



**CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS  
DEL INSTITUTO POLITÉCNICO NACIONAL**

**UNIDAD IRAPUATO  
UNIDAD DE GENÓMICA AVANZADA-LANGEBIO**

**“Identificación de lncRNAs durante el fenómeno de proximidad  
vecina en *A. thaliana*”**

**Tesis que presenta  
Irving Jair García López**

**Para Obtener el Grado de  
Maestro en Ciencias**

**En la Especialidad de  
Biotecnología de Plantas**

**Directora de la Tesis: Dra. Selene Lizbeth Fernández Valverde**

**Irapuato, Guanajuato**

**Mayo 2019**

## **Agradecimientos**

Principalmente quiero agradecer al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca proporcionada (No. de beca 615597) y al Cinvestav-Langebio Unidad Irapuato por las instalaciones, infraestructura y personal que contribuyeron en gran medida a este proyecto.

A la Dra. Selene Lizbeth Fernández Valverde por su aceptación en el laboratorio, instrucción y atención a lo largo del proyecto, al igual que a los miembros del comité, el Dr. Alfredo Cruz Ramírez y el Dr. Charles Stewart Gillmor III por su interés y aportaciones al mismo.

Gracias a todo el laboratorio de Genómica Evolutiva y Funcional del ARN regulador por sus comentarios y recomendaciones hacia el proyecto.

A todos mis amigos que he conocido antes y durante la realización de mi maestría por su amistad.

Por último y no menos importante, gracias a mi familia por todo su apoyo constante para poder concluir este logro.

# Índice

<b>Resumen</b>	<b>8</b>
<b>I Introducción</b>	<b>9</b>
<b>II Antecedentes</b>	<b>11</b>
2.1 Definición y descubrimiento de los lncRNAs	11
2.1.1 Características de un lncRNA y clasificación	11
2.1.2 Funciones biológicas de los lncRNAs	14
2.1.3 Descubrimiento genómico de los lncRNAs	19
2.2 lncRNAs en plantas	19
2.2.1 Funciones de lncRNAs en plantas	20
2.3 Interacción y competencia planta-planta durante la proximidad vecina	23
2.3.1 Fenómeno de sombra y su prevención por SAS (Syndrome Shade Avoidance)	24
2.3.2 lncRNAs en estímulos de luz y fotomorfogénesis	30
<b>III Hipótesis</b>	<b>31</b>
<b>IV Objetivos</b>	<b>31</b>
4.1 Objetivo general	31
4.2 Objetivos específicos	31
<b>V Metodología</b>	<b>32</b>
5.1 Diseño Experimental	32
5.2 Preparación de librerías para ensamblado	33
5.3 Ensamble de transcriptomas	33
5.3.1 Ensamble <i>de novo</i> (Trinity)	33
5.3.2 Ensamble utilizando un genoma de referencia (Cufflinks)	34
5.4 Análisis de la calidad de los transcriptomas	35
5.5 Identificación de lncRNAs	35
5.5.1 CPC (Coding Potential Calculator)	35
5.5.2 Pipeline estricto (Cabili)	36
5.6 Clasificación de lncRNAs	39
5.7 Análisis de expresión diferencial de genes codificantes y lncRNAs	40
5.8 TopGO	41
5.9 Construcción de la red de co-expresión (WGCNA) <i>A. thaliana</i>	41
5.10 Generación gráfica de redes para cada módulo	44
<b>VI Resultados</b>	<b>45</b>
6.1 Ensamble de transcriptomas	45
6.2 Identificación y clasificación de lncRNAs	48

6.3 Expresión diferencial de genes codificantes y no codificantes _____	51
6.5 Construcción de redes de co-expresión y anotación funcional para la identificación de lncRNAs candidatos _____	63
<b>VII Discusión de resultados _____</b>	<b>75</b>
<b>IX Bibliografía _____</b>	<b>86</b>

## Índice de tablas y figuras

<b>Figura 1.</b> Localización genómica y contexto de los lncRNAs. ....	13
<b>Figura 2.</b> Funciones de lncRNAs en el núcleo y citosol.....	18
<b>Figura 3.</b> COOLAIR y COLDAIR, lncRNAs expresados durante la vernalización. <b>A)</b> COOLAIR se transcribe de forma antisentido a lo largo de FLC.....	21
<b>Figura 4.</b> Las interacciones entre plantas durante la competencia.....	23
<b>Figura 5.</b> Diferencias entre la luz solar y la luz de sombra.. ....	24
<b>Figura 6.</b> La respuesta fenotípica por SAS según la relación R:FR. El fenotipo de <b>a)</b> <i>A. thaliana</i> y <b>b)</b> <i>Brassica rapa</i> crecidas bajo condiciones de alta R:FR y baja R:FR.. ....	26
<b>Figura 7</b> Las reacciones de luz (líneas sólidas) y oscuridad (línea punteada), definen la abundancia de phyB Pfr activo (circulo amarillo) en <i>Arabidopsis</i> . ....	27
<b>Figura 8 A)</b> En alta incidencia de luz, phyB se encuentra en la forma Pfr.....	28
<b>Figura 9.</b> Representación esquemática de la extracción de ARN para el análisis mediante RNA-seq.....	32
<b>Figura 10</b> Pipeline general para anotar y clasificar lncRNAs en datos de RNA-seq. ....	38
<b>Figura 11.</b> Análisis de la topología de red para varias potencias de umbral suave. ....	42
<b>Figura 12.</b> Diagrama de flujo para encontrar lncRNAs candidatos.....	43
<b>Figura 13.</b> lncRNAs identificados en el transcriptoma.....	49
<b>Figura 14.</b> Clasificación de lncRNAs de acuerdo a su contexto genómico.....	49
<b>Figura 15.</b> Distribución de los lncRNAs a lo largo del genoma clasificados de acuerdo a su posición genómica. ....	50
<b>Figura 16.</b> Histogramas del número de exones en lncRNAs para cada cromosoma de <i>A. thaliana</i> , donde chr es cromosoma. ....	51
<b>Figura 17.</b> PCA para la visualización de la distancia entre muestras Tejidos de Cotiledón e Hipocótilo.. ....	52
<b>Figura 18.</b> PCA de tejidos independientes.....	52
<b>Figura 19.</b> Total de genes inducidos y reprimidos en cada contraste de tiempo en <b>A)</b> Cotiledón e <b>B)</b> Hipocótilo durante el tratamiento de sombra.). ....	54
<b>Figura 20.</b> Boxplot de conteos por millón (cpmlog2) entre transcritos codificantes (cajas azules) y no codificantes (cajas grises) en <b>A)</b> cotiledón e <b>B)</b> hipocótilo.. ....	54
<b>Figura 21.</b> Diagrama de Venn de lncRNAs que se comparten en ambos tejidos. ....	55
<b>Figura 22.</b> Heatmap de todos los genes expresados diferencialmente para <b>A)</b> Cotiledón e <b>B)</b> Hipocótilo durante el fenómeno de luz roja lejana.....	56
<b>Figura 23.</b> Heatmap de todos los lncRNAs contrastados expresados diferencialmente para <b>A)</b> Cotiledón e <b>B)</b> Hipocótilo durante el fenómeno de luz roja lejana.1.....	57
<b>Figura 24.</b> Enriquecimiento funcional de términos GO de los genes diferencialmente expresados en <b>Cotiledón</b> durante el tratamiento de sombra.....	61

<b>Figura 25.</b> Enriquecimiento funcional de términos GO de los genes diferencialmente expresados en <b>Hipocótilo</b> durante el tratamiento de sombra.....	62
<b>Figura 26.</b> Módulos co-expresados con detección de lncRNAs diferencialmente expresados durante el tratamiento de luz roja lejana para <b>A)</b> cotiledón e <b>B)</b> hipocótilo. ....	65
<b>Figura 27.</b> Heatmap de los módulos coexpresados de los términos GO con funciones biológicas de interés implicados en el fenómeno de sombra.. ....	67
<b>Figura 28.</b> Módulos de coexpresión relacionados al fenómeno de sombra.....	69
<b>Figura 29.</b> Heatmap de los genes co-expresados diferencialmente en módulos interesantes durante el fenómeno de luz roja lejana que contienen al menos un lncRNA para <b>A)</b> Cotiledón e <b>B)</b> Hipocótilo.....	70
<b>Figura 30.</b> Redes de coexpresión de los módulos con funciones relevantes en la condición de sombra. <b>A)</b> Para el módulo rojo, <b>B)</b> para el amarillo y <b>C)</b> el azul cielo3.. ....	72
<b>Figura 31.</b> Locus de transcritos no codificantes con inserción de TDNA en alguna parte del cuerpo del gen.....	74
<b>Tabla 1</b> Métricas de BUSCO para Trinity y Cufflinks .....	45
<b>Tabla 2.</b> Métricas de ensamble por RNAQuast para Trinity y Cufflinks .....	47
<b>Tabla 3.</b> Lista de genes candidatos (lncRNAs) expresados diferencialmente en módulos de interés (respuesta a SAS).....	73

