



**Universidad Andrés Bello**  
Facultad de Ciencias Biológicas  
Escuela de Ingeniería en Biotecnología

# **“Caracterización del transcriptoma del erizo chileno (*Loxechinus albus*) utilizando secuenciación masiva”**

Tesis presentada como parte de los requisitos para optar al Grado de **Magíster en Biotecnología**.

**Director de Tesis: Dr. Juan Antonio Valdés**

Facultad de Ciencias Biológicas

Universidad Andrés Bello

**Co-Director de tesis: Dr. Claudio Meneses**

Facultad de Ciencias Biológicas

Universidad Andrés Bello

**Paulette Valesca Antiqueo Aravena**

**Santiago, Chile**

**Agosto, 2017**



**Universidad  
Andrés Bello**

**UNIVERSIDAD ANDRÉS BELLO**

Facultad de Ciencias Biológicas

Ingeniería en Biotecnología

**“Caracterización del transcriptoma del erizo chileno (*Loxechinus albus*)  
utilizando secuenciación masiva”**

**Paulette Valesca Antiqueo Aravena**

Este trabajo fue elaborado bajo la supervisión de los Directores de Tesis Dr. Juan Antonio Valdés y el Dr. Claudio Meneses, en el laboratorio de Bioquímica Celular, Universidad Andrés Bello, aprobado por los miembros de la Comisión de evaluación.

---

**Dr. Juan Antonio Valdés.**

Director de Tesis.

---

**Dr. Claudio Meneses.**

Co-Director de Tesis.

---

**Dr. Eduardo Castro.**

Comisión de Tesis.

---

**Dr. Rodolfo Farlora.**

Comisión de Tesis.

**SANTIAGO-CHILE  
Año 2017**

## Índice

1. Resumen .....	5
2. Introducción .....	7
2.1 Erizo chileno como recurso económico .....	7
2.2 Características de <i>Loxechinus albus</i> .....	9
2.3 Reproducción y crecimiento de <i>Loxechinus albus</i> .....	10
2.4 Importancia investigativa de los equinodermos. ....	12
2.5 Mitogenoma de <i>Loxechinus albus</i> y su posición filogenética.....	13
2.6 RNA-seq como herramienta para el análisis de la expresión génica.....	13
2.7 Hipótesis.....	15
2.8 Objetivos.....	15
2.8.1 Objetivo General. ....	15
2.8.2 Objetivos específicos.....	15
3. Materiales y métodos .....	16
3.1 Animales y toma de muestras.....	16
3.2 Extracción de RNA. ....	16
3.3 Construcción de librerías de cDNA y secuenciación. ....	17
3.4 Filtro de secuencias, ensamble de novo. ....	17
3.5 Filtro de contigs y validación del transcriptoma. ....	17
3.6 Anotación funcional y análisis ontológico. ....	18
3.7 Identificación de unigenes y redes de similitud de secuencia.....	18
3.8 Análisis de expresión diferencial y ontología génica tejido específico. ....	19
3.9 Validación expresión diferencial por qPCR.....	19
4. Resultados .....	22
4.1 Secuenciación, ensamble y filtro del transcriptoma de <i>L. albus</i> . ....	22
4.2 Validación y anotación funcional del transcriptoma total.....	25

4.3 Unigenes y anotación con redes de similitud de secuencia (RSS).....	29
4.4 Expresión diferencial.....	31
4.5 Enriquecimiento ontológico y rutas. ....	34
4.6 Validación transcritos tejido específico por qPCR .....	45
5. Discusión.....	48
6. Conclusión y proyecciones.....	56
7. Referencias .....	57
8. Anexo .....	66

## 1. Resumen

Chile es uno de los principales exportadores de gónadas de erizo en el mundo. Sin embargo, este recurso ha sido sobreexplotado en los últimos años producto del incremento en sus tasas de captura. El erizo chileno (*Loxechinus albus*), se distribuye a lo largo del litoral de todo el país, posee un lento crecimiento y se alimenta principalmente de algas. Este organismo no posee dimorfismo sexual externo y su forma de reproducción es exclusivamente sexual a través de fecundación externa. El erizo chileno tiene su mitogenoma secuenciado, con lo cual se estableció la relación filogenética con otros equinodermos, grupo que es ampliamente estudiado debido a su relación filogenética con los cordados. Complementariamente se encuentra disponible el genoma secuenciado del erizo modelo *Strongylocentrotus purpuratus*, en donde se puede observar la gran similitud de genes con los vertebrados, destacando un sofisticado sistema inmune innato, siendo los celomocitos las células con capacidades fagocíticas que comandan su respuesta inmune. La información génica y genómica disponible sobre *L. albus* es muy escasa, encontrándose solo las secuencias pertenecientes a los estudios del mitogenoma y microsatélites. Con el objetivo de “Caracterizar el transcriptoma de celomocitos, intestino y gónada del erizo chileno”, se realizó la secuenciación del ARN de estos tres tejidos. Las lecturas resultantes fueron filtradas con Flexbar y ensambladas con Trinity, encontrando alrededor de ciento ochenta mil contigs, de los cuales fueron anotados el 31% de estos con la herramienta Blast. Adicionalmente, se realizó una segunda anotación utilizando redes de similitud de secuencia, logrando anotar un 75% de los unigenes. Con los resultados obtenidos a partir de la anotación y usando la herramienta DAVID, fue posible determinar los términos ontológicos asociados a cada uno de los tejidos, destacando en celomocitos e intestino su rol en el sistema inmune del erizo. Por otra parte, el intestino cumpliría un rol clave en la homeostasis general de este organismo, y se encontró que las gónadas serían un órgano con un comportamiento dual, relacionado a reproducción y almacenamiento de nutrientes. Este trabajo sirve de base para la realización de futuros estudios en esta especie, no solo para conocer la biología básica de este organismo, sino que también asociado a la producción de este recurso pesquero.

## Abstract

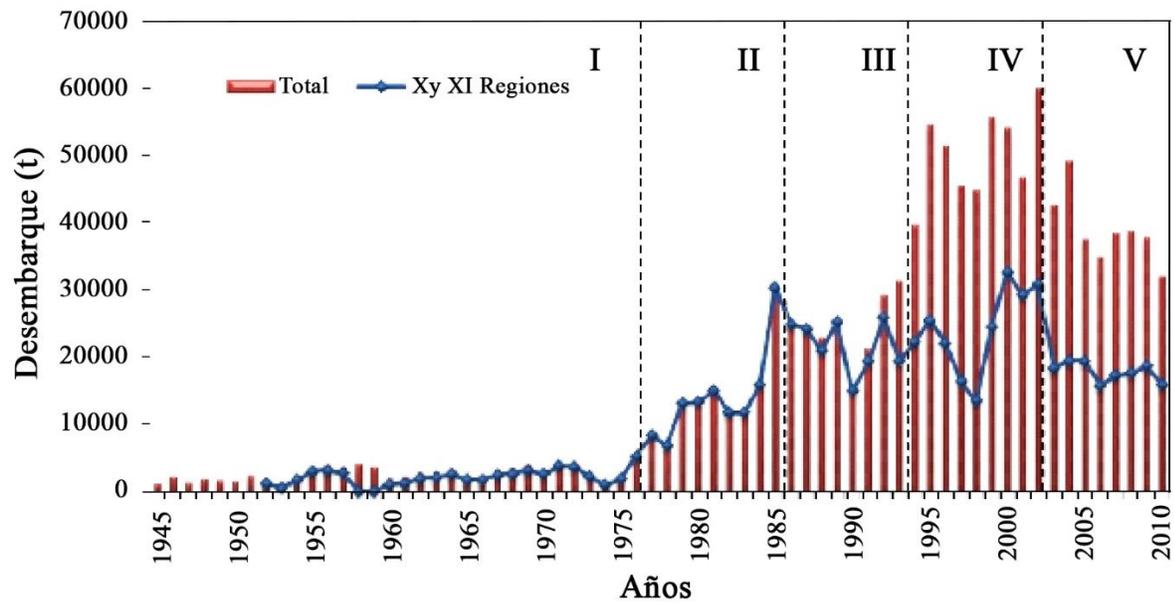
Chile is the most important producer of raw sea urchin in the world. However, this resource has been over harvested in recent years depleting their natural populations. Red sea urchin (*Loxechinus albus*) distributed along the Chilean coast, has a slow growth and feeds algae. Sea urchin sexual dimorphism is externally indistinguishable and has sexual reproduction through external fertilization. The mitogenome of Chilean sea urchin is sequenced, which established the phylogenetic relationship with other echinoderms. Echinoderms are widely studied due to its phylogenetic relationship with the chordates. In addition, the sequenced genome of the model sea urchin *Strongylocentrotus purpuratus* is available, showing a great genes similarity with vertebrates, standing out a sophisticated innate immune system, being coelomocytes cells with phagocytic capacities command immune response. The *L. albus* genomic information available is limited, consists in a complete mitogenome and microsatellites studies. The objective of study is "Characterize the transcriptome of coelomocytes, intestine and gonad of Chilean sea urchin", the RNA sequencing of three tissues was performed. Resulting reads were filtered with flexbar and assembled with Trinity, finding 185.239 contigs, of which 31% were annotated with Blast tool. In addition, a second annotation was made using sequence similarity networks, 75% of unigenes were identified. Results obtained from the annotation alongside with DAVID tool determined ontological terms associated to each tissue, highlighting in coelomocytes and intestine a role in the immune system. Intestine would play a key role in the homeostasis; also, we found that gonads could have a dual role in reproduction and storage of nutrients. This work provides a resource for future studies, either to increase its commercial value or study basic biology in this organism.

## 2. Introducción

### 2.1 Erizo chileno como recurso económico

El mercado chileno ha tenido históricamente grandes hitos en relación con la extracción del erizo (*Loxechinus albus*) a nivel mundial, siendo el mayor generador de gónadas de erizo por mucho tiempo, alcanzando en 1995 su primer punto más alto de producción con 54.609 toneladas del producto pesquero, lo que representó el 45% de la producción mundial (Andrew *et al.* 2002). Los mayores aportes en la captura del erizo provienen de las regiones X y XII, las cuales aportan respectivamente un 33% y 19% en la captura anual total del recurso (SERNAPESCA, 2016). En 1986 la pesquería del erizo alcanzó el “régimen de plena explotación”, lo que quiere decir que la cantidad de recurso extraído es mayor o igual a los excedentes productivos del recurso. Este periodo se ha caracterizado por la sobreexplotación de las áreas ubicadas en la X región y una disminución drástica del rendimiento de la XI región, lo que se traduce en una disminución cercana a las 22.300 toneladas al año (Figura 1) (Molinet *et al.* 2005).

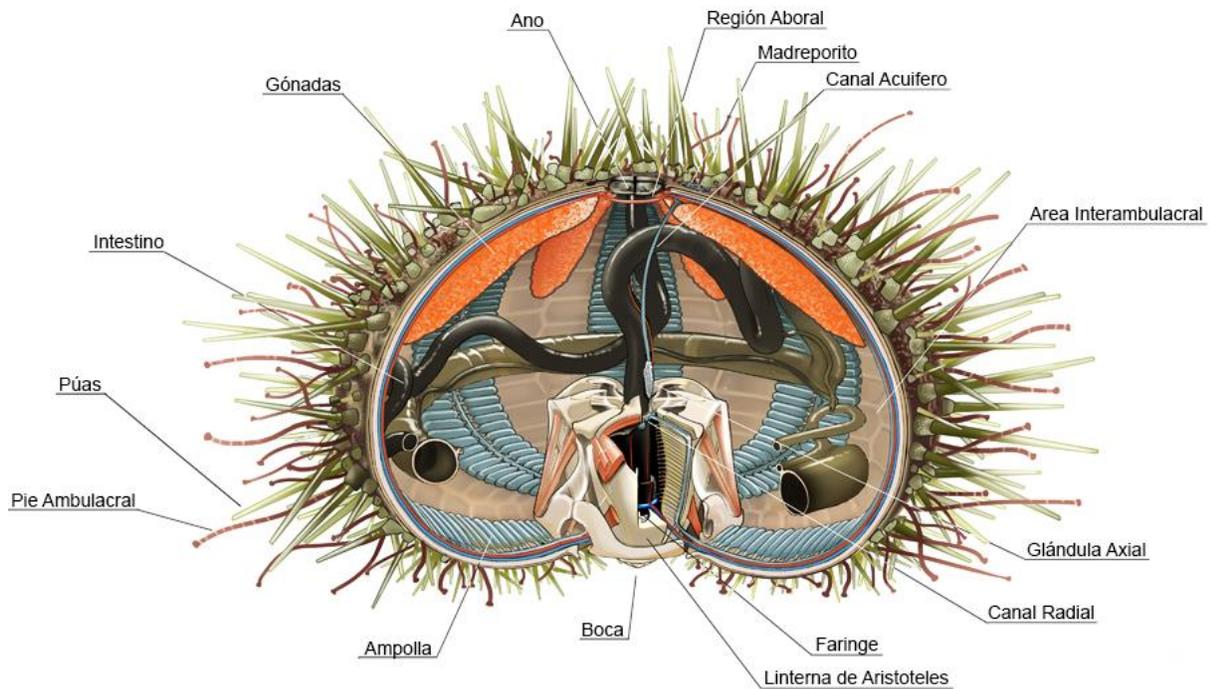
Debido al alto valor económico del recurso, se ha motivado el desarrollo de la investigación en esta especie. Con el objetivo de hacer sustentable la producción del erizo en las áreas de producción, la autoridad pesquera comenzó en 2001 el informe “Bases biológicas para rotación de áreas en el recurso erizo” (Barahona *et al.* 2003). Este documento recopila datos tanto de investigaciones nacionales como extranjeras, así como datos empíricos consultando a pescadores e industria. Según este estudio, existen a lo menos cuatro factores que estarían relacionados con la calidad del recurso siendo el más relevante el alimento que disponen los erizos para su nutrición (Barahona *et al.* 2003; Bustos y Olave, 2001; Kino y Agatsuma, 2007). Este parámetro es también mencionado en otros estudios en relación al erizo utilizado como organismo modelo *S. purpuratus* (Ebert, 1958). La calidad del producto (gónada) se califica según el tamaño, talla del ejemplar, color y sabor (Barahona *et al.* 2003).



**Figura 1. Desembarque nacional del recurso erizo entre 1945 y 2010 en toneladas.** En el eje de las ordenadas se indican los intervalos de años desde 1945 al 2010, en las barras rojas se indica el total del desembarque de erizo en toneladas. Los puntos y líneas continuas azules muestran el desembarque de las regiones X y XI y en las líneas verticales punteadas se delimitan las cinco etapas de evolución extractiva del recurso. Etapa I: inicial; Etapa II: crecimiento; Etapa III: régimen de plena explotación; Etapa IV: crecimiento sostenido; Etapa V: decrecimiento (IFOP, 2010).

## 2.2 Características de *Loxechinus albus*.

*L. albus* es detritívoro en etapas larvales y herbívoro en etapas juveniles y adultas en donde se alimenta principalmente por ramoneo de algas. No posee dimorfismo sexual externo (Contreras y Castilla, 1987). Se encuentra distribuido a lo largo de la costa Chilena, desde Arica (18°S) hasta tierra del Fuego (55°S), en la región de Magallanes (XII Región) (Oyarzún *et al.* 1999), y tiene una distribución batimétrica hasta los 106 metros de profundidad (Moreno y Molinet, 2013). Posee un caparazón semiesférico de un color verde-rojizo rodeado de púas de distinto tamaño. Este caparazón está formado por diversas placas cohesionadas entre sí. En términos anatómicos, el erizo se puede dividir en dos regiones, una inferior llamada región oral y otra superior o región aboral. En la región oral se encuentra la boca del erizo, en la cual se encuentra la Linterna de Aristóteles, desde la cual nace el tubo digestivo. En la región aboral se encuentran internamente las gónadas y por el exterior el ano, el cual se encuentra rodeado por cinco placas perforadas, cuatro de ellas son placas genitales y una placa llamada placa madreporita, la cual es un poco más grande que el resto, y es en donde se regula el intercambio de líquido (Figura 2).



**Figura 2. Anatomía interna general de un erizo de mar.** Imagen representativa de la anatomía interna general de un erizo de mar (Echinodermata: Echinozoa) (IFOP, 1998).

### **2.3 Reproducción y crecimiento de *Loxechinus albus*.**

Su reproducción en las costas chilenas ocurre una vez al año. Los meses de desove son distintos dependiendo de la localización, ya que a medida que se va avanzando hacia el sur, el desove es cada vez más tardío, no así en las costas de Punta Arenas en donde las corrientes no son las mismas que se encuentran en el resto del país, por lo que el comportamiento de desfase latitudinal no se observa (Zamora y Stotz, 1992). Según los datos recopilados por Barahona (2003), las épocas de desove se inician en los meses de julio - agosto para las regiones del extremo norte; mayo – junio para las regiones centro y centro norte; diciembre-enero para las regiones centro sur; octubre para las regiones del sur y septiembre para las regiones del extremo sur.

El desove y la fecundación ocurren en el mar (fecundación externa) seguido de esto comienza la segmentación del huevo. En un período de 24 horas este pasa de estados sucesivos de división a gástrula, dentro de las 48 horas siguientes se convierte en prisma y pasados 3 a 4 días de ese proceso se origina el estado larval, el cual se caracteriza por la formación y elongación del esqueleto de la larva, formación de órganos de locomoción, formación y desarrollo del celoma, el cual está ligado a la formación del sistema acuífero. Finalmente, entre los días 20 y 24 se produce la metamorfosis de la larva dando origen a un individuo juvenil el cual se adhiere al sustrato (Arrau, 1958; Bustos y Olave, 2001). El erizo chileno alcanza su primera madurez al alcanzar 40 a 50 mm de diámetro en bancos naturales (Figura 3; IFOP, 1998) y puede vivir hasta los 20 años. Tiene un crecimiento lento alcanzando un tamaño de 130 milímetros de diámetro como talla máxima (Andrew *et al.* 2002).

