGDK
SIQTNFYETSATLTVLPGYNYTVMVSVTAYGLNSVIQTMKNTKFNVMIKSRINNVNGVIENSDFTELYKTNAGPEEPN
ITQPVNTGSTSNMNTITIILPANTFNDKQGP I EAFGVYITKETQKKPGQPDLRSDNSTNLMTCEVAVWYTPAGVANTPS
TSSRKRKSYDNPIGNLITPVISSGSESTISPKWTLFVNLPLAVDTPYVVAVAAKTSNNVPLVSTNWSQPIRTEPVPAPNV
CLIVGLTVACVVIVVLIIVVWAFYRKRKESKDDVSSENQIPLQQRKTKVIQLAEFLDLLKVMKADSDFKFSEEEYEFKTV
GRDQATVAALLPENRGNKRYTNILPYDATRVKLSAIDDEQGTDYINANFIPGNMNRQREYIATQGPLGKEDFRMVMWEQ
NSRNIYVMVTQTVBERGKIKCDHYNPFNDNEITVADYTLQMTSESILPEWTRIFEFKITHGSDTRRIRQPHYTVWPHGVPDPA
ETLVKFIYRVVRRITDREAKHSGPTVVHCSAGVGRGTGTFIAMDRLLQHLDPNNYVDIFGIVHQR IHRVFMVQTESQYILIH
QMVQDILNRVYDDEDDDDQEPVYENTTISDPIYENAEFNGKQKNGGVNVALISPSEDDQNKEDSEEEGEESEEEGEEES
EEEGNEENDQNESKLISGSGGNHDPVSIKQDAVEETSKERGSKATVV

Ciona2 XP_002123247.2 phosphatidylinositol phosphatase PTPRQ, partial [Ciona
intestinalis]

LFPHPVADYTNVKNKIFLVLNQLPEGAGYIIRIAAINSAGLGTIFIKDELKIVMEETPGPVNQLQVSSYNATTVTIISWQP
PLEPNGEIIYYKLVYKQNLVATIINFEEKQENENQFNCLLDNAPRDLVSSFMQLLLPNSLLQHNPNSTTLTKESGDVTIQA
LSEIITSQTDNNSTKKGKQRTLSRIVTEITNEYFEPLCPVATIVTSLKPFTNYRFNVVGTAVGEGIGVDVSHTMPQEIPE
GPLNVRVDLATDSTIRVMWSSPAKPNQIVTYTINIELGNWVYQSNNTTYVDVNLKPFSTYAIKVKAIKTVGEGPWSNWI
IANTLEGSPESVNSFNVTSKTSRNITLWVEPPLTPNGVLLGYLIVEENNHIVRSYWLAEENNGATREIPSQSDDVIQFDD
DILGGKLVDSYMKIKTKTISDLTPHTAYAFNIFAATVGGNGSQTLTYDVTLEDKDPSPPNLVSYSNITSTSVNIWMEPL
KANGVITKYTVYITKDKLQETSVTERYAVISGLNIYTTYEVYVRATTKVGDGPPSSMKQIHTDEDDVPGSPVLVDVNFQNIS
SSVVKVTSWSPPLVTVNGVILGYTIFYSNFTTYNTTSKTTSVYLEKLMFANYTVRIASQTSKGLSPMLSGAITIQTMGQP
STPPFNLSYNTTSSVIQLTWLKLPLVNGIIQFYEITLQTNNTIRKTTQSSSETSKTIKGLKPFSLYLANVRANKYGD
GGQASATYLIYTMEDAPGSGVMNVSYVNLSTSVNVSWLQPAEPNGRITKYKIESKLYTEYIISISAYTMAGEGPEVV
KICTDEDEPPSPPDGIFFHQHNSTISILTNWNPPLTPNGIITLYSVHYRHGNKSLIRTTATPGITLNNLKKFTSYDVYIRA
STKFGDGNQTSKIKSFKTLHDAADPPPHDVRVAASSTSNVWSPPTPNGLIQFYTVYQHNPNSTVQTKNVTKGMQV
IGNLRKFTNYLVWVTSSTALGDGNQMSDAIKVQTLLEDVPEQKQVQELSPRVNNSAIRVTVLPGQPLTGVTYFMVNVSTKT
SPYTEYISVTPHVGVDTPGPTSLVTVTTDEAKPSSPVRNISIIITITQTVTITFLPPTKSNGVITYSIMTSLDHKQL
VNGSHTKIKLENLKPYYVNVTVMPYTKMGKPKAAILTFQTLGVPEDSPQKVQITNITDQSASLHWEPTPLNPIIRTY
TISYQNESQRNFNSKQVQLTNLAPFTNYSLQLYAWTTRGMGKLP SRVLYFTTHESEPSAPTNLCKIYDNYKILVTV
DLPLLSNGRVIAYKLSHSLVNDTQTYFVKPPQINQTLQHLQPFTEYFISVNAKTSAGWGPSATCNITGEGYATAPQDLN
INKTSDVTITIQWNKPNVNGRLNGYVIMYIPFDACQPNSGILNKRILSTNKTWTLIGMKQHTYKISVFAATGRGL
GAPSNISVRTLVQVPEHVENLQAIIVTSSTTVNVWESPEFAGPTTYKVEAFHSTMQRTVSLPSSSTSSYVMSGLDE
DSIYGTVTSTMTSVGLQSSKPVVIMTYEDVPSDFPRDVNVAIVMGNSSVRVLFKPPRPDNGVLTNYTIQFKRLENSESR
NAIIPVSDLVPEPSSNNVYVVTINGLGGRLYKFRVRAATAVGAGPWSLWTTDVLQIPITAPPVPALLPEVLIQDQKAIIVT
SSSVIVKPKCVFSDDDNGLKLSVIVAQEGATLDAEPTYWAKAYSEEPSPPYKIVITEEPQSYCNSRNRKSTQTNMGFVI
GTSDCRRSLTSQCNGLKSNTEYRFKRAEANNGLMTDTEYSEVFRTNPSFIEAHMPTLLVSIQALGIFIIILLTISVALL
RRRRRNQIPNGNESLKTMGPIETMGPIIEEKSKVPVMESSVHNSVAFVMDSDRMKDTARSPSINEFSRPVAVTDFSLLV
EKLSONKNGKGFSSFPDDIRGIPYAGTTSVAEKACNKTKNRTAKLVPPFDHCRVKLEGI PAVGSGSNYINASYVPGPHSPBQY
VATQTPLDHTKDFWRLIWETGNSITVMLCNVMEGGKRCSEYWPKKQTEYFGLNAIQMKTQDVCCKDWIVRHSVTMRDK
VRHVTQFMFNWSVVEQGNPLVRFIKSYRTTRHNEYKPTIVMCSNGSGRSGVFIGLDKIVDTMTSYIDVFGTVASLRK
YRFHMVQTLTEYIVMHQCNVNFIDDDVTIEQPTGDGLEETIF

CionaNoCat.1 Ciona |XP_002120279.2| ENSCINT00000036281 (ESTs:BW474929.1|
 FF941901.1| FF962764.1| FF924695.1| BW485618.1| BW182207.1) PREDICTED:
 phosphatidylinositol phosphatase PTPRQ-like NO catalitic domain[Ciona
 intestinalis]

MVSSLVSSSRFDQIDVWTWGAPTMPNGVITQYQITWSPGNGMMNFSSTTTTHSITPLTSLNLDYTVTVYAYTAIGRVNS
 AVQRTLESVPIDPVTSVVIATAATS NVNVSFNQPATVTGMLSIEVKYVYVTSAGPSTAVEQNFTSTADTITGLMPNTAYTVI
 VTPQTTAGKGGAAATVVTQTFMTDEAAAAAVASVTEGKTADSITVNWAPPVNVNGVITQYRVTWNSGAGNQVVTSGTSHT
 ITGLSAFTDYTVTVYASTAVGEGPGVNVQVIRTNEAVPAQIGAFVAVGRNTNIRVTITPPVPANGVLLRYEATLNQDQAE
 TRFGDTSCTQNRVQSVTANGNTPGVIVFNLTSFRSYEITVRASTSAGFGPVSIVLTSRTTISIAPQGVPTNVQITSLSDT
 SAIVTWDSAKPVPNGSTYIVLTLAVSNSSDVRITITIPSTVSLNNSVVFNTLQAGHGYSYSGVQSTVSGMSDPSTSATLY
 RAFTLIAGAAPPATAGPTTALGVSPVAFQVENPCLRKAQLFSEIGGTINQIQILVWQIQAPTPTTANVWVANVINQNP
 VAYQAGTITCTGTASGRKKRALFANNDGYVVGADSTCTTTTAVCNGPLPSGRQFNVSYIGVNGGGQSTGMTAPGGPFST
 STPTGLEAGEADAVYRCKGVNRRRQYVYVVKRRRASGMHTVSGRDNRSFDDTQKQRYQVEASETPTSTTYSTQKQK
 I

CionaNoCat2 Ciona savigny protein profiling

SNLALVNSFDTVQVWAAPIMPNGVITAYRVEWTPGGGGVDLGGSTATHTHTISSLTAYTNYTVSVVASTAIGQGVVSDT
 IATVESVPTDQVTSINLIAQATSVGVSVFVPPATVTVGVYTYQVRYVYVTSAGSTTAMTQQFTTTSDSITGLLPHTAYTIEVR
 DTFMKFTDLPTQVSPQTTAGIGPATTMTITTAEEAAPVAVASVNSGGSTFNSLTVNWPAPMPPNGVITGYRVTVWPATGGG
 SAVVSTGTSHTISGLTAYTEYTMVMPASTAAGEGPGMSTTLRTAEAVPAKITTASLSTATSIAVVTITPPNPANGVLVAY
 MVSTRYLVLTIRANTSVRGYAVSDPFFTRLATTPQGSATNVNVLNATSALVNWMSLFPVPGNSSYITLTLALTNPNI
 NTVSWIFSIKQTQTVVSMYVNNVAVFTNLLTGHGYSVVTIQSTSSSGSTASSPFLLYRPTIISGAPGPIILSPTQALFG
 ASPTHAFQIENPCNRKSSSLFSEIGGVIATAEIVVWQTAASSKWEAATPAIWADAINQNPPIPPYVAGTIQCTASGRKKR
 ALTPNGSGYVIGAECTAANRVTNCGPLISGRQYNVAYRGNLGTSPDMTASNGPFSTSTQRGLEAGEIPLVYSSLY
 VYRTRKQYVYVVKRRRSSGAHTVSGRDNRSFDDSQKQRYKVEAATTDEAATTYSQKQYAKRL

DEUTEROSTOMAS (VERTEBRADOS)

PTPRQ.human gi|222537743|ref|NP_001138498.1| phosphatidylinositol phosphatase
 PTPRQ precursor [Homo sapiens]

MDFLIIFLLFFIGTSETQVDVSNVVPGRTRYDITISSISTYTSVPVTRIVTNTVTKPGPPVFLAGERVSGAGILLSWNTPPN
 PNGRIISYIVKYKEVCPMWMQTVYTVQVRSKPDLSLEVLTLNLPNGTTYEIKVAAENSAGIGVSDPFLFQTAESAPGKVVNLT
 VEAYNASAVKLIWYLPQNGKITSPKISVKHARSGIVVKDVSIRVEDILTGKLEPCNENSESFLWSTASPPTLGRVTPP
 SRTHSSSTLTQNEISSVWKEPISFVVTHLRPYTTYLFVESAATTEAGYIDSTIVRTPESVPEGPPQNCVGNITGKSFSI
 LWDPTTIVTGKFSYRVELYGPSGRILDNSTKDLKFAFTNLPTFTMYDVYIAETSAGTGPKSNISVFTPPDPVGAVFDLQL
 AEVSTQVRIITWKKPRQNGIINQYRVKVLVPEGTIILENTLLTGNNIYNDMPAPEIIVNIEPVMVGLYEGSAMESDLSH
 LATFIYNHPDKNFPARNRAEDQTSVVTTRNQYITDIAAEQLSYVIRRLVFPFTEHMI SVSAFTIMGEGPPTVLSVRTRQQ
 VPSSIKIINYKNISSSSILLYWDPPEYPNGKITHYTIYAMELDTNRAFQITITDINSPLITGLKKYTKYKMRVAASTHVGES
 SLSEENDIFVRTSEDEPESSQDVEVIDVTADEIRLKWSPPEKPNGII IAYEVLYKNIDTLYMKNTSTTDIILRNLRPHTL
 YNISVRSYTRFGHGNQVSSLSVTRTSETVSDAPENITYKNISSGIELSFLPSSPNGIIQYTIYLRKNSNGNEERTINT
 TSLTQNIKVLKKYTYIIEVSASTLKGEGVRSAPISILTEEDAPDSDPPQDFSVKQLSGVTVKLSWQPPELNGIILYYTVY
 VNRSSLKTINVTETSLELSDLDYNEVYSAYVTASTRFGDGKTRSNII SFQTPGAPSDPPKDVYYANLSSSSIILFWTPP
 SKPNGIIQYYSVYRNTSGTFMQNFTLHEVTNDFNMVSTIIDKLTIFSYYTFLWTASTSVGNGNKSSDIIEVYTDQDIP
 EGFVGNLTYESISSSTAINVSWVPPAQPNGLVFFYVSLILQQTPRHVRPPLVTYERSIYFDNLEKYTDYILKITPSTEKGF
 DTYTAQLYIKTEEDVPEPSPPIINTFKNLSSTSVLLSWDPPVKPNGAII SYDLTLQGPENIYSFITSNDNYIILEELSPFTLY
 SFFAAARTRKGLGPSSILFFYTDSEVPLAPPQNLTLINCTSDFWLKWSPSPLPGGIVKVSFKIHEHETDIYKKNLSGF
 KTEAKLVGLEPVSTYSIRVSAFTKVGNGNQFSNVVFTTQESVDPDVQNMQCMATSWQSVLVKWDPPKANGIITQYVMVT
 ERNSTKVPQDHYMTIKLLANTSIFYVKVRASTSAGEBDESTCHVSTLPEVPSVPTNIAFSDVQSTASATLTIWRPDTILG
 YFQNYKITTLQRAQKCKEWESEBCVEYKIQYLYEAHLTEETVYGLKFRWYRFQVAASTNAGYGNASNWI STKTLPGPPD
 GPPENVHVAVTSPFISISWSEPAVITGPTCYLIDVKSVDNDEFNISFIKSNEENKTIIEIKLEIFTRYSVVITAFGTNIS
 AAYVEGKSSAEMIVTTLESAPKDPNNMNTFQKIPDEVTKFLQTLFPPSQPNNGIQVYQALVYREDDPTAVQIHNLSIIQKT
 NTFVIAMLEGLKGGHTYINSVYAVNSAGAPKVPMRIITMDIKAFARPKTKPTPIYDATGKLLVSTTITIRMPICYSDDH
 GPKVNVQVLTETGAQNDGNVTKWYDAYFNKARPYFTNEGFNPPCTEGKTKFSGNEEYIYIGADNACMIPGNEDKICNGP
 LKPKKQYLPKFRATNMGQFTSDSDYSDPVKTLCEGLSERTVEIIEVSTLQLESILDLGTAIFARIRQKQKEGGTYSYQD
 AEIIDTKLKLQDLITVADLELKDRLRLLSYRKSIPKSKSFLQHVEELCTNNSLKFQEEFSELKFLQDLSSTADLDP