## Abstract

Long non-coding RNAs (lncRNAs) are RNA molecules >200 nt long with very low coding potential that have been shown to be widespread across all eukaryotes. Many lncRNAs have now been demonstrated to play gene regulatory roles, including participating in development and the establishment of chromatin structure in animals and plants. However, only a dozen lncRNAs have been functionally characterized in plants. In this work, we focused on the identification and annotation of lncRNAs that display differential expression in the response to neighbor proximity phenomenon induced by light source competition in *Arabidopsis thaliana* seedlings. We assembled all transcripts across four timepoints post shade treatment (low-red light) in cotyledons and hypocotyls and identified 2,388 lncRNAs, including 159 that were not previously annotated in public databases. Of these, 83 were significantly differentially expressed in cotyledons and 514 in hypocotyls in response to shade treatment. To infer potential lncRNA functions, we functionally annotated 29 groups of genes co-expressed with our lncRNAs, identifying functions related to hormone response and cell wall modification. With this work, we have expanded the list of known lncRNAs as well as identified candidate lncRNAs which may participate in the regulation of the response to neighbor proximity in plants.

## Resumen

Los ARN largos (lncRNAs) no codificantes son moléculas mayores a 200 nucleótidos con muy bajo potencial para codificar proteínas que se encuentran a lo largo de todos los eucariontes. Se ha demostrado que muchos lncRNAs juegan roles importantes en la regulación génica, incluyendo la participación en el desarrollo y el establecimiento de la estructura de la cromatina en animales y plantas. Sin embargo, solo una docena de lncRNAs han sido funcionalmente caracterizados en plantas. En este trabajo, nos enfocamos en la identificación y anotación de lncRNAs que presenten expresión diferencial en la respuesta al fenómeno de proximidad vecina inducida por la competencia de luz en plántulas de *Arabidopsis thaliana*. Se ensamblaron transcriptomas a través de cuatro tiempos después de un tratamiento de sombra (baja R:FR) en cotiledones e hipocótilos, y se identificaron 2,388 lncRNAs incluyendo 159 que no han sido previamente anotados en bases de datos públicas. De esos, 83 fueron diferencialmente expresados significativamente en cotiledones y 514 en hipocótilos en respuesta al tratamiento de sombra. Para inferir la función de los lncRNAs, anotamos funcionalmente 29 grupos estrechamente co-expresados con nuestros lncRNAs, identificando funciones relacionadas a la respuesta a hormonas y modificación de pared celular. Con este trabajo hemos expandido la lista de lncRNAs conocidos, así como la identificación de lncRNAs candidatos los cuales podrían participar en la regulación de la respuesta al fenómeno de proximidad vecina en plantas.



## I Introducción

El genoma de eucariotas había sido una misteriosa caja de Pandora hasta antes de la llegada de la era de secuenciación. La historia nos pone a principios del siglo XXI, en donde los genetistas predecían que el genoma humano contenía aproximadamente 100,000 genes que codificaban proteínas, de acuerdo a la hipótesis de que organismos más complejos podían albergar un mayor número de genes (Wright & Bruford, 2011). Tras el desarrollo de tecnologías de secuenciación masiva (NGS, por sus siglas en inglés) con proyectos financiados por el Instituto Nacional de Investigación del genoma humano, el panorama cambió en 2004 (Metzker, 2010), con la disposición de una gran cantidad de datos genómicos masivos de una extensa variedad de organismos y una mejor calidad de un genoma humano bien anotado, la cifra de genes bajó a 20,000 (Wright & Bruford, 2011). Previamente a esto, se pensaba que una gran parte del genoma humano era ADN basura (junk DNA) sin ninguna función o propósito, o simplemente el resto de ADN representaba artefactos de secuenciación (Wahlestedt, 2013; Wright & Bruford, 2011). Según el consorcio FANTOM, más del 70% del genoma humano se transcribe a ARN, sin embargo, solo el 2% se traduce a proteínas (Wahlestedt, 2013). En concordancia con lo anterior, el consorcio ENCODE ha indicado que el 80% del genoma humano se transcribe, resultando en miles de transcritos de ARN no codificantes (Wahlestedt, 2013). Consecuentemente, ha surgido un gran interés para abordar los papeles funcionales para el 98% del genoma humano restante que no codifica proteínas (Wilusz, Sunwoo, & Spector, 2009). Se ha descrito que estos transcritos no codificantes pueden tener funciones importantes en la regulación de la expresión génica, las cuales son presentadas por varios subtipos diferentes. Entre ncRNAs (ARNs no codificantes) bastante estudiados se encuentran: el ARN ribosomal, el cual tiene una estructura de dos subunidades (la grande y pequeña); por otro lado está el ARN de transferencia (tRNA), el cual es el responsable de tomar los aminoácidos (aa) y llevarlos a la cadena polipeptídica naciente (Wahlestedt, 2013). No obstante, la lista de ncRNAs se ha diversificado más, incluyendo los ARNs nucleares pequeños (sRNA), ARNs nucleolares pequeños (snoRNA), micro ARNs (miRNA), ARNs que interactúan con PIWI (piRNA), ARNs del spliceosoma,

genes de la RNAsa P/MRP y ARNs largos no codificantes (lncRNAs); la principal tarea de los ARNs anteriores, es la regulación a nivel de expresión génica (Wahlestedt, 2013; Wright & Bruford, 2011). En el caso de los lncRNAs, muchas de sus características se desconocen, sin embargo cumplen con un amplio abanico de funciones en la regulación transcripcional, post-transcripcional y epigenética, aunque también trabajos recientes resaltan sus funciones como andamios (scaffolding) de complejos proteicos ribonucleares (Chen, Chen, & Zhang, 2015).

## **II Antecedentes**

### **2.1 Definición y descubrimiento de los lncRNAs**

Más allá de su característica de carecer potencial codificante, no existe una clara definición de un lncRNA basándonos en una argumentación biológica y que sea aceptada por la comunidad científica (Ma, Bajic, & Zhang, 2013). Sin embargo, hay una definición respaldada por varios autores basada en su longitud, la cual nos dice que un lncRNA tiene que estar por arriba de un umbral de 200 nucleótidos (nt) (Ma et al., 2013), esto para diferenciarlos de los ARNs pequeños.

Los lncRNAs fueron descritos como una clase transcripcional durante la secuenciación de librerías de cDNA de longitud completa en ratón (Mercer, Dinger, & Mattick, 2009). Pero fue que por primera vez en el año 1993, el término de lncRNA funcional se introdujo con el descubrimiento del gen *Xist* (Wahlestedt, 2013). *Xist* es un lncRNA que mediante el reclutamiento de *PCR2* (Polycomb Repressive complex 2) que dirige la inactivación del cromosoma X en mamíferos (Chen et al., 2015; Heo, Lee, & Sung, 2013; Kim & Sung, 2012; Wahlestedt, 2013).