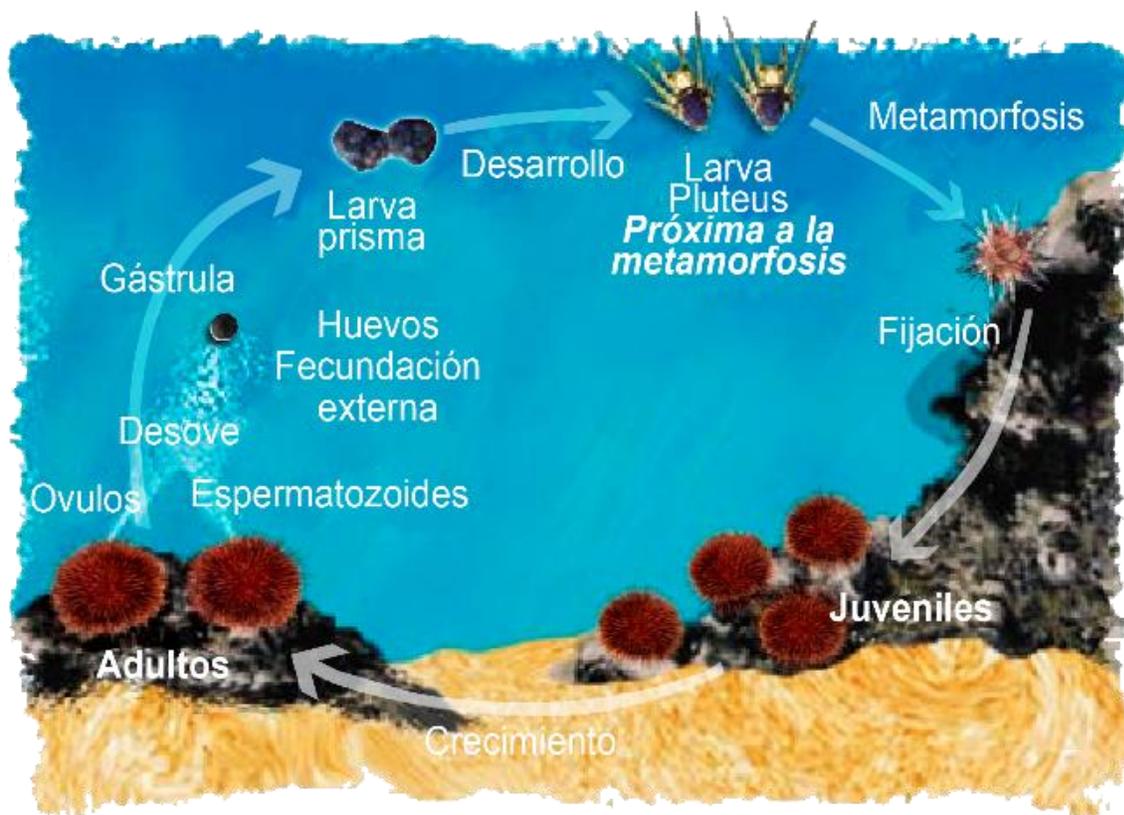
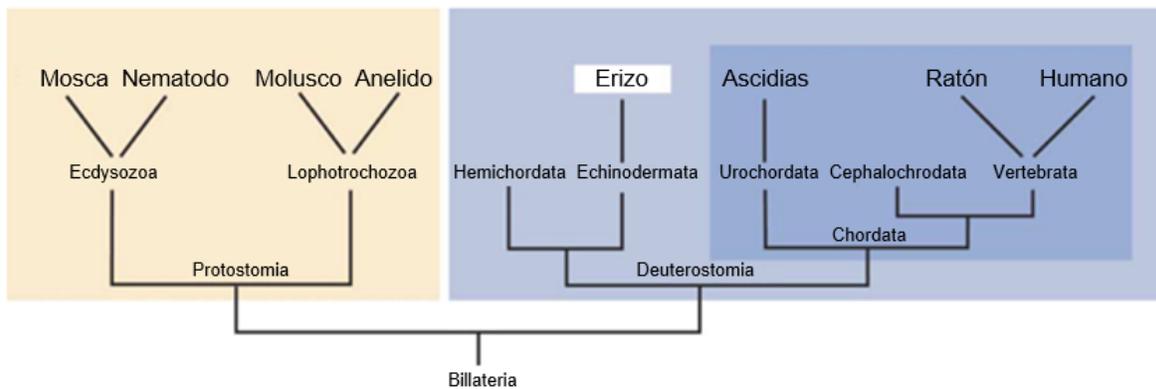


Figura 3. Ciclo de vida de *Loxechinus albus* (IFOP, 1998, Modificado de Bustos y Olave, 2001).

## 2.4 Importancia investigativa de los equinodermos.

El erizo pertenece al filo de los equinodermos (Echinodermata), los cuales han sido ampliamente estudiados debido a su estrecha y cercana posición filogenética con los vertebrados (Figura 4) ( Du *et al.* 2012; Mashanov *et al.* 2014). Presentan una simetría penta-radial y tienen capacidad regenerativa. Uno de los equinodermos más estudiados es el organismo modelo *Strongylocentrotus purpuratus* (Wygoda *et al.* 2014), especie que ya tiene su genoma secuenciado con un tamaño estimado de 814 Mbp y alrededor de 23.300 genes anotados (número similar a los estimados en otros vertebrados). (Sodergren *et al.* 2006). Estos genes son específicos de vertebrados y otros solo han sido encontrados en los erizos, lo que sugiere que fueron perdidos en los vertebrados durante la evolución. También, se menciona que tienen un diverso y sofisticado sistema inmune innato mediado por un gran número de proteínas que reconocen patógenos. El tejido encargado de la respuesta inmune son los celomocitos, los cuales son células con capacidades fagocitarias (Johnson, 1969). Se encontraron ortólogos génicos asociados con visión, audición, equilibrio y quimiotaxis en vertebrados, lo que sugiere capacidades sensoriales (Sodergren *et al.* 2006).



**Figura 4. Posición filogenética de los erizos, en relación a otros modelos y humanos (Sodergren *et al.* 2006).**

## **2.5 Mitogenoma de *Loxechinus albus* y su posición filogenética.**

La información génica encontrada en las bases de datos en relación con *L. albus* es muy escasa, siendo en su mayoría representada en dos estudios. El primero se centra en la búsqueda de microsatélites por medio de la secuenciación masiva del ADN, describiéndose 17 nuevos microsatélites, los cuales son útiles para evaluar los niveles de diversidad genética entre poblaciones (Arevalo *et al.* 2014). Cea *et al.* (2015) realizó la secuenciación del mtGenoma (Código NCBI: JX888466), describiendo 37 secuencias codificantes de las cuales 13 son genes codificantes para proteínas, 2 son rRNA y 22 tRNA. Estas secuencias se encuentran disponibles tanto en NCBI como en Uniprot, estando anotadas, pero no revisadas. Este mismo artículo estableció que el genoma mitocondrial tiene un tamaño total de 15.737 pb, y su comparación con todos los mitogenomas disponibles para otras especies de erizo, logró establecer de mejor forma la relación filogenética entre especies. Se reafirmó la posición filogenética del erizo *L. albus*, encontrando que tanto la estructura, el tamaño y la composición de los genes del genoma mitocondrial son altamente conservados, y se estableció que el erizo chileno (*L. albus*) está estrechamente relacionado con el erizo común *Paracentrotus lividus*, el cual está distribuido en las aguas del Mediterráneo y noreste del Atlántico (Rahman *et al.* 2014).

## **2.6 RNA-seq como herramienta para el análisis de la expresión génica.**

El análisis de la expresión génica es ampliamente usado para identificar los mecanismos que controlan procesos celulares o fisiológicos en plantas, animales y microorganismos (Van Verk *et al.* 2013).

El avance de la tecnología ha llevado al desarrollo de sistemas de alto rendimiento o tecnologías de Secuenciación de la Nueva Generación (NGS; del inglés “Next Generation Sequencing”), como por ejemplo la piro-secuenciación en la plataforma 454 GS FLX de Roche o la secuenciación por síntesis Miseq/Hiseq en Illumina (Qian *et al.* 2014).

La transcriptómica funcional ha progresado vertiginosamente gracias a las herramientas de microarreglos y la secuenciación masiva del RNA (RNA-seq). Los microarreglos tienen la desventaja de necesitar información previa del genoma y que se pueden observar una cantidad limitada de genes. En contraste con las tecnologías NGS, las cuales permiten realizar estudios transcriptómicos en especies donde no existe información genómica y para cualquier modelo, lo

que permite identificar nuevos genes o variantes de empalme de exones (Mutz *et al.* 2013). El hecho de que el RNA-seq permite conocer la complejidad total de organismos poco estudiados sin tener información previa (McGettigan 2013; Van Verk *et al.* 2013), ha permitido en los últimos años, la obtención de información muy importante en términos de anotación, determinación de la estructura de los genes, y análisis del total del transcriptoma en distintos procesos biológicos (Qian *et al.* 2014).

El RNA-seq también ha sido utilizado para realizar estudios en equinodermos (Evans *et al.* 2013) (Dong *et al.* 2014; Mashanov *et al.* 2014; Purushothaman *et al.* 2014; Richardson y Sherman, 2015). Ejemplo de esto es la secuenciación masiva del RNA de tejidos como gónadas, celomocitos, intestino y músculo del erizo endémico de Nueva Zelanda *Evechinus chloroticus*, la cual arrojó un total de 75.002 transcritos, de los cuales 24.655 fueron funcionalmente anotados (32%) (Gillard *et al.* 2014). Un estudio similar se realizó en el erizo *Strongylocentrotus intermedius*, en donde se secuenciaron tejidos de celomocitos, tubo digestivo y membrana peristomial, en el cual se obtuvieron 138.421 contigs de los cuales fueron anotados 69,9% contra la base de datos *non redundant* de NCBI (Chen *et al.* 2015). Estudios más completos se han realizado en *S. purpuratus*, uno de ellos relacionó 22 transcriptomas del erizo en distintos estados (embrión, larva, juvenil y adulto), encontrando 21.092 genes en total (Tu *et al.* 2012). Estos artículos han aportado gran cantidad de información de secuencias, y han relacionado los distintos tejidos a rutas o vías como la vía del complemento y coagulación en celomocitos, vía de señalización Notch en membrana peristomial y vías relacionadas con digestión y absorción en tubo digestivo. Además, numerosos marcadores genéticos que pueden ser utilizados como recurso en acuicultura e investigación han sido identificados (Chen *et al.* 2015; Gillard *et al.* 2014; Tu *et al.* 2012).

Por todos los antecedentes mencionados anteriormente relacionados con la importancia económica del erizo chileno y por ser un interesante modelo de estudio debido a su cercanía a los vertebrados, resulta necesario su estudio para conocer los genes que se transcriben en este organismo. Para esto se secuenciará el mRNA total de tres tejidos (tubo digestivo, gónadas y celomocito), los cuales podrían proporcionar información importante relacionada con procesos diversos como la nutrición, crecimiento, reproducción y respuesta inmune.

## **2.7 Hipótesis.**

Mediante la secuenciación masiva del RNA es posible identificar genes y patrones de expresión asociados a procesos como reproducción, sistema inmune innato y nutrición en tres tejidos de importancia fisiológica en el erizo chileno *Loxechinus albus*.

## **2.8 Objetivos**

### **2.8.1 Objetivo General.**

Caracterizar el transcriptoma de celomocitos, intestino y gónadas del erizo chileno (*Loxechinus albus*).

### **2.8.2 Objetivos específicos.**

1. Secuenciar el transcriptoma del erizo *Loxechinus albus* provenientes desde celomocitos, intestino y gónadas.
2. Generar un transcriptoma de referencia para *L. albus*.
3. Realizar un análisis de transcriptómica comparativa entre tejidos y construir redes de similitud de secuencias.
4. Validar la identificación de genes expresados diferencialmente entre tejidos por qPCR.

### **3. Materiales y métodos**

#### **3.1 Animales y toma de muestras.**

Los erizos (*Loxechinus albus*) fueron obtenidos desde el Centro de Investigación Marina de Quintay (CIMARQ) de la Universidad Andrés Bello, ubicada al sur de la Región de Valparaíso, Chile (33° 13'S, 71° 38'O), los cuales fueron alimentados con algas extraídas directamente de la costa de Quintay hasta el momento de la disección.

Los experimentos se realizaron con tres ejemplares juveniles de *Loxechinus albus* con peso promedio de  $32 \pm 0,36$  g y talla promedio de 5 cm de diámetro. Estos fueron transportados a Santiago en un acuario con flujo de aire constante y agua de mar, y se mantuvieron en una incubadora a 4°C hasta la disección, la que consistió en abrir el erizo con material quirúrgico estéril por medio de un corte latitudinal a lo largo de él, recolectando el contenido acuífero liberado, además de los tejidos correspondientes.

#### **3.2 Extracción de RNA.**

La extracción de RNA total se realizó inmediatamente después a la disección y recolección de muestras, mediante el kit “Mini Kit RNeasy” de Qiagen (Hilden, Alemania) según las instrucciones del fabricante. Para el caso de los celomocitos, previa extracción de RNA, se centrifugó todo el contenido acuífero de los erizos por 5 minutos a 5.000 rpm, con el fin de concentrar todas las células presentes. Se descartó el sobrenadante y el pellet fue re-suspendido en el buffer provisto por el kit de extracción.

Posteriormente el ARN fue cuantificado, utilizando 2 µL de muestra (en duplicado) en el equipo Epoch multi-volumen (BioTek, Winooski, VT, EE.UU.). Mediante el equipo Fragment Analyzer de AATI (Advanced Analytical Technologies, inc., Ankeny, EE.UU.) se midió el RQN (del inglés RNA Quality Number) de cada una de las muestras con el fin de ver la integridad del RNA. Este valor se encuentra en una escala de 0 a 10, en donde 10 es el RNA totalmente integro. Valores sobre 6,5 fueron considerados como RNA con integridad adecuada para la secuenciación.

### **3.3 Construcción de librerías de cDNA y secuenciación.**

Las librerías se generaron a partir de un pool de RNA por tejido de los tres individuos en concentraciones iguales y se utilizó para su construcción el kit “TruSeq RNA Library Preparation Kit v2” (Illumina Inc.). La secuenciación del RNA se realizó por medio de la tecnología Illumina en la plataforma Miseq en pair-end 2x300 pb en el Centro de Biotecnología Vegetal (Facultad de Ciencias Biológicas, Universidad Andrés Bello) en dos carreras de secuenciación (réplicas técnicas).

### **3.4 Filtro de secuencias, ensamble *de novo*.**

La secuenciación en ocasiones entrega errores de lectura por lo que deben ser removidas las que tengan baja calidad. Esto último sucede debido a que la enzima que agrega los nucleótidos pierde precisión en los 100 pb finales del extremo 3'. Para esto, se aplicó un filtro de las secuencias con el programa FlexBar v2.5 (Dodt *et al.* 2012; <http://sourceforge.net/projects/flexbar/>), en este caso, se realizó un filtro de calidad con Q (phred) promedio sobre 38 y un largo de lectura de 250 pb y también se hizo la remoción de adaptadores con las secuencias estándar reportadas por Illumina.

El ensamblaje *de novo* de las lecturas se realizó con el programa Trinity (Haas *et al.* 2014; <https://github.com/trinityrnaseq/trinityrnaseq/>), combinando las lecturas provenientes de las seis librerías, se utilizaron las librerías en formato fastq, en pair-end, además de utilizar una cobertura mínima de kmers = 2.

### **3.5 Filtro de contigs y validación del transcriptoma.**

Se realizó un filtro de contigs de baja cobertura de lecturas con la herramienta Corset (Davidson y Oshlack, 2014; <https://github.com/Oshlack/Corset/>), el cual agrupa contigs ensamblados *de novo* y analiza las cuentas de estos. Para utilizar la herramienta Corset se necesitan preliminarmente las lecturas mapeadas, lo que fue realizado con el software Bowtie2 para todos los tejidos junto con sus réplicas.

Para la validación del transcriptoma se ocuparon los programas CEGMA (Parra *et al.* 2007; [http://korflab.ucdavis.edu/Datasets/genome\\_completeness/](http://korflab.ucdavis.edu/Datasets/genome_completeness/)) y BUSCO (Simão *et al.* 2015;

<http://busco.ezlab.org/>), los cuales comparan las secuencias del transcriptoma con una base de datos de genes ultraconservados en eucariontes, con el fin de validar el ensamble de contigs.

### **3.6 Anotación funcional y análisis ontológico.**

Para la anotación de los contigs se realizó un Blastx contra la base de datos no redundante (nr) de NCBI y uniprot de los erizos *S. purpuratus* y *L. albus* con un e-value bajo 1E-6 y un Blastx contra la base de datos de uniprot de *Homo sapiens* con los mismos parámetros. También se realizaron dos Blastn contra la base de datos de RNA no codificante (<http://www.noncode.org/>) y RNA no codificante del erizo *S. purpuratus*. La identificación de los procesos biológicos, componentes celulares, funciones moleculares y rutas (Kegg) del transcriptoma total se realizó con la herramienta DAVID (<https://david.ncifcrf.gov/tools.jsp>). Para esto se tomó la lista de transcritos anotados con uniprot, con los parámetros requeridos “UNIPROT\_ACCESION” y “Gene List”. Para el análisis y construcción de los gráficos se utilizó la información entregada en: “GOTERM\_BP\_FAT”; “GOTERM\_CC\_FAT”; “GOTERM\_MF\_FAT” y “KEGG\_PATHWAY”.

### **3.7 Identificación de unigenes y redes de similitud de secuencia.**

Utilizando el software CD-HIT-EST (Li y Godzik, 2006) al 90% de similitud se identificaron los unigenes, estos fueron anotados contra las mismas bases de datos mencionadas anteriormente y sumado a eso se realizaron las redes de similitud de secuencia (RSS). Las RSS son una herramienta frecuentemente utilizada para analizar e identificar familias de proteínas (Atkinson *et al.* 2009) en relación a su secuencia. Para la construcción de dichas redes se tomó como entrada una comparación de las secuencias del transcriptoma con Blastn; a partir de esta comparación es posible obtener por medio de un script proporcionado por D. Almonacid un archivo \*.sif y otro \*.eda los cuales entregan la información necesaria para construir y filtrar la red. Para asociar nombres de proteínas es necesario construir una tabla que contenga los atributos necesarios para realizar el análisis, esta tabla fue realizada por medio de la herramienta de Uniprot “Retrieve/ID mapping” con los datos de proteínas provenientes del Blastx. Para la visualización de la red se utilizó el programa Cytoscape 2.8 (<http://www.cytoscape.org>), en el cual es posible filtrar y

adicionar los atributos necesarios para la mejor comprensión de los datos entregados. Para la organización de los datos se utilizó la opción “Organic”.

### **3.8 Análisis de expresión diferencial y ontología génica tejido específico.**

Para las cuentas y análisis de expresión diferencial se utilizó el pipeline Bowtie2/RSEM/edgeR. En primer lugar, se realizó el mapeo de lecturas de cada una de las librerías (tejidos) por separado contra el transcriptoma, utilizando el programa Bowtie2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>), se continuó con la estimación de la abundancia de transcritos de cada una de las librerías con RSEM (<http://deweylab.biostat.wisc.edu/rsem/>). Obtenidos estos datos se realizó la normalización de estos para obtener finalmente las cuentas normalizadas por TMM-FPKM, los cuales sirven de entrada tanto en la identificación de los transcritos diferencialmente expresados con edgeR ([https:// bioconductor.org/packages/release /bioc/html/edgeR.html](https://bioconductor.org/packages/release/bioc/html/edgeR.html)), como para el análisis de componentes principales. Para la expresión diferencial se utilizaron los parámetros de FDR (del inglés “False Discovery Rate”) de  $1e-3$  y FC (del inglés “Fold change” o veces de cambio) de 2. A partir de los datos de expresión diferencial se construyeron los mapas de calor utilizando el paquete de Trinity.

Para el análisis de enriquecimiento ontológico, se seleccionaron los transcritos únicos para cada tejido, esto se realizó seleccionando los transcritos que tenían valores de FPKM  $> 0$ , en uno de los tres tejidos. Por medio de la herramienta DAVID, se identificaron los procesos biológicos, funciones moleculares, componentes celulares y rutas (Kegg) involucrados para cada tejido utilizando como “Gene list” la anotación por Uniprot de los transcritos únicos por tejido, seleccionando la opción “UNIPROT\_ACCESION” y “Gene List”. Para el análisis y construcción de los gráficos se utilizó la información entregada en: “GOTERM\_BP\_FAT”; “GOTERM\_CC\_FAT”; “GOTERM\_MF\_FAT” y “KEGG\_PATHWAY”.