WNRakNRPfNkIPYnnnrVklIAdsvPgsdyInAsyISgylcPnefiATqGplPgTVGDFWRmVwetrakTLvmlTQcfe
KGRIRChQYwPdnkPvTVfGdIvITklMEDVQIDWTIRDLKIERHGDCMTVRQCnftAwPEHGVPENsAPLIHFVKLVRA
sRAhdTTPMIVhCSAGvGRtGVfIAlDhLTQHihDhDFVDIYGLVAELRSERmCMVQnLAQYIFLHQcILDLLSNKGSnQP
ICFVnYSALQkMdsLDAMEGDVELEWEETM

PTPRQ.mouse gi|124487427|ref|NP_001074901.1| ENSMUST0000050702 phosphatidy
l inositol phosphatase PTPRQ precursor [Mus musculus]

MDFLFFFLSLIGTSESQVDVSGSFDdTVYDITLSSISATTYSSPVSRtLATNVSKPGPPVFLAGERVGSAGILLSWNTPP
NPNGRIISYVVKYKEVCPWMQTAYTRVRakPdsLeVLLtNLnPGtTYEIKVAENsAGIGVfSDPFLFQTAEsAPGKVVNL
TVEALNYSAVNLIWYLPQPNGKITSFKISVkhARsGIVVKDVSikVEDLLSGKLPEcNENsDSfLWStTSpsPtlSRATP
PLRtTHLSnTLARnkISsVwKEPISfVvTHLRpYtTYLFEVSAvtTEAGYIDStIVRtPESVPEgppQNCITGNVTGKAfS
ISWDPpAIvTGKfSYRvELyGpTGRILdNSTKDLRFVfVfThLTPfTMdVYVAABTsAGvGPKsNLsVfTPpDVPgAVFDLQ
IvEVEATEIRvSWRkPQRPNVEDISQYRVKvSVLESgVILEntLLtGQDeYIINNpMTPeIMnLDPMIGfYEGSGEMSSDLH
SLASfIYNshPHdPArTnVgDQRSPVvATRnQYMTDIAAEHLsYIRRLVpPTEHTISVSAFTVMGEGPpTVLTvTRtREQ
VPSSiQIINyKNISSSSiLlyWDPPEYPNGKIthYtIYAMEldTNRAFQMTtVDNSFLITGLKkYtRYKMRVAAsThVGES
SLSEENDLfvrtPeDEPESSpQDVkVtDvSPSELsLTwSPPEKPNgiIIAYEVfYQnADALFvKNTStTnITLSDLkPYtL
YNIISIqSYrLrGHnQSSiLlSVrTsetVpDSAPENITyKNISSEIEIEIFFLPPRSPNGiIQkYtIYLKRSNSHEARTIET
tSLTLTIgGLKkYthYIeVSAstLkGEGSMPISILTEEDApsDPpQnFsvKQLSGVtVmlSWpPLePngIILyYtVY
VwDKVSLKtINATEVSLSDLDYHADYSAYvTASTRFgDgKTRSSVINFRtPEGEpsDPPKDVHYVnLSSSSiILfWtPp
VKPNgiIQYsvYyQNTSSTFVQnFTLLeVtQEPGNvTVsARiYkLAVfSYtFWLTAStLVGNGKSSDVIHvYtDQDIp
EGGvGNLTyESLsStAINSWtPpSQPNGLVfYVvSLNLQqSPpRHRRPpLTYENSiYFDNLEkYtDYIffkITpStEKGF
SEtYtQAHLkTEEDVpDTPPIINTFKNLSSiLLSWDpPLKPNgALsYHLTLQgThANrTfTtSGNHVLESGHESpTL
YsFLAAARtMkGLGpSSiLfyTDESAPLAPPQnLTLINyTSDfVWLTWSPSPpLGGiVKVYSfKIEHEHtDtvfYkNIsG
fQTDakLAGLEPvStYsISVsAftKvGNGNqfSNvKfTtQESVpDAVQNIACvARdWQSVsVMWDPpRKANGiIIHYMIT
VEGNStKvSPRdPMYtFtKLLANtSYIFeVRAStSAGEGNEsQCNvStLPETvPSvPNTAFSNvQStSVTLRwIKpDtl
GfQNYkIITQLRAQCREWEPEECVEHQEVQYLYEANQTEdTVRGLKkFQwYRFQVAAsTnAGYGNASSWiStQTLpGPp
DGPPENVRvVATSPFGINISWNEPAIITGPTfYLIDvKSVDNDNFNISfVKSNEENtTEINDLEVPtRyTVtAFVGNV
SGAYTDGKSSAEvIITLESvPKDPPNMtFQkIPDEvTKfQLSFLPpSQPNgiIQvYqALvYREDDpTAVQIHnLSiIQk
TDTsvIAMLEGLKGGHTYnISvYAINsAGAPkVQMRITMDIKAPARPKTKPIIHdATGKLlVtStTITIRMPICyYNDd
HGPIrNVQVLVAEAGAQDGNvTKWYDAYFNkARPYfTNEGPNPBCIEGkTKfSGNEEIVVIGADnACMIpGNEEKICNG
PLKPKKQYLFKFRATNVMGQFTDSEYSDPICKTLQGLSERTVEIILSVTLCLISIIILGTAFARIRQKVEGGTYSPR
DAEIIDTKFKLDQLITVADELKDLERLRLLSYRKSikPVSKKSLfQHVEELCTNNLkFQEEPSelPKfLQDLSStDADL
PWNRAkNRPfNkIPYnnnrVklIAdsvIPgsdyInAsyVSGylcPnefiATqGplPgTVGDFWRmVwetrakTLvmlTQcfe
EKGRIRChQYwPdnkPvTVfGdIvITklMEDiQIDWTIRDLKIERHGDCMTVRQCnftGWPEHGVPENtTPLIHfVKLVr
TSRAhdATPMVhCSAGvGRtGVfIAlDhLTQHihDhDFVDIYGLVAELRSERmCMVQnLAQYIFLHQcILDLLSNKGGHq
PVCfVnYSfLQkMdsLDAMEGDVELEWEETM

PTPRQ.chicken XP_001235338.2| PREDICTED: phosphatidylinositol phosphatase PTPRQ
ENSGALT0000017794[Gallus gallus]

MPDTGVfDGIYvTtNGGNATfSLKSDgKlTvenLTPGTEYDFCVfTKSREMLSSSYrVtGVkTCLAAPLNIREGNVtDTS
VQIawDRAEGDFQYEVtCTNCAsAFrvQkVQETATfSNLVPgkLYSfTVrTEKEGFRDSVLvAKEIETVPSAVKYLNYS
RDSEsITvTWPpAQnKfDgYVLSIKsIFNKENMLSSGVRMYKAECllPGTDfLISIVtTSGLKRshPTfLkISTCPDPPS
DLQVLGQEENTVYLSWKLPRGGFDKfQLSYCLMNNEKpFTRTVYDSRTvVKNLAPGMEYtFQLRTIKGWDSsVAveKNvIT
KPAGICNLALKMvNTSSATLMWNPtKNtNSYkASLSntTfISKFMIPgAVSKfSVnTLtaggiYnFtlQrLQGNIEGSPA
fLEIVAEpAKPEGLKfFNvSSNSfLYwRLPYGHVDRfCVdLIPDHGSvVIsDLGvREYQADfYNTpGTtYnVtVStVSS
StYSSpASrTvtTnVtNPGPPAFLAGERVGSAGILLSWNTPQHfPnGRILSYIVKYKEVCPWMQTAYtTPGTtYnVtVStVSS
NLNPGtTYEIKVAENsAGIGVfSAPFLFQTAEsAPGKVVNLtVEALNYSAVNLIWfLPQPNGKITSFKISVkhARsGIV
VKDVLVKVEDLLSGRLPEcNDNSeSLWStTTPStTfGKStIPsStTvtASTEASSQMSAVWNEPISfVITNLrPYtTYL
ELVSAvtTEAGYIDStIVRtPESVPEdPPQnFAkVNITSKfSfVMWDPpTIVtGKfSYRvELyGpSGHILdNSTKDLKfVfH
nsvFNtMYDvYGAETsAGvGPKtNLtVfTpsDVPgAVSDLHLVEVEATYIKIvWRKfPQPNgiITQYrVkvHQEKEVTL
ENIILIGKNHLIDSLEPYMINENIEPSPTWNSIETANELYEGSGEMfSSIQTvSPVILtTVSDdHLPAINGAAELHStLD
DQYATIDILGEELSYIKGLVPfDYtISVSAftAIGEGPPSVLTVrTREQVPSsVQsISyKNISSSSVLLYwDPPANpNGK
IIHYtYAMEldTKRAFHTtSSNLSLMTGLKkYtNYKMRVAAsTVIGESALSEENDIFvrtPeDEpDspPQnVEIINvTA
TEILNVLWPEEQNGLIthYEVLYSDSdLfvRntStnISLTEMpPYtLYNIslTEMpPYtLYNIslTEMpPYtLYNIslTEMpPYtLYNI
SAPENITyWNISStEIElSfFPpSPNGiIQYtIYLKringTEERvINTHlVLRITDLKkYtEYMIEVSAstMLGEGlR
SAPLHLLtDEDApSSPPEsLsvKQLSGvTVKLSWkPpLePngIILyYtVYvWnkTSKRsvNvtETSLEfTDLennSEySAY
VAAsTRfGDgNIKSDTIIFRTSgAPGDpPKDIvYKNLTSStIMLFWSPpKPNginRyYsvYFRnDSGIFIQNTfSHDND
SDVnMSPsAVLDDLAKYshYtLWLTASTAVGDGnKtSEIIDVYtDQIDpGPVENLVYQNISSSSVnVSWLPPSQPNGLVf

FHVSLSLQLGNTKILSFLTYNTSIIIFDNLEKYTDYILKITPATDKGSSSELHALSLHIRTDEDPV**PESAPIIKTFSNLSTTS**
VMSLWDPVPVKPSGIIISYDLNLPFGSEKNIISFSTTNNFIILEDLVPSTLYSIYTAARTMKGGSPSVLHFYTDSEVPLAPPQ
NLITITNYTADSVLWKWEPSPQNGVITRYNLKIYQNDTEKIFYQ**NISGSNNEAKLDGLKPFSTYFISVSAPTKLGNQFS**
NAVQFT**TTMESV**DPVVFQNVHCIAITSWESIFMQWPEPPASSNGVITYYIVTVEGSSTNFSSDTLHFTFQNLNLSISYQFKIKAA
TSAGDGEEQICNASTLPEEV**PSAPRDIVFSNVQSTSVTLNWRSPKSIPIGYFQNYKITTLQLSIYCSNWETKECIEDEIHQY**
LYEKGVNAQIEETVYGLKRYWRVFAVAASTNAGYGGSSPWISTQTLPGS**PDGPPENVTVLATS**PHSINISWSEPVVITGP
TCYLIDITSVDNENYKAQFLRTNDEGKILEISDLKAFTRYSVVIA**IFTGDVSAAAIEGKASSPVI**VSTFEAV**PEDPPNNVT**
FQKIPDEVTKFQVTFVPPSEPNNGIQQVYQAMVYNEDDPAAIRIHNLSVIDKTDQSVTAMIEGLKGGHTYNSVYAINGAGA
GPKIQLKITMDIK**EP**PRPKK**PAPVYD**TNGALLV**TATITIRMPV**CY**YSDDHGPIK**KIQV**LVVEAGA****QHDGNVTKWHDAYF**
NRPRPYFTNEGFPNPPC**IEGKEDLSGKEE**LYVIGAD**TCMISGSQDKICNGPLK**PK**QYL**FKFRAT**NVKGQFTDSDYSDPV**
KTLGEGRS**SGSVEV****ILAVTLCILSVVLLVA**AVY**AFAR**IRQKQKEGGTYSRDAEIIDTKFKLDQLITVADLELDERF**TRY**
SSFFRRRKEIFV**IQ**LLSYRKS**IKPISK**SFLQ**HVEELCTNNNLKQFE**EFSELPKFLEDLASTDADLWNRSKNRF**PNIKP**
NNNRVKLMPDAGIPGSDYINASVSYGLKCPNEFIATQGPLPGTVGDFWRM**VWETRAKTL**DEMT**CFE**KGR**IRCHQY**W**PEDN**
KPVTVPGDIVITKLVEDIQIDWTIRDLKIERHGDCMMVRQCNFTSWPEHGVPET**TAP**IIHFVKLIRASRAHDNT**PMVVHCS**
AGVGR**TGVYALDHLTQH**INDHD**FVDIYGLVAELR**SERMCMVQ**NLAQYIFLHQCVL**DL**LLTSR**SSQPICFVNYSALQK**MDS**
LDAME**GDVELEWEETM**

PTPRQ.fish |XP_009291679.PREDICTED: phosphatidylinositol phosphatase PTPRQ
ptprq-001 A8DZA4 ENSDART00000143740 [Danio rerio]