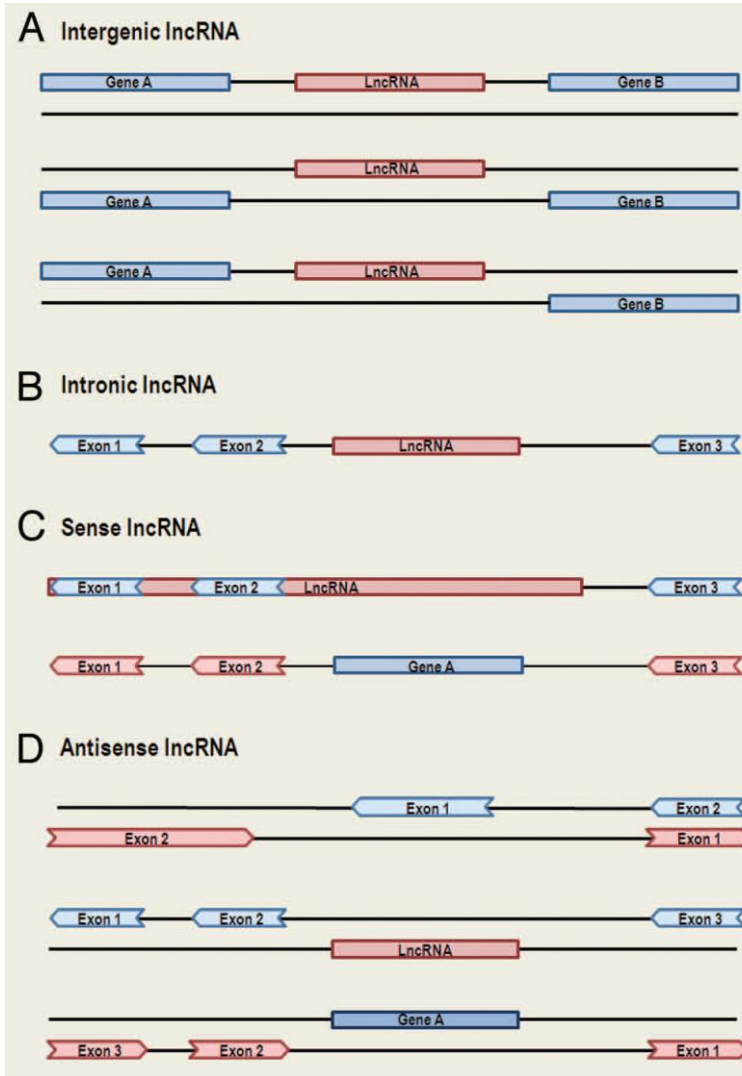
#### **2.1.1 Características de un lncRNA y clasificación**

Los lncRNAs comparten características con los ARNm y debido a esto, el proceso para clasificarlos no es tan trivial. Algunas de estas características compartidas son que la mayoría de lncRNAs son poliadenilados, al igual poseen la capucha 5' metilguanosa, también sufren de splicing alternativo y en general tienen longitudes similares (Chekanova, 2015; Tuck & Tollervey, 2013; Y. C. Zhang & Chen, 2013). También se transcriben principalmente por la Polimerasa II y mayoritariamente se expresan en tejidos específicos (Chekanova, 2015). Por otro lado también resaltan diferencias con los ARNm: la mayoría de los lncRNAs tienen una expresión substancialmente más baja que los ARNm, por lo cual se pensaba que podría tratarse de ruido transcripcional, así como una pobre conservación en su secuencia entre diferentes especies (Costa, 2005; Ma et al., 2013; Y. C. Zhang & Chen, 2013).

El genoma posee una característica modular. Es decir, contiene secuencias únicas incorporadas, tanto de transcritos codificantes y no codificantes, en dirección sentido y antisentido (Mercer et al., 2009). Gracias a las distintas combinaciones dadas por un splicing alternativo, iniciación y terminación de la transcripción, existe una diversificación de transcritos en el genoma (Mercer et al., 2009). Esta complejidad transcripcional también se observa en un sinnúmero de isoformas que presentan los genes de lncRNAs (Mercer et al., 2009).

Los lncRNAs a menudo se asocian con los genes codificantes y de acuerdo a esto pueden clasificarse en contexto a su posición genómica, relativo a los genes que codifican proteínas (Chekanova, 2015; Mercer et al., 2009). Existe un gran número de regiones no codificantes insertadas entre los genes codificantes en el genoma humano (98-99%) (Ma et al., 2013). En el caso de los lncRNAs que se localizan y se transcriben en regiones intergénicas, se definen como lncRNAs intergénicos (lincRNAs) (Figura 1A) mientras que los que completamente surgen de los intrones de genes codificantes se denominan como lncRNAs intrónicos (Figura 1B) (Ma et al., 2013).

Por otro lado, los lncRNAs sentido (Figura 1C) que se transcriben desde la hebra sentido del gen codificante, los cuales contienen exones de los genes codificantes (Ma et al., 2013). Estos pueden superponerse con gran parte del gen codificante o cubrir completamente la secuencia del gen codificante (Ma et al., 2013). Los lncRNAs antisentido, en contraste se transcriben de la hebra antisentido del gen codificante (Figura 1D) (Ma et al., 2013). Estos pueden presentarse en tres escenarios: 1) un transcrito derivado la hebra antisentido del gen codificante, sobrepuesto con un exón de un gen sentido a través de los exones de un lncRNA, 2) un transcrito desde un intrón de un gen sentido, que no resulta en una sobreposición exón-exón con este gen sentido y 3) un transcrito cubre en su totalidad la secuencia de un gen sentido con un intrón (Ma et al., 2013).



**Figura 1.** Localización genómica y contexto de los lncRNAs. Genes codificantes y sus exones son representados por cajas azules, mientras que los lncRNAs son representados por cajas rojas. **A)** lncRNAs intergénicos, se transcriben de regiones entre dos genes codificantes de ambas hebras. **B)** lncRNAs intrónicos se transcriben completamente de intrones de genes codificantes. **C)** lncRNAs sentido se transcriben desde la hebra sentido del gen codificante y contienen exones de genes codificantes que superponen con parte de genes codificantes o cubren totalmente la secuencia de un gen codificante a través de un intrón. **D)** lncRNAs antisentido, se transcriben desde la hebra antisentido del gen codificante, superponiendo con regiones exónicas o intrónicas o cubriendo completamente la secuencia del gen codificante a través de un intrón (Ma et al., 2013).

## 2.1.2 Funciones biológicas de los lncRNAs

La abundante transcripción de los lncRNAs nos puede dar indicio de que tienen un amplio abanico de funciones (Chen et al., 2015). Aunque existe una gran discusión sobre si esta transcripción a lo largo del genoma representa una transcripción en su mayoría inútil (ruido transcripcional) o si estos lncRNAs si tienen funciones aún no descubiertas (Ponting, Oliver, & Reik, 2009; Wilusz et al., 2009). A pesar de lo anterior se han descrito diversas funciones en múltiples procesos celulares como impronta genómica, diferenciación y desarrollo y respuesta antiviral (Marchese, Raimondi, & Huarte, 2017).

Mientras que muchos miRNAs se han estudiado arduamente para entender su función en la regulación génica, muchos lncRNAs permanecen con un funciones desconocidas (Fang & Fullwood, 2016). Esto es debido a que la función de los lncRNAs no puede ser inferida si solo nos enfocamos en su secuencia o estructura como sucede en los miRNAs (Mercer et al., 2009).

En general existe una pequeña clasificación o subgrupo de lncRNAs que debe ser refinada, sin embargo, se puede distinguir mecanismos moleculares de acuerdo con su localización celular; es decir, si su localización es nuclear, es posible que el lncRNA interactúe sobre la cromatina, participe en regulación transcripcional o procesamiento de ARN (splicing), mientras que si su localización celular es citoplasmática pueden modular la estabilidad de ARNm o traducción (Schmitt & Chang, 2016). Al igual que las proteínas, los lncRNAs albergan secuencias endógenas análogas a péptidos señal en proteínas, que determinaran su localización subcelular. Por ejemplo, *BORG* es un lncRNA que posee un motivo pentámero mientras que *FIRRE* es un lncRNA que posee un motivo dentro del dominio repetido RRD, ambos motivos para la retención nuclear (Goff & Rinn, 2015). De acuerdo con su interacción con copartícipes, también se pueden definir diversos mecanismos, específicamente si la regulación es a nivel cromatina se pueden observar mecanismos tales como modificación de cromatina, ya sea en forma *cis* o *trans* (modificación de histonas, reposicionamiento de nucleosomas), regulación transcripcional (represión o activación), formación de horquillas (loops) y metilación

(Schmitt & Chang, 2016; Taylor, Chu, Spektor, & Soloway, 2015). Interactuando con proteínas algunos lncRNAs pueden servir como andamios (scaffolding) para el ensamble de moléculas reguladoras en un mismo locus o reclutadores de FT (Goff & Rinn, 2015; Taylor et al., 2015). En interacciones con ARN, algunos lncRNAs funcionan principalmente como señuelos para miRNAs (Figura 2) (Taylor et al., 2015). Por último se pueden diferenciar por sus propiedades estructurales o por la secuencia de donde se originan (enhancers activos, promotores o regiones intergénicas) (Schmitt & Chang, 2016; Taylor et al., 2015).