### **3.9 Validación expresión diferencial por qPCR.**

Posterior a la identificación de los transcritos diferencialmente expresados y únicos en cada uno de los tres tejidos se evaluó un sub-set de cuatro transcritos por tejido más un normalizador, el cual se comportaba de manera estable en los tres tejidos. Para esto se hizo la extracción de RNA

de cada uno de los tejidos (tubo digestivo, gónada y celomocitos) para tres erizos distintos con las mismas características que los utilizados en la secuenciación, utilizando el kit “Mini Kit RNeasy” (Qiagen). Luego se realizó la síntesis del cDNA usando el kit “Reverse Transcription kit” (Qiagen) siguiendo las instrucciones del fabricante.

Con la información entregada por el transcriptoma, se diseñaron los partidores (Tabla 1) utilizando los programas Primer 3 (<http://bioinfo.ut.ee/primer3-0.4.0/>) y OligoAnalyzer 3.1 (Integrated DNA technologies; <https://www.idtdna.com/calc/analyzer>). La reacción de qPCR contiene 7,5 µL de Brilliant® II SYBR Green QPCR Master Mix (Agilent Technologies), 6 µL de cDNA (dilución 1/40), 1,5 µL de partidores y agua libre de nucleasas necesaria para llegar a 15 µl. El qPCR se realizó en el equipo Mx3000P de GeneXpress (Life Science Business). El programa de amplificación fue de tres etapas, una inicial de 10 min a 95°C, seguido de una intermedia de 40 ciclos de 30 s. a 95°C, 30 s a T<sub>m</sub> (temperatura de fusión de la pareja de partidores), 32 s a 95°C y una etapa final de 30 s a 72°C. No solo se evaluó el comportamiento de los transcritos, sino que también si estos eran efectivamente tejidos específico, para eso, y sumado a las réplicas del tejido a evaluar, se adicionó un pool de cDNA de los tejidos restantes. Para la normalización de la expresión se usó como referencia el transcrito de 18S.

**Tabla 1. Lista de partidores usados en la validación.** Lista completa de los partidores que se diseñaron para realizar los qPCR.

Contig target	Genes	Secuencia 5'a 3' (F/R)
DN29540_c0_g1_i1	Proteína serina/treonina específica de testículo 3 (TSSK3)	CGGTGGTTGGGTTTCGTA CTACTGGCACATCTTGGGCG
DN17270_c0_g1_i1	Centrina 2 (CETN2)	CAGCGACGGAGGAAGAAAT TGTGCGGGTAAGTGGATGTC
DN14854_c0_g1_i1	Proteína canal de calcio asociada a espermatozoides 3 (CATSPER3)	CCCCAGTGCTCTTTGTCTGA ATCTGAACGCTCTGAGGACG
DN34598_c0_g2_i3	Proteína de superficie del espermatozoide Sp17 (SPA17)	AAGGAGGGGGAGACACTGAC CGGCTACATCGGTCGTA
DN22347_c0_g3_i1	Proteína de choque térmico 70 kDa 1A (HSP70)	CGTTCTCGTCTGTGAATCGT AGGAAAAGCACCCAGCCATC
DN19486_c0_g10_i1	Ubiquitina del residuo A-52 del ribosoma (UBA52)	TTGTTGTTGCTGTTGCTGCT CGTATCCTCGCCCAGAAGTA
DN44721_c0_g2_i4	Regulador del tráfico lisosomal (LYST)	GGGCATCAGCATCTTTGTCT GTGAAGGAGGGAACACTGGA
DN42711_c2_g1_i2	Célula B linfoma 2 (BCL2)	GAGGATGGGTGGGACTTGT CGGTGTTGGGTAAGTGTCTT
DN23637_c0_g1_i2	Proteína homóloga a notch 1 (NOTCH1)	ACTTCCAACCCGATGATGAC ATCCATTACCCCGTCTACCA
DN21465_c0_g1_i1	Receptor tipo toll 3 (TLR3)	ATCTGGAATGGGGATGATGA TTCTGGGGGCACTAAATGAT
DN39186_c0_g1_i2	Caspasa 3 (CASP3)	ATGGGTAGCGTGTGAAGGAC TGTCAGAAGGGTGGGACTCT
DN15322_c0_g1_i1	Glutación S-transferasa theta-1 (GSTT1)	CTCTCCCAGCCTTGTCGT GCAAAAGTCTCCGTCTTGA
Normalizador	18S	GTGGAGCGATTTGTCTGGTAAAG AAGGGCATCACAGACCTGTT

## 4. Resultados

### 4.1 Secuenciación, ensamble y filtro del transcriptoma de *L. albus*.

Para obtener las secuencias provenientes de los tres tejidos, se realizó la extracción de RNA de tres individuos juveniles de *L. albus*, la calidad de estos RNA fue medida por electroforesis capilar obteniendo valores de RQN  $\geq 6.8$  (Tabla A1). A partir de estas muestras, se construyeron las librerías de cDNA las cuales fueron secuenciadas en la plataforma Miseq de Illumina. Se realizaron dos carreras de secuenciación obteniendo alrededor 96 millones de lecturas (Tabla 2). Posterior a los filtros de lecturas por calidad, tamaño y la remoción de adaptadores, se obtuvieron aproximadamente 91 millones de lecturas, es decir, el 95% de las secuencias totales (Tabla 3).

**Tabla 2. Número de lecturas de la secuenciación por librería.** Tabla con el detalle del número de lecturas para cada uno de los tejidos (ambas réplicas) y total antes del filtro.

<b>Librería</b>	<b>Número de lecturas secuenciación</b>	<b>Número de lecturas secuenciación (Replica técnica)</b>	<b>Número de lecturas total</b>
<b>Total</b>	51.323.260	44.422.380	95.745.640
<b>Celomocitos</b>	20.682.190	16.865.448	37.547.638
<b>Intestino</b>	12.145.212	11.348.164	23.493.376
<b>Gónada</b>	18.495.858	16.208.768	34.704.626

**Tabla 3. Número de lecturas por librería posterior al filtro.** Tabla con el detalle del número de lecturas para cada uno de los tejidos (ambas replicas) y total después de aplicar el filtro de lecturas por calidad, largo y remoción de adaptadores.

	Número de lecturas post-filtro	Número de lecturas post-filtro (Replica técnica)	Número de lecturas total post-filtro
<b>Total</b>	49.599.274	41.520.026	91.119.300
<b>Celomocitos</b>	19.948.624	15.699.186	35.647.810
<b>Intestino</b>	11.748.696	10.620.928	22.369.624
<b>Gónada</b>	17.901.954	15.199.912	33.101.866

Para continuar, se realizó el ensamble *de novo* con el software Trinity con la totalidad de las librerías y sus respectivas réplicas técnicas, arrojando **278.803 contigs**, con un largo promedio de 326 pb y un %GC de 38,52%, un N50 de 1.418 pb. En total se ensamblaron 197.480.887 bases (Tabla 4). Posteriormente, se realizó un filtro de contigs de baja cobertura con Corset, el cual realiza un conteo de lecturas mapeadas contra los contigs, eliminando aquellos que están bajo el umbral. Para el filtro de contigs, se necesitó un mapeo preliminar de las lecturas con bowtie2, el que arrojó 87.126.452 lecturas mapeadas, lo que representa un 95,6% de las lecturas totales y un 4,4% de lecturas no mapeadas, es decir, 3.992.848 lecturas. El análisis por Corset arrojó una disminución de contigs con baja cobertura de lecturas, quedando **185.239 contigs** totales, agrupados en 185.226 grupos. Este nuevo transcriptoma tiene un porcentaje de GC de 38,81%, un N50 de 1.719 y un número de bases ensambladas de 172.122.579 bases (Tabla 4).

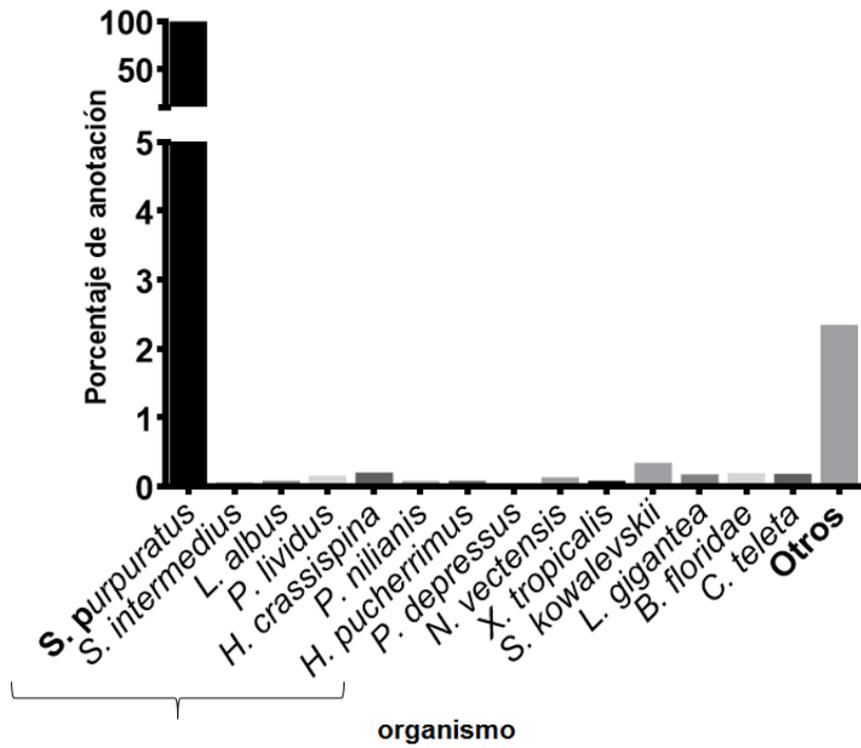
**Tabla 4. Detalle del transcriptoma antes y después del filtro de secuencias poco representadas.** Tabla resumen del ensamble del transcriptoma de novo y el transcriptoma final, posterior al filtro de contigs por Corset.

<b>Transcriptoma</b>	<b>Ensamble <i>de novo</i></b>	<b>Posterior al filtro</b>
<b>Número total contigs</b>	278.803	185.239
<b>Promedio largo de contig (pb)</b>	326	929
<b>Cobertura contig</b>	708,32	-
<b>%GC</b>	38,2	38,81
<b>N10 (pb)</b>	5.015	5.328
<b>N30 (pb)</b>	2.645	2.945
<b>N50 (pb)</b>	1.418	1.769
<b>Número total de bases</b>	197.480.887	172.122.576

## 4.2 Validación y anotación funcional del transcriptoma total.

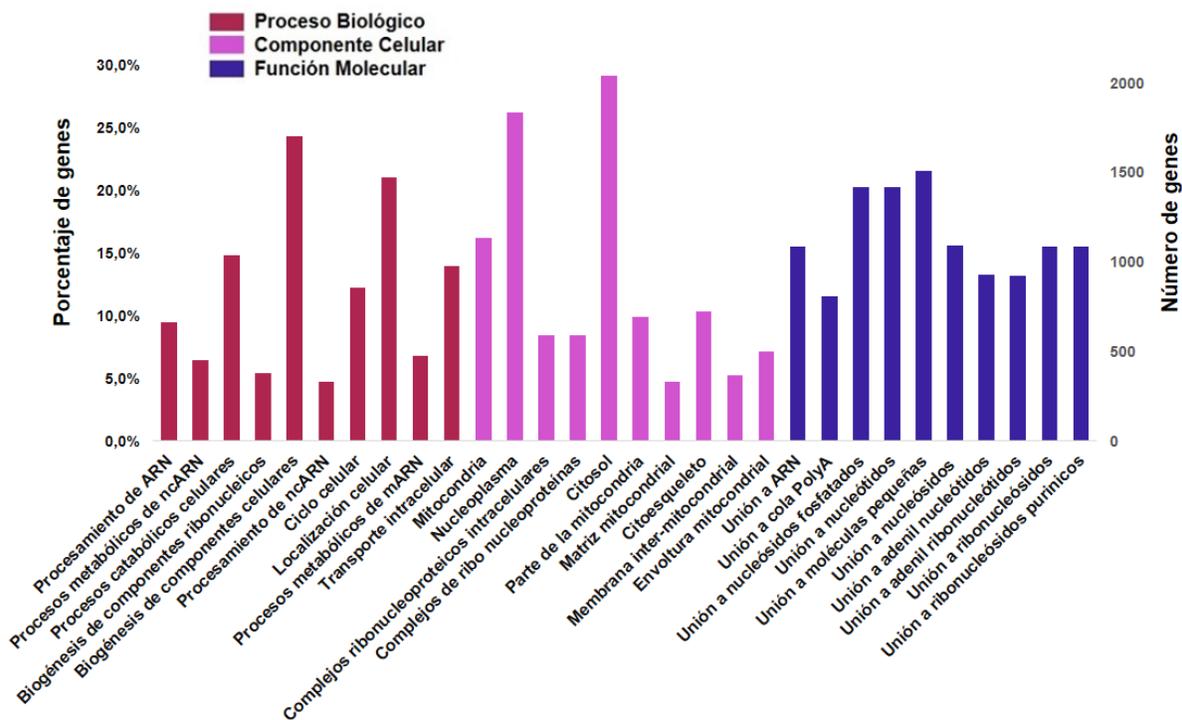
Para la validación del ensamble del transcriptoma, se realizó una búsqueda de secuencias conservadas con las herramientas CEGMA y BUSCO, las cuales permiten obtener un porcentaje del número de secuencias totales encontradas en cada una de las bases de datos que ocupan estas herramientas. Estos porcentajes indican que tan representadas se encuentran estas secuencias ultra-conservadas en nuestro transcriptoma, siendo de utilidad para resolver la calidad del ensamble. Los resultados con CEGMA arrojaron que, de 248 transcritos ultra conservados en eucariontes, 240 transcritos fueron encontrados completos, lo que representa un 96,77%, mientras que 245 transcritos fueron identificados parcialmente, lo que representa un 98,70% de las secuencias totales que tiene la base de datos de CEGMA. Para el caso de BUSCO se encontraron 89,5% de las secuencias completas, 44% para los completos duplicados, 3% para los fragmentos y un 7,5% de las secuencias no se lograron encontrar.

Para continuar con la identificación de los transcritos se realizó un Blast contra las bases de datos de proteínas de NCBI, Uniprot y transcritos no codificantes. Los resultados de la anotación con Blastx contra las bases de datos de proteínas arrojaron una anotación cercana al 31% (57.106 transcritos) con un e-value  $\leq$  a  $1E-6$ , el organismo con mayor porcentaje de anotados fue *S. purpuratus* con un 96,4%, seguido de *Saccoglossus kowalevskii* con un 0,3%, entre otros (Figura 5). La búsqueda de transcritos no codificantes dio un 0,19% (345 transcritos) con la base de datos de *S. purpuratus* y de 0,56% (1.033 transcritos) con la base de datos *noncode*. Con el objetivo de comparar lo conocido en humano con la nueva información del erizo *L. albus*, es que se realizó un Blastx contra la base de datos de Uniprot/*Homo sapiens*, identificando 33.217 transcritos, es decir, un 17,93% de las secuencias totales.



**Figura 5. Anotación contra las bases de datos.** Porcentaje de anotación por organismo, incluyendo lo ya conocido en las bases de datos de *L. albus*. Se señala con una llave las especies de erizos de mar encontradas en la anotación.

Para identificar la ontología génica y rutas representadas en el transcriptoma se asignaron los términos ontológicos a partir de la identificación de ortólogos encontrados anteriormente contra la base de datos de Uniprot/*Homo sapiens* y la herramienta DAVID. Entre los procesos biológicos más representados se encuentran: “Biogénesis de componentes celulares” (22,3%), “Localización celular” (19,3%) y “Procesos celulares catabólicos” (13,6%). En componentes celulares se encuentran: “Citosol” (28,9%), “Nucleoplasma” (26,0%) y “Mitocondria” (16,1%) y en funciones moleculares se encuentran: “Unión a moléculas pequeñas” (21,5%), “Unión a nucleótidos” (20,1%) y “Unión a nucleósidos fosfatados” (20,1%) (Figura 6). En relación con las rutas, el 24,6% de los transcritos encontrados pertenecen a Rutas involucradas en el metabolismo (Tabla 5).



**Figura 6. Anotación de ontología génica (OG) del transcriptoma.** Detalle de los 10 primeros términos OG de cada categoría; Proceso biológico, Componente celular y Función molecular anotados con DAVID para el transcriptoma total, utilizando la base de datos de humano.

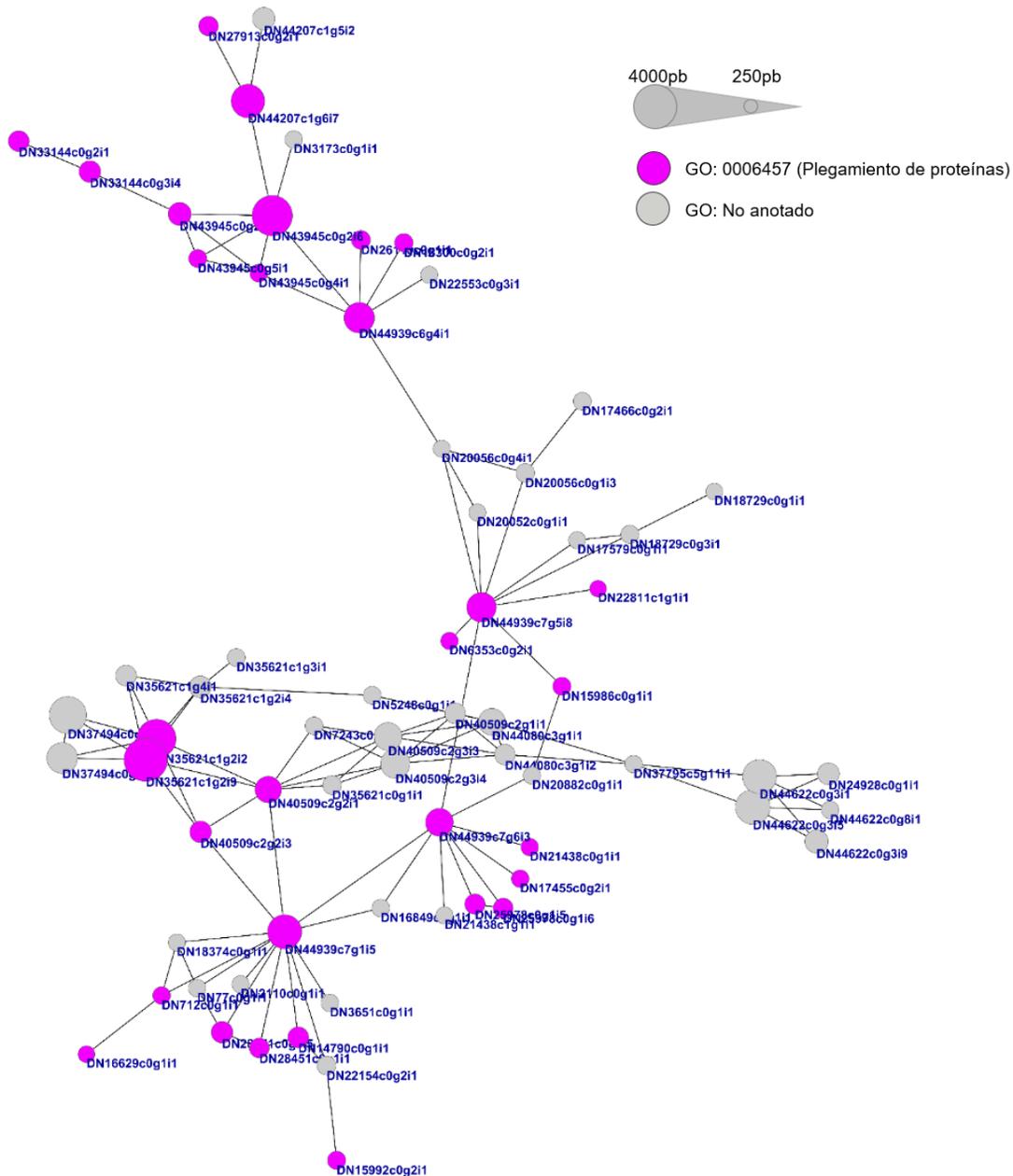
**Tabla 5. Rutas metabólicas.** Rutas metabólicas involucradas en el transcriptoma con el número de transcritos encontrados por vía y el porcentaje del total.