MTTAA**AQD**FRYGL**TDHHP**LQH**QDL**H**ESSVSI**HD**LTPGSDY**S**FDIQS**FLGSD**SQTT**SKNI**STR**PAGVCSLFVSDV**NVT**
SATVSWAA**AGGDF**YRCEIVR**NS**SHKW**TIDAS**PHLQ**EVTV**SGLWSGCSY**NTTV**QRF**RNTIS**GAA**ATINI**HS**VPAS**PEGLRV
ISV**SRFS**LHWLAS**PGCEK**TSY**QV**LYPDHG**NI**NT**TDAD**NNVQAQVSV**TPGTSY**TVTV**NA**VAAS**SGF**SPSV**RLVTTI**
ES**SAP**SK**PEGER**V**SGT**IL**SWRMP**PLDPS**IHS**FVIRYK**EMCP**Y**DDPS**FE**IT**KS**LDI**PET**LLNT**LTPGATY**NIK**VAAVN
KAGV**GF**PSQ**SLY**KT**AEAP**PGLV**NS**L**TAF**AE**DHTSV**IV**TW**FL**PVR**INGLIT**KFAV**K**VKHART**GQ**IV**RT**KEL**NA**EDIM**NGAL
PH**CND**AADIL**SRGT**SP**SQ**TS**AVL**PPITLS**AV**PPAS**IWS**VPIS**VK**IDE**L**RPY**TPY**VE**SA**FT**SD**GE**Q**AST**MVRM**
PE**AA**PE**DD**PPQ**NV**LR**NI**TSK**SV**SL**TWE**PK**II**TGR**FS**Y**VI**QL**H**SE**GL**ISE**NS**TID**QM**FI**Y**TGL**TP**Y**TY**I**H**VM**AK**SAGA
AG**PA**AV**IN**IT**L**AE**APS**AV**SL**L**KAE**AV**D**STS**VR**LS**WR**SP**I**Q**PN**GL**I**THYR**LL**V**Y**HD**TL**VQ**DI**TR**LG**Q**N**L**S**PL**RR**NA**RS**L
D**FT**TES**V**FT**L**SH**ET**FT**SL**GR**D**FT**NT**D**E**AP**S**AL**PL**TP**ST**PH**TP**PP**WL**TV**TH**DR**TD**AS**V**Q**T**DA**VT**ED**L**TE**HP**VP**TV**L
P**ATE**V**S**AD**L**SS**DQ**I**TH**V**RR**LS**P**FE**TY**E**K**FS**V**SA**K**NI**I**GE**GP**ST**EV**TV**K**T**NE**Q**V**PS**S**Q**V**Q**NS**Y**Q**NS**TS**IR**V**SE**PL**NP**N**
G**K**IT**H**Y**AV**YA**Q**NL**T**Q**EL**RQ**MD**TT**TA**VL**TGL**D**K**Y**SS**Y**K**VR**VA**AS**TAV**GES**PL**T**NE**D**YI**V**V**TL**ED**V**PD**SP**PR**GL**T**IV**KT**
TS**ST**AT**L**SW**SP**E**K**PN**GI**RM**YE**IS**Y**NG**T**YS**N**TV**N**STS**AS**AT**L**RY**L**K**PH**TH**Y**N**V**TV**RA**FT**LL**GH**GE**Q**I**SE**TL**Q**ML**SG**ED**V
PG**S**PP**Y**D**LS**Y**ES**IG**SS**SV**N**WS**P**LL**ANG**V**IL**F**YN**VE**Y**W**N**AT**Q**SL**N**Q**T**TH**MP**Y**IV**LS**N**L**R**K**Y**A**H**Y**R**IS**V**Q**AA**T**Q**V**G**M**GN**H
T**SE**IL**N**IT**TL**ED**V**PS**SP**PK**FL**I**ARK**LSD**TE**VEL**S**WE**A**PE**AN**SE**IL**Y**IV**RV**W**N**L**ST**EF**VA**N**VT**ETS**V**V**V**VD**GG**Q**Y**N**AS
V**SS**W**TR**L**GD**GG**LL**I**Y**IT**FT**S**IES**AP**SD**PP**Q**D**V**Y**TL**LT**I**ST**VR**LS**WR**PP**NE**PN**GI**I**Q**Y**TY**I**Y**TD**NN**E**Y**TER**VP**GE**H**Q**L**
LL**SD**L**Q**AG**Q**D**YS**VM**SS**TS**VD**GG**L**F**SP**SL**N**FT**LED**V**PS**GP**V**HN**LS**AT**IS**Y**ST**AV**VI**SW**DP**LS**PN**GR**V**Y**Y**Q**LS**L**Q**E**AG**
I**TH**PS**IN**RT**V**N**T**IT**KT**TT**DI**Y**L**FT**K**L**R**K**Y**FP**Y**I**NV**TP**AT**S**AG**SV**N**HT**ML**HL**RT**DE**V**PS**SP**PL**SG**SN**K**LS**TS**I**Q**
V**SW**FP**P**VE**ANGE**I**IE**Y**AV**N**L**Q**GP**ST**SN**K**Y**T**SE**TH**L**V**L**SE**L**TP**FP**Y**N**LS**IA**AV**TR**V**GT**GP**P**V**LV**SL**HT**DE**AG**PM**S**PP**R**N**L**
T**IF**N**H**T**AN**SV**W**L**Q**W**EP**SP**EN**GV**Q**LY**G**FR**IL**EL**NT**DS**FR**Y**Q**NS**SD**AST**Q**AE**L**GG**FK**PH**NS**Y**E**IS**VC**TF**TR**AG**ND**Q**Y**SL**P**
V**LF**TT**NE**S**V**SD**AV**GN**LS**CS**GL**D**W**DS**V**Y**ME**W**EL**PD**HP**NG**E**IL**Y**LR**IS**GD**LE**DE**A**HP**V**VL**TH**VA**HT**LT**GL**S**PH**AF**YL**IT**VA**
AV**NS**AG**S**EE**AN**CS**HT**PP**ES**V**PG**PS**H**SL**TV**DT**VD**TS**D**SV**SW**Q**R**P**V**H**V**GL**L**Q**Y**N**VE**I**V**HC**S**AG**R**NC**ER**Q**Q**Q**Q**V**ES**C**ME**AG**V**
I**LE**W**EG**ET**TR**AT**L**TS**LL**K**Y**Q**Y**R**IR**V**AF**TR**AG**AG**EP**SE**WI**Y**T**Q**T**LAG**NP**D**AP**PGA**IS**V**VP**S**AN**GL**K**I**E**W**D**K**PS**VI**SG**PT
SY**I**ID**IT**AL**D**GS**G**Y**NI**TL**VR**H**SE**IR**TV**IG**N**LS**AF**TL**HS**VT**IT**AF**T**GPL**S**N**ARR**D**G**K**ASE**P**V**L**IR**TE**DE**EP**K**DP**PK**N**V**L
TV**IP**EE**V**TR**V**Y**TF**SP**DE**PN**GI**SA**Y**R**VD**I**Y**R**NG**QL**DF**I**NS**LS**VI**SN**PN**MT**AI**ID**GL**K**GG**F**N**YS**IR**IA**AV**NG**AP**GP**G**
PS**SE**V**H**TT**GV**K**APP**K**PK**T**PR**A**AL**NS**AG**VI**IST**SK**T**IT**I**EM**PE**CF**FT**DD**H**GP**I**Q**K**V**Q**VI**SE**PA**V**MD**Y**GN**L**SN**K**SV**FL**H
PT**AP**Y**L**TD**DD**G**FL**NP**EC**PK**NS**ER**SS**ST**K**TY**VI**GE**DE**GC**LS**ED**AE**TL**CNGP**L**K**PK**TH**Y**V**FK**FR**AT**N**IR**G**Q**FT**D**SE**Y**SD**K**V**RT
AD**DR**LL**TR**DE**Q**II**LG**V**LL**S**F**L**AL**L**LI**I**Y**GS**VK**I**H**RR**K**EG**GT**YS**PR**Q**AE**II**ET**K**FK**LD**QL**I**AV**AD**LE**L**K**E**K**IN**R**LL**S**
Y**R**KS**L**K**P**IS**K**S**FL**Q**H**VE**DL**C**AND**NA**K**F**Q**EE**FA**EL**PK**L**Q**DL**AT**SD**AD**L**P**NR**S**K**N**R**F**T**NI**K**P**Y**NN**SR**V**K**L**SE**PG**MP**GS**D
Y**IN**AS**F**VS**G**Y**L**CP**NE**F**I**AT**Q**GL**P**ST**VA**DF**WR**MI**W**ET**GT**K**I**V**ML**T**QC**FE**K**GR**IR**CH**Q**Y**W**ED**KN**P**V**TV**F**AD**II**I**IK**LT**ED**
VR**PD**WT**VR**AL**K**VER**H**GS**Y**MI**V**H**FN**Y**T**SW**PE**H**GV**PE**SS**ST**L**V**Q**F**V**K**AV**R**SN**R**GH**ENT**Y**IV**VC**S**AG**VR**T**GV**F**I**AL**D**HL**I**Q**
HL**R**D**H**EF**VD**I**Y**GL**VA**EL**R**SERM**CM**V**Q**N**L**A**Q**Y**M**FL**H**Q**ST**LD**LL**SA**K**NS**Q**SI**W**F**V**N**Y**SA**L**E**K**MD**S**L**D**AME**GD**V**E**L**E**W**E**ET**M**

PTPRB.human PTPRB-001 ENST000003344142200 [Homo sapiens]

MEAEFY**M**VIL**T**CLIF**R**NS**E**GF**Q**IV**H**V**Q**K**Q**CL**F**K**NE**K**V**V**G**SC**N**RT**I**Q**N**Q**Q**W**M**W**T**E**D**E**K**L**L**H**V**K**S**A**L**C**L**A**I**S**N**SS**R**GP**S**R**S**
A**IL**DR**C**S**Q**AP**R**W**T**C**Y**D**Q**E**G**L**E**V**E**NAS**L**F**L**Q**Q**GS**R**V**V**V**K**K**AR**K**Y**L**H**S**W**M**K**ID**V**N**K**E**G**L**V**NE**S**L**C**L**Q**A**GL**GA**E**V**S**R**ST**
R**NT**AP**P**Q**I**LT**T**FA**VP**D**GL**V**FL**IR**NT**TE**AF**IR**NA**AE**NS**Q**NS**SE**R**Q**HP**N**L**H**MT**G**IT**DT**S**W**VL**ST**T**Q**P**FS**ST**TE**E**T**GL**AE**PE**
R**C**N**FT**LA**ES**K**AS**H**S**V**SI**Q**WR**IL**G**SP**C**N**F**SL**I**Y**SS**DT**L**GA**AL**C**P**T**FR**ID**NT**TY**G**C**N**L**Q**DL**Q**AG**T**I**Y**FR**I**IS**L**DE**ERT**V**V**L
Q**TD**PL**P**PAR**F**GV**S**KE**K**TT**ST**SL**H**V**W**TP**SS**G**K**VT**SY**EV**Q**L**F**DE**NN**Q**K**I**Q**V**Q**I**Q**EST**S**W**NE**Y**TF**F**N**L**T**AG**S**K**Y**NI**A**IT**AV**S
G**G**K**R**S**F**SV**Y**T**NG**ST**V**PS**P**V**K**D**I**G**I**ST**K**AN**SL**LS**W**SH**G**SG**N**V**ERY**RL**ML**MD**K**G**IL**V**H**GG**V**DK**H**ATS**Y**AF**H**GL**T**PG**Y**LY**N**L

TVMTEAAGLQNYRWKLVRTAPMEVSNLKVNTDGSLSLSLKVWQRPFGNVDSYNITLSHGKTIKESRVLAPWITETHFKELV
PGRLYQVTVSCVSGELSAQKMAVGRTPDKVANLEANNGRMRSLLVSVSWSPAGDWEQYRILLFNDSVLLNITVKGEEKQ
YVMDTGLVPGRQYEVIVESEGNLKNSERCQGRTPVLAVLQLRVKHANETSLSIMWQTPVAEWEKYIISLADRDLIIHK
SLSKDAKEFTFTDLVPGRMATVTSISGLDKNSSSVKGRTPAQVTDLHVANQGMSSLFNTWTAQGDVEFYQVLLIHE
NVVINKNESISSETSRYSFHLSKSGSLYSVVVTVSGGISSRQVVEGRTPSSVSGVTVNNSGRNDYLSVSWLLAPGDVDN
YEVTLSHDGKVVQSLVIAKSVRECSFSSLPGRLYTVTITRSGKYENHSFSQERTVDPKVQGVSVNSARSVDYLRVSWVH
ATGDFDHYEVTIKNNNFIQTKSIKSENECVFVLVPGRLYSVTVTTKSGQYEAENEQNGRTIPEPVKDLTLNRSTEDL
HVTWSGANGDQYIEIQLLFNDMKVFPFHLVNTATEYRFTSLTPGRQYKILLVTISGDVQQSFAIEGFTVPSAVKNIHIS
PNGATDSLTVNWTGGGDVDSYTVSAFRHSQKVDSTQIPKHVFEHTFHRLEAGEQYQIMIASVSGSLKNQINVVGRTPVPS
VQGVIAADNAYSSSLIVSWQKAAGVAERYDILLTENGILLRNTSEPAATTKQHKFEDLTPGKKYKIQILLTVSGGLFQSKAQ
TEGRTPAAVTDLRITENSTRHLSFRWTASEGELSWYNIFLYNPDGNLQERAQVDPLVQSFQNLQGRMYKMVIVTHSG
ELSNESFIFGRTPASVSHRNSNRNTDLSLWFWNSPASGDFDFYELIYNPNGTKKENKDKDLTEWRFQGLVPRGRKYL
WVYTHSGDLSNKVTVAESRTAPSPSPLMSFDIANTSLAITWKGPPDWTVDYDFELQWLPRKDLVTFNPNRKSSEGRIVYG
LRPGRSYQFNVTVSGDSWKTYSKPIFGSVRTKPKDKIQNLHCRPQNSTAIACSWIPDSDFDGYSIECRKMDTQEVFESRK
LEKEKSLNIMMLVPHKRYLVS IKVQSAGMTSEVVEDSTITMIDRPPPPPHIRVNEKDVLSKSSINFTVNCSEWFSDTNG
AVKYFTVVVREADGSEDELKPEQQHPLPSYLEYRHNASIRVYQNTYFASKCAENPNSNSKSFNKLGAEMESLGGKCDPTQQ
KPCDGLPKPHTAYRISIRAFDTQLFDEDLKEFTKPLYSDFPFSLPITTESPLFGAIEGVSAGLPLIQLMLVALVAFV
VSHGRERPSARLSIRDRPLSVHLNLGQKGNRKTSCPIKINQFEGHFMKLQADSNYLLSKEYEELKDVGRNQSCDIALLE
NRGKNRYNNILPYDATRVKLSNVDVDDPCSDYINASYIPGNFRREYIVTQGPLPGTKDDFWKMWQNVHNIVMVTQCVEK
GRVKCDHYWPAQDQSLYGLDILQMLSESVLPEWITREFKICGEEQLDAHRLIRHFHYTVWPDHGVPETTQSLIQFVRTVR
DYINRSPGAGPTVVHCSAGVGRGTGFIALDRILQQLDSKSDSDIYGAVHDLRLHRVHMVQTECCQYVYLHQCVRDVLRARKL
RSEQENPLFPYIENVNPEYHRDPVYSRH

PTPRB.mouse 201 ENSMUST0000092167 [Mus musculus]

MLRHGALTALWITLSVQTVGVAEQVKCNFTLLESRVSSLSASIQWRTFASPCNFSLIYSSDTSQPMWCHPIRIDNFTYGCN
PKDLQAGTVYNFRIVSLDGEESTLVLQTDPLPPARFEVNRKASTTQLQVRWTPSSGKSWYEVQLFDHNNQKIQEVQVQE
STWVSQYTLNLLTEGNSYKVAITAVSGEKRSFPVYINGSTVPSFVKDLGISPNNSLIWSRGSNGVNEQLVRLVMDKGA
VQDNTVDRRDTSYAFHELTGPHLYNLTIVTMASGLQNSRWKLVRTAPMEVSNLKVNTDGRITSLNVKQKPPGDVDSYSIT
LSHQGTIKESKTLAPPVTEQFQDLVPGRLYQVTISICISGELSAEKSAAAGRTVEKVRNLVSYNEIWMKSFVNWTTPAGD
WEHYRIVLFNLSVLLNNTTVGKEETHYALDGLELIPGRQYIEVIVESEGNLNRNSERCQGRTPVLAVLQLRVKHANETS
LGIWTRAPLGEWEKYIISLMDRELLVYHKSLSKDAKEFTFTDMLPGRNYKATVTSMSGDLKQSSSIKGRTPAQVTDLHVNNQ
MTSSLFNTWKALGDVEFYQVLLIHENVVVKNESVSSDTSRYSFRAKPGSLYSVVVTVSGGISSRQVVAEGRTVPSVS
GVTVNNSGRNDYLSVSWLPAPGEVDHYVVSLSHSGKVDQFLIIAKSVSECSFSSLPGRLYNVTVTTKSGNYASHSFTTEER
TVDPKVQGISVNSARSVDYLVKSVWHATGDFDHYEVTIKNRESFIQTKTIKSENECEFIELVPGRLYSVTVSTKSGQYEA
SEQGTGRTIPEPVKDLTLNRSTEDLHVTWSRANGDQYEVQLLFNDMKVFPFHLVNTATEYKFTALTTPGRHYKILLVLT
ISGDVQQSFAIEGLTVPSTVKNIHISANGATDRMLVSDTWSPPGGGDVDSYVVSFAFRQDEKVDSTQIPKHASEHTFHRLEAGAK
YRIAIIVSVSGLRNQIDALGQTVPASVQGVVAANAYSSNSLTVSWQKALGAERYDILLNENGLLSNVSEPARAQHKF
EDLTPGKKYKMLLTVSGGLFSKESQAEGRTPAAVNTLRIENSSRYLSFGWTASEGELSWYNIFLYNPDRTLQERAQVD
PLVQSFSPQNLQGRMYKMVIVTHSGELSNESFIFGRTPAAVNHKGSNRNTDLSLWFSWSPASGDFDFYELIYNPNGT
KKNENKEDVTEWRFQGLVPGRKYTYLVVTVHSGDLSNKVTGEGRTAPSPSLLSFADVANTSLAITWKGPPDWTVDYDFEL
QWFPGDALTI FNPYSSRKSEGRIVYGLHPGRSYQFSVKTVSGDSWKTYSKPIFGSVRTKPKDKIQNLHCRPQNSTAIACSWI
PPDSDFDGYSIECRKMDTQIEFESRLEKEKSLNIMMLVPHKRYLVS IKVQSAGMTSEVVEDSTITMIDRPPQPPHIRV
NEKDVLSKSSINFTVNCSEWFSDTNGAVKYFAVVVREADSMDELKPEQQHPLPSYLEYRHNASIRVYQNTYFASKCAESPD
SSSKSFNIKLGAEEMDSLGGKCDPSQKFKCDGPLKPHAYRISIRAFDTQLFDEDLKEFTKPLYSDFPFSLPITTESPLFGV
IEGVSAGLPLIQLMLVALVAFVICRQKASHSRERPSARLSIRDRPLSVHLNLGQKGNRKTSCPIKINQFEGHFMKLQADSN
YLLSKEYEELKDVGRNQSCDIALLEPNRGNRYNNILPYDASRVKLCNVDDVDDPCSDYINASYIPGNFRREYIATQGPLG
TKDDFWKMAWEQNVHNIVMVTQCVEKGRVKCDHYWPAQDQSLYGLDILQMVSESVLPEWITREFKICSEEQDLAHLRIRH
FHYTVWPDHGVPETTQSLIQFVRTVRDYINRSPGAGPTVVHCSAGVGRGTGFIALDRILQQLDSKSDSDIYGAVHDLRLHR
VHMVQTECCQYVYLHQCVRDVLRARKLRNEQENPLFPYIENVNPEYHRDAIYSRH