Sólo unos cuantos lncRNAs predichos han sido caracterizados experimentalmente (Li, Wu, Fu, & Han, 2014). Un ejemplo muy citado de regulación en *cis* es el regulador maestro *Xist* (X-chromosome inactive-specific transcript), que se transcribe desde el locus *Xic* (X inactivation center) para llevar a cabo la compensación de dosis (Dosage Compensation) mediante la inactivación del cromosoma X (Penny, Kay, Sheardown, Rastan, & Brockdorff, 1996; Taylor et al., 2015). Existe un lncRNA que ejemplifica la regulación tanto en *cis* como en *trans*. La transcripción del lncRNA intergénico *Dali* requerido durante la formación de células neuronales *in vitro*, que regula la expresión de su gen vecino *Pou3f3* además de interactuar con la proteína y regular la expresión de genes lejanos involucrados en el programa de diferenciación celular en *trans* (Chalei et al., 2014).

Juntos con muchos genes codificantes, los lncRNAs se expresan de manera específica de un tipo celular durante la diferenciación (Delás & Hannon, 2017). En el sistema nervioso, durante el desarrollo de células madre neuronales se expresa el lncRNA nuclear *Gomafu*, el cual posee un mecanismo post-transcripcional uniéndose a *SF1* (Splicing Factor 1) a través de secuencias repetidas en tándem (Yoshida et al., 2011). *TERRAs* (Long non coding Telomeric Repeat-containing RNAs) son transcritos que su expresión es regulada durante el ciclo celular y se transcriben en los telómeros para ayudar el mantenimiento telomérico (Azzalin & Lingner, 2015).

Un sello distintivo de la función de los lncRNAs es su capacidad para regular la regulación epigenética (Li et al., 2014). A grandes rasgos, “epigenética” se refiere a los cambios que pueden afectar la regulación génica, esto incluye cambios en la

metilación del ADN, reposicionamiento de los nucleosomas y expresión de miRNAs así como la modificación de histonas y remodelamiento en orden superior de la cromatina (Fang & Fullwood, 2016; Li et al., 2014; Taylor et al., 2015). La metilación juega un rol importante en el desarrollo normal, mientras que aberraciones en las modificaciones epigenéticas pueden ocasionar enfermedades como el cáncer (Li et al., 2014). El silenciamiento del locus *HOXD* de 2.2 kb mediante la interacción del lncRNA *HOTAIR* y *PCR2* donde el principal complejo H3K27 que contiene la histona metiltransferasa necesaria para la deposición de la marca epigenética H3K27me3 (tri-metilación en la lisina 27 en la histona 3) a lo largo del locus *HOXD* (Rinn et al., 2007). Otro ncRNA que interactúa con *PCR2*, es un repetido dentro de *Xist* de 1.6 kb (*RepA*), identificado como blanco principal por *PRC2* en ratón (J. Zhao, Sun, Erwin, Song, & Lee, 2008). La reducción de *RepA* mediada por el ARN antisentido *Tsix* afecta la interacción *RepA-PC2*, suprimiendo la inducción de *Xist* de longitud completa y la marca H3K27m3 del cromosoma X (J. Zhao et al., 2008). Como se hacía mención, los lncRNAs pueden formar complejos de ARN-proteína con factores de transcripción y regular la actividad de factores de transcripción (FT), varios trabajos sugieren que al menos algunos lncRNAs pueden regular la transcripción sirviendo como ligandos para FT (Li et al., 2014). Debido a su amplia distribución, los lncRNAs pueden afectar la actividad de FT específicos y polimerasas (Li et al., 2014). También pueden actuar como co-factores para modular la actividad de un FT (Mercer et al., 2009). Por ejemplo, en ratón, el ncRNA *Evf2* se transcribe desde un enhancer distante muy conservado, reclutando la interacción y la acción del FT *DLX2* a el mismo enhancer para inducir la expresión de los genes codificantes vecinos (Mercer et al., 2009).

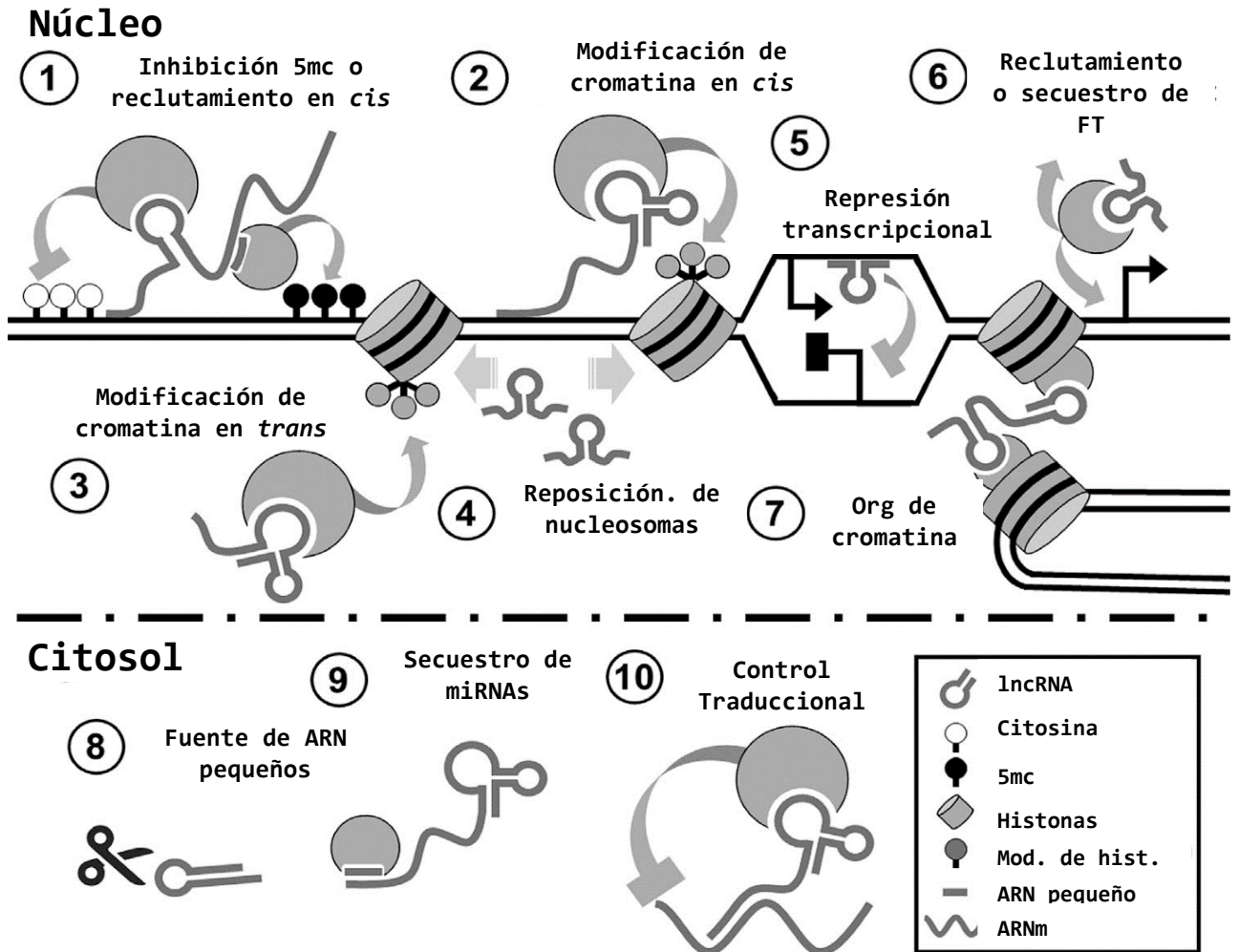
La capacidad de los ncRNAs para reconocer secuencias complementarias permite interacciones que son responsables para la regulación del proceso post-transcripcional de mRNAs, lo que conlleva su splicing, edición, transporte, traducción y degradación (Mercer et al., 2009). Los lncRNAs que ayudan al splicing de mRNAs pueden actuar a través de la interacción o modulación de factores de splicing, o directamente mediante la hibridación o fusión con secuencias de mRNA para parar el splicing (Ma et al., 2013). En el caso de la regulación por control



traduccional, algunos lncRNAs pueden interactuar con factores de traducción o directamente con los ribosomas (Ma et al., 2013). Por otro lado, en mamíferos pueden actuar como inhibidores antisentido naturales para promover la degradación del mRNA (Ma et al., 2013; Mercer et al., 2009).