<b>Ruta principal</b>	<b>ID Kegg</b>	<b>Nombre sub ruta</b>	<b>Número de genes</b>	<b>Porcentaje de genes</b>
hsa01100: Rutas metabólicas	230	Metabolismo de purinas	113	3,50%
	1200	Metabolismo del carbono	84	2,60%
	240	Metabolismo de pirimidinas	74	2,30%
	1212	Metabolismo de ácidos grasos	40	1,20%
	520	Metabolismo de aminoazúcares y nucleótidos	36	1,10%
	600	Metabolismo de esfingolipidos	33	1,00%
	480	Metabolismo de glutatión	33	1,00%
	860	Metabolismo de porfirinas	31	1,00%
	380	Metabolismo de triptófano	30	0,90%
	270	Metabolismo de cisteína y metionina	29	0,90%
	260	Metabolismo de glicina, serina y treonina	29	0,90%
	620	Metabolismo de piruvato	26	0,80%
	640	Metabolismo de propanoato	24	0,70%
	250	Metabolismo de alanina, aspartato y glutamato	22	0,70%
	51	Metabolismo de fructosa y manosa	21	0,60%
	53	Metabolismo de ascorbato y aldarato	20	0,60%
	410	Metabolismo de beta-alanina	20	0,60%
	760	Metabolismo de nicotinato y nicotinamida	19	0,60%
1210	Metabolismo de ácido 2-Oxocarboxílico	15	0,5%	

### 4.3 Unigenes y anotación con redes de similitud de secuencia (RSS).

Para conocer el número posible de genes (unigenes) de los cuales provienen estos transcritos es que se realiza una agrupación de contigs por similitud de secuencia. A partir de la agrupación de contigs con CD-HIT-EST al 90%, se identificaron 81.539 unigenes, de los cuales se logró anotar 30.202 contigs (37%) con las bases de datos, nr, *S. purpuratus* y *L. albus* y 17.885 contigs (22%) con la base de datos de humano.

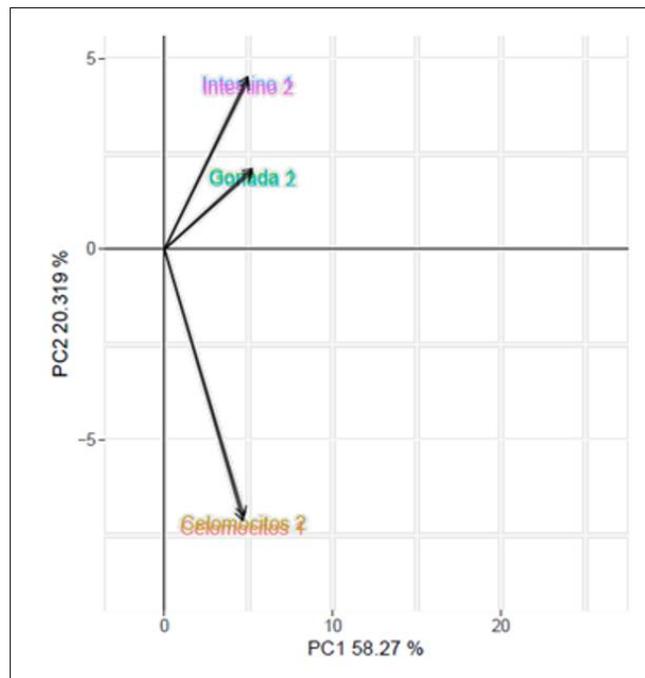
Con el objetivo de aumentar el número de transcritos o contigs identificados, se utilizó una aproximación usada frecuentemente para el estudio y análisis de familias de proteínas. En este caso se usó como entrada una comparación por homología de secuencia utilizando Blastn de todas las secuencias provenientes del transcriptoma de *L. albus*, entregando como parámetro a filtrar el e-value, junto con esto se construyó una tabla con los “atributos” de cada uno de los contigs previamente anotados, en la cual se incluyen en detalle lo conocido en estas secuencias. Debido a la gran cantidad de datos provenientes de la secuenciación, se realizaron tres redes separadas por intervalos de e-value, las cuales contenían 49.067, 36.301 y 28.892 nodos cada una (Figura A1, A2, A3). A partir de este análisis fue posible aislar familias de proteínas presentes en el transcriptoma de *L. albus*, ejemplo de esto es la familia de proteínas a las cual pertenece “Sacsin”, la cual es una co-chaperona de HSP70, en la que por medio de la anotación convencional por Blast fue posible anotar 48 de estas y adicionalmente con RSS fue posible identificar 21 secuencias más (Figura 7 y tabla A2). En conjunto se estimó que con esta herramienta fue posible identificar aproximadamente un 64% de los contigs, por lo que sumado a la anotación contra las bases de datos dio un 75% aproximadamente de unigenes anotados. Un grupo de contigs (22.956 contigs) no fueron incluidos en la red debido a que no agrupó con ninguna otra secuencia, de estos fueron identificados contra las bases de no codificantes el 0,03% (28 contigs).



**Figura 7. Red de similitud de secuencia de la familia “Sacsin”.** Visualización de parte de una red perteneciente a la familia de proteínas Sacsin, con un filtro de e-value entre 0 a 1e-50. Los nodos representan cada secuencia del transcriptoma, en donde el tamaño es proporcional al largo de la secuencia y el color representa el proceso biológico al cual están involucrados, en gris se encuentran las secuencias no anotadas y el largo de las conexiones o “edge” es proporcional al e-value (Ver tabla A2).

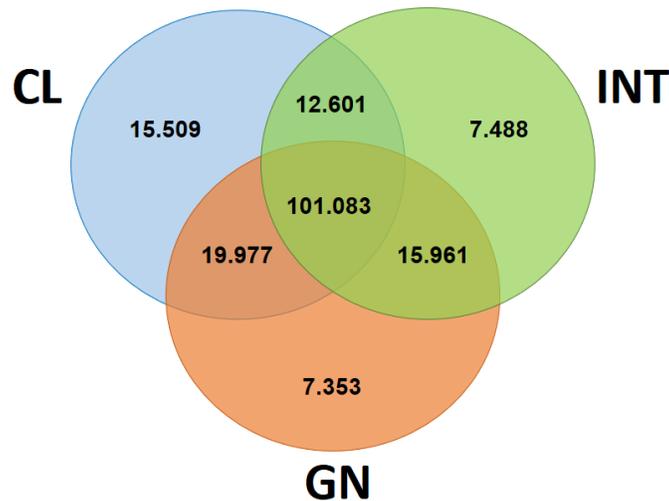
#### 4.4 Expresión diferencial.

Para analizar la expresión diferencial de los datos se realizó un mapeo de las lecturas con Bowtie2 contra el transcriptoma final, es decir, el transcriptoma posterior a la eliminación de contigs con baja cobertura. Este mapeo a diferencia del anterior se realizó con las réplicas por separado, el programa arrojó un total de 87.027.081 lecturas mapeadas, correspondiente al 95% de las lecturas totales. Las cuentas fueron calculadas utilizando RSEM, posterior a esto se realizó la normalización de datos TMM-FPKM, a los cuales se les realiza una normalización adicional para construir el gráfico de componentes principales (PCA), el cual indica la correlación entre los tejidos. Esto es posible de determinar por la lejanía entre los vectores que representan cada una de las librerías. En el análisis de PCA fue posible determinar que gónada e intestino se encuentran más relacionados entre sí que los celomocitos. Adicionalmente, se desprende del gráfico la alta correlación entre las librerías y su réplica técnica, las cuales se encuentra casi sobrepuestas entre sí (Figura 8).



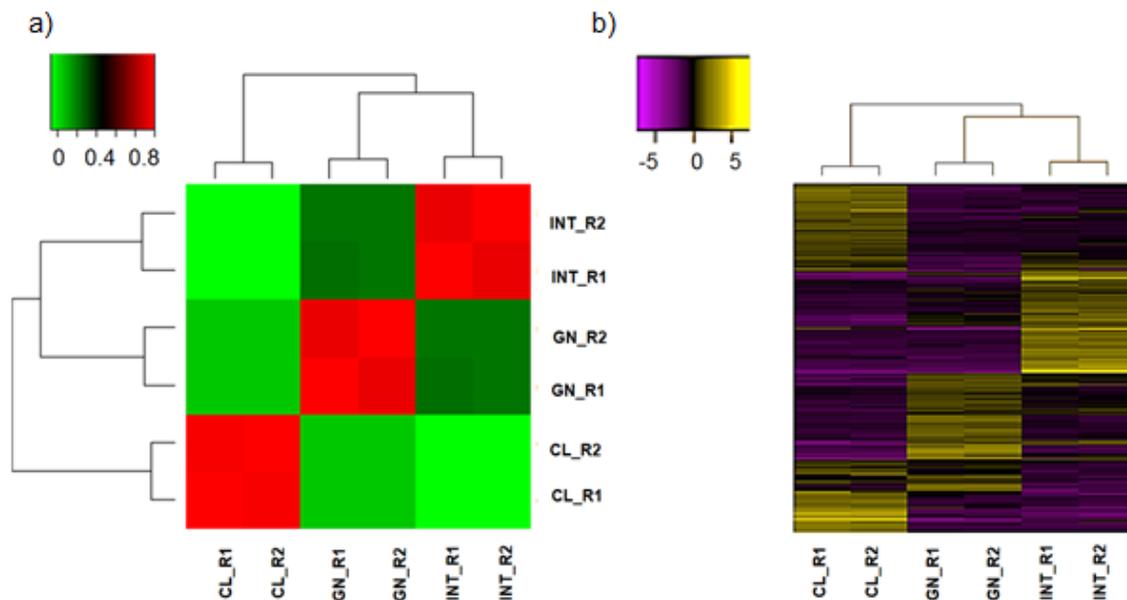
**Figura 8. Análisis de PCA de la expresión de transcritos entre tejidos.** Gráfico de componentes principales que muestra la correlación entre tejidos y su réplica técnica de la expresión de transcritos a partir de la normalización de datos TMM-FPKM.

Para continuar se determinó el número de transcritos compartidos entre tejidos, además de los transcritos tejido específico. Esto se realizó analizando el número de FPKM de cada uno de los transcritos. A partir de esto se obtuvo que 101.083 transcritos estaban presentes en los tres tejidos, sumado a eso 15.961 transcritos compartidos entre gónada e intestino, 19.977 transcritos compartidos entre gónada y celomocito y 12.601 transcritos compartidos entre celomocitos e intestino. Además, se encontraron en total 30.350 transcritos específicos para uno de los tres tejidos (15.509 en celomocitos, 7.488 en intestino y 7.353 en gónada) (Figura 9).



**Figura 9. Número de transcritos por tejido.** Diagrama de Venn con los transcritos expresados únicos y comunes entre los tres tejidos con FPKM > 0. CL: Celomocitos; GN: Gónada; INT: Intestino.

De todos estos transcritos y por medio de edgeR se detectaron que **34.158** transcritos mostraban una expresión diferencial estadísticamente significativa (ED) con un filtro de FDR < 0,001 y un FC > 2, los cuales son mostrados en los mapas de calor de la Figura 10, se puede ver en el primer caso la correlación de los tejidos entre ellos, basado en el comportamiento global de los transcritos ED y en el segundo caso el comportamiento de cada uno de los 34.158 transcritos con ED, en los tres tejidos con su respectiva réplica, aquí se puede observar una leve correlación entre gónada e intestino, la cual es respaldada por lo encontrado con la globalidad de transcritos en el análisis de componentes principales (Ver figura 8).



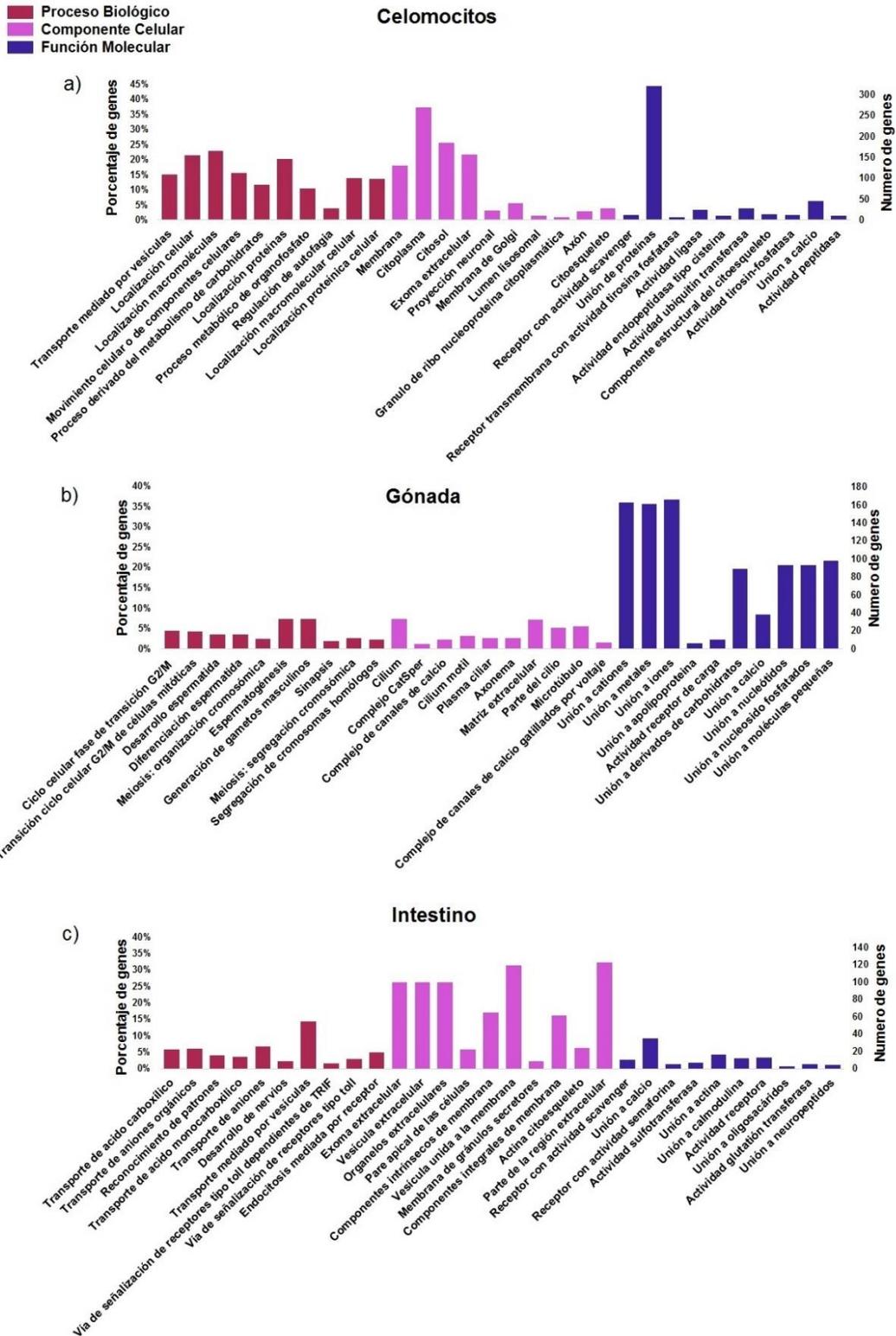
**Figura 10. Comportamiento de los transcritos diferencialmente expresados.** a) Mapa de calor de la correlación entre tejidos. Los colores indican la correlación entre los tejidos con su réplica. En escala de color de verde (no correlacionado) a rojo (altamente correlacionado). b) Mapa de calor basado en los transcritos con expresión diferencial entre los tres tejidos con su réplica. Los colores indican el logFC de los transcritos en escala de color de morado (regulación negativa) a amarillo (regulación positiva) CL: Celomocitos, GN: Gónada y INT: Intestino. R1: Replica uno y R2: Replica dos.

#### 4.5 Enriquecimiento ontológico y rutas.

Para la identificación de los términos de ontología génica asociados a las tres categorías (procesos biológicos, funciones moleculares y componentes celulares), involucrados en cada uno de tejidos se utilizaron los transcritos propios de cada uno de ellos (tejido específico) (Figura 11), los cuales fueron identificados previamente. Los resultados muestran que en celomocitos los términos más representados en la categoría de procesos biológicos son “Localización macromolecular” (22,9%), “Localización celular” (21,2%) y “Localización proteínas” (20,1%), para componente celular son “Citoplasma” (37,2%), “Citosol” (25,5%) y “Exosoma extracelular” (21,5%) y para función molecular son “Unión a proteínas” (44,3%), “Unión a calcio” (6,2%) y “Actividad ubiquitin transferasa” (3,7%). Para gónada los términos más representados en procesos biológicos son “Generación de gametos masculinos” (7,3%), “Espermatogénesis” (7,3%) y “Ciclo celular: fase de transición G2/M” (4,4%), para componente celular son “Cilium” (7,3%), “Matriz extracelular” (7,0%) y “Microtúbulo” (5,5%) y para función molecular son “Unión a iones” (36,6%), “Unión a cationes” (35,9%) y “Unión a metales” (35,5%). Finalmente, para intestino se observó que los procesos biológicos más representados son “Transporte mediado por vesículas” (14,5%), “Transporte de aniones” (6,8%) y “Transporte de aniones orgánicos” (6,1%), para componente celular son “Parte de la región extracelular” (32,4%), “Vesícula unida a la membrana” (31,3%), “Exosoma extracelular” (26,3%), “Vesícula extracelular” (26,3%) y “Organelos extracelulares” (26,3%) y para función molecular son “Unión a calcio” (9,2%), “Unión a actina” (4,2%) y “Actividad receptora” (3,4%).

Con la herramienta DAVID es posible también determinar las categorías más enriquecidas en cada tejido, la cual nos permite entender de mejor forma lo que está sucediendo. Los primeros tres grupos de anotación más enriquecidos (con mayor puntaje de enriquecimiento o PE) muestran que para el primer grupo de anotación en celomocitos (PE=4,47), están involucrados funciones moleculares y procesos biológicos asociados a la actividad de algunos receptores. Para el grupo dos (PE=3,54), la regulación y autofagia en general y para el grupo tres (PE=2,91), la actividad de enzimas participes de la formación de ATP. En el caso de gónada se ve que en el grupo de anotación uno (PE=7,10) se encuentra la unión a cationes, metales y aniones; para el grupo dos (PE=3,94), procesos relacionados a la generación de espermatozoides y reproducción y para el grupo tres (PE=3,10), unión a nucleótidos y partes de estos. Finalmente, para intestino se ve que para el grupo

uno (PE=4,39), existen funciones y procesos involucrados con receptores y endocitosis. Para el grupo dos y tres (PE=3,30; PE=2,49), componentes celulares como vesículas, región extracelular y región apical de la célula (Tabla 6).