PTPRB.gallus PTPRB-201 ENSGALT00000016330 [Gallus gallus]

MSSPNQWQWAMDDKLFHMKSGQCLSISTRALSRSIIINCSQAVRWTCHEEDGLLKVANSSFLTKQGGKVMKQSKKY
LHTWMQLEASETGKPVYVNLCSQKAHNSLSTGLPLNFPDLWGMNTSTTTSSPLNNTTSTRPNDLPASSTQMSIRNVT
EHTKNIENLYFDLKSVAITEKPTVPTTPPYSSSTTEETKSAYLLTDRGKTHCSFLYALFQSPNTHQLWQRGMQVPAP
YTEETNSHESVPPVFRFHVDKPLQSLNLRANLASSLAKLQSVSVKDCCTHLLKAHLFSTTLEIIMLTILYTYITNITSKS
ALLRWSAAAHACNFNL TGRSEDRPAGCRPASAGNSYGCIVRDEAGTWYHLRIEPLADGEEANVTLQTDPLPPRFEIN
KEKTTTSLQVWRDPSPGKVDLYNIVVFDHNSKQLQESSIPGRFSKTEETFTGLVPGNKYNIIVLTVAGNKSTPEFRINGS
TAVSPVKNIQVIVKTDSIHASWSPGSGHVDLYKLVDFDNHGTVHESHQENPLTSYIFSGLTAGHLYNLSVITQAAELEST

AFRIVRTAPAEVLDLTVTNDSDFTLTKVWRPSSGGVDFYNITLSHLRSVKEVKTLPQAVTETHFDKLTTPGRLYQITARTI
SGELFTDKMATGRTPQKVSSELKAGGGGWLRLTLRVNWLPPAGDLERYHLLLNWNSALVLTNTLTKGNVTEYLVRVGLIPGR
QYEVVVVESEGDLQSRASCTGRTAPEFVLQLRVKHANESSLSVMWVTPVAEWDSEYVVSGLDGRDLTVIKKGLGKEAKEFTFN
ELVPGKRYTATVTTISGILSNWTIECEGRTPVAQVTLGLTVASQGSTNSLFTNWTKALGDVDYVQVLLIHENVIKNETVPS
TNEYHYFPLKPGGLYSVVVTVSSGGIISRQRTIAEGRTVPSVSTGVTVNNSGRSDYLSVSWLPASGDVDSYLVTLSSHDDQII
QTLTISKSLSECSFSSLTPTGLYNVMTTKSGKYENHSFRERTVPSGGVQLTVSNSARSYDLKVSWSLHAGSGSFDSEYV
KNNNDFIQTKSPVKDENECVFTNLVPGRQYSVTVSTRSGKYETSERVFGRTMPESVKGLTSLNRSTEDLRTVTSKAEAGD
KYEIQLLYNDMKIFPTIFLGNITIECWFALTTPGRLYKIVVLTISGDAQRATFIEGLTIPSAVRNINVSNGMDSLKVSW
TPGGGDVDSYTVTIFQQNHQLDSRSVSKDVSEHTFHLEAGEQYRVVVQSNNGALHNSLAAGFRITPASVQELLADHAYSS
HSLLVSWQKAPGVAERYDILLNLEQGVLLSNKSEPAKQHKFEDLLPGKKYKIHILTVSGGLFSKRAETVGRTPAAVTN
LKVTKNTDQLSFSWTTSSQGELESDYIFLYNPKLSLHDIRISGEQHLQCCSFQNLRQGRMYRMVIVTHSGDLTVNESSVFGRT
APAPVVGKASNRNMTDRLWFTPNKVAAGDVFYEFNLYPNNGTQKESLQKDLKKEWYFQGLVPGRKYTVLTVNTHSGDLINT
ANAEGRTAPSPNPTVDFWANTSLISITWLGPPDWDYDDFELQWLKPKDPLTVFNPYSSSKSVRIIYGLRPGRLYKFSV
TVSGDSWKTYSQSQSESVRTRPDKIQSLHCRPQTSTAIACSWTPPDSDFDGYSECKKLDREVEFSKRIEKDRTLNIMT
LVPHKRYLVSIVKHSADMTSEVVEDSTITMIDRPPQPPDIRVNNKKEVLTIKSSINFTFNC~~SWFSDTNGAVKYFTVVVREA~~
~~CGSEGPKEPDEQHPLPSLEYKHNDRIYQTNFYASRCAENPDSYKSFDIKLGEMENLGGKCDPQKQKCDGGLKPRTA~~
~~YRISIRAFPTQLFSEDPKELPOPLFADTFFSLPITTEAEPLFGVIEGV~~~~SAGLFLTVMLVAVTALE~~VCRQKVSNGHERPTARL
SIRRDKPLAVHLNLGQKGNRKTSSPIKVSFHEAHFTKLQADSNYLLSKEYEDLKDVGGRNQTCDIALLPENRGNRYNNILP
YDTSRVKLSNVDDPCSDYINASYIPGNFRREYIATQGPLPGTKDEFWKMMAWEQNVHNI VMVTQCVEKGRVKCDHYWPLD
QDSLYYGLLIVEMLSVLPWEVIREFKICSEEDQLDSTRILRHFFHYTVWPDHGVPEPQTQSLIQFVRTVDRYINRTPDTGPT
VVHCSAGVGRGTGFIALDRILQQLDSDKTDVDIYAAVHDLRLHRVHMVQTECQYVYLHQVCRVDLRLRARKLRSEQENPLFPIY
ENVNPEYHRDAVYSRH

PTPRB.frog xenopus ptprb-201 ENSKETT0000024858 Xenopus tropicalis

MLRYAFGLSLWIVLNLKINVEPVKCTINNTDVTVTSQSIYLEWTSPPGSHCNFSLTCCSSHNFEITCNP IQKSNDSECTLS
GLEAGTVFDLRIVSLQDGEKMLSLQTDPLPPSNFVNKNSVTSSSQANWAPSSGKVDLYKVKLLDSKKEIVQIEIVSGI
SLTKEVTFKNLTPGTIKSYIEITAIISGNKRSPSTVINASTVPSVTKSIDANSKTDITDASWKPGHGNIDSYKLLILLNDRKI
TEVVLEQNITSYSFHHGLPGYVYNLTVISEAAGLRTWNSKLIRTAPEVQAISVQSDGRDALKVQWKEPRGKLDYFNVTL
SDTGSIKYSRTPHGGPTPEITFTGLTPGHLYQVGI STIAGELSTSRMATGQTIPEKVADLKAVSNGRLSLRLSWVPPSGDW
DKYNI VLYNNTAVLVNTTIDKKYREYVIQDIGLIPGRQYEAUVIVESGKFKNLARCKGRTPAPAVMDLRLKHANESSLITM
WMPVAEWDNYVVSADRLDTPVKNVLSKEAKEFTFNHLAQGRKYIATVTSISGDLSSRASVEGRTPVAQVTNLNVENQGT
TSSLHADWTKAPGDVDSYQVLLIHEKIVIKNETVSNEINKYSFHSKPGGLYSVVVTVSSGGIISRQTVSEGRTRKLLSI
ISSNTGSLTFPFLHQSSNSRLRYIGWTSLICQKFLAKDLFFAWPGIECSFPVCVIGKALYFVRMDYQQTGVTNRYPRDNEA
IPLVVNTCCINCLSSYRKLVSIESISPINKRNNIKRVSLSVSSEYTHPRLRHSGCVTTTRANSGMLKQLPQAGLLVPE
AVKLLTAEASSEELHVTWSRAGDGDVHYEIQLLFNDMKVFPPI TLNNTADEYKFTSLTPGRLYKIVVLTQFSGEAQRAMFV
EALTVPSTVKNLHISPNGMTNSLKVNWTPGGGDVDSYTVSIFHNSPIGSHHTSKHIFEHTFQNLNPGELLYKVDVQNTSGS
LHKSLESFGRTPATVHSLSDADNAYSHELLIVSWQSAGKGVTDRIYDILLSDHGIVITNKSVPIMAKSHKFEDLVPGKRYI
QVFTVSGGLFSAAVETEGQTPAAVSDVKITGNSTDSLSFTWNSSEGELESDYIFLYNPNDSLHDKRSGNPDRLRCSFQGL
LPGRLYKMYIVTHSGSLTNESSIHGRTPAPVSSQLVSNRNTSESLWFTWAPALGDVDIYELIILYNPNNGTEKEIKQDKLTL
ESHQNLVPGRLYTLVIFTHSGDLFNKATAVGRTEAPQPNSVSVFADVKINLSIMWLGPPDWDYDDFELQWSPKDLLLV
VNPYSIGRSKGR IINGLYPGRQYTFGVQSVSGNTGKTLSPAIFGTVRTKPKDIHHLHCRPQTSTAIACSWTPPDSDFDGY
VECRNGSDVVEFSKRLEKDKSSITITSLVPHKRYVVSIVKHSADMTSQVIEDSAITMID~~PPPPPLLRVNNKDKTFFISK~~
~~SIHFRFNC~~~~SWFSDTNGAVKYFTVIVSEADGNDNQRPEASLPLPSYADYKTNKSTKIYQTSYFSPRC~~~~AENPDYNIQSYKIKI~~
~~STGMELLGGKCDQONENKYCDGPLSPRTSYRISVRAFTQLFTEEMRTFPEPLYSDFTFSLPITTEAGSLFFNKNITD~~~~LPLQI~~
~~FPQTKAMILV~~VLCRYEQDPMTHLSSQIERIPSVHLNVGHIQIGDRISSRPIILTAQFEHFHFKLQDTSNYLLSREYENLKD
FGRDQSSDTALLPENRGNRYSNILPYDSTRVKLANVDDPCSDYINASYMGPINFREYIATQGPLPATKDDFWKMWVEQ
NVHIIVMVTQCTERGRACKDHYWMPDQDSYYYGDLIVQMLSESVLPWEVIREFKICSEEDQIDAPRLVRHFHYTVWPDHGV
ETTQSLIQFVRTVDRYINRTPGSGPTVVHCSAGVGRGTGFIVLDRMLQVQDVTVDSDVIFGAVRDLRIHRMYMVQTECQYVY
LYQCVRDVLRLRARKLRNEQDNPLFPIYENVNPEYHRDAVYLRH

PTPRB.fish Zebrafish ptprb-202 ENSDART00000127299 NP_001186063.1| FN428722 Danio rerio

MLRNKAFHAAVWAACGILSTMVDATECSIEIVKVISSTESIRLTLDNANVKQYTVSVKDRHTYDTSKGRQDREHSCQIE
SLDPGTWYHLDLVISTLDEKQQSQRAVTLQTRPSAVENLQVSGDANSLDVSWQPGKGTERTYWIIVLIDSSGRDSAWNSTVAS
TATSYTMKGLISGRLYNIIVTVEGELQNSASTQAQTA PASVSKLRTENNGDRNSIRVLWDKASGDVDSYLVSLTTPGSNS
IEKVLPPDDTYVFNLSPGKGYQVSVSTRSGALSNTWITGKADPGKVSDDLVIENLSVRGALKISWTPPSEGWEHIRVAL
SNGSEVLRNQTQVGRTAKEILLSGNLNLPGRVYRMAVSVENGLANTVIEYEGELGPPVSQLVHRVHSDETSLSALWAHAASS

SSRDGYI IQLFQSNSTSVIQRNLSRDMRECTFNVLTPGRLYDITVTTAKIIRSSATLQGRTPLKVNHKLKLSNKGSTDS
LNASWEKPLGDDLFYHMLLRDQTVHNITASANTTSSLRLLPFLRPGALHKILVTTISGSQTSKLAEAECRTPVPAAVSDITV
TNSGPDFLNVSWKAAEGVDVNVVMLKQEKIVHTLATSATTECVFRSLVSGRLYSISIAHSGSYRNQTLQERTKPKST
VQNPTAIHSARDDFLKLYWNHAGSDYDYVVTIEHNGTRLQSQKLNRTQSDCAFSDVLPGRLYNLTVSTWSGQYSAVSIY
GRTFPGAAGVNLSTLTERGGTFLRVSWTSAPGDVDHYELQILFNDTQVSAAVNLSSAIGEHLFSALTPGRLYKIVLSTHSGSY
QRAEILEGRTPVSPQVSVHLSAGTADGSLRVSSADGDDLFYSVSLFQETHLQDSRRVPKHITQAEFYNLI PGQLYSVTV
QSVSGSQTNNSTTTGRTEPSTVTGLRADNELSTHSLVSNWPAVGVHDGYRLWLLDDGGTNIHNASVPAASRHHLFENLTP
GRRYQAHVYRSLSGTAEKSDVVAEGQTRPSAVSALHVCNTSSDLSCFASAVGWVDGYELYLYDRDETLHYHTLGS DALG
WSFTLLQPQGHYKMTITSKSGKLSNQSSVWAHTAPASVPELHVENQGTDSLSSWTPAGGLTGYSVTVDGREQRVGPVEV
TQVVFHSLVAGRLYLATVQWSEDLNSSTTAVGRTVPAAPSSVSVSCSGSVDMMKWHVPDGDYEDFEVTFWFPQD TLYISH
LRPTQRILEGLHPGRLYNISLRTVSGKRHSPVITYSRPVYHTIRTPPLPAPSIHCFPLSSTSVSCSWTPPHSDYDGFVVQCH
RQSGKAVYTHTLGNHTLSQQFDRLEFPFKNYTIYVAVMSGDKQSSVTKNVITMIDRPPVPPVTVRVNHSASAVITHTFIRE
KPNCSWFSANGAIRYFTI IATESNDV DNGLEPQRHPLPSYLDYRQNHSLKAYOTGFYHSTCAEGSDGKQVQVFEINLGAGM
KHLGGACKLDPESIQHGSHLCDGFLRSRTSYRSLVRAFTQLFDEENREFPHPPLYTDTYLSLPLLTQSA PRSLGTGCTAA
PAGVPPGALPAAALYRKRRAHKIAVQESPVMMCKKWKALPTSQMCLIRSPVQAANFESHAKLQSDSSYLLSEFEGLKDV
GRIQTQNAARLLGNRNKNRYNNILPYDSTRVRLSCLLEDDPCSDYINANYIPGNFRWEYIATQGPLPGTKDDFRMVMWEQN
VHSLVMVTQCVERGMKCDHYWPTDSEPLCYGDI VVQLLSEKVFPEWTIRDFKISCESQLRYPRMVRQFHYTIWPDHGVPD
TTQSLIQFVRTVRDFINRTNSPGISVVHCSAGVGRGTGTFIVLDRALQQLDRNCTVDIYGCVFDLRLHRSYVMVQTECQYAYI
HQCVRDVLRARKLGCEERDNLPLFIYNNII