Otro mecanismo de regulación de los lncRNAs es que se desempeñan como señuelos o esponjas debido a su capacidad para aparear sus bases con pequeños ARNs para modular sus actividades (Wilusz et al., 2009). Por ejemplo, pueden interactuar con miRNAs, donde inhiben competitivamente la capacidad de los miRNAs para interactuar con sus mRNAs blancos regulando su actividad y controlando la traducción de estos (Chen et al., 2015; Wilusz et al., 2009).

Se ha descubierto que los lncRNAs también pueden servir como ARNs estructurales (Wilusz et al., 2009). En las células de mamíferos, el núcleo no está solo compartimentado de tal forma que está separado del citoplasma, sino que existen muchos compartimentos membranosos que tienen funciones especializadas (Wilusz et al., 2009). Por ejemplo, el lncRNA *NEAT1* es esencial para la formación de cuerpos subnucleares encontrados en el espacio de la intercromatina o paraspeckles (Fox & Lamond, 2010; Lin et al., 2018). *NEAT1* tiene una organización espacial de orden superior dentro el paraspeckle, de tal manera que sus extremos 5' y 3' se localizan en la periferia del paraspeckle, mientras que las secuencias centrales de *NEAT1* se localizan en el núcleo (Lin et al., 2018).



**Figura 2.** Funciones de lncRNAs en el núcleo y citosol. **1)** Los lncRNAs pueden llevar a cabo una variedad de modificaciones de la cromatina, como la metilación del ADN. Modificación de las histonas en **2) cis** o en **3) trans**. **4)** Pueden reclutar remodeladores de cromatina y controlar las posiciones de los nucleosomas. **5)** Transcritos antisentido pueden inhibir la expresión génica. **6)** lncRNAs pueden interactuar con factores de transcripción afectando esta. **7)** La organización tridimensional del núcleo es moldeada por lncRNAs donde regiones distantes del ADN pueden interactuar formando loops. **8)** Algunos lncRNAs pueden ser fuente de ARNs pequeños, como por ejemplo: los miRNAs. **9)** Por otro lado pueden inhibir la expresión de ciertos miRNAs. **10)** Por último pueden regular la acumulación de proteínas, inhibiendo la traducción de los mRNAs (Taylor et al, 2015).

### **2.1.3 Descubrimiento genómico de los lncRNAs**

El poder de los microarreglos y proyectos de secuencias genómicas proporcionó el primer vistazo a la transcripción generalizada de los ncRNAs. En estudios previos, se utilizaron microarreglos de ADN con secuencias blanco anidadas que comprenden toda una secuencia de ADN cromosómico y nos permite estudiar de manera imparcial las regiones transcritas (Rinn & Chang, 2012). Sin embargo, existen limitaciones en los ensayos con microarreglos, como lo es el potencial para hibridación cruzada, la falta de información hebra-específica si el cDNA se hibridó al ensayo, y por último se desconoce la conectividad entre regiones transcritas (Rinn & Chang, 2012)

Tenemos por otro lado la tecnología por secuenciación masiva que nos da la oportunidad para secuenciar cDNA a una escala nunca vista y de alto rendimiento (RNA-Seq) (Mortazavi et al., 2008). Con la ayuda de enfoques computacionales, se puede realizar la reconstrucción de transcritos e isoformas a resolución de un nucleótido, lo cual ha ayudado a la identificación de transcritos no codificantes a través de muchos tipos celulares y tejidos (Rinn & Chang, 2012).

## **2.2 lncRNAs en plantas**

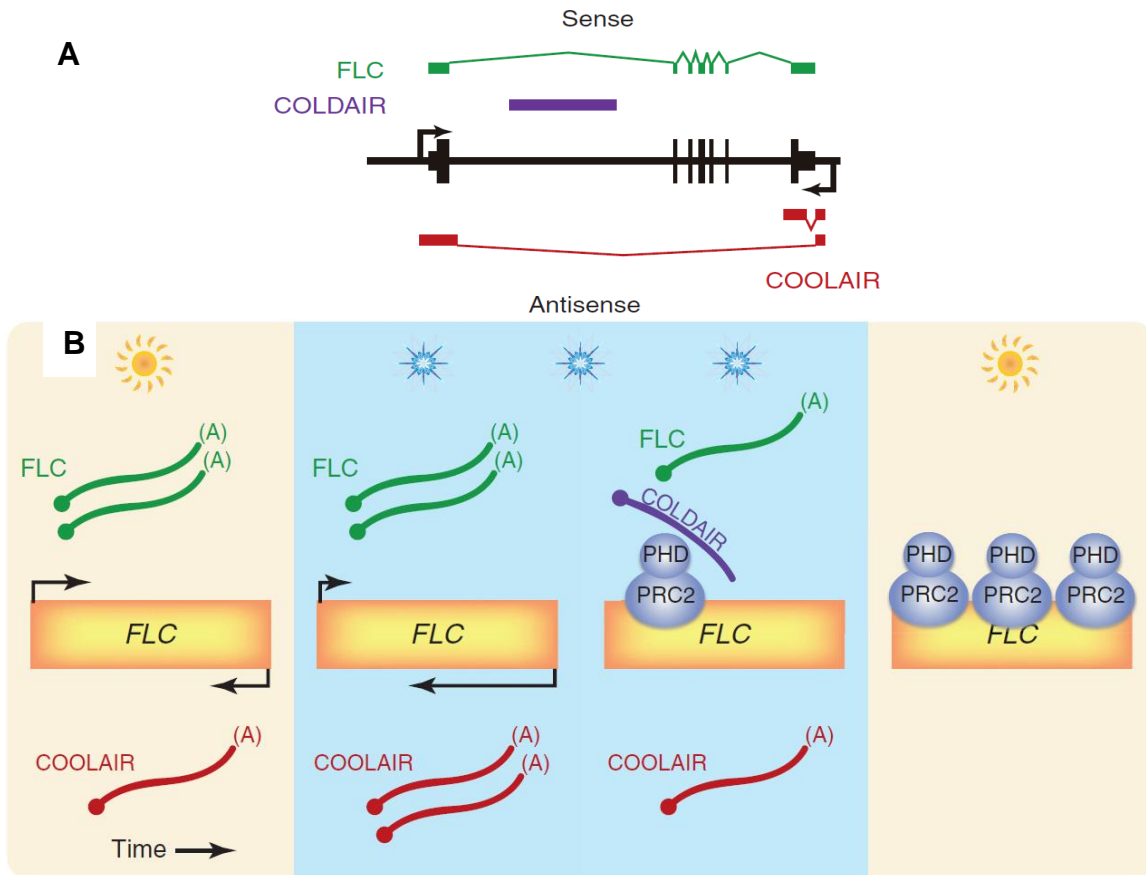
Los lncRNAs se han descubierto desde levaduras hasta organismos eucariotas más complejos (Liu et al., 2015). En un estudio transcriptómico en *A. thaliana* se identificaron aproximadamente 6,480 transcritos largos no codificantes, ya sean perfiles de órganos específicos o expresión inducida por estrés (Liu et al., 2015). Por otro lado, en un análisis exhaustivo de 200 sets de datos de *A. thaliana* se identificaron ~40,000 lncRNAs putativos, de los cuales más de 30,000 eran NAT (Naturales antisentido) y arriba de 6,000 eran lincRNAs (lncRNAs intergénicos) (Chekanova, 2015). Sin embargo, aún se conoce muy poco en plantas.

## 2.2.1 Funciones de lncRNAs en plantas

Principalmente los lncRNAs son regulados durante el desarrollo en plantas y por las condiciones ambientales, por lo que probablemente representen elementos funcionales del transcriptoma (Chekanova, 2015).

Algunos roles importantes descubiertos en plantas involucran el tiempo de floración, silenciamiento génico, organogénesis de raíz, fotomorfogénesis de plántulas y reproducción (Chekanova, 2015).