**Figura 11. Anotación ontología génica (OG) por tejido.** Gráfico que muestra los 10 primeros términos OG de cada categoría; Proceso biológico, Componente celular y Función molecular anotados con DAVID.

**Tabla 6. Enriquecimiento ontológico.** Categorías y términos de ontología génica asociados a los cinco primeros grupos de anotación con mayor puntaje de enriquecimiento (PE) para cada uno de los tres tejidos junto con el número de genes en total que pertenecen a ese término, porcentaje de transcritos del total que pertenecen a ese término (%) y el p-value de cada uno de ellos.

<b>Celomocitos</b>					
	<b>Categoría</b>	<b>Término</b>	<b>Número de genes</b>	<b>Porcentaje de genes (%)</b>	<b>p-value</b>
<b>Clúster de anotación 1</b>	<i>Función Molecular</i>	Receptor con actividad scavenger	12	1,8	3,65E-08
	<i>Función Molecular</i>	Receptor con actividad implicado en la captación de ligando	13	1,95	2,91E-11
	<i>Proceso Biológico</i>	Endocitosis mediada por receptor	29	4	3,59E-12
<b>Clúster de anotación 2</b>	<i>Proceso Biológico</i>	Regulación de la autofagia	27	3,72	1,23E-12
	<i>Proceso Biológico</i>	Regulación de la macroautofagia	16	2,21	3,16E-11
	<i>Proceso Biológico</i>	Autofagia	43	5,93	6,09E-11
<b>Clúster de anotación 3</b>	<i>Función Molecular</i>	Actividad pirofosfatasa	59	8,87	1,39E-12
	<i>Función Molecular</i>	Actividad hidrolasa, actuando en ácidos anhídros que contengan fosforo	59	8,87	1,47E-12

	<i>Función Molecular</i>	Actividad hidrolasa, actuando en ácidos anhidros	59	8,87	1,57E-12
	<i>Función Molecular</i>	Actividad nucleosido trifosfatasa	53	7,97	1,04E-03
	<i>Función Molecular</i>	Función acoplada a actividad ATPasa	23	3,46	2,45E-02
	<i>Función Molecular</i>	Actividad ATPasa	28	4,21	4,09E-02
<hr/>					
<b>Clúster de anotación</b>	<i>Función Molecular</i>	Unión a iones	226	33,98	1,43E-03
<b>4</b>					
	<i>Función Molecular</i>	Unión a metales	214	32,18	2,20E-03
<b>PE: 2,71</b>					
	<i>Función Molecular</i>	Unión a cationes	216	32,48	2,36E-03
<hr/>					
<b>Clúster de anotación</b>	<i>Función Molecular</i>	Aparato de Golgi	93	13,94	1,87E-03
<b>5</b>					
	<i>Componente Celular</i>	Parte del aparato de Golgi	62	9,3	2,07E-03
<b>PE: 2,42</b>					
	<i>Componente Celular</i>	Membrana de Golgi	46	6,9	1,43E-02
<hr/>					
<b>Gónada</b>					
<hr/>					
	<b>Categoría</b>	<b>Término</b>	<b>Número de genes</b>	<b>Porcentaje de genes (%)</b>	<b>p-value</b>
<hr/>					
<b>Clúster de anotación</b>	<i>Función Molecular</i>	Unión a cationes	163	39,76	4,68E-08
<b>1</b>					
	<i>Función Molecular</i>	Unión a metales	161	39,27	6,35E-07
<b>PE: 7,10</b>					
	<i>Función Molecular</i>	Unión a iones	166	40,49	1,67E-09

<b>Clúster de anotación</b> <b>2</b>  <b>PE: 3,94</b>	<i>Proceso Biológico</i>	Desarrollo espermatidas	16	3,52	2,80E-09
	<i>Proceso Biológico</i>	Diferenciación espermatidas	16	3,52	4,47E-09
	<i>Proceso Biológico</i>	Espermatogénesis	33	7,27	1,11E-11
	<i>Proceso Biológico</i>	Generación de gametos masculinos	33	7,27	1,15E-11
	<i>Proceso Biológico</i>	Proceso reproductivo de multi-organismos	49	10,79	4,08E-10
	<i>Proceso Biológico</i>	Reproducción sexual	42	9,25	5,64E-10
	<i>Proceso Biológico</i>	Generación de gametos	36	7,93	7,23E-10
	<i>Proceso Biológico</i>	Proceso reproductivo	62	13,66	1,91E-12
	<i>Proceso Biológico</i>	Reproducción	62	13,66	2,01E-12
	<i>Proceso Biológico</i>	Proceso reproductivo de organismos multicelulares	41	9,03	2,34E-12
	<i>Proceso Biológico</i>	Reproducción de organismos multicelulares	41	9,03	3,00E-12
	<i>Proceso Biológico</i>	Proceso reproductivo de un organismo	56	12,33	3,43E-11
	<i>Proceso Biológico</i>	Desarrollo de células germinales	17	3,74	6,36E-11
	<i>Proceso Biológico</i>	Proceso celular involucrado en la reproducción de organismos multicelulares	18	3,96	1,02E-03
	<i>Proceso Biológico</i>	Desarrollo del proceso involucrado en reproducción	26	5,73	5,58E-02
<b>Clúster de anotación</b> <b>3</b>	<i>Función Molecular</i>	Unión a derivados de carbohidratos	89	21,71	7,31E-09

**PE: 3,10**

<i>Función Molecular</i>	Unión a nucleótidos	93	22,68	1,06E-12
<i>Función Molecular</i>	Unión a nucleósidos fosfatados	93	22,68	1,07E-12
<i>Función Molecular</i>	Unión a moléculas pequeñas	98	23,9	1,46E-12
<i>Función Molecular</i>	Unión a adenil ribonucleótido	64	15,61	2,83E-11
<i>Función Molecular</i>	Unión a adenil nucleótido	64	15,61	3,43E-11
<i>Función Molecular</i>	Unión a ribonucleótido	74	18,05	6,56E-11
<i>Función Molecular</i>	Unión a ATP	61	14,88	7,27E-11
<i>Función Molecular</i>	Unión a ribonucleótido purina	73	17,8	8,31E-11
<i>Función Molecular</i>	Unión a nucleótido purina	73	17,8	1,01E-03
<i>Función Molecular</i>	Unión a ribonucleósido purina	71	17,32	1,29E-03
<i>Función Molecular</i>	Unión a nucleosido purina	71	17,32	1,35E-03
<i>Función Molecular</i>	Unión a ribonucleósidos	71	17,32	1,35E-03
<i>Función Molecular</i>	Unión a nucleósidos	71	17,32	1,50E-03
<i>Función Molecular</i>	Unión a ribonucleósidos purina trifosfatados	70	17,07	1,77E-03
<i>Función Molecular</i>	Unión a componentes orgánicos cíclicos	181	44,15	2,55E-02
<i>Función Molecular</i>	Unión a componentes heterocíclicos	178	43,41	3,04E-02

**Clúster de**


---

<b>anotación</b>	<i>Componente Celular</i> Cilium motil	14	3,33	9,14E-10
------------------	--	----	------	----------

<b>PE: 2,89</b>	<i>Componente Celular</i>	Parte del espermatozoide	13	3,1	2,02E-03
	<i>Componente Celular</i>	Flagelo espermatozoide	7	1,67	1,18E-02
<b>Clúster de anotación 5</b>	<i>Proceso Biológico</i>	Transición fase G2/M ciclo celular	20	4,41	1,90E-10
		Transición G2/M de ciclo celular mitótico	19	4,19	2,47E-10
		Proceso Ciclo celular	58	12,78	6,71E-11
		Transición fase G2/M ciclo celular	26	5,73	3,53E-03
		Ciclo celular	65	14,32	3,55E-03
		Transición fase ciclo celular	27	5,95	3,82E-03
		Proceso Ciclo celular mitótico	38	8,37	6,61E-03
		Ciclo celular mitótico	39	8,59	1,41E-02
		Regulación del ciclo celular	34	7,49	1,33E-01
		División celular	21	4,63	1,52E-01
		Regulación del proceso del ciclo celular	21	4,63	1,91E-01

### Intestino

<b>Clúster de anotación</b>	<b>Categoría</b>	<b>Término</b>	<b>Número de genes</b>	<b>Porcentaje de genes (%)</b>	<b>p-value</b>
<b>1</b>	<i>Función Molecular</i>	Receptor con actividad scavenger	10	2,82	1,00E-10

		Receptor con actividad			
<b>PE: 4,39</b>	<i>Función Molecular</i>	implicado en la captación de ligando	11	3,1	3,17E-09
	<i>Función Molecular</i>	Endocitosis mediada por receptor	19	5	4,10E-12
	<i>Proceso Biológico</i>	Endocitosis	28	7,37	2,08E-03
<hr/>					
<b>Clúster de anotación 2</b>	<i>Componente Celular</i>	Exosoma extracelular	100	27,4	1,29E-11
<b>PE: 3,30</b>	<i>Componente Celular</i>	Vesícula extracelular	100	27,4	1,56E-11
	<i>Componente Celular</i>	Organelos extracelulares	100	27,4	1,58E-10
	<i>Componente Celular</i>	Vesículas unidas a la membrana	119	32,6	6,26E-11
	<i>Componente Celular</i>	Parte de la región extracelular	123	33,7	2,13E-03
	<i>Componente Celular</i>	Región extracelular	141	38,63	3,60E-03
<hr/>					
<b>Clúster de anotación 3</b>	<i>Componente Celular</i>	Parte apical de la célula	22	6,03	5,44E-11
<b>PE: 2,49</b>	<i>Componente Celular</i>	Región de la membrana plasmática	38	10,41	6,27E-03
	<i>Componente Celular</i>	Membrana plasmática apical	16	4,38	9,74E-03
<hr/>					
<b>Clúster de anotación 4</b>	<i>Proceso Biológico</i>	Transporte de ácidos grasos	9	2,37	8,17E-11
<b>PE: 2,46</b>	<i>Proceso Biológico</i>	Transporte de icosanoides	6	1,58	5,92E-03
	<i>Proceso Biológico</i>	Transporte de prostaglandinas	4	1,05	8,65E-03

<b>Clúster de anotación</b> <b>5</b>  <b>PE: 2,27</b>	<i>Proceso Biológico</i>	Homeostasis tejido	13	3,42	2,54E-03
	<i>Proceso Biológico</i>	Homeostasis organismo multicelular	16	4,21	7,33E-03
	<i>Proceso Biológico</i>	Homeostasis de estructura anatómica	17	4,47	8,41E-03

---

Adicionalmente, al análisis de enriquecimiento ontológico se realizó una búsqueda de rutas Kegg representadas en cada uno de los tejidos. Para esto se realizó el mismo procedimiento que en la búsqueda de términos ontológicos, pero esta vez con la base de datos de Kegg, mostrando que las rutas estadísticamente más representadas son “Infección por *Vibrio cholerae*” (ruta:05110) en celomocitos, “Lisosoma” (ruta:04142) en intestino y “Secreción de Renina” (ruta: 0494) en gónada (Tabla 7).

**Tabla 7. Rutas metabólicas tejido específicas.** Rutas metabólicas involucradas en cada uno de los tejidos, con el número de transcritos encontrados por cada vía y el porcentaje del total, filtradas a un p-value < 1E-2.

<b>Celomocitos</b>				
<b>ID Kegg</b>	<b>Termino Kegg</b>	<b>Número de genes</b>	<b>Porcentaje de genes</b>	<b>p-value</b>
5110	Infección por <i>Vibrio cholerae</i>	10	3,2%	5,70E-04
4142	Lisosoma	13	4,1%	8,80E-03
<b>Intestino</b>				
<b>ID Kegg</b>	<b>Termino Kegg</b>	<b>Número de genes</b>	<b>Porcentaje de genes</b>	<b>p-value</b>
4142	Lisosoma	10	2,4%	3,00E-03
860	Metabolismo de porfirina y clorofila	6	1,4%	3,50E-03
982	Metabolismo de drogas - citocromo P450	8	1,9%	6,50E-03
480	Metabolismo de glutatión	7	1,7%	8,20E-03
<b>Gónada</b>				
<b>ID Kegg</b>	<b>Termino Kegg</b>	<b>Número de genes</b>	<b>Porcentaje de genes</b>	<b>p-value</b>
4924	Secreción de renina	8	4,2%	1,80E-03
330	Metabolismo de arginina y prolina	7	3,6%	2,30E-03
71	Degradación de ácidos grasos	6	3,1%	4,20E-03
1212	Metabolismo de ácidos grasos	6	3,1%	4,5E-03

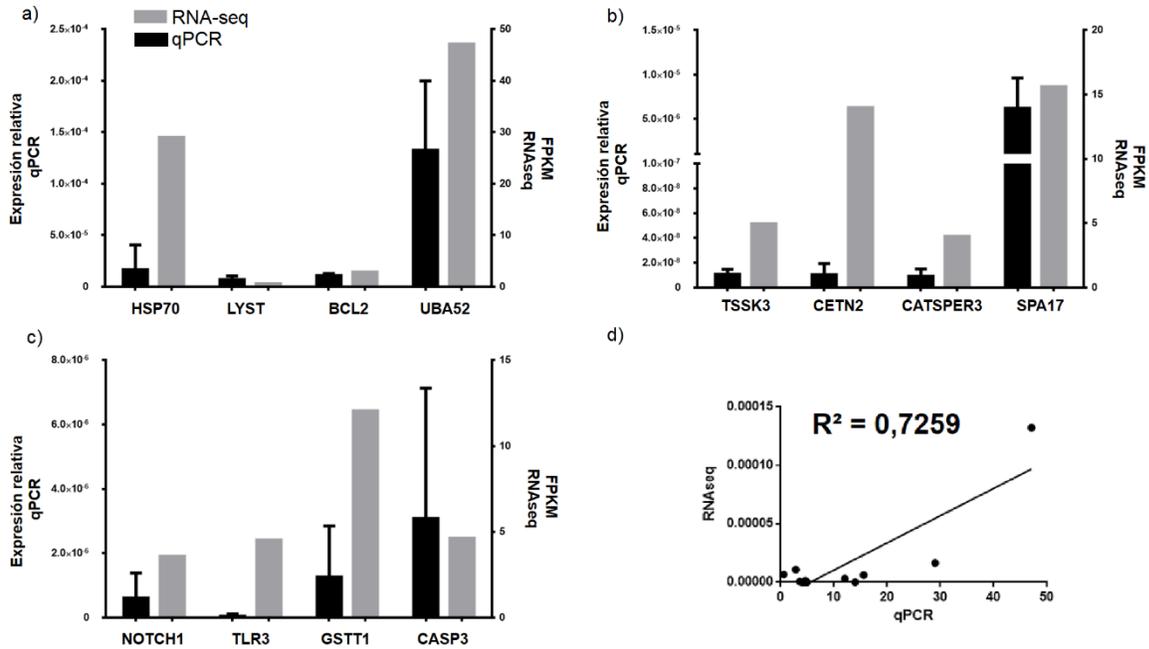
#### 4.6 Validación transcritos tejido específico por qPCR

La elección de los 12 transcritos candidatos (Tabla 8) a validar se realizó según la función y el comportamiento que tenía entre los tejidos (tejido específico), estos fueron HSP70, LYST, BCL2 y UBA52 para celomocitos, TSSK3, CETN2, CATSPER3 y SPA17 en gónada y NOTCH1, TLR3, GSTT1 y CASP3 en intestino. De los transcritos elegidos, seis de ellos (50%) fueron detectados de manera tejido específico por qPCR (TSSK3, CETN2, CATSPER3, NOTCH1, TLR3 y GSTT1). El transcrito de CASP3 fue encontrado además de celomocitos en intestino. En relación con la expresión diferencial se observó una correlación del 0,73 incluyendo los 12 transcritos validados (Figura 12).

**Tabla 8. Genes candidatos a validar por qPCR.** Tabla con los contigs seleccionados para validar por qPCR con su función o rol biológico.

Tejido	Nombre proteína	Rol biológico
Gónada	Proteína serina/treonina específica de testículo 3 (TSSK3)	Puede estar involucrado en una vía de señalización durante el desarrollo o maduración de espermatozoides.
	Centrina 2 (CETN2)	Juega un rol fundamenta en organización de los microtúbulos, es requerido por los centriolos para una correcta formación del huso.
	Proteína canal de calcio asociada a espermatozoides 3 (CATSPER3)	Canal de calcio gatillado por voltaje, juega un rol en la respuesta fisiológica dependiente de calcio esencial para la fertilización.
	Proteína de superficie del espermatozoide Sp17 (SPA17)	Proteína de unión de la superficie del espermatozoide a la zona pelucida. Ayuda a la unión del espermatozoide a la zona pelucida con una alta afinidad.
Intestino	Proteína homologa a notch 1 (NOTCH1)	Afecta la implementación de la diferenciación, proliferación y programas apoptóticos.

	Receptor tipo toll 3 (TLR3)	Control de la respuesta inmune del hospedero a través del reconocimiento de patrones moleculares específicos en los microorganismos.
	Caspasa 3 (CASP3)	Involucrado en la activación de la cascada de señalización responsable de la ejecución de la apoptosis.
	Glutación S-transferasa theta-1 (GSTT1)	Conjugación de glutatión reducido a un amplio número de electrófilos exógenos y endógenos hidrofóbicos.
Celomocitos	Proteína de choque térmico 70 kDa 1A (HSP70)	En cooperación con otras chaperonas, Hsp70 estabiliza proteínas preexistentes y media el correcto plegamiento de ellas.
	Ubiquitina del residuo A-52 del ribosoma (UBA52)	Marcaje de proteínas para la degradación en el proteosoma. Involucrado en la mantención de la estructura de la cromatina, regulación de la expresión génica y repuesta a estrés.
	Regulador del tráfico lisosomal (LYST)	Puede ser requerido para clasificar las proteínas endosómicas residentes en los endosomas multivesiculares tardíos mediante un mecanismo que involucra microtúbulos.
	Célula B linfoma 2 (BCL2)	Regulación de la apoptosis



**Figura 12. Validación por qPCR de los resultados de RNA-seq.** Expresión relativa de los transcritos seleccionados versus los FPKM obtenidos por RNA-seq para a) Celomocitos, b) Gónada y c) Intestino. Las barras grises indican el número FPKM y las barras en negro los datos de expresión relativa de los transcritos con su desviación estándar. d) Correlación de los niveles de expresión de todos los transcritos elegidos.

## 5. Discusión

En este estudio se realizó la secuenciación masiva del transcriptoma de gónada, intestino y celomocitos en el erizo chileno *L. albus*, generando información genómica de esta especie. Además, se identificaron transcritos utilizando análisis de redes que no había sido utilizada para este tipo de datos en esta especie, aportando aún más en la identificación o anotación de las secuencias.