PTPRJ.human ENSG00000149177 [Homo sapiens]

MKPAAREARLPPRSPGLRWALPLLLLLLRLLGQILCAGGTPSPPIPDPSVATVATGENTGIISSSTAESFHKQNGTGPQVET
NTSEDESSGANDSLRTEQGSNGTDGASQKTPSSSTGPPVFDIKAVSISPTNVILTWKSNDDTAASEYKYVVVKHMENEKT
ITVVHQPCNITGLRPATSYVFSITPGIGNETWGDPRVIVKITEPIPVSDLRVALTGVRKAALSWSNNGNTASCRVLES I
GSHEELTQDSRLQVNI SGLKPGVQYININPYLLQSNKTKGDPGTEGGLDASNTERSRAGSPTAPVHDESLVGPVDPSSGQQ
SRDTEVLLVGLEPGRYRNATVYSQAANGTEGQPAIEFRNTAIQVFDVAVNISATSLTLIWKVSDNESSNYTYKIHVAG
ETDSSNLNVSEPRAVIPGLRSSTFYNIITVCPVLGDI EGTGFLQVHTPPVPSDFRVTVVSTTEIGLAWSSHDAESFQMHI
TQEGAGNSRVEITTNQSI IIGLFPGTYKCFEIVPKGPNTEGASRTVCNRTVPSAVFDIHHVYVTTTEMWLDWKS PDGAS
EYVYHLVIESKHSNHTSTYDKAITLQGLIPGTYNITISPEVDHVGDPNSTAQYTRPSNVSNIDVSTNTTAATLSWQNF
DDASPTYSYCLLIEKAGNSNATQVVTDIGITDATVTELIPGSSYTFEIPAQVGDGKLSLEPGRKSFCTDPASMASFCEV
VPKEPALVLKWTCPAGNAFLEVEVSSGAWNNATHLESCSSENGTEYRTEVYLNFTSYNISITTVSCGKMAAPTRNTCT
TGITDPPPPDGSPTNITSVSHNSVKVFGSGFEASHGPIKAYAVILTTGEAGHPSADVLKTYTEDPKKQASDYYVYLYLIRTEH
KGRSQSLSEVLKYEIDVGNESTTLGYINGRLEPLGSYRACVAGFTNITFHPQNKGLIDGAESYVSFRYSDAVSLPQDPGV
VQAVPGCIPGALVIVTVGCFE FWRKRKDAKNNEVSFSQIKPKKSKLIRVENFEAYFKKQADSNCGFAEYEDLKLIVGI
SQPKYAAELAENRGKRNRYNNVLPYDISRVKLSVQTHSTDDYINANYMPGYHKKDFIATQGPLPNTLKDFFRMMVWEKNVYA
IIMLTKCVEQGRTKCEEYWPSKQAQDYGDITVAMTSEIVLPEWTIRDFVTKNIQTSSEHPLRQFHFTSWPDHGVPDITDLDL
INFRYLVRDYMKQSPPELIVHCSAGVGRGTGTFIAIDRLIYQIENENTVDVYGIYVYDLRMLRPLMVQTEQYVFLNQCVL
DIVRSQKDSKVDLIYQNTTAMTIYENLAPVTTFGKTNGYIA

PTPRJ.mouse Ptpj-201 ENSMUST00000168621 [Mus musculus]

MTRGGGRSSRGRSRELGATRGGWAPLAPPREAPASLRPRPLRARRARLRRVAAAAAAMSPGKPGAGGAGTRRTGWRRR
RRRRLETETRAPFGHTAGRVPPTFGQAQGMKPAARETRTPRSPGLRWALLPLLLLLLRLLGQVLCAGAAPNPIFDIEAVV
SPTSLLTWKHNDSGASECRIENKMESNLTFPVKNQTSNITGLSPGTSYTFSIISVTTNETLNKTIITTEPWPVSDLHVTS
VGVTQARLWTSNANGTASYRMLIELTTHSSVNI SGLKPGTNNFAPFESNETQADFAVAEEVVDANGTKRIPVITNLSQLH
KNSLVSDVPPSGQDPLSLEILLTDLKPDQYQYNTIYSQAANGTEGQPRNKVFKTNTSQVSDVRAMNISASSMTLTKWSNYD
GSRTSIVYKIHVAGGTHSVNQTVNKTEAILGLSSSTLYNITVHFFLQGTGTEGFLQVYTSQVSDFRVTNVSTRAIGL
AWRSNDKSFEIFIKQDGGKHRNASTGNQSYMVEDLKPGETSYHFEIIPRGPDTEGLSSTVNGSTDP SAVTDI RVNRI ST
TEMQLEWQNTDDASGYTHLVLESKSGSIRTNSQKWITVGSLLTPGTYLVNVTI FPEVDQIQGISNSITQYTRPSSVSHIE
VNTTTTTAARLWKNEDAASASYAVSVLILKTGDGNSVTSNFTKDP SILPELIPGVSYTVKILTQVGDGTTSLVPGWNLFC
TEPEPVTSFHCVEVVPKEPALVLKWCAPFGMYTGFELGVRSDSWDNMTRLENCTSDDDTECRTEVAYLNFSTSYNISIAITLS
CGKMLPAQNICCTGITDPTPDGSPNITSVSHNSVKVFGSGFEASHGPIKAYAVILTTGEAAQPSADVLKTYTEDPKRGA
SDTYVYLYLIRIEKQSQGLSEVLNVEIDVGNQSTTLGYINGRLEPLGSYRACVAGFTNITYNLQNDGLINGDESIVSFSFSE
YSEAVFLPQD PTPRNSSTHCGALVAVVGNVWVNRKRKRTDAKNNEVSFSQIKPKKSKLIRVENFEAYFKKQADSNCG
FAEYEDLKLIGISLPKYTAEIAENRGKRNRYNNVLPYDISRVKLSVQTHSTDDYINANYMPGYHKKDFIATQGPLPNTLKD
DFRMMVWEKNVYAVMLTKCVEQGRTKCEEYWPSKQAQDYGDITVAMTSEVVLPEWTIRDFVVKNMQNSSEHPLRQFHFTS

WPDHGVDPDITDLLINFRYLVRDYMKQIPPESPILVHCSAGVGRGTGTFIAIDRLIYQIENENTVDVYGVIVYDLRMRPLMVQ
TEDQYVFLNQCVLDIIRAQKDSKVDLIYQNTTAMTIYENLEPVSMFGKTNGYIA

PTRJ.chiken PTRJ-202 ENSGALT0000045335 NP_989952.1 [Gallus gallus]

MRRLLPPLPPCLLLLLLLPAEVRCTTACTDDCSLKNVTEEMGTSSNDELSVNATSGNRRLSSEVSLPGRAMSDQNSVAQPR
AVLDLKTVEYGVTSVNLWTVNDTASDSYTYRIEVRNGSSINNKTSIDITDAEITGLIPGTLTYTFVFAVAADGQTAGEGAS
ISLYTKPSVLDLKAEYVGVTSVNLWTVNDTASASYTYRIEVTSDSSIDSLTSSVTMAEITGLIPGTLTYTFVFAVAADN
RTEADGASISLYTKPSPVLDLKAEYVGVTSVNLWTVNDTASASYTYRIEVTSDSSIDSLTSSVTMAEITGLIPGTLTYNFT
VFAVAADNRTEADGAFSTLYTKPTPVLDLKAHEGVTSSVSLNWMVNDTASDSYTYRIEVRNGHSSVNNKTSNIPETETGLNLP
GTLTYTFVFAVAADGETEGEGASISVYTKPRAVLHLKTEYVGVTSVNLWTVNDTASASYTYRIEVRNGSSINNKTSIDITD
AEITGLDPGTLTYTFVFAVAADGQTAGEGASISLYTKPSMVLNLKAEYVMTSVNLWTMVMNDAESASYTYRIEVAHESLIN
ETMSNVTKSIVTYLIPGTSYNFTVFAIAADNQTEGEGASISQNTVPSVNAFQCEAVANMSYLTLLKWNCPYGGYSGPDIIE
FNGTWTKKQQSQFCREGSEIEFKTEPLDYKYTYTVSVTTVSDGLTSLPVQKICKTSITDPPVFNKAPLVKAVSHNSLSVE
FPDFESVNGPLKAYAVMIVTEAEGCLPSKSDLDYTYNDFKQKMTATYVTVIVDVEEISSSSHQNGHNVVDVGRKNTMYGY
ENGLIPLHSYRASVAGFTNINFTVANKIMGEQSYVSFSPCSEVVSLLPQDPGVASHTDGLALALPAVAFQVLRWRRR
KDKRNTVEVSFSPKSKMIKVENFESYFKKQADSNCGFAEYEELKSAGVHQPKFAAEIAENRGNKRYNNVLPYDISRVKL
SNPSCITDDYINANYMPGYSSKAFIAAQPLPNTIEDFWRMIWEKNIYSIVMLTKVEQARTKCEQYWPDKQSKSYGDI
VTMVSEVVLPEWTIRDFNVENADTMESHSTRVQFHFTSWPDHGVPEITDILLNFRHLVHEYSQNPIDSPILVHCSAGVGR
GTFIAIDRLIQQIEMENTVDVYGVVYDLRMRPLMVQTEQYVFLNQCVMDIIRSQKEKKTDLIYQNTTAMAIYENFTPGP
AFGKANGYHA

PTRJ.fish FN428723 NP_001129714.1 [Danio rerio]

MGQHYRLQAIIVSRITVFLVVLVSIHHGHCTDCSLCQYKSTTTTSEINVFVGNSSFCIAQNASFTNSSQNTINVTGLLPNT
YFLIISCTSCCENFSSPAQVQNIIVVNYTNSSVTVAVTVPDGHVDSYVNCSTQSIPIINDNNTISSLPPGCLYNLTV
TISIGVLTNVSNIQFATKPNPEAYRVSQTFSSISLLWSTPLSMSGVAVSYGVSYPVQPGVNLPLTQTSSNMNITN
LLSGTYITVVTNGVNNLSSSFINLIGYTVPKIVQNLAVSNVSTNSVSLTWLPPNGTSSLSFNINISLQGTFTYTSNS
YQITQIQPQTQYNCYTTLIVAGNISGSPQIQCNTKPLPVSNLKAIPVGTIVILLSWTPSTALYLYNVSANGWSTYNSS
EGANVTGLTPGNNYTFVITVANGLSSEPVTISAFGLGKAIISAIAGNTQTMVQVTPPAGTVSLSARILLMDGTVLNT
KNQTKTSMFSDLLPQTQYSVIVSSISGPKQDSDSVTNATLPTPGEISTTQSTNLTNLSWAQPLNMSVSYFLLQYN
NSTLINCSQNTSLTGLNAGFPYNIIMWTVGAMNYSKSLTAFTNPSVTVNVTNVEFTETSVTLSWQQNDPQQSGYSYL
LNYKTINGTSNQTNNNTVKLEQLQSASQYNIISIALTPGGTKSDPQFITACSRPNRVNVSVALNVSTVRLSWARPLQ
YNQLVSYQVLVSNCTKNSMNMSTSTEMITVTLNLPQGLTLCQISIYSVACGLLGPVNTSVLTMPTSTVQPVVNSQGSNNLLV
LWSHPDGGLDMYILNIISSDGWSDSFLNSTENNYTFTQLKAAALTYTLTTLTVKGSFQETSGAVVSAIYPNSPGMITVLFKN
THSVLLHWDIAQNMTPGSFNYSLSFWSKNSLYLTPNNTLLLDGLQSGTPYNVSLATVGPLNLQSESMRYNVTMPDPVK
NLKVVSTTTNTISVTKMVDVPKYVSVNNAVKITSSNNMTIHNLPSTQYNIISVQSTFDNTEGEAVCLQDCTDAAPVEN
FVTCVGNLTLPLMLNLFWNHPLGGMNFVNLSSISEKTKNLNLYNFTGLKYNKTYSVTLRTLGCCKSTVKNISCTTGFT
VVVPKITAASVSEQQYNKFAVLVQSDAFNDSNGTVCSYGLLVSSGSPDCLDNNQNYNCCLLTKYDDWAKQSDTFLAVVKI
NVSKSELETIIIGDGSKWNKYTNGELNAKGTYNFAIVAFTLLEVKDGKVDVSKSYYSVSAYQSVTLPENVIIGGAAAGVG
VTVIIVVIIIGIVACTRKKRKEVNSVPIHSLSDPIKVEDYEAAYKRQRADSFSGFAEEFEDLRPVGINQSKTVAVFPENK
AKNRYNNVLPYDSSRVKLSVLSPPDDYINANYMPGYITKKEFIAAQGPLPVTNDFWRLIWEKNVHTIVMLTKCNEQGRV
KCEEYWAEMKTFSNLTVTTISDIPLEDWTLRDFEVKNMKTAEIRSVRHFHFTAWPDHGVPEITELLINFRHLVREHMDEY
SRHSPTLVHCSAGVGRGTGTFIAIDRLIFQIERDGVVDVYGI IHDLRMRPLMVQTEQYVFLNQCAMDIKSRGTNNVDLI
YQNTAALTIYENFEPLKKGKNGYHKA

PTRJ.human ENSP00000281171 [Homo sapiens]

MGLHPTGIHGARLLPPLLWFLVLFKNATAFHVTVQDDNNIVSLEASDVISPASVYVVKITGESKNYFFEFENSTLPP
VIFKASYHGLYIITLVVNGVNVTKPSRSITVLTKPLPVTSVSIYDYKPSPETGVLFELHYPKYNVFTVNIISYWGKD
FRTMLYKDFFKGKTVFNHNLPGMCYSNITFQLVSEATFNKSTLVEYSGVSHPEKQHRAPYPQNIISVRIVNLKNWEEQ
SGNFPEESFMRSQDTIGKEKLFHFTEETPEIPSGNISSGWPDFNSSDYETTSQPYWWSASAAPESEDEFVSVLPMEYENN
STLSETEKSTSGSFFFPQMILTWLPPKPTAFDGFHIEREENFTYELMVDEEAHEFVAELKEPKGYKLSVTFSSSG
SCETRKSQSAKLSFYISPSGEWIEELTEKPHVSVHVLSTTALMSWTSQENYNTISVSVSLTCQKQKESQRLEKQYC
TQVNSSKPIIENLVPGAQYQVVIYLRKGLIPLGPPSDPVTFAIVPTGIKDLMLYPLGPTAVVLSWTRPYLGVFRKYVEMFY
FNPATMTSEWTTYEIAATVSLTASVRIANLLPAWYNYFRVTMTWGDPELSCDSSSTISFITAPVAPEITSVEYFNLLY
ISWTYGDDTTLDSHSMHVMVAEGKIKKSVTRNVMTAILSLPPGDIYNLSVTACTERGNSMLRLVLEPAPPKSL
FAVNKTQTSVTLWVEEGVADDFEVFCQQVGSQKTKLQEPVAVSSHVVTISSLPATAYNCSVTSFSDSPSVPTFAVS