Entre algunos lncRNAs muy bien estudiados en plantas se encuentran *COOLAIR* y *COLD AIR* (Figura 3A). *COOLAIR* es un lncRNA antisentido que participa en el proceso de vernalización en plantas, involucrado en el silenciamiento epigenético del locus *FLC* (*FLOWERING LOCUS C*) inducido por frío, donde el lncRNA se asocia físicamente con el locus de *FLC* y acelera el apagado transcripcional de *FLC* (Csorba et al., 2014). *COOLAIR* incrementa de forma transitoria en etapas tempranas de vernalización, y variantes distales poliadeniladas de *COOLAIR* se superponen parcialmente con el extremo 5' del *FLC*. En el mismo contexto, *COLD AIR* es un transcrito intrónico no codificante de 1.1 kb de longitud, que es inducido transitoriamente durante la vernalización, donde en experimentos *in vitro* se comprobó que *COLD AIR* interactúa con la proteína CLF (*CURLY LEAF*), un homólogo del componente E(z) del complejo modificador de cromatina *PRC2* (Heo et al., 2013). Líneas en las que *COLD AIR* está deletado demostraron que el enriquecimiento de CLF y la marca H3K27me3 en la cromatina del *FLC* locus, se redujo durante la vernalización, indicando que este lncRNA es necesario para el reclutamiento del complejo *PRC2* hacia la cromatina de *FLC* (Figura 3B)(Heo et al., 2013).



**Figura 3.** *COOLAIR* y *COLDAIR*, lncRNAs expresados durante la vernalización. **A)** *COOLAIR* se transcribe de forma antisentido a lo largo de *FLC*. Es alternativamente poliadenilado, con un sitio de poliadenilación en sentido al sexto intrón y un sitio alejado de poliadenilación en sentido a la región promotora de *FLC*, expresado en condiciones de calor y frío. *COLDAIR* a diferencia solo se expresa en dirección sentido en el primer intrón en un ambiente frío, tiene cap pero no es poliadenilado. **B)** Silenciamiento epigenético de *FLC* durante vernalización. *FLC* se expresa en grandes cantidades que *COOLAIR*. Durante las primeras dos semanas de frío, *COOLAIR* es inducido, mientras que el ARN *FLC* no tiene splicing, comúnmente se piensa que refleja el sentido de la transcripción, desincrementa. Sin embargo, los niveles de ARNm siguen siendo similares a las condiciones no vernalizadas, posiblemente a las diferencias en los tiempos de vida entre los ARNs nacientes y procesados. Durante fríos prolongados *FLC* es silenciado epigenéticamente mediante el mecanismo dependiente de polycomb con una correlación de la expresión de *COLDAIR*. Ligeramente río abajo en sentido del promotor, el complejo *PHD-PCR2* reclutado se asocia a esta región conduciendo a una inducción de la marca de represión H3K27me3 en ese mismo sitio. Retornando a condiciones cálidas, el complejo *PHD-PCR2* se asocia a través del locus entero, como consecuencia, altos niveles de la marca represora se expanden a través del gen. Un alto nivel de marca H3K27me3 es necesaria para la estabilidad epigenética de *FLC* a través del resto del desarrollo (Chekanova, 2015 ; letswart et al 2012).

En plantas existe también el mecanismo de mimetismo o esponja para la regulación de funciones de miRNAs. Durante este mecanismo, las interacciones entre los miRNAs y sus blancos verdaderos son bloqueados por la unión de un lncRNA al miRNA, mediante complementariedad parcial de secuencias. Un clásico ejemplo de este mecanismo es el lncRNA endógeno *IPS1* (Induced by Phosphate Starvation 1), el cual se une a una secuencia complementaria de miR399 suprimiendo el corte y evitando la inhibición de los genes blanco miR399 (X. Liu et al., 2015a).

Durante estrés biótico se caracterizó el lncRNA *ELENA1* (ELF18-INDUCED LONG-NONCODING RNA1), involucrado en la respuesta inmune de la planta, mejorando la resistencia contra *Pseudomonas syringae*, donde las líneas deletadas de *ELENA1* redujeron la expresión del *PR1* (Pathogenesis-related gene1) siendo más susceptibles al patógeno (Seo et al., 2017). *ELENA1* interactúa con la subunidad *MED19a* (19a mediadora) y afecta el enriquecimiento de *MED19a* en el promotor de *PR1*, indicando que *MED19a* regula la expresión de *PR1* a través de *ELENA1* (Seo et al., 2017).

Otro lncRNA caracterizado en arroz fue el lncRNA *LDMAR* (long-day-specific male-fertility-associated RNA), de 1,236 bases de longitud, que participa en la regulación de la esterilidad masculina que es sensible al fotoperiodo (Ding et al., 2012). En este trabajo se encontró que una cantidad suficiente del transcrito *LDMAR* se requiere para el desarrollo normal del polen, en plantas que crecen bajo condiciones de días largos, donde una mutación espontánea (G-C), altera la estructura secundaria de *LDMAR* incrementando la metilación en su región promotora, lo cual a su vez ocasiona una reducción del transcrito provocando una prematura muerte celular programada en anteras bajo días prolongados, lo que finalmente causa esterilidad (Ding et al., 2012).

## 2.3 Interacción y competencia planta-planta durante la proximidad vecina

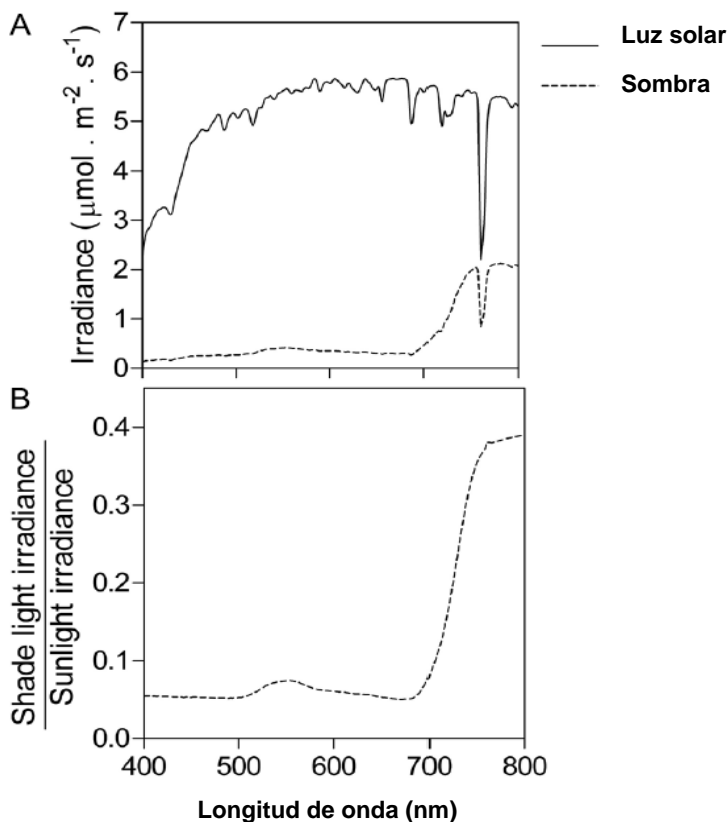
La competencia entre plantas determina la diversidad y abundancia de especies de las comunidades naturales, así como un buen rendimiento en sistemas agrícolas (Pierik, Mommer, & Voeselek, 2013). Se presume que la competencia por recursos es uno de los principales determinantes de la abundancia y distribución de especies en la vegetación natural, comunidades y ecosistemas (Pierik et al., 2013). El desempeño de la competitividad en plantas individuales se potencia una vez los vecinos próximos se han detectado (Pierik et al., 2013). Como organismos sésiles las plantas compiten a diario por los recursos de manera simultánea, lo cual es probable que diferentes señales sean percibidas, que necesitan ser integradas para producir un resultado óptimo multifacético (Pierik et al., 2013). Existen muchos mecanismos de detección a su disposición, tanto en la parte aérea y la parte subterránea de la planta para identificar vecinos próximos y su impacto en la disponibilidad de recursos. Un vasto número de ejemplos muy estudiados incluyen la calidad de luz, nutrientes y sustancias químicas tales como exudados de raíz y componentes volátiles, sin embargo, probablemente existan más señales, como lo es el mecanismo de detección de vecinos (Figura 4).



**Figura 4.** Las interacciones entre plantas durante la competencia. Entre estas conllevan múltiples señales, como la reflexión o transmitancia de luz roja lejana (FR), compuestos volátiles biogénicos, sombra, variación en la concentración de nutrientes y exudados solubles de raíz (Pierik et al., 2013).