### **Ensamble *de novo* y validación del transcriptoma.**

En contraste a lo reportado en otras especies de erizos en estudios similares (Chen *et al.* 2015; Pérez-Portela *et al.* 2015), el ensamble *de novo* de las lecturas de alta calidad para los tres tejidos del erizo chileno *L. albus*, entregó un número mayor de contigs. Posterior al filtro de contigs de baja cobertura, el número de estos fue reducido a 185.239 contigs. Esto es comparable con lo encontrado en un estudio similar realizado en el erizo *Evechinus chloroticus*, en donde ensamblaron 209.654 contigs, y después de la eliminación de estos por baja cobertura quedaron 186.427 contigs (Gillard *et al.* 2014). La composición nucleotídica del transcriptoma compuesto por los tres tejidos de *L. albus* (38,81% GC), es similar a lo encontrado en otras especies (Dilly *et al.* 2014; Gillard *et al.* 2014), lo que indica la semejanza en la composición de nucleótidos entre especies de erizo.

Por mucho tiempo el valor de N50 por sí solo se usó para medir la calidad del ensamble. Este se ha ido dejando de lado debido a que muchas veces no solo muestra que existe un alto número de transcritos posiblemente completos, sino que puede significar un alto número de quimeras (Li *et al.* 2014). Por esto, para determinar de manera más completa la calidad del ensamble, no solo se debe considerar el N50, sino que también otros parámetros como lo son el número de lecturas mapeadas, la búsqueda de genes conservados y el número total de unigenes (Honaas *et al.* 2016).

Siguiendo en orden los parámetros a considerar para evaluar la calidad del ensamble tenemos que el N50 observado a partir del ensamble *de novo* de *L. albus* es más alto que lo encontrado en trabajos similares. En el caso de las lecturas, el 95% de estas fueron mapeadas. Además, el resultado de la búsqueda de secuencias conservadas con las herramientas CEGMA y BUSCO, ambas utilizadas actualmente para evaluar el ensamble de secuencias (Rana *et al.* 2016),

arrojó alrededor del 90% de las secuencias completas encontradas. El número de unigenes determinados no es comparable con otros estudios en erizos debido a que no son considerados, lo más próximo es lo encontrado en la secuenciación del pepino de mar *Apostichopus japonicus* con 94.704 unigenes (Zhou *et al.* 2014). En resumen, todos los parámetros considerados anteriormente apuntan a la alta calidad del ensamble *de novo* del transcriptoma de *L. albus*.

### **Anotación transcriptoma y redes de similitud de secuencia.**

Los resultados de la anotación son concordantes al número de secuencias encontradas en las bases de datos. *S. purpuratus* lidera la lista de especies con la cual fue anotado el transcriptoma de *L. albus*, seguido por *S. kowalevskii*. Esto es debido a que *S. purpuratus* es el equinodermo con mayor número de secuencias en las bases de datos, teniendo alrededor de 30.000 proteínas identificadas en NCBI y Uniprot, en contraste con especies como el erizo *P. lividus*, que tiene sólo alrededor de 1.000 y 400 proteínas identificadas en NCBI y Uniprot, respectivamente. *S. kowalevskii*, segunda especie con mayor porcentaje de anotación, es mencionado en otros estudios de equinodermos (Vaughn 2012; Applebaum *et al.* 2013; Gillard *et al.* 2014) debido a su similitud con ellos, encontrándose filogenéticamente cercano (Cameron *et al.* 2000; Simakov *et al.* 2015).

Las redes de similitud de secuencia (RSS), han sido ya validadas como una herramienta que permite predecir la función de una proteína en base a su secuencia (Atkinson *et al.* 2009; Janga *et al.* 2011; Zhao *et al.* 2014), con el fundamento de que la función de una proteína viene dada por la secuencia que esta tenga, la cual es conservada entre las proteínas que comparten una función (Tian y Skolnick, 2003). Esto lo hace relacionando datos de información de secuencia, procesos involucrados, entre otros, con información nucleotídica de un polipéptido, de manera de asociarlo con una función similar en base a un puntaje de corte determinado (por ejemplo, e-value, porcentaje de identidad, momento de expresión) (Rhee y Mutwil, 2013). Actualmente en la bibliografía no se encuentran estudios transcriptómicos que utilizan estas herramientas, sino que la mayor parte de la literatura asociada es a la identificación y análisis de familias de proteínas en específico, por lo que este es el primer trabajo que utiliza esta técnica para identificar secuencias a partir de la secuenciación masiva del ARN.

En nuestro caso, en el transcriptoma de *L. albus*, las tres redes incluyeron más del 70% de las secuencias, con lo cual se estimó un 64% de anotación utilizando esta herramienta. Esto, sumado a la anotación convencional por Blastx, da como resultado un 75% de anotación, lo cual permite reconocer un mayor número de transcritos que en otros trabajos similares, en donde las técnicas convencionales de identificación de transcritos entregan en promedio un 35% de anotación (Dilly *et al.* 2014; Gillard *et al.* 2014; Perez-Portela *et al.* 2015).

El análisis exhaustivo de las redes de similitud nos permite reconocer la totalidad de familias involucradas en un proceso, momento o tejido en específico, lo que hace más alcanzable el análisis y comparación de un gran número de datos. Las RSS nos permiten aumentar la data disponible de cualquier especie, haciendo más fácil sobre todo el estudio de especies no modelo, en donde la información es escasa.

## **Enriquecimiento ontológico y rutas Kegg**

### ***Celomocitos***

Los celomocitos son las células encargadas de la inmunidad en erizos, de los cuales el 50% al 70% son células motiles con un alto requerimiento energético (De Faria y Da Silva 2008) y considerados como un equivalente de los macrófagos humanos, las cuales son células inmunes con una alta actividad catabólica y son parte de la respuesta inmune innata involucrada en digestión de patógenos y autofagia (Amer *et al.* 2005).

Esto mismo se ve reflejado en la combinación de resultados encontrados en el análisis de enriquecimiento ontológico y rutas Kegg, en donde se observa en primer lugar en las rutas Kegg la vía involucrada en la infección por *Vibrio*. Esto además se une a lo encontrado en el enriquecimiento ontológico en donde se observan los receptores de tipo *Scavenger* (SSCR; del inglés “Scavenger receptor cysteine-rich”) los cuales son un sub-grupo de receptores reconocedores de patrones capaces de identificar lipoproteínas LDL modificadas provenientes de la oxidación de las LDL nativas (Krieger, 1997); además de los “Cargo Receptor” (receptores con actividad de reconocimiento de ligando), que tienen la capacidad de reconocer proteínas ubiquitinadas, patógenos intracelulares y daño mitocondrial, gatillando la autofagia (Stolz *et al.* 2014). Esto muestra que en erizos existe una inmunidad basal activa en celomocitos, con la

capacidad de reconocer tempranamente la invasión o daño celular con el fin de llevar a cabo el proceso de respuesta inmune.

El proceso de autofagia es una de las mayores herramientas que tiene el sistema inmune frente a la invasión de una bacteria intracelular, siendo muchas veces modulado por los mismos patógenos en la persistencia de la infección (Campoy y Colombo, 2009). La autofagia es un proceso homeostático, en donde la célula descompone sus propios componentes celulares, respondiendo a diversos aspectos como lo son una infección (Levine *et al.* 2011) o el estrés (Branco *et al.* 2014). Este proceso se ve reflejado en el grupo de anotación dos del análisis de enriquecimiento ontológico, dando cuenta que podría ser el mecanismo de respuesta gatillado por los receptores en erizo.

En el grupo de anotación tres se observan funciones moleculares encargadas del gasto de ATP. El ATP es considerado en inmunidad en mamíferos como un marcador de actividad inmune, favoreciendo la actividad antibactericida de los macrófagos, induciendo la fusión del fagosoma-lisosoma por la vía dependiente de calcio y la secreción de IL-1 $\beta$  (Kusner y Barton 2001; Fossati *et al.* 2008; Cohen *et al.* 2013) reafirmando el rol de los celomocitos en inmunidad.

En el grupo de anotación cuatro, se encuentra representado el componente celular “Aparato de Golgi”, el cual es el organelo encargado de procesar y enviar las proteínas hacia otros lugares de la célula. Se ha descrito que en celomocitos desafiados con LPS existe un aumento de los niveles de mensajeros involucrados en la síntesis, procesamiento y degradación, lo que se relaciona con una mayor exportación de proteínas relacionadas con la vía del complemento. Esta vía es parte del sistema inmune innato y está compuesta de múltiples proteínas encargadas de la opsonización del patógeno (Gross *et al.* 2000; Janeway *et al.* 2002; Smith *et al.* 2006).

Tanto en el grupo de anotación cuatro del enriquecimiento ontológico en celomocitos como el primero en gónada muestran las mismas funciones moleculares asociadas a la unión de cationes, aniones y metales, en ambos casos con un alto número de cuentas. Esto es debido a que la mayor parte de estas funciones moleculares se relacionan con factores enzimáticos, los cuales son requeridos por un alto número de proteínas para formar su estructura y llevar a cabo la reacción (Wittung-stafshede 2002).

Considerando todos los antecedentes y resultados, sumando además que los celomocitos extraídos provenían de erizos en condiciones normales, se sugiere que este tejido en *L. albus* estaría constantemente expresando moléculas y proteínas encargadas de la respuesta y detección temprana de un patógeno o estresor, gatillando la autofagia.

### ***Gónada***

En erizos las gónadas son consideradas como un órgano dual, ya que es el órgano encargado de la reproducción y del almacenaje de nutrientes (Walker *et al.* 2007; Gaitán-espitia *et al.* 2016) Estas funciones son llevadas a cabo por dos poblaciones celulares, las células germinales y las células somáticas, estas últimas también llamadas fagocitos nutritivos (NP; del inglés “Nutritive phagocytes”) (Walker *et al.* 2005; Walker *et al.* 2007).

Los resultados del análisis de enriquecimiento ontológico reflejan que los erizos secuenciados correspondían a machos, debido a que gran parte de los procesos se dirigen hacia la generación de espermatozoides. Esto sumado al grupo de anotación tres y cuatro, los cuales reflejan la motilidad flagelar, esto viene dado a que el flagelo está compuesto por el axonema el cual a su vez está compuesto por microtúbulos y dineínas, las cuales tienen la capacidad de moverlos por medio de la hidrólisis de ATP (Summer y Gibbons, 1971; Gibbons y Fronk, 1972; Inaba 2011). Esto indica que los espermatozoides se encuentran activos, dando cuenta que los erizos se encontrarían maduros sexualmente, con la capacidad de reproducirse.

Dentro de los resultados encontrados en las rutas Kegg podemos ver la relación de las gónadas (testículos) con la ruta de secreción de Renina. Esta proteína es una proteasa encargada de cortar a la proteína Angiotensina, la cual está involucrada en el mantenimiento de la homeostasis. En mamíferos se sabe que Renina puede encontrarse en las glándulas renales, estómago y testículo (Gomez y Sequeira-Lopez, 2016), en donde en este último cumpliría un rol en regulación de la secreción hormonal (Leung *et al.* 2000).

Como se mencionó anteriormente, dentro de la gónada existen mayoritariamente dos poblaciones celulares, las cuales son las células somáticas almacenadoras de nutrientes o NP y las células germinales (Walker *et al.* 2005; Walker *et al.* 2007). Dentro del análisis de rutas es que se

pudo observar la característica de las gónadas con relación al almacenaje de nutrientes, ya que se pudieron observar rutas involucradas en el metabolismo de aminoácidos y ácidos grasos.

Tanto las células somáticas como las células germinales (espermatogonia) indiferenciadas se multiplican por mitosis. Al momento de que las espermatogonias se diferencian a espermatocito I se comienzan a dividir por meiosis hasta llegar a espermátidas, las cuales se diferencian finalmente a espermatozoides (Mc Laren, 1984). El grupo de anotación cinco, muestra principalmente el proceso de mitosis, lo que podría dar cuenta del alto estado de proliferación celular, tanto de las células somáticas como de las células germinales, dado por la condición normal del tejido (alta proliferación de gametos) o por el estado de crecimiento del erizo (erizos juveniles).

### ***Intestino***

El intestino es el órgano encargado de la digestión del alimento (Farmanfarmaian y Phillips, 1816). En equinodermos ha tenido un rol clave en la respuesta inmune, en donde se ha descrito que presenta un alto número de secuencias relacionados a este proceso (Ramírez-gómez *et al.* 2008; Yang *et al.* 2009; Chen *et al.* 2015).

Es en esta línea que se agrupan los resultados, en donde en primer lugar vemos las mismas funciones observadas en celomocitos, mostrando un alto enriquecimiento en las secuencias relacionadas a SSCR y “Cargo Receptor”. Esto tiene concordancia con la importancia del intestino en la respuesta inmune, por lo que se ha postulado que el control de la flora normal o la respuesta del intestino asociada a patógenos es considerado como responsable del aumento de la expresión de NLR (“Nod-like receptor”) y posiblemente TLR (“Toll-like receptor”) (Hibino *et al.* 2006), siendo ambos receptores reconocedores de patrones (Kufer y Sansonetti, 2011). La misma situación se observa en la búsqueda de rutas, ya que tanto en celomocitos como en intestino las rutas lisosomales están presentes, lo que es también encontrado en la secuenciación de genes relacionados con inmunidad en el erizo *S. intermedius* (Chen *et al.* 2015).

Tanto en el grupo dos y tres se mencionan componentes celulares parte del epitelio intestinal. Los exosomas en mamíferos son vesículas que funcionan como vehículo intracelular de péptidos y proteínas como HSP70 frente a un estrés (Van Niel y Heyman 2002; Lancaster y Febbraio 2005; Browne *et al.* 2007). En erizos se ha sugerido que podrían ser partícipes de la

exportación de *Pl-GAL-8*, la cual se cree que podría estar contenida en la musculatura intestinal (Karakostis *et al.* 2015).

En el grupo de anotación cuatro se observa el proceso biológico asociado al transporte de prostaglandinas. Las prostaglandinas son ácidos grasos del tipo eicosanoides o icosanoides (Mathias y Dupont, 1985), que juegan un rol en el control de la motilidad intestinal en humanos (Misiewicz *et al.* 1969), inhibiendo la contracción del musculo gastrointestinal (Thor *et al.* 1985).

En el grupo de anotación cinco se observa el proceso de homeostasis. En el intestino la homeostasis depende la interacción entre la microbiota, el epitelio intestinal y el sistema inmune (Maloy y Powrie, 2011; Canton *et al.* 2013); en mamíferos se conoce que un desbalance de la microbiota puede desencadenar finalmente un proceso inflamatorio en el intestino (Garrett *et al.* 2011).

Considerando la totalidad de los procesos involucrados y los antecedentes en intestino, es que se podría mencionar que en *L. albus* el intestino es el primer órgano que tiene contacto con los patógenos. Por lo tanto, debe estar constantemente sensando el medio, considerando la posibilidad de un cambio en la homeostasis interna. Esto puede ocurrir tanto como por un desbalance iónico como por un desbalance en la microbiota bacteriana, además del constante transporte de nutrientes y desechos desde la región oral a la región aboral del erizo.

## **Validación**

Dentro de los seis genes que se comportaban por qPCR y RNA-seq de manera tejido específico, en gónada se encuentran TSSK3, CETN2 y CATSPER3, estos tienen roles en diferenciación, fertilidad y división celular en testículos en mamíferos (Hart *et al.* 2001; Lobleby *et al.* 2003; Li *et al.* 2010). El gen SPA17, el cual fue encontrado en los tres tejidos, es al igual que los anteriores considerado propio de espermatozoides y testículo, aunque también se ha visto en el surgimiento de diversos tipos de cáncer en mamíferos (Kong *et al.* 1995; Andrade *et al.* 2008). En equinodermos el proceso de regeneración es comandado por diversas vías, siendo una de ellas las vías relacionadas con cáncer (Mashanov *et al.* 2014), por lo que SPA17 podría estar involucrado.

En intestino se encontró el mismo escenario en donde NOTCH1, TLR3 y GSTT1 fueron encontrados de manera específica en este tejido, los cuales son participes en diferenciación y

proliferación celular, absorción, inmunidad y biotransformación de compuestos (Ria Vos, 1990; Gordon, 2002; Heath, 2010). CASP3 fue encontrado tanto en intestino como con celomocitos, esto es debido a que cumple un rol en inmunidad (Hibino *et al.* 2006), y como se mencionó anteriormente el intestino es un órgano clave en este proceso (Ramírez-gómez *et al.* 2008; Yang *et al.* 2009; Chen *et al.* 2015).

En celomocitos ninguno de los cuatro transcritos evaluados por qPCR fue tejido específico (HSP70, LYST, BCL2 y UBA52). Esto puede deberse a la profundidad de la secuenciación, por lo que por medio del RNA-seq estos contigs no fueron encontrados dentro de las otras librerías correspondientes a los otros tejidos. El qPCR, al ser una técnica distinta y más sensible si los detecta, aunque en estudios similares se encontró que existía un bajo número de contigs únicos de celomocitos detectados en los otros tejidos, debido a que estas células se encuentran presentes en todos los tejidos y órganos de los erizos (Smith *et al.* 2006; Chen *et al.* 2015).

En un estudio similar reciente de Gaitán-Espitia *et al.* (2016), se realizó el ensamble *de novo* y mapeo de lecturas contra una referencia (*S. purpuratus*) de gónadas de macho del erizo *L. albus*. En este estudio lograron ensamblar de manera exitosa 42.530 contigs con un N50 de 645 pb, además realizaron la validación de este con la herramienta CEGMA, encontrando 176 de 248 transcritos (71%), sumado a la anotación de 15.544 secuencias (36,6%) contra las bases de datos de Uniprot y NCBI.

Este artículo viene a reafirmar lo encontrado en esta tesis, debido a que en el transcriptoma compuesto por tres tejidos, se logró anotar un mayor porcentaje de transcritos por herramientas convencionales (Blast) y por RSS. Esto sumado a que los datos antes mencionados dan cuenta de la mayor calidad de ensamble que tiene en comparación al reciente transcriptoma de gónada, validando de mejor forma la técnica elegida. Aun así, el artículo de Gaitán-Espitia, logró encontrar a través de ontología génica y rutas (Kegg) procesos relacionados con reproducción, generación de gametos, fertilización, desarrollo de espermátidas, espermatogénesis, diferenciación celular, motilidad flagelar, entre otros, resultados similares a los encontrados en este trabajo, aunque estos no se encontraban enriquecidos en el artículo.

## **6. Conclusión y proyecciones**

Se logró secuenciar y ensamblar *de novo* el transcriptoma de *L. albus* proveniente de tres tejidos (celomocitos, gónada e intestino), con una alta calidad, anotando un alto número de contigs utilizando herramientas convencionales (Blast) y no convencionales (RSS).

En base a lo encontrado y a la expresión diferencial de genes únicos en cada tejido se encontró que cada uno de los tejidos secuenciados cumple un rol importante en la fisiología del erizo, en donde se destacan el rol en reproducción y almacenamiento de nutrientes de las gónadas, inmunidad en celomocitos e intestino, además de homeostasis en este último.

Este trabajo sirve de base para nuevos estudios en este organismo, no solo para potenciar el conocimiento básico a nivel biológico, sino que también abriendo posibilidades para su aplicación al cultivo y crecimiento de este producto de importancia comercial para el país.