IPGYSAAQEYIATQGPLETRNDFWKMILQQKQVIVMLTQCNEKRRIKCDHYWPFTEPEVNYGDITVEMASEEQSDWAQ
RVFRVSYADETQCVTHTFNFTAWPDHGVPTVNAAESVLQFVQVVRQKALKSKGPPVTVHCSAGVGRGTGTFIALDWMQHIRDH
EFVDILGLVSELRSRMSMVQTEEQVFVHQCAQLMWKKRQQRFRISDVIYENVVSKS

PTPRO.fish zebrafish ptpro-204 ptpro-204 ENSDART0000122021 Danio rerio

MSSARARFFFIIFIFIQSAAGFRVQLQDGVRLMSLDDGDIQPGVSEYAVRVSGEPLTHLTLFQQAADAHTLPEPLLFNAN
YHGLLYSISLMTDGRTHTRTSIITQPLPLDSVEMWDYAPSPETAFFVQIRSPDRNIFTRVNIYSYTEGHQRRYMLYKDFLHG
KTVFKHWLQSGVCYSNITFQLISEASVNRSALLSRSDITHNPQQHRTVNPPLNVSLKILHLSGRGPPGVLVTGAILKNSHN
ASVVRERDLPEYEEQEPTSDTLNTEENLTHTQLNQSESENESESEPEVTAEPPLNSSTQSLWAWQTVSPAPTEEEEEGFVN
ALVPEYEDSNEPQSALGIPLEAVPMPVMPPLAPVLLQLRWSPAPHTAYDAFNIYIYRNGNSTETATVDENTHEFLAEL
SESGTYRIHVTTLSSAGDCEARESSANTAFTFYLSPSGEWMEQPQERPQAVSVKMLDSSAAVSWAPSTHTYNGSLISVQS
LTLCLRPSISQRMELNYCSEENITSDIISLTPGAQYRVVYHTNGPLVSPASEPVIIDIEPTGVRDLVVYPLSPSAVILLSW
QRPYNAVFRKYVLQTFFFNSATQTAQWSTYYEIAATASVIASRVTDLLPAWFYFRVTVMTWGDPPPLSCDDTSTVSFITA
PEAPHISSEVFSHGSVFRWRYGDLPTDLTHSRMLHWQVVAEGKKSARRRFVSDVTRSVMKASALAPAGDIYNLTVTACTE
RSCNTISAPVHIDPPAPRSLYAVNASDTSVTLIWAEEGVVDHYLITCRALGAHAEQKVKREPLVSAHVLTVSGLQASTT
YNCVSRSSSYSSSDPWHITVSTTVREMNPVVAISALAVLSVLLISLLVFLVLRKHKHMLRESRECAETVFNFAFERD
GKLPYWRRLSFAFLTLPLSCLWTDYLLAFYINPWSKTALKKRKLTSPVQLSDFEAYLKMKGDSAYKFSLQFELKSVGL
DLSHEAADLPINRPKNRYTNILPYDFSRVKLISLHNDGSDYINANYIPGYNPREYIATQGPLDTRNDFWKMVQLQKVE
IIVMLTQCNERRRVKCDHYWPFSDPEVAYGEISVEMLAETDSPAETIRSFRLAYADETQDVLHFNYTSWPDHGVPTVNAE
SILQFVQIVRQVNRSKGPIVHCSAGVGRGTGTFISLDRMLQHIQEHEYVDVGLVSDMRSHRMSMVQTEEQYVFIHQCVL
LMWKKKQQSHTSDVIYENVVSKS

PTPERO.shark gi|632971848|ref|XP_007902372.1| PREDICTED: receptor-type tyrosine-
protein phosphatase O [Callorhinchus milii]

MHWDLHRGAVCPMVLFITCFLFALKSATAITATVDGDNNIIVSLQPSDVTNGSSLYYVKITQSVNLTNYKIQFKEFNSTLP
PPVVFPAAYLHGIYLLVTLVGNITISVTKPVHLVAVLTKPLPVSNVAVDYKLFPEIGVLPKETTFFSRLLKLSYFK
GEEYQSSFFHKDFLKGKLIFFNWNVPGTCYSNITLQLVAEANVRVPLMEYSAVGHEPKHRTVPPYPPGNVSMKIVYGRKNS
SVEVHAGREESFVDTNRTADETTTINVQVNNLTAKESLHWGDYNTSGQNDENIMNWTPEFQQDSREISASGSDYFDIIN
VSEEDQSIYITTEPELSETNAEEVWLVLWELWLPKPPPTVYDHFNIYVQKEGNSTEIFTAENGSYAFKMELOGGPAKYEASV
TVSSSGICGTRESSSRKTLTFYLSPTGKWFEEITKRPENIVVKVNLNSTAAKVSWKHSQETSYSIVSVIISQCQKFGAQ
RTESHYCTEVLNLSSTVIPHILIPGGYQVIVYLVKNGPAIGPPSLPVTFLGDPGSKVDSLPLGSAIILSRPYSVVFDDK
YVVMYHHDVFNQNSKWKTIQDIPATIALTTSMTVGNLLPAARYEFRVTMVSGGEPHSCCDNSSASYTAPLAPEITRLE
YSKNILSLSWKYGDTVMGESNLTLPYLVVAVGRII KKSVPVRSVMTTVLRLLPBGDIYNLTVVAYTEGSSNTSAPRLVQLD
PAPPKSLFAVNKTQTSVTLWVEEGITDHYEVICKQIGSSVKGKKEATVYSTIVTLDNLVPSTSYNCVSVSVSYDSYSQP
VPIAVSTTVELNSMNVISALALSLLLTGLLLITLFLVLRKHLQHTRNCGAGIFVNFATLERDGLKIPFNWRRSIFAVLA
LLPSCWTDYLLAFGINPWNKYALKRRLTNPVQLDDFEMYLKDMARDSDYKFSLQFEDLKLVAQDVSHDTAELPVNRCKN
RYTNILPYDFSRVKLISLHNDGSDYINANYIPGYNPREYIATQGPLAETRDFWRMLVQKSHIIVMLTQCNERRRVK
DHYWPFNKEPVCYGDITVEMVFESELPDQSEDSEQEQPEWIIREFNVSYADAVQRVHYNYMAWPDHGVPNADTAENILQ
FVQMIHQQTVKRPGSIIVHCSAGVGRGTGTLIALDRMLQHIQEHEYVDILGLVSEMRMYRTSMVQTEEQYIPIHRCVQLMWK
KKRQQYNSSDVIYENVVSPQL

PTPRO.lamprey gi|577010973|ref|XP_006006115.1| PREDICTED: receptor-type tyrosine-
protein phosphatase O, partial [Latimeria chalumnae]

NVATFLVNVNEDNCIIIFLEPSDITSPSSSTYIVKITGESKNYFFQFEDFNDTLPDPIVFNASYHGLYLIISLKVLSNMVMS
KPAKSLAVLTKPLPVNSVFIIDYKSSPETGVIFIEIQYPEKYNVFRVNIYSLEGGQFRSMLYKDFRGTTFVFNHWLPGTCY
SNITFQLVSEVNVNKTTLMESSGVGHEPKQHRVTPYPPSNISMKISKINKGYSVSGSLLFGSEQLQQLQRLNNGIQEAP
SIQPFEDPESLTVNASYSWIDYNYNGTDYETTSQPYWTSIDATTHESEDFVNKLYSGYDDENMGMGMESSEVSPFLPVEL
FLMWSPPKPTSYDGNFIYIQREGNSTEVSTVDEETHVFTAELKEPKYKVSVTTLSSSGVCPRESTSDTSFTFFISPTG
EWFQELTEKPSNVSVKLLSSTMAIVSWSSLMGITNDTIVSVVSLTCKKQKQQRRIEKEHCSETNSSSGIIGNLIPGAQYRV
TVYLVKNGPLVGPSPDAVTFGTAPTQVGRDLVLYPLGPTTVILSWNRPFYVFKYVVMEMFFNPIITLSEWMTMYEITATVS
LSASVRVINLLPAWYVFRVMTVWGDPELSCDSSSTISFITAPLHPEIRSVYFNLLYVSWTYGDDTTDLSHSRMLHWL
VIADGKRVKVSVTRNVMTVTLNLPBGDTYNLSVTACTERSSNTSKPYVVTLEPAPPKSLFAVNKTQTSVTLWVEEGVAD
FYEYVCKQAASSQEKRLQEPISVTSHTVITISLLPSTSYNCVSTVSHNSPSTPAFIAVSTLVAEMNPVNVISVLLALS
LLTGLLLITLILRKHLMQARECGAGTFVNFASFERDGLKLPYWRRLSIFAFLLTLPLSCLWTDYLLALYINPWSKNALKKR
KLTSPQLDDFDAYIKDMAKSDYKFSLQFEELKLVGLDIPHVAADLPVNRCKNRYTNILPYDFSRVKLVSMDHDEGLDYI
NANYIPGYTSPQYIATQGPLPETERNDFWKMVQLQNSIIIVMLTQCNEKRRVKCDHYWPFTEDEPVSYGDITVEMLSEEEQT
EWWIREFRVSYADEVQDVLHFNFTAWPDHGVPTTNAAESILQFVQMIHQQRGRKRGKPMIIVHCSAGVGRGTGTYIALDRLLQY
IQDHEFVDILGLVAELRSYRTSMVQTEEQYIPIHRCVQLLWSKKKQRFCSIDVIYENVVSKS

PTPRH.human ENSP0000036528 [Homo sapiens]

MAGAGGGLGVWGNLVLGLCSWTGARAPAPNPGRNLTVETQTTSSISLSWEVPDGLDSQNSNYWVQCTGDGGTTETRNTTA
TNVTVDGLGPGSLYTCVSWVEKDGVNSSVGTVTATAPNPVRNLRVEAQNTSSIALTWEVPDGPDPQNSTYGVVEYTGDDGR
AGTRSTAHTNITVDGLEPGCLYAFSMWVGKNGINSSRETRNATTAHNPVRNLRVEAQNTSSISLSWEVPDGTDPQNSTYCV
QCTGDGGRTETRNTTDRVTVDGLGPGSLYTCVSWVEKDGVNSSVEIVTSATAPNPVRNLRVEAQNTSSIALTWEVPDGPDP
QPNSTYGVVEYTGDDGRAGTRSTAHTNITVDRLPEPGCLYVFSVWVGKNGINSSRETRNATTAHNPVRNLRHMETQNTSSIALC
WEVPDGPYPQDYTYWVEYTGDDGGTETRNTTNTSVTAERLEPGTLTYFSVWAEKNGARGSRQNSISITVPAVNTSLSKQDW
TNSTIALRWTAPOGPGQSSSYVWVWREGMTDPTQSTSGTDITLKELEAGSLYHLTVWAERNEVRGYNSTLTAATAAPNE
VTDLQNETQTKNSVMLWKKAPGDPHSQLYVYVWQWASKGHPRRGQDPQANVWVQTSRNTNETWYKVEALEPGTLNFTVWAE
RNDVASSTQSLCASTYPTDTVTITSVSTASAGYVNLISWPPQGGYEAFELEVGGRGSGDRSSCEAVSVLGLGPARSYPA
TITTIWDMKVVSHSVCHTESAGVIAAPVQILLPLLVGLLPLKRRNKKKQKPELRDLVFSPPGDI PAEDFADHVR
KNERDSNCGFADKIQQLSLVGHSSQSMVASASENNAKNRVYRNLVYDWSRVLPKPIHEEPGSDYINASFMPGLWSPQEFIA
TQGLPLQTVGDFWRLVVEQQSHTLVMLTNCMEAGRKCEHYWPLDSQPCTHGLRVLTVLGEVEMENVTWRELLLLQVEEQK
TLVSRQFHYQAWPDHGVSPDPTLLAFWRMLRQWLDQTMEGGPPIVHCSAGVGRGTGLIALDVLRLQLQSEGLLGPFSFVR
KMRESRLMVQTEAQYVFLHQCILRFLQSSAQAPAEKEVPYEDVENLIYENVAAIQAHKLEV

PTPRH.mousePptrh-202 ENSMUST00000166650 [Mus musculus]

MARAGNCGVWRSLVLLGLYGCSSVRAAGTSTVVDHRHAPASSYEFMSWVEKDGVSSSPQIPVTTAAAPNPVRNLRVEGQNNI
SISLSWEPPDQSSLQGLTYWTVQCSRHHGQTETRNNTDTSVTVDGLDGPSSYECSSVWVEKDGLYSKNETLNTSTAPNPVRNLR
RVEGQNNISISLSWEPPDQPSLQGLTYWAQCSRHHGQTETRNNTADTSVTVDGLDGPSSYECSSVWVEKDGVYSTNETLNTST
APNPVRNLRVEGQNNISISLSWEPPDQPSLQGLTYWAQCSRHHGQTETRNNTDTSITVDGLDGPSSYECSSVWVEKDGVYST
NETLSNNTAPNPVRNLRVKSQNNFISISLSWEPPDQPSLQGLTYWAQCSRHHGQTETRNNTDTSVTVDGLDGPFLYKCSVWV
EKDGVYSTNETLNTSTVPIASANPVRNLRVEGQNNFISISLSWEPPDQSSLQGLTYWAQCSRHHGQTETRNNTADTSVTVDGL
DGPSSYECSSVWVEKDGVYSTNETLNTSTVPAAVNITSISTSGGYVLLTWSPPSGGYESFEVKVGRKWRSENGSLCKGKV
TVSDLEPAQSYTATVTVFKDLKAQSLSTCHTESAAAPGAVVGGTADGTFNGCAVWPKRRRKRKQPKVEVPKDLVCS
GDILAKDFADHVRENEKDSNCGFAEEYQQLALEGGQSQITASALENRSKNRYRNLVYDWSRVLPQLQEPGSDYINAS
FMPGLWSPKFEIATQGPLPNTVGDVWVWVEQQSHTLVMLTNCMESGRVKEHYWPLDQAQPCIHGQLQVMLISEASENWT
VRHLQLFHMEKQQLSLRQFHYLAWPDHGVYSPDPLLAFRKMLRQWMDQTTDGGPPIVHCSAGVGRGTGLIALDVLRLQL
ECEGLVGPFSFVKMRESRLMVQTEAQYVFLHQCILKSLQKAPALVPEEAMYENVASLVIYENASAIMAHESEFSASGC

PTPRH.xenopus XP_004916623.1 [Xenopus (Silurana) tropicalis]