### 2.3.1 Fenómeno de sombra y su prevención por SAS (Shade Avoidance Syndrome)

Cuando las plantas crecen en altas densidades, estas pueden detectar vecinos incrementando las señales de sombra proporcionadas por plantas, en donde las hojas superiores interceptan la radiación fotosintética que de otra manera tendría que llegar a las hojas inferiores (Casal, 2013; Roig-Villanova & Martínez-García, 2016). Las plantas dependen de captación de la radiación fotosintéticamente activa (RFA) que va de los 400-700 nm (Figura 5) para producir su alimento (carbohidratos), sin embargo, esta reducción por debajo de sus niveles de saturación compromete la fotosíntesis y disminuye su desempeño (Casal, 2012). Por lo tanto, esto ha proporcionado la fuerza evolutiva para activar las respuestas contra la sombra.



**Figura 5.** Diferencias entre la luz solar y la luz de sombra. **A)** Distribución del espectro de luz solar alcanzando el top o la base de una vegetación densa del dosel. **B)** Espectro de la proporción entre la irradiancia alcanzando la base del dosel (luz de sombra) y la irradiancia de la luz solar antes de hacer contacto con la vegetación (calculada después de a) (Casal 2012).



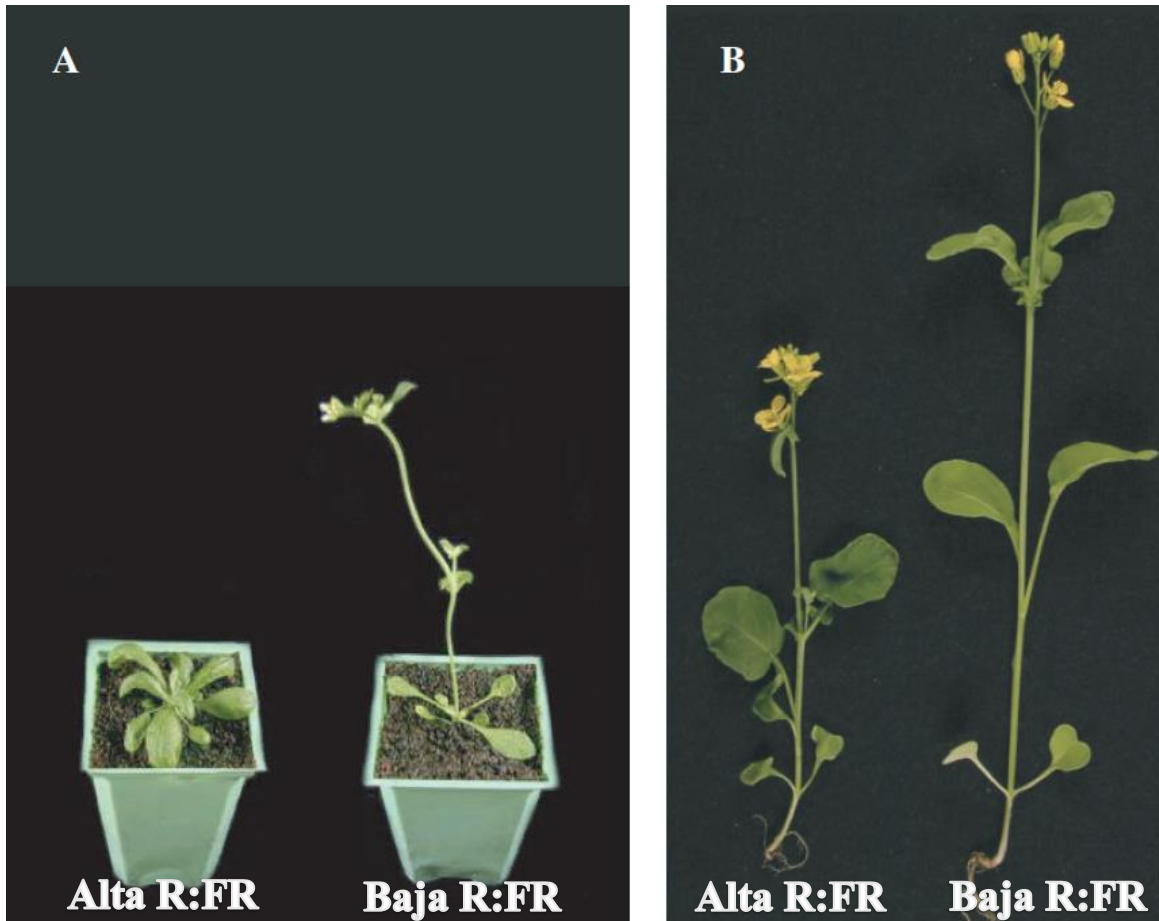
En condiciones de sombra, los niveles de luz roja y azul disminuyen, mientras que la luz roja lejana (FR) (700-800 nm), al no poder ser absorbida por la clorofila, es transmitida o reflejada (Pierik et al., 2013). Como consecuencia, la proporción luz roja/luz roja lejana (R/FR) es reducida en rodales densos (Pierik et al., 2013)

El nivel de sombra va a impactar varios procesos fisiológicos en la planta (Casal, 2013). El cambio de la composición del espectro de luz visible altera el balance entre los fotosistemas I y II (Casal, 2013). Los cambios en la temperatura afectan las reacciones metabólicas y reducen la demanda de agua (Casal, 2013).

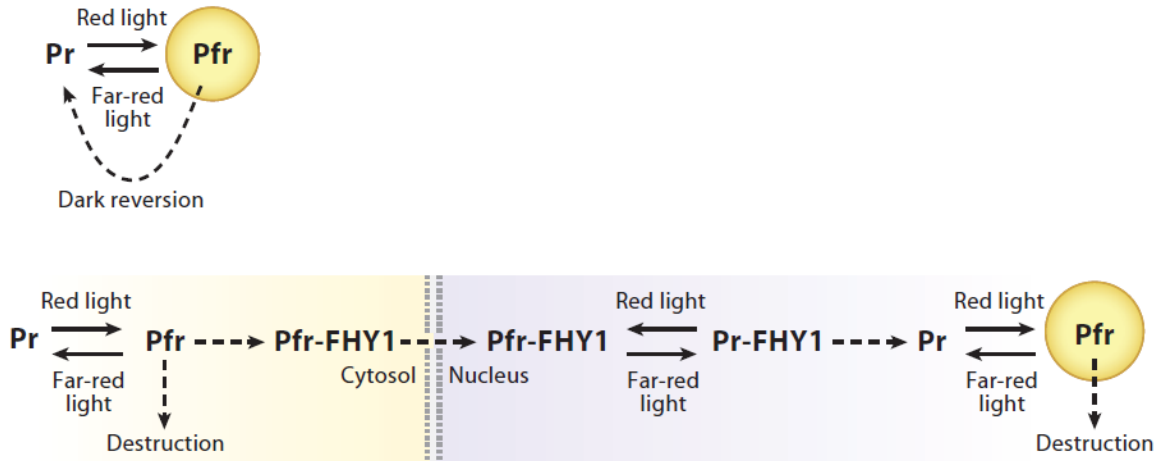
Debido a lo anterior algunas especies de plantas (intolerantes a la sombra) han desarrollado mecanismos de plasticidad fenotípica denominados SAS (Shade Avoidance Syndrome), que le han permitido responder a la presencia de vecinos y escapar de la sombra (Pierik et al., 2013). Estas especies, antes de ser alcanzadas por la sombra directamente, desencadenan respuestas de crecimiento enfocadas a alcanzar la luz solar no filtrada, escapando de un ambiente no propicio (Kohnen et al., 2016). El SAS es caracterizado por respuestas en el crecimiento y desarrollo de distintos órganos de la planta (Kohnen et al., 2016). Los órganos que demuestran un fenotipo marcado durante SAS, son los tallos o en su defecto hipocótilos, los cuales demuestran una elongación, en el caso de peciolo, estos adoptan una posición vertical (hiponastía) y se alargan, mientras que en hojas laminares o cotiledones hay una reducción y se inhibe la ramificación (Figura 6) (Kohnen et al., 2016; Franklin & Whitelam, 2005).

Esta percepción de distintas longitudes de onda es gracias a que la planta posee fotorreceptores que le permiten percibir los cambios en la R/FR lo cual indica que hay proximidad vecina o sombra, existiendo una familia pequeña de genes que codifican cinco miembros (*phyA-phyE*) en *A. thaliana* (Roig-Villanova & Martínez-García, 2016). Los fitocromos existen en dos formas fotoconvertibles en una reacción reversible: una forma inactiva (Pr) que absorbe luz roja y otra activa (Pfr) que absorbe luz roja lejana (Figura 7) (Roig-Villanova & Martínez-García, 2016). Durante una alta R/FR (densidad vegetal baja), el fotoequilibrio es desplazado hacia la forma Pfr y la respuesta por SAS es suprimida, mientras que en baja R/FR (alta

densidad y sombra) el fotoequilibrio es desplazado hacia la forma inactiva induciéndose SAS (Roig-Villanova & Martínez-García, 2016). Los fitocromos A y B (*phyA* y *phyB*) desempeñan los papeles más importantes que el resto de los fitocromos en la señalización por sombra (Duek & Fankhauser, 2005).