## 7. Referencias

- Amer A, Byrne B, Swanson M (2005) Macrophages rapidly transfer pathogens from lipid raft vacuoles to autophagosomes. *Autophagy*, 1: 53-58
- Andrade V, Vettore A, Felix R, Almeida M, Carvalho F, Salvador R, Oliveira De, *et al.* (2008). Prognostic impact of cancer / testis antigen expression in advanced stage multiple myeloma patients. *Cancer immunity* 8: 1–8.
- Andrew N, Agatsuma Y, Ballesteros E, Bazhin G, Creaser E, Barnes D, Botsford L, *et al.* (2002). Status and management of world sea urchin fisheries. *Oceanography And Marine Biology An Annual Review*. 343-425
- Applebaum Scott L, Ginsburg David W, Capron Charles S, Manahan Donal T. (2013). Expression of amino acid transporter genes in developmental stages and adult tissues of antarctic echinoderms. *Polar Biology* 36: 1257–67.
- Arevalo A, Bruning P, Sanchez R, Cárdenas L. (2014). Development and characterization of est-microsatellites for the edible sea urchin *Loxechinus albus* using next generation sequencing. *Conservation Genetics Resources* 6: 433–35.
- Arrau L. (1958). Desarrollo del erizo comestible de Chile, *Loxechinus albus* (Molina, 1782). *Biología marina Chile* 7: 39-62.
- Atkinson H, Morris J, Ferrin T, Babbitt P. (2009). Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS ONE* 4: e4345.
- Barahona N, Orensanz J, Parma A, Jerez G, Romero C, Miranda H, Zuleta A, Catasti V, Galvez P. (2003). Informe final FIP 2000-18: bases biológicas para rotación de áreas en el recurso erizo. Informe FIP No 2000-18, 197.
- Levine B, Mizushima N, Virgin H. (2011). Autophagy in immunity and inflammation. *Nature* 469: 323–35.
- Branco P, Figueiredo D, Silva J. (2014). New insights into innate immune system of sea urchin: coelomocytes as biosensors for environmental stress. *OA Biology* 2: 1–7.

- Browne C, Swan J, Rankin E, Calvert H, Griffiths S, Tytell M. (2007). Extracellular heat shock protein 70 has novel functional effects on sea urchin eggs and coelomocytes. *The journal of experimental biology* 210: 1275–87.
- Bustos E, Olave S. (2001). Manual: el cultivo del erizo (*Loxechinus albus*). Diversificación de La Acuicultura En La X Región. Instituto del fomento pesquero. 1-24
- Cameron C, Garey J, Swalla B. (2000). Evolution of the chordate body plan: new insights from phylogenetic analyses of deuterostome phyla. *Proceedings of the National Academy of Science of the United States of America* 97: 4467-74
- Campoy E, Colombo M. (2009). Autophagy in intracellular bacterial infection. *Biochimica et biophysica acta* 1793:1465–77.
- Canton J, Neculai D, Grinstein S. (2013). Scavenger receptors in homeostasis. *Nature Reviews Immunology* 13: 621–34.
- Cea G, Gaitán-Espitia J, Cárdenas L. (2015). Complete mitogenome of the edible sea urchin *Loxechinus albus*: genetic structure and comparative genomics within echinozoa. *Molecular Biology Reports* 42: 1081–89.
- Chen Y, Chang Y, Wang X, Qiu X, Liu Y. (2015). De novo assembly and analysis of tissue-specific transcriptomes revealed the tissue-specific genes and profile of immunity from *Strongylocentrotus intermedius*. *Fish & Shellfish Immunology* 46: 723–36.
- Cohen H, Briggs K, Marino J, Ravid K, Robson S, Mosser D. (2013). TLR stimulation initiates a CD39-based autoregulatory mechanism that limits macrophage inflammatory responses. *Blood* 122: 1935-45
- Contreras S, Castilla J. (1987). Feeding behavior and morphological adaptations in two sympatric sea urchin species in central Chile. *Marine Ecology Progress Series* 38: 217–24.
- Davidson N, Oshlack A. (2014). Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biology* 15: 410.
- De Faria MT, da Silva JRMC (2008) Innate immune response in the sea urchin *Echinometra lucunter* (Echinodermata). *Journal of Invertebrate Pathology*, 98, 58-62

Doty M, Roehr J, Ahmed R, Dieterich C. (2012). Flexbar—flexible barcode and adapter processing for next-generation sequencing platforms. *Biology* 1: 895–905.

Dong Y, Sun H, Zhou Z, Yang A, Chen Z, Guan X, Gao S, Wang B, Jiang B, Jiang J. (2014). Expression analysis of immune related genes identified from the coelomocytes of sea cucumber (*Apostichopus japonicus*) in response to LPS challenge. *International Journal of Molecular Sciences* 15: 19472–86.

Du H, Bao Z, Hou R, Wang S, Su H, Yan J, Tian M, *et al.* (2012). Transcriptome sequencing and characterization for the sea cucumber *Apostichopus japonicus* (Selenka, 1867). *PLoS One* 7: e33311.

Ebert T. (2015). Growth rates of the sea urchin *Strongylocentrotus purpuratus* related to food availability and spine abrasion. *Ecology* 49: 1075–91.

Evans T, Chan F, Menge B, Hofmann G. (2013). Transcriptomic responses to ocean acidification in larval sea urchins from a naturally variable pH environment. *Molecular Ecology* 22: 1609–25.

Farmanfarmanian A, Phillips J. (1816). Digestion, storage, and translocation of nutrients in the purple sea urchin (*Strongylocentrotus purpuratus*). *Biological Bulletin* 123: 105-20

Fossati G, Rubartelli A, Piccini A, Carta S, Tassi S, Lasiglie D. (2008). ATP is released by monocytes stimulated with pathogen-sensing receptor ligands and induces IL-1 and IL-18 secretion in an autocrine way. *Proceeding of the National Academy of Science of the United States of America* 105: 8067–72.

Gaitán-Espitia J, Sánchez R, Bruning P, Cárdenas L. (2016). Functional insights into the testis transcriptome of the edible sea urchin *Loxechinus albus*. *Nature Scientific Reports* 6: 1–10.

Garrett W, Gordon J, Glimcher L. (2011). Homeostasis and Inflammation in the intestine. *Cell* 140: 859–70.

Dilly G, Gaitán-Espitia J, Hofmann G. (2014). Characterization of the antarctic sea urchin (*Sterechinus neumayeri*) transcriptome and mitogenome: a molecular resource for phylogenetics, ecophysiology and global change biology. *Molecular Ecology Resources* 15: 425-39.

Gibbons I, Fronk E. (1972). Some properties of bound and soluble from sea urchin sperm flagella. *The journal of cell biology* 9: 365–81.

Gillard G, Garama D, Brown C. (2014). The transcriptome of the nz endemic sea urchin kina (*vechinus chloroticus*). *BMC Genomics* 15: 15-45.

Gomez R, Sequeira-lopez M. (2016). Reviews novel functions of renin precursors in homeostasis and disease. *Physiology* 31: 25–33.

Gordon S. (2002). Pattern recognition receptors: doubling up for the innate immune response. *Cell* 111: 927–30.

Gross P, Clow L, Smith L. (2000). Spc3, the complement homologue from the purple sea urchin, *strongylocentrotus purpuratus*, is expressed in two subpopulations of the phagocytic coelomocytes. *Immunogenetics* 51: 1034–44.

Haas B, Papanicolaou A, Yassour M, Grabherr M, Philip D, Bowden J, Couger M, *et al.* (2014). De Novo Transcript Sequence Reconstruction from RNA-Seq: Reference Generation and Analysis with Trinity. *Nature Protocols* 8: 1494-1512.

Hart P, Poynter G, Whitehead C, Orth J, Glantz J, Busby R, Barrett S, Salisbury J. (2001). Characterization of the x-linked murine centrin *cetn2*. *Gene* 264: 205-13

Heath J. (2010). Transcriptional Networks and Signaling Pathways That Govern Vertebrate Intestinal Development. *Current Topics in Developmental Biology* 90: 159-92.

Hibino T, Loza-coll M, Messier C, Majeske A, Cohen A, Terwilliger D, Buckley K, *et al.* (2006). The immune gene repertoire encoded in the purple sea urchin genome. *Developmental biology* 300: 349–65.

Honaas L, Wafula E, Wickett N, Der J. (2016). Selecting superior de novo transcriptome assemblies: lessons learned by leveraging the best plant genome. *PLoS One* 11: e0146062.

Instituto de Fomento Pesquero. (1998). Erizo *loxechinus albus*: recursos objetivo áreas de manejo pesca artesanal. Chile, 1-5

Instituto de Fomento Pesquero. (2010). Erizo *loxechinus albus*: recursos objetivo areas de manejo pesca artesanal. Chile, 1–4.

- Inaba K. (2011). Sperm flagella: comparative and phylogenetic perspectives of protein components unicellular algae *Chlamydomonas*. *Molecular Human Reproduction* 17: 524–38.
- Janeway C, Medzhitov R. (2002). Innate immune recognition. *Annual Review Immunology*. 20: 197-216.
- Janga S, Diaz-Mejia JJ, Moreno-Hagelsieb G. (2011). Network-based function prediction and interactomics: the case for metabolic enzymes. *Metabolomic Engineering* 13: 1–10.
- Johnson P. (1969). The coelomic elements of sea urchins (*Strongylocentrotus*). *Histochemie* 17: 213–31.
- Karakostis K, Cost C, Zito F, Matranga V. (2015). Heterologous expression of newly identified galectin-8 from sea urchin embryos produces recombinant protein with lactose binding specificity and anti-adhesive activity. *Scientific Report* 7: 1–14.
- Kino S, Agatsuma Y. (2007). Reproduction of sea urchin *Loxechinus albus* in Chiloé Island, Chile. *Fisheries Science* 73: 1265–73.
- Kong M, Richardson R, Widgren E, Rand M, Hill C, Carolina N. (1995). Sequence and localization of the mouse sperm autoantigenic protein, sp17. *Biology of Reproduction* 53: 579–90.
- Krieger M. (1997). The other side of scavenger receptors: pattern recognition for host defense. *Current Opinion in Lipidology*. 8: 275-280
- Kufer T, Sansonetti P. (2011). Perspective NLR functions beyond pathogen recognition. *Nature Immunology* 12: 121–28.
- Kusner D, Barton J. (2001). ATP stimulates human macrophages to kill intracellular virulent *Mycobacterium tuberculosis* via calcium-dependent phagosome-lysosome fusion. *Journal of Immunology* 167: 3308-15.
- Lancaster, G. I. and Febbraio, M. A. (2005). Exosome-dependent trafficking of HSP70: a novel secretory pathway for cellular stress proteins. *The Journal of Biology Chemistry*. 280: 23349-55.
- Leung P, Wong T, Lam S, Chan P, Wong P. (2000). Testicular hormonal regulation of the renin-angiotensin system in the rat epididymis. *Life Science* 66: 1317–24.

- Li B, Fillmore N, Bai Y, Collins M, Thomson J, Stewart R, Dewey C. (2014). Evaluation of de novo transcriptome assemblies from rna-seq data. *Genome Biology* 15: 1–21.
- Li W, Godzik A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658–59.
- Li Y, Sosnik J, Brassard L, Reese M, Spiridonov N, Bates T, Johnson G, Anguita J, Visconti P, Salicioni A. (2010). Expression and localization of five members of the testis-specific serine kinase ( tssk ) family in mouse and human sperm and testis. *Molecular human reproduction* 17: 42–56.
- Lobley A, Pierron V, Reynolds L, Allen L, Michalovich D. (2003). Identification of human and mouse catsper3 and catsper4 genes: characterisation of a common interaction domain and evidence for expression in testis. *Reproductive biology and endocrinology* 15: 1–15.
- Maloy K, Powrie F. (2011). Intestinal homeostasis and its breakdown in inflammatory bowel disease. *Nature* 474: 298-306.
- Mashanov V, Zueva O, García-Arrarás J. (2014). Transcriptomic changes during regeneration of the central nervous system in an echinoderm. *BMC Genomics* 15: 357.
- Mathias M, Dupont J. (1985). Dietary factors affecting lipid metabolism. *Lipids* 20: 791-800.
- McGettigan P. (2013). Transcriptomics in the rna-seq era. *Current Opinion in Chemical Biology* 17: 4–11.
- McLaren (1984). Meiosis and differentiation of mouse germ cells. *Symposia of the Society for Experimental Biology*. 38:7-23
- Misiewicz J, Waller S, Kiley N. (1969). Effect of oral prostaglandin E1 on intestinal transir in man. *Lancet* 1: 648–51.
- Molinet C, Rubilar P, Zuleta A, Rosales S, Ariz L, Barahona N, Yung Z, *et al.* (2005). Bases biológicas para rotación de áreas en el recurso erizo (fase ii). *FIP 2003* 13: 276.
- Moreno C, Molinet C. (2013). El paradigma de la distribución batimétrica de *Loxechinus albus* (molina) en Chile. *Revista Chilena de Historia Natural* 86: 225–27.
- Mutz K, Heilkenbrinker A, Lönne M, Walter J, Stahl F. (2013). Transcriptome analysis using next-generation sequencing. *Current Opinion in Biotechnology* 24: 22–30.

- Oyarzún S, Marín S, Valladares C, Iriarte J. (1999). Reproductive cycle of *Loxechinus albus* (Echinodermata: Echinoidea) in two areas of the Magellan region (53° S, 70-72° W), Chile. *Research Gate* 63: 439–49.
- Parra G, Bradnam K, Korf I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23: 1061–67.
- Pérez-Portela R, Riesgo X. (2015). Characterization of the transcriptome and gene expression of four different tissues in the ecologically relevant sea urchin *Arbacia lixula* using RNA-seq. *Molecular Ecology Resource* 16:794-808.
- Purushothaman S, Saxena S, Meghah V, Swamy C, Ortega-Martinez O, Dupont S, Idris M. (2014). Transcriptomic and proteomic analyses of *Amphiura filiformis* arm tissue-undergoing regeneration. *Journal of Proteomics* 112: 113–24.
- Qian X, Ba Y, Zhuang Q, Zhong G. (2014). RNA-seq technology and its application in fish transcriptomics. *OMICS: A Journal of Integrative Biology* 18: 98–110.
- Rahman M, Arshad A, Yusoff F. (2014). Sea urchins (Echinodermata: Echinoidea): their biology, culture and bioactive compounds. *International Conference on Agricultural*. 39-48
- Ramírez-gómez F, Ortíz-pineda P, Rojas-cartagena C, Suárez-castillo E, García-ararrás J. (2008). Immune-related genes associated with intestinal tissue in the sea cucumber *Holothuria glaberrima*. *Immunogenetics* 60: 57–71.
- Rana S, Zadlock F, Zhang Z, Murphy W. (2016). Comparison of de novo transcriptome assemblers and k-mer strategies using the killifish, *Fundulus heteroclitus*. *PLoS One* 11: e0153104
- Rhee S, Mutwil M. (2013). Towards revealing the functions of all genes in plants. *Trends in Plant Science* 19: 1–10.
- Richardson M, Sherman C. (2015). De novo assembly and characterization of the invasive northern Pacific seastar transcriptome. *PLoS One* 10: e0142003.
- Sernapesca. (2016). Antecedentes preliminares Erizo *Loxechinus albus*: recursos objetivo áreas de manejo pesca artesanal, Chile: 1-4.

Simakov O, Kawashima T, Marlétaz F, Jenkins J, Koyanagi R, Mitros T, Hisata K, *et al.* (2015). Article hemichordate genomes and deuterostome origins. *Nature* 527: 459–65.

Simão F, Waterhouse R, Ioannidis P, Kriventseva E. (2015). Busco: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31: 3210-2.

Smith LC, Rast JP, Brockton V, Terwilliger DP, Nair S V, Buckley KM, Majeske AJ, Smith L Courtney. (2006). The sea urchin immune system. *Invertebrate Survival Journal* 114: 25–39.

Sodergren E. *et al.* (2006). The genome of the sea urchin. *Science* 314: 941–52.

Stolz A, Ernst A, Dikic I. (2014). Cargo recognition and trafficking in selective autophagy. *Nature* 16: 495–501.

Summer K, Gibbons I. (1971). Adenosine triphosphate-induced sliding of tubules in trypsin-treated flagella of sea-urchin sperm. *Proceedings of the National Academy of Science of the United States of America* 68: 3092–96.

Thor P, Konturek J, Konturek S, Anderson J (1985). Role of prostaglandins in control of intestinal motility. *The American Journal Physiology* 248: G353-9.

Tian W, Skolnick J. (2003). How well is enzyme function conserved as a function of pairwise sequence identity?. *Journal of Molecular Biology* 333: 863–82.

Tu Q, Cameron R, Worley K. (2012). Gene structure in the sea urchin *strongylocentrotus purpuratus* based on transcriptome analysis. *Genome Research* 22: 2079–87.

Van Niel G, M Heyman. (2002). The epithelial cell cytoskeleton and intracellular trafficking ii. intestinal epithelial cell exosomes: perspectives on their structure and function. *American journal of physiology, Gastrointestinal and liver physiology* 283: G251–55.

Van Bladeren R. (1990). Review article glutathiones-transferases in relation to their role in thebiotransformation of xenobiotics. *Chemico-biological interactions* 75: 241–65.

Van Verk M, Hickman R, Pieterse C, Van Wees S. (2013). Rna-seq: revelation of the messengers. *Trends in Plant Science* 18: 175–79.

Vaughn Roy. (2012). Comparative developmental transcriptomics of echinoderms. Graduate Theses and Dissertation. University of South Florida.

Walker CW, Harrington LM, Lesser MP, Fagerberg WR (2005) Nutritive phagocyte incubation chambers provide a structural and nutritive microenvironment for germ cells of *Strongylocentrotus droebachiensis*, the green sea urchin. *The Biological Bulletin* 209: 31–48

Walker C, Unuma T, Lesser M. (2007). Gametogenesis and reproduction of sea urchins. *Biology and Ecology* 2:11-33

Wittung-stafshede P. (2002). Role of cofactors in protein folding. *Accounts of chemical research* 35: 201–8.

Wygoda JA, Yang Y, Byrne M, Wray GA. (2014). Transcriptomic analysis of the highly derived radial body plan of a sea urchin. *Genome Biology and Evolution* 6: 964–73.

Yang A, Zhou Z, He C, Hu J, Chen Z, Gao X, Dong Y, *et al.* (2009). Analysis of expressed sequence tags from body wall, intestine and respiratory tree of sea cucumber (*apostichopus japonicus*). *Aquaculture* 296: 193–99.

Zamora S, Stotz W. (1992). Ciclo reproductivo de *Loxechinus albus* (Molina 1782) (Echinodermata: Echinoidea) en Punta Lagunillas, IV región, Coquimbo, Chile. *Revista Chilena de Historia Natural* 65: 121–33.

Zhao S, Sakai A, Zhang X, Vetting M, Kumar R, Hillerich B, Francisco B, *et al.* (2014). Prediction and characterization of enzymatic activities guided by sequence similarity and genome neighborhood networks. *Elife* 3: e03275

Zhou ZC, Dong Y, Sun HJ, Yang a F, Chen Z, Gao S, Jiang JW, Guan XY, Jiang B, Wang B. (2014). Transcriptome sequencing of sea cucumber (*apostichopus japonicus*) and the identification of gene-associated markers. *Molecular Ecology Resources* 14: 127–38.