MEYDEALTGICNVQILLKAMRTGCVTLGYVPIRSHRFQKRLLSGVPSVPQSVSVSNITNTNVLTSWVPPNDINKDITYTYS
ITVYNGSSPDWPKNISGTSTQVTGLIPGVYTFVSVFTVTSNGTMSSQAGSTSGTTPSPVQSVSISINIMTNTVLSWVPP
NDINKDITYTYSITVYNGSTSWNGPTNISGTSTQVTGLIPGVYTFVSVFTVTSNGTKSSQAGSTSGTTPSPVQSVSISINIM
TNTVLSWVPPNDINKDITYTYSITVYNGSSPWNPKNISGTSTLVTGLIPGVYTFVSVFTVTSNGTMSSQNGSTSGTTPGS
LVTTVSIKQTNNSVTISWTPPDQINQATYMSINVTNSAGSNVNSSTNVSSKEVLDLPGVYTFVSVFTVTSNGTKSSD
CSPDSSSTRPTLVKDVTKSQTTDSVTLSTWTPNDINKGSYYLITACGTTTCNVNSTVNANGATVASLLPGLTYQFSVY
TVTSDGTNSSAFASTSGTAPNAPTSVSGNPVNTTALS VNWTPNDLNKNNTYTVFVQRNNEQSLNQSITSNTIIGGLTA
GNQYNSVIVSINNVPVSVKAVAYLQTNPEVPPGGFKIVTITNSSVNVSWTIPSGGFSFGIEINAVSGTSLNLTQFTGSPSE
ILDGLIPGTAYNLILRSFSSYPSGSSGRISRRATTNNIITYSNPIIEKVQLDPNTVTGICTKVGGGYQLKVAFPVCSGSY
SSIKILADGNEKETIPAQNCQSGATVQSLQPATKYIITVETVADSRQVSDFIICYTDNVGVIVGSIPLVGLLFLVGLIAY
FLRKRRGSKNPPFKKERGNMQKINIVRMPVMSKSAFPDYYQRQHADSDGFAEEYQQLSNVGINQSKLAAELSENRSKNRF
TNVLPYDHSRVRNLNRIADETSIDYINANYMPGYNSSKEFIASQGPLPNTSADFWRMWENQVSTIVMLTNCMENGRVKEH
YWPLDYTPCTYGDITVTVTSEMILPDWTVRDFTLKHAQQGNKHARHFHTVWPDHGVPENTTTIVEFRNLVREYMDLKR
SGPTVHVCSAGVGRGTGLIALDYLQKMEKEQRIGIYSFVQKMRQNRPLMVQTESQYVFLNKCMMLDLIQPNNDIENIENQI
GGDLIYENAS

PTPRH.fish ptrh-202 ENSDART00000114768 zebrafish Danio rerio

MELVLAFALLCVLVSVSQATLPLDVNDVRIHTRTELETFQWDKAGGSNSYSYILRRDGDSDIITAGTVNQVTKVSDLS
GTYNNYTLFTVFGDQRSEKGNFSAATVPPNVEVSVSIDQSETTITLQWKKDAEYSYNTLYGDKVDTPPSPQSTVEYTVTS
LSPGTYNFTLYTVFEGLKSSGFSFAFVTPSKITGFSVIGRSETALTFQWTKKEDDEYTELKSDANTQISFTDITVVC
ITLSLPGTYSYNYLAVFEGVKQAQKISEVTVPSNVENVSVEANDTDVILQWNAVPEQNKYNYILKYNDVEESDFTT
ENEIKHHVSLIPATNYNFTLHTEFYSONSTGYNFNHTFLSSVTEVRVNRSLTELVWKNLQNNVNYTLRSDGAEI
DFTGSAVVDVVKHKYSKLPKGTVYSFTLFTVSNVRSSEHSFKSITALDCASFNWVNTNSIVADVYGCTVYTAQNSGSS
GKNASIEGDKVNLRLDLYPGESYTVSLFYNLESKMLTQCSQRLTLVPYVVSNLKCEYHSGGYGLAVMWDYDPGVVDVVEVDV
NKSFNHSGSDEPTQLVFTGLQAAQWYKVKATSFSGARRSETILINCQTPDPAQVIGVAVQVFLVILCAGVAVNRIRYKS
ARNNGPDASADLVKSKNYKIACDKFPEHFRNMSRDDNRAFSAEYDLSVGVQSKVAALLPENKDKNRFVNLVYDTS
RVHLLINKAGSDSYINANYMPGYGNASREYIAAQGLPLSTVNDVFRMIWEKSSITVMTNCTEGBGRVKCEQYWPLDYTPC

LYENLLVTVKSENKSSQSWTLREFNVKNKMTSETRTVVRHFHTAWPDHGVPRGTEELIQFRDLIRQHIESHSTGPTVVHCS
AGVGRGTLLIALDVLQLNREKAVGVAAFVQEMRLNRLPLMVQTESQYVFLHQCILDSLQTKFVQQSEPIYQNTDTTIYVNA
MALKDFENSSHT

PTPRG.Hs ENSG00000144724

MRRLEPCWWILFLKITSSVLHYVVCFPALTEGYVVALHENRHGSVAQIRRRKASGDPYWAYS GAYGPEHWVTSVSSCGG
RHQSPIDILDQYARVGEYQELQLDGFNDNESSNKTWMKNTGKTVAAILLKDDYFVSGAGLPGRFKAKEKVEFHGHSNGSAG
SEHSINGRRFPVEMQIFFYNPDDFDSFQTAISENRIIGAMAIFQVSPRDNALDPIIHGLKGVVHHEKETFLDPFVLRD
LLPASLGSYYRYTGSLLTPPCSEIWEWVFRFPVPISYHQLEAFYSIFTTTEQQDHVKSVEYLRRNFRPQQRLHDRVVSKS
AVRDSWNHDMTDFLENPLGTEASKVCSSPPIHMKVQPLNQALQVSWSQPETIYHPPIMNYMISYSWTKNEDEKEKFTFK
DSDKDKLKAISHVSPDSLFLFRVQAVCRNDRSDFSQTMFLQANTTRIFQGTRIVKTVPTASPASSADMAPISGSSSTW
TSSGIFPFSVSMATGMGPPSSSGQATVASVVTSTLLAGLFGGGGSISSPSTVWPTLRPTAASASKQARVPLATTEALA
SPGPDGDSSTPKDGEGETEEGKDEKSESEDEGEREHEDEKDEKSEKKEKSGVTHAAERNQTEPSPPTSSPNTAEAGHQQT
IPGHEQDHTAVPTDQTGGRRDAGPLDDPMVSTQVPPATEEQYAGSDPKRPEMPSKKPMRGRDRFSEDSRFITVNPAAE
KNTSGMISRPAGRMEW **IPDLIVSALTFVQVILLIAVLV** RGCNKIKSKGFPFRFREVPSGGERGKGRKCFQTAHF
YVEDSSSPRVVNPESIP IIPIDDMDEAI PVKQFVKHIGELYSNNQHGFSEDFEVEVQRCTADMNITAEHSNHPENKHKHRY
INILAYDHSRVLKRLPLPGKDSKHSYINANYVDGYNKAKAYIATQGPLKSTFEDFWRMIWEQNTGIIVMITNLVEKGRK
CDQYWPTEENSEEYGNII VTLKSTKIHACYTVRRFSIRNTKVKKGQKGNPKGRQNERVVIQYHYTQWPDMDGVPEYALPVLT
FVRRSSAARMPETGPVLVHCSAGVGRGTGYIVIDSMLQQIKDKSTVNVLGLFKHIRTQRNYLVQTEEQYIF IHDALLEAI
LGKETEVSNQLHSYVNSLLIPGVGKTRLEKQFKLVTQCNAYVECFSAQKECNKKNRNSVVPSEERARVGLAPLPGM
KGTDYINASYIMGYRSNEFIITQHPHPTTKDFWRMIWDHNAQIIVMLPDNQSLAEDFVYWPSSREESMCEAFTVTLI
SKDRCLSNEEQII IHHDFILEATQDDYVLEVRHFQCPKWPNDAPISSTFELINVIKKEALTRDGPPIVAHDEYGAVSAGM
LCALTTLSQLEENENAVDVFQVAKMINLMRPGVFTDIEQYQFIYKAMLSLVSTKENGNGPMTVDKNGAVLIAEDSDPAES
MESLV

PTPRF.Hs ENSG00000142949

MAPEPAPGRMTMVLVLPALVMLGLVAGAHGDSKPVFIKVPEDQTLGSGGVASFVCQATGEPKPRITWMKKGKVVSSQRFVE
IEFDGAGSVLRIQPLRVQRDEAIYECTATNSLGEINTSAKLSVLEEEQLPPGFPSIDMGPQLKVVEKARTATMLCAAGG
NPDPEISWFKDFLPVDPATSNRIKQLRS GALQIESSESDQGYECVATNSAGTRY SAPANLVYRVRVAPRFSIPSS
QVVMPPGSSVNLTCVAVGAMPYVKWMMGAELTKEDEMPVGRNVLELSNVVRSANYTCVAISSLGMIEATAQVTVKALPK
PPIDLVTETTATSVTLTWDSGNSEPVTYIGIQYRAAGTEGPFQEVGDVATTRYISIGGLSPFSEYAFRVLAVNSIGRGGP
SEAVRARTGEQAPSSPRRVQARMSASTMLVQWPEPEEPNGLVGRYRVYTPDSRRPPNAWHKHNTDAGLLTTVGSLLP
GITYSLRVLAFSTAVGDGPPSPPTIQVKTQQGVPAQPADFQAEVESDTRIQLSWLLPPQERIIMYELVYAAEDEDQHKVT
FDPTSSYTLLEDLKPDTLYRFQLAARSDMGVGVFTPTIEARTAQSTPSAPPQKVMCVSMGSTTVRVSWVPPADSRNGVIT
QYSVA YEAVDGEDRGRHVVDGISREHSSWDLVGLKWTYRVVWRAHTDVGGPESPVLVRTDEDVPSGPPRKVEVEPL
NSTAVHVYWKLPVPSKQHQGIRGYQVTVRLENGEPRGLPIIQDVMLAEAQRPEESEDYETTISGLTPETTYSVTVAA
TTKGDGARSKPKIVTTTGA VPGRPMTMI STTAMNTALLQWHPPKELPGELLGYRLQYCRADARPNTIDFGKDDQHFTVT
GLHKGTTYIFRLAAKNRAGLGEFEKEIRTPEDLPSGFPQNLHVTLTSTTELAWDPPVLAERNRIISYTVVFRDINS
QQELQNIITDTRFTLTGLKPDTTYDIKVRAWTSKGGSPSPSIQSRTPMVEQVFAKNFRVAAMKTSVLLSWEVPSYKS
AVPFFKILYNGQSVVDGHSRMLIADLPNTEYSFVLMNRGSSAGGLQHLVSI RTAPDLLPHKPLPASAYIEDGRFDLSM
PHVQDPSLVRFYIVVVPIDRVGSSMLTPRWSTPEELEDELLEAIEQGGEQRRRRRQAERLKYVAAQLDVLPEFTTL
GDKNRYRGFYNRPLSPDLSYQCFVLASLKEPMDQKRYASSPYSDEIVVQVTPAQQQEPEMLWVTG **PEAVIILLLIVT**
ELK KRKRTHSPSSKDEQSIGLKDLSLAHSSDPVEMRRLNYQTPGMRDHPPIPIITDLADNLERLKANDGLKFSQYESID
PGQQFTWENSNLEVNPKPNRYANVIAYDHSRVILTSIDGVPGSYINANYIDGYRKNAYIATQGPLPETMGDFWRMVE
QRTATVMMTRLEEKSRVKCDQYWPARGTETCGLIQVTLTLDVTELATYTVRTFALHKS GSSEKRELRQFQFMAWPDHGV
EYPTPILAFLLRRVACNPLDAGPMVHCSAGVGRGTGCFIVIDAMLERMKHEKTVDIYGHVTCMRSQRNYMVQTEDQYVFI
HEALLEAATCGHTEVPARNLYAHIQKLGQVPPGESVTAMELEFKLLASSKAHTSRFISANLPCNKFKNRLVNIIMPYELTR
VCLQPIRGVEGSDYINASP LDGRRQKAYIATQGPLAESEDFWRMLWEHNSTIIVMLTKLREMGREKCHQYWPAERSAR
YQYFVVDPAEYNNMPQYILREFKVTDARDGQSRTIRQFQFTDWPEQGVPKTGEGFIDFIGVHKTEQFGQDGPITVHCS
AGVGRGTGVFTILSIVLERMRYEGVVDMFQTVKTLRTQRPA MVQTEDQYQLCYRAALEYLGSDHYAT

Figura Suplementaria 2: Alineamiento de secuencias representativas de UPK2a, UPK3a, UPK3b, PTPRQ y PTPRB. Grupo de secuencias representativas de UPK2a, UPK3a, UPK3b, PTPRQ y PTPRB se alinearon utilizando el servidor MAFFT. Las secuencias abarcaron desde el segundo exón hasta el final del cuarto exón. Las fases de los intrones están marcadas de color rojo o azul si corresponden a las fases 1 o 2 respectivamente. Las Cys (C₁₋₄) están subrayadas de color amarillo y el sitio de corte de furina está representado con una flecha vertical roja. En las secuencias donde se llega a producir el corte por parte de la furina, se ha representado una barra vertical de color rojo.