## 8. Anexo

**Tabla A 1. Calidad de RNA (RQN) usado en la construcción de las librerías.** Tabla con el RQN entregado a partir de la medición de la calidad del RNA por electroforesis capilar.

Erizo	Gónada	Celomocito	Intestino
1	8,2	9,8	9,6
2	6,8	9,5	9,2
3	7,1	9,1	9,1

**Tabla A 2. Ejemplo de resultado anotación con redes de similitud de secuencia.** Tabla detalle con las secuencias anotadas de manera convencional y con RSS a partir del filtro de e-value de 0 – 1e-50, las secuencias anotadas por RSS son representadas por el nombre “PREDICTED SSR: saccin-like”.

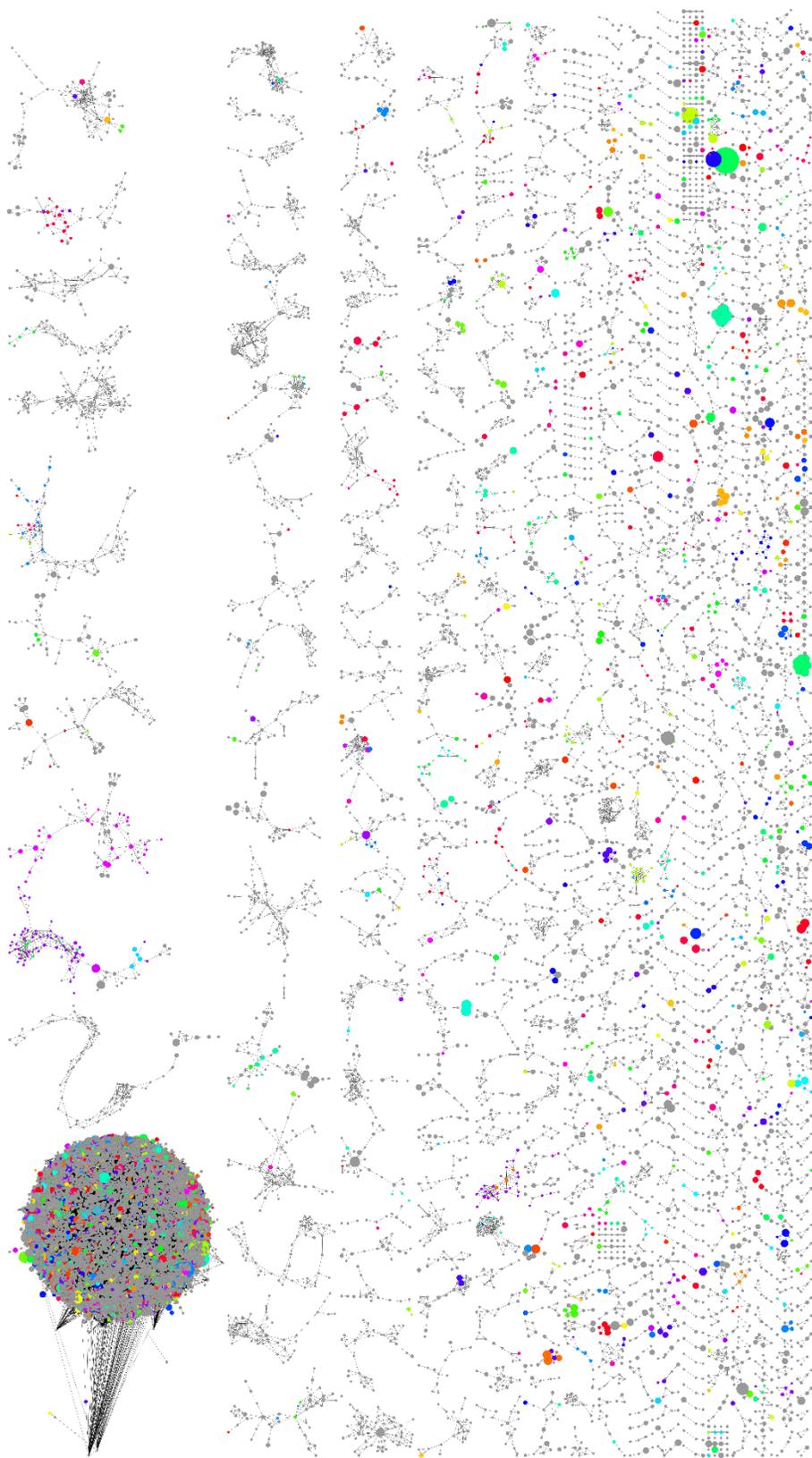
ID contig	Nombre proteína	Largo
DN3173c0gli1	tr W4YGZ7 W4YGZ7STRPUUncharacterizedproteinOS=[Stro ngylocentrotuspurpuratus]GN=Sp-SacS16PE=4SV=1	351
DN77c0gli1	tr W4XTA6 W4XTA6STRPUUncharacterizedproteinOS=[Stro ngylocentrotuspurpuratus]GN=Sp-SacS18PE=4SV=1	353
DN44207c1g5i2	Saccin (Fragment)	1043
DN35621c1g2i2	Saccin (DnaJ homolog subfamily C member 29) (DNAJC29)	3391
DN22811c1gli1	Saccin (DnaJ homolog subfamily C member 29) (DNAJC29)	225
DN28451c0gli5	Saccin (DnaJ homolog subfamily C member 29) (DNAJC29)	929
DN14790c0gli1	Saccin (DnaJ homolog subfamily C member 29) (DNAJC29)	807
DN43945c0g5i1	Saccin (DnaJ homolog subfamily C member 29) (DNAJC29)	425
DN44939c7g6i3	Saccin (DnaJ homolog subfamily C member 29) (DNAJC29)	1803
DN40509c2g2i1	Saccin (DnaJ homolog subfamily C member 29) (DNAJC29)	1614
DN15986c0gli1	Saccin (DnaJ homolog subfamily C member 29) (DNAJC29)	450
DN43945c0g4i1	Saccin (DnaJ homolog subfamily C member 29) (DNAJC29)	389
DN25978c0gli5	Saccin (DnaJ homolog subfamily C member 29) (DNAJC29)	748
DN26110c0gli1	Saccin (DnaJ homolog subfamily C member 29) (DNAJC29)	543

<b>DN44207c1g6i7</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	2606
<b>DN33144c0g3i4</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	928
<b>DN44939c7g1i5</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	2695
<b>DN17455c0g2i1</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	353
<b>DN712c0gli1</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	376
<b>DN43945c0g2i6</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	3576
<b>DN25978c0g1i6</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	675
<b>DN21438c0gli1</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	352
<b>DN43945c0g2i10</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	1134
<b>DN16629c0gli1</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	249
<b>DN27913c0g2i1</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	641
<b>DN40509c2g2i3</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	934
<b>DN44939c6g4i1</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	2186
<b>DN15992c0g2i1</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	458
<b>DN28451c0gli1</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	677
<b>DN33144c0g2i1</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	810
<b>DN44939c7g5i8</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	2052
<b>DN18300c0g2i1</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	470
<b>DN6353c0g2i1</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	328
<b>DN35621c1g2i9</b>	Sacsin (DnaJ homolog subfamily C member 29) (DNAJC29)	4005
<b>DN17466c0g2i1</b>	>ref XP011682937.1  PREDICTED: saccin-like [Strongylocentrotus purpuratus]	385
<b>DN2110c0gli1</b>	>ref XP011682937.1  PREDICTED: saccin-like [Strongylocentrotus purpuratus]	308
<b>DN20056c0gli3</b>	>ref XP011682937.1  PREDICTED: saccin-like [Strongylocentrotus purpuratus]	513
<b>DN20056c0g4i1</b>	>ref XP011682937.1  PREDICTED: saccin-like [Strongylocentrotus purpuratus]	288
<b>DN20052c0gli1</b>	>ref XP011682937.1  PREDICTED: saccin-like [Strongylocentrotus purpuratus]	298

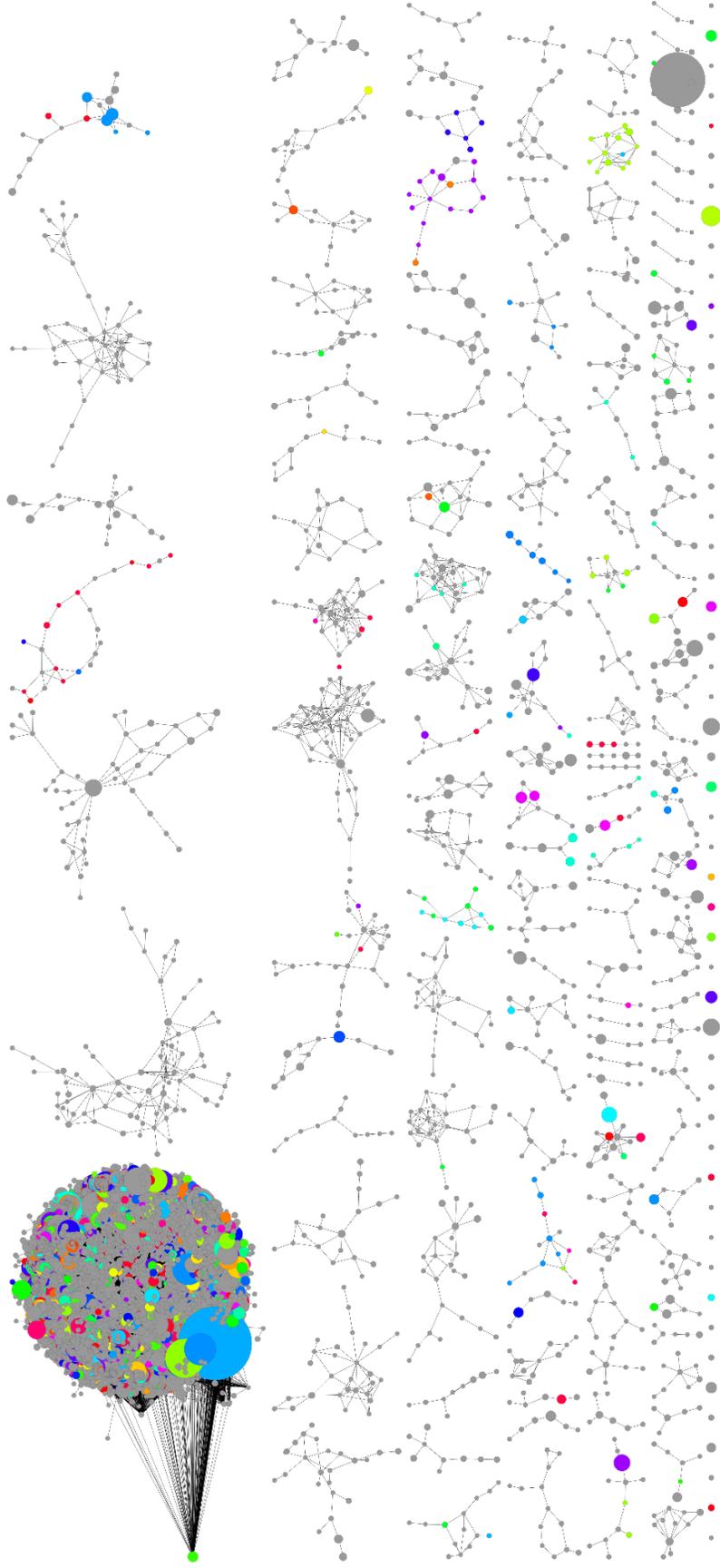
<b>DN16849c0gli1</b>	>ref XP011682937.1  PREDICTED: saccin-like [Strongylocentrotus purpuratus]	378
<b>DN18374c0gli1</b>	>ref XP011682935.1  PREDICTED: saccin-like [Strongylocentrotus purpuratus]	282
<b>DN3651c0gli1</b>	>ref XP011682935.1  PREDICTED: saccin-like [Strongylocentrotus purpuratus]	346
<b>DN22154c0g2i1</b>	>ref XP011682935.1  PREDICTED: saccin-like [Strongylocentrotus purpuratus]	545
<b>DN17579c0gli1</b>	>ref XP011677160.1  PREDICTED: saccin-like [Strongylocentrotus purpuratus]	286
<b>DN20882c0gli1</b>	>ref XP011677160.1  PREDICTED: saccin-like [Strongylocentrotus purpuratus]	339
<b>DN21438c1gli1</b>	>ref XP011677160.1  PREDICTED: saccin-like [Strongylocentrotus purpuratus]	296
<b>DN22553c0g3i1</b>	>ref XP011674925.1  PREDICTED: saccin-like [Strongylocentrotus purpuratus]	296
<b>DN18729c0gli1</b>	>ref XP011674925.1  PREDICTED: saccin-like [Strongylocentrotus purpuratus]	250
<b>DN18729c0g3i1</b>	>ref XP011674925.1  PREDICTED: saccin-like [Strongylocentrotus purpuratus]	428
<b>DN37494c0g2i4</b>	PREDICTED SSR: saccin-like	3111
<b>DN44622c0g3i5</b>	PREDICTED SSR: saccin-like	2771
<b>DN44622c0g3i1</b>	PREDICTED SSR: saccin-like	2675
<b>DN37494c0g2i1</b>	PREDICTED SSR: saccin-like	2217
<b>DN40509c2g3i4</b>	PREDICTED SSR: saccin-like	2027
<b>DN40509c2g3i3</b>	PREDICTED SSR: saccin-like	1805
<b>DN44080c3gli1</b>	PREDICTED SSR: saccin-like	1654
<b>DN44622c0g3i9</b>	PREDICTED SSR: saccin-like	1170
<b>DN35621c1g2i4</b>	PREDICTED SSR: saccin-like	962
<b>DN24928c0gli1</b>	PREDICTED SSR: saccin-like	953

<b>DN40509c2g1i1</b>	PREDICTED SSR: saccin-like	810
<b>DN35621c1g4i1</b>	PREDICTED SSR: saccin-like	809
<b>DN44080c3g1i2</b>	PREDICTED SSR: saccin-like	786
<b>DN35621c0g1i1</b>	PREDICTED SSR: saccin-like	537
<b>DN5248c0g1i1</b>	PREDICTED SSR: saccin-like	399
<b>DN35621c1g3i1</b>	PREDICTED SSR: saccin-like	386
<b>DN7243c0g1i1</b>	PREDICTED SSR: saccin-like	368
<b>DN37795c5g1i1</b>	PREDICTED SSR: saccin-like	360
<b>DN44622c0g8i1</b>	PREDICTED SSR: saccin-like	350

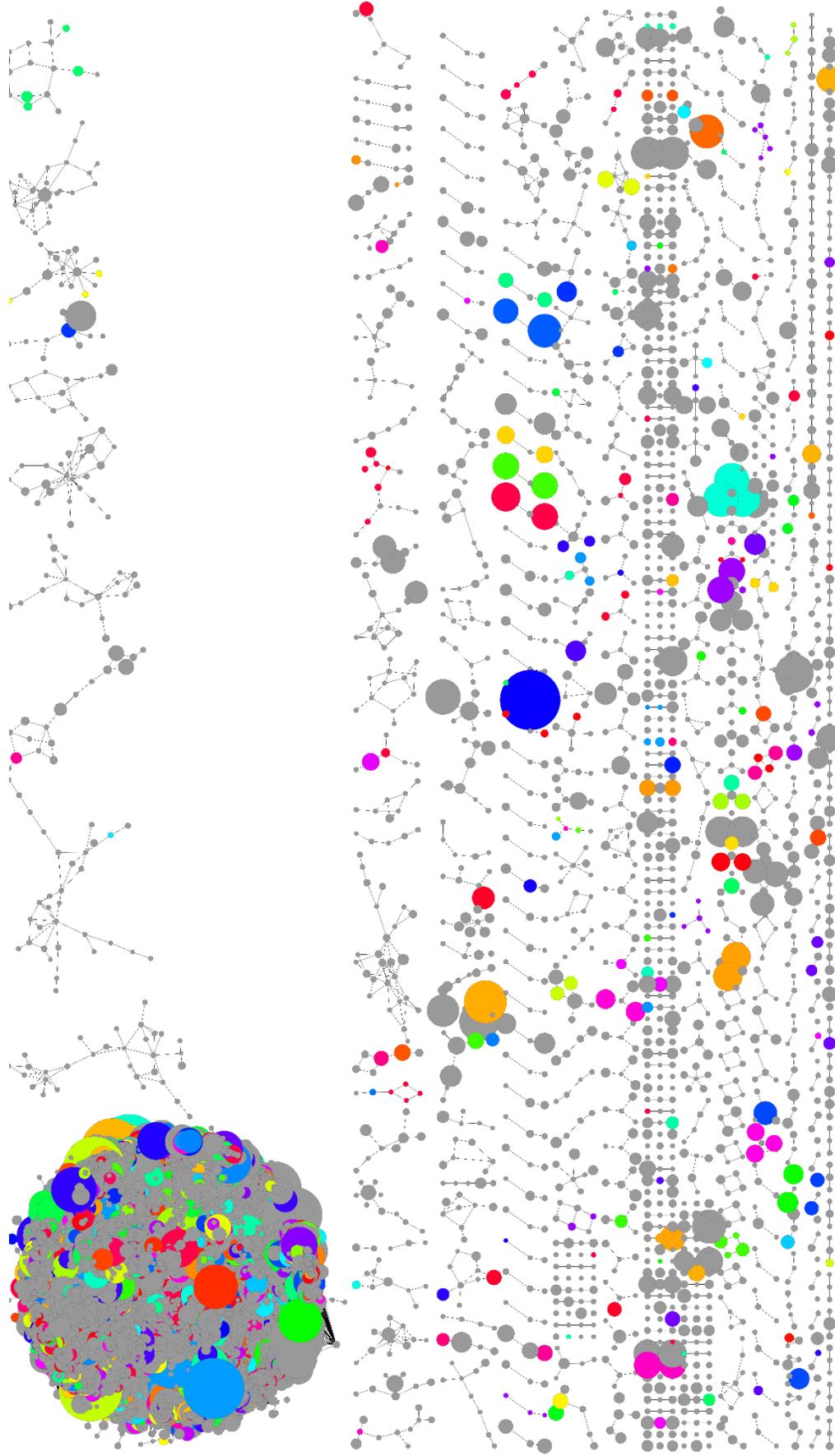
## Anexo: Redes de Similitud de secuencia.



**Figura A1. Visualización de red n°1.** Imagen de una porción de la red n°1 a partir del filtro por e-value entre 0 y 1E-50. Cada uno de los círculos corresponde a un unigen, en donde los tamaños son proporcionales al largo de la secuencia y los colores indican distintos procesos biológicos.



**Figura A2. Visualización de red n<sup>o</sup> 2.** Imagen de una porción de la red n<sup>o</sup> 2 a partir del filtro por evalúes entre 1E<sup>-50</sup> y 1E<sup>-30</sup>. Cada uno de los círculos corresponde a un unigen, en donde los tamaños son proporcionales al largo de la secuencia y los colores indican distintos procesos biológicos



**Figura A3. Visualización de red n°3.** Imagen de una porción de la red n°3 a partir del filtro por evalués entre  $1E^{-30}$  y  $1E^{-20}$ . Cada uno de los círculos corresponde a un unigen, en donde los tamaños son proporcionales al largo de la secuencia y los colores indican distintos procesos biológicos