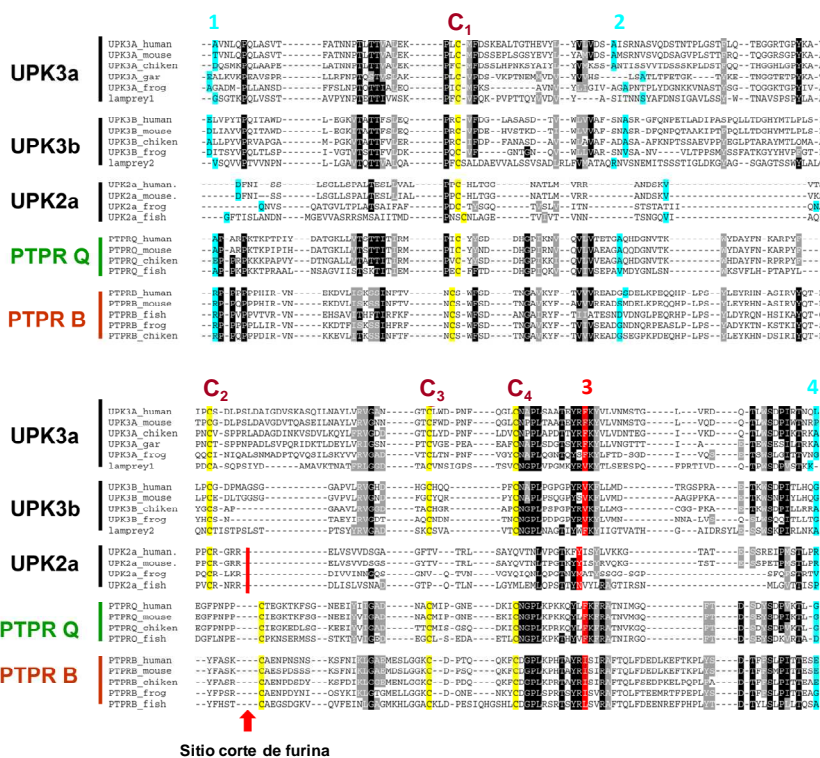
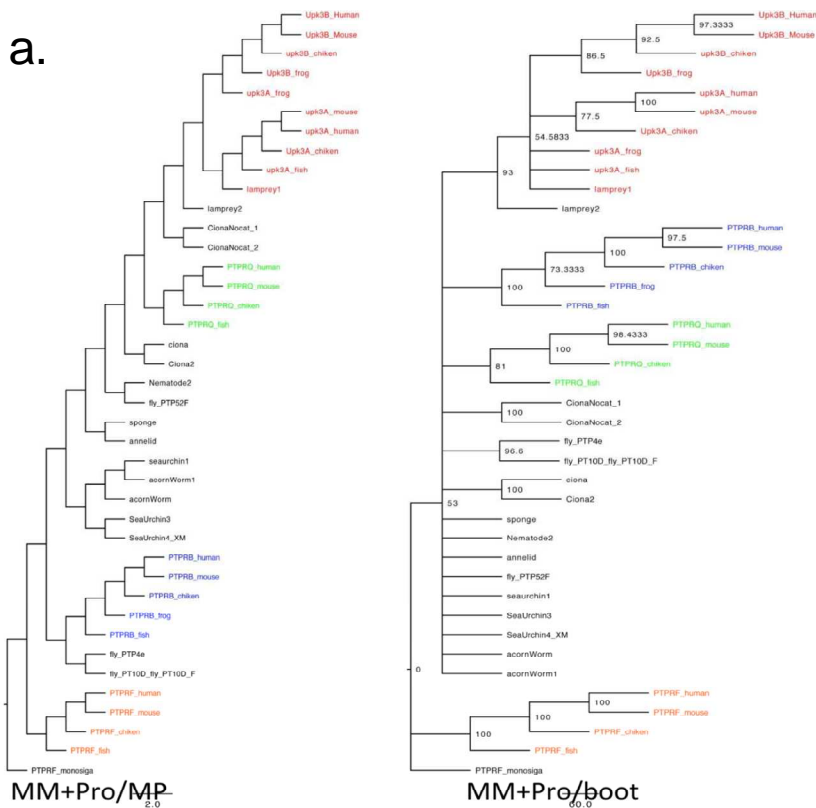
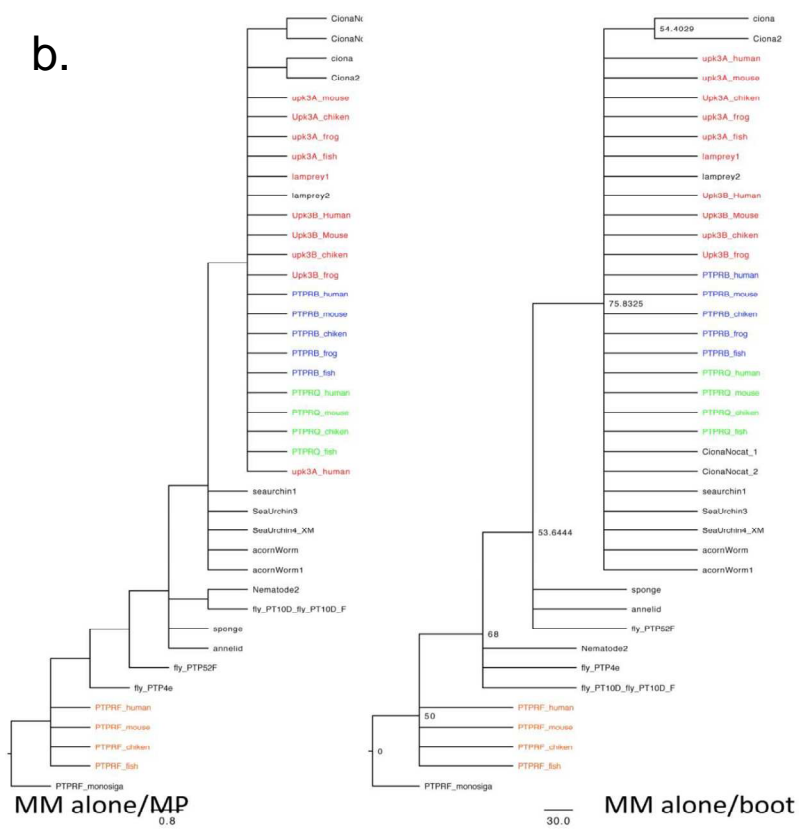
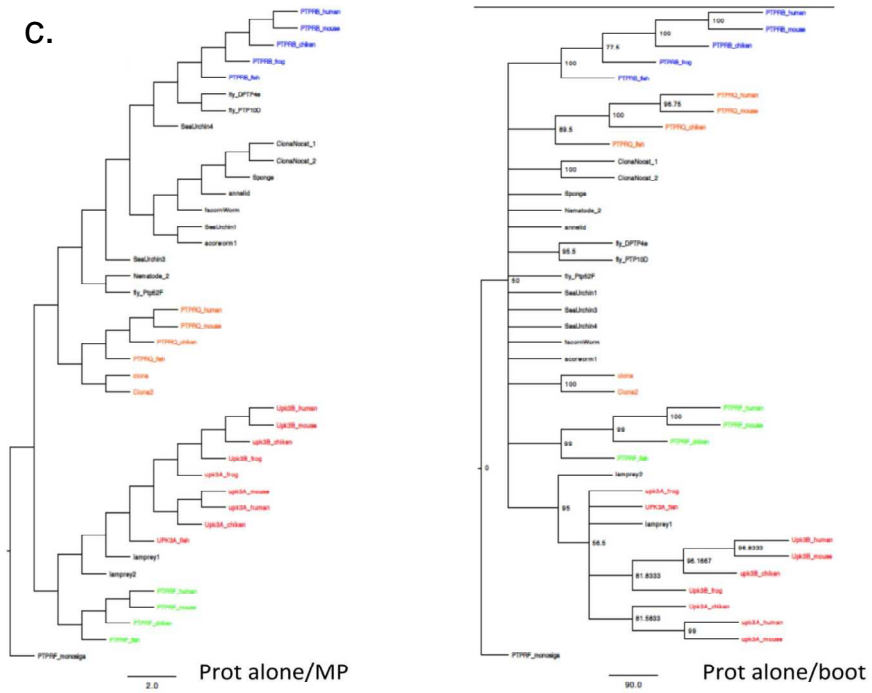


Figura Suplementaria 3: Árboles filogenéticos de la parte de la secuencia que alberga las regiones similares (dominio urofibrina y DTM) de las secuencias proteicas de PTPRS R3 y UPK2/3s usando análisis de parsimonia y Bayesian. (a.) Árboles generados a partir de la protein y de la matriz MM utilizando máxima parsimonia. El árbol de la izquierda es el árbol de máxima parsimonia y el de la derecha es el análisis de bootstrap utilizando máxima parsimonia. (b.) Árboles generados solamente a partir de la matriz de MM. El árbol de la izquierda es el árbol de máxima parsimonia y el de la derecha es el análisis de bootstrap utilizando máxima parsimonia. (c.) Árboles generados solamente a partir de la proteína. El árbol de la izquierda es el árbol de máxima parsimonia y el de la derecha es el análisis de bootstrap utilizando máxima parsimonia. (d.) Árboles generados utilizando inferencia Bayesiana. El árbol de la izquierda es solamente la inferencia Bayesiana de la proteína y el de la derecha es el árbol de la inferencia Bayesiana de la proteína y de MM.







d.

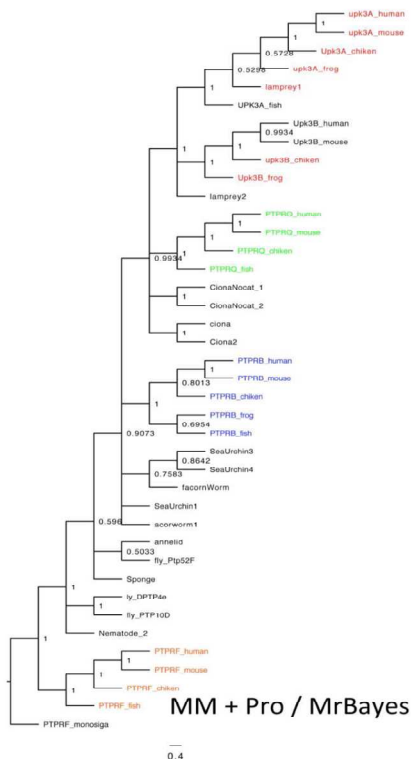
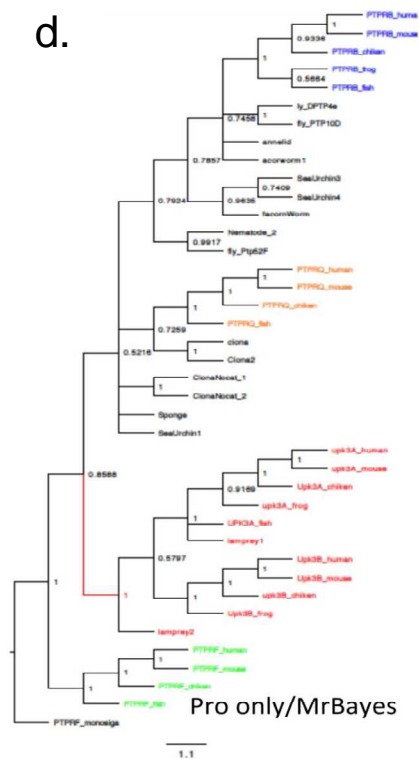



Figura Suplementaria 4: Conservación de los extremos DS2 y 7p13 en los grandes primates. Alineamiento generado con el servidor MAFFT y los parámetros establecidos de serie de las zonas de unión de DS2 y 7p13 donde se observa su conservación. Las secuencias resaltadas de color rojo y verde corresponden al fragmento A y B de RASA 4 respectivamente. Las letras resaltadas de azul clarito reflejan las 13 pb añadidas, las de color rosa la zona de microhomología y la flecha roja la zona donde se produjo la rotura. La identidad viene reflejada en la última línea como idéntico (*), similar (.) o sin identidad (.).

ZONA DE UNIÓN EN EXTREMO 5' ENTRE DS2 Y 7P13

```

Human      tgaagttacaaaagttatactcctatgcaaa-----agaggtactttcaatttttc
Chimpanzee -----agcctgattggttgtggaaagggaccaatcagaggtactttcaatttttc
Gorilla    -----agcctgattggttgtggaaagggaccaatcagaggtactttcaatttttc
Orangutan  -----agcctgattggttgtgggaagggaccaatcagaggtactttcaatttttc
            *. *. . . *.** .**                *****
            

Human      atctaccatgcagaaaaaggt-tgggggtgggggttgcaaaggaagtagccaacaactc
Chimpanzee atctaccatgcagaaaaaggt-tgggggtgggggttgcaaaggaagtagccaacaactc
Gorilla    atctaccatgcagaaaaaggt-tgggggtgggggttgcaaaggaagtagccaacaactc
Orangutan  atctaccatgcagaaaaaggttgggggtgggggttgcaaaggaagtagcc-----
            *****

Human      tgactacaaacagaggtccccagtgggtagggggctgggacaaaagcttttgggggagggg
Chimpanzee taactacaaacagaggtccaaagtgggtagggggctgggacaaaagcttttgggggagggg
Gorilla    tgactacaaacagaggtccccagtgggtagggggctgggacaaaagcttttgggggagggg
Orangutan  -----

Human      aggacagaggacccttggggctgggggatct
Chimpanzee aggacagaggacccttggggctgggggatct
Gorilla    aggacagaggacccttggggctgggggatct
Orangutan  -----
    
```

ZONA DE UNIÓN EN EXTREMO 3' ENTRE DS2 Y 7P13

```

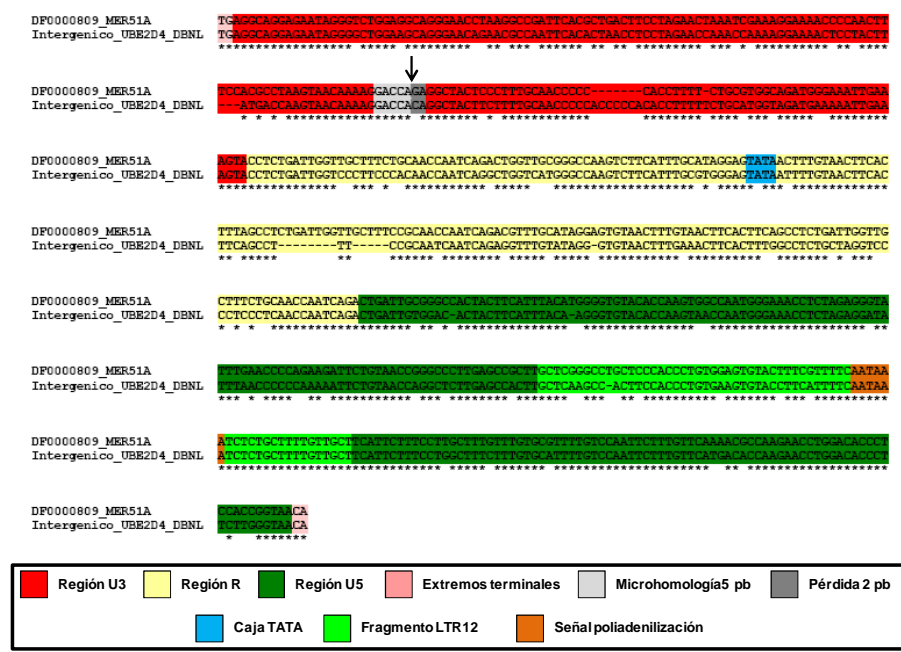
Human      tcatttgcaagatgaggacaatatagcatgcaacctcaaccttcagaaaagactccatgga
Chimpanzee tcatttgcaagatgaggacaatatagcatgcaacctcaaccttcagaaaagactccatgga
Gorilla    tcatttgcaagatgaggacaatatagcatgcaacctcaaccttcagaaaagactccatgga
Orangutan  -----
            

Human      ctttgaggccagagagttgctccggaacaggggtgttggtcttttgtttacttgggcataag
Chimpanzee ctttgaggccagagagttgctccggaacaggggtgttggtcttttgtttacttgggcataag
Gorilla    ctttgagggtcagagagttgctccggaacaggggtgttggtcttttgtttacttgggcataag
Orangutan  -----
            tgtggtcttttgtttacttgggcataag
            *** * .*****

Human      taggcgttttccttttgatttgggttctaggaggttagtgtgcattggcgttctgttgcct
Chimpanzee taggcgttttccttttgatttgggttctaggaggttagtgtgcattggcgttctgttgcct
Gorilla    taggcgttttccttttgatttgggttctaggaggttagtgtgcattggcgttctgttgcct
Orangutan  taggagttttccttttgatttgggttctaggaggttagtgtgaattggcgttctgttgcct
            *** *****

Human      gttccagcccctattctcc
Chimpanzee gttccagcccctattctcc
Gorilla    gttccagcccctattctcc
Orangutan  gttccagcccctattctcc
            *****
    
```


Figura Suplementaria 6: La DS2 se insertó en la región U3 de un MER51A. Alineamiento del MER51A del locus genético 7p13 de orangután y de la secuencia consenso de MER51A DF000809. Las letras resaltadas de rojo, amarillo y verde reflejan las regiones U3, R y U5 del MER51A respectivamente. Las letras resaltadas de rosa, extremos terminales, las resaltadas de color azul la caja TATA, las resaltadas de verde claro el fragmento LTR12 típico del MER51A y las naranjas la señal de poliadenilización. La flecha indica el lugar donde es insertado el IC y las letras subrayadas de gris oscuro, los 2 pb perdidos durante la inserción y las de gris clarito las de la microhomología.



UNIVERSITAT ROVIRA I VIRGILI

EVOLUTION AND MOLECULAR CHARACTERIZATION OF UROPLAKIN 3C, A NOVEL HUMAN TETRASPANIN ASSOCIATED UROPLAKIN

Javier Ugarte Chicote

UNIVERSITAT ROVIRA I VIRGILI

EVOLUTION AND MOLECULAR CHARACTERIZATION OF UROPLAKIN 3C, A NOVEL HUMAN TETRASPANIN ASSOCIATED UROPLAKIN

Javier Ugarte Chicote