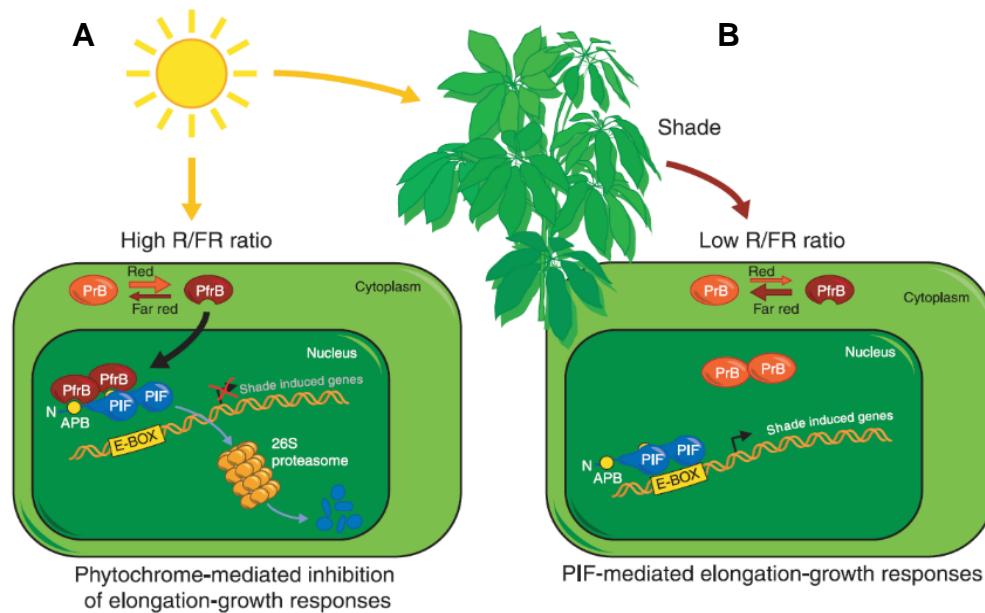
**Figura 6.** La respuesta fenotípica por SAS según la relación R:FR. El fenotipo de **A)** *A. thaliana* y **B)** *Brassica rapa* crecidas bajo condiciones de alta R:FR y baja R:FR. Las plantas se crecieron bajo luz blanca fluorescente proporcionándole radiación fotosintéticamente activa (400-700 nm). A cada especie de la derecha se le proporciono de FR (luz roja lejana), para disminuir su relación R:FR (Franklin & Whitelam, 2005).



**Figura 7.** Las reacciones de luz (líneas sólidas) y oscuridad (línea punteada), definen la abundancia de *phyB* Pfr activo (círculo amarillo) en Arabidopsis. Las reacciones de luz (líneas sólidas) y oscuridad (línea punteada), definen la abundancia de *phyA* Pfr activo (círculo amarillo) en el núcleo en Arabidopsis (Casal, 2013).

Antes de la foto excitación, la proteína *phyB* es soluble en el citoplasma, sin embargo en su conformación Pfr, *phyB* se acumula en el núcleo, donde sugieren que regula la expresión génica (Allen, Duek, Whitelam, Fankhauser, & Building, 2008). En su estado Pfr se ha reportado que *phyB* se puede unir físicamente para interactuar con un grupo de factores de transcripción conocidos como factores de transcripción PIF (Phytochrome Interacting Factors) del tipo bHLH (Basic Helix Loop Helix) que interactúan con *phyB* actuando como reguladores negativos de la vía *phyB* (Allen et al., 2008). La conformación Pfr ya sea del *phyA* y/o *phyB* interactúa con varias proteínas PIFs (1,2,4 y 5), induciendo una rápida fosforilación y degradación de los factores de transcripción (Figura 8A) (Leivar et al., 2008).

Hasta que los fotorreceptores son inactivados por la detección de una baja R/FR, los PIFs actúan colectivamente de una forma redundante y ahora pueden unirse a las regiones promotoras de genes que codifican factores que promueven el crecimiento para mejorar su transcripción, y así constitutivamente promover la skotomorfogénesis (represión de la fotomorfogénesis) en plántulas que crecen en la oscuridad (Figura 8)(Leivar et al., 2012).



**Figura 8. a)** En alta incidencia de luz, *phyB* se encuentra en la forma Pfr. Esta se acumula en el núcleo donde interactúa con PIFs mediante un dominio de unión del fitocromo activo. La interacción da lugar a la degradación de los factores PIF mediante el proteosoma, desactivando a los genes inducidos por sombra. **b)** Durante la sombra, la vegetación es enriquecida en luz FR, y *phyB* es convertido en Pr, el cual ya no más interactúa con PIF. Estas proteínas se reacumulan, promoviendo la expresión de genes tales como *PIL1*, *ATHB2* y *HFR1* (Lorrain, 2008).

Plantas en estadio de plántulas han demostrado responder a la sombra a nivel de hipocótilos/epicótilos, cotiledones y hojas primarias (Roig-Villanova & Martínez-García, 2016). Se ha concluido que la respuesta por SAS es regulada al menos en parte por la interacción de los *phy* activos con los PIFs, los cuales resultan en cambios rápidos en la expresión de docenas de genes *PAR* (Phytochrome rapidly regulated), necesarios para la implementación de las respuestas SAS (Roig-Villanova & Martínez-García, 2016).

Se han reportado múltiples respuestas por hormonas involucradas en la morfogénesis regulada por sombra. Tratamientos donde simulaban condiciones de sombra mostraron un rápido y transitorio incremento en los niveles endógenos de auxina (Roig-Villanova & Martínez-García, 2016). Entre *PIF4*, *PIF5* y *PIF6* se reportó que regulan de manera directa la expresión inducida por sombra de algunos genes *YUC*, que codifican para proteínas parecidas a una flavin monooxigenasa y catalizan el paso de la tasa limitante en la biosíntesis de IAA; en donde además el incremento

regulado por sombra de auxinas se redujo en dobles mutantes de *pif4pif5* y en la mutante sencilla *pif7* de plántulas (Roig-Villanova & Martínez-García, 2016). En un experimento de plántulas de *Arabidopsis thaliana* expuestas a sombra simulada espaciotemporalmente, durante cuatro tiempos (15,45,90 y 180 min), se enfocaron a términos de GO (Gene Ontology) asociados con hormonas (Kohnen et al., 2016). De acuerdo con la presencia temprana de auxinas, se reportó que los términos de respuesta a auxinas fueron representados significativamente entre genes sobre expresados tanto en hipocótilos y cotiledones durante los cuatro tiempos (Kohnen et al., 2016). Las categorías GO de respuesta a brasinoesteroides mostraron una sobrerrepresentación similar en los dos tejidos durante los últimos tres tiempos (Kohnen et al., 2016). Cuatro genes *YUC* incrementaron durante los primeros 15 minutos de sombra en cotiledones, mientras que la expresión de *YUC3* y *YUC8* incrementó en hipocótilos después de la exposición a sombra (Kohnen et al., 2016).

Para una respuesta inducida por sombra para un crecimiento total en hipocótilo se requieren de los brasinoesteroides. Durante el crecimiento de una planta bajo el dosel, se interrumpió la biosíntesis de brasinoesteroides con la mutante *diminuto/dwarf1*, en donde el hipocótilo fue incapaz de mostrar un fenotipo elongado bajo circunstancias de sombra. Enzimas *XTH* (Xiloglucano Endotransglucosilasa/Hidrolasa) representan un punto de control para la elongación celular y la expresión de varios genes *XTH* promovida por baja luz azul y la respuesta de algunos de estos genes requieren de la percepción de auxinas y síntesis de brasinoesteroides (Casal, 2012).

Por otra parte, las giberelinas también intervienen en la promoción del crecimiento de hipocótilos durante baja R/FR, en donde la biosíntesis de giberelinas se ha visto afectada al usar paclobutrazol (inhibidor), mientras que por otro lado se han hecho ensayos con mutantes deficientes en la biosíntesis de esta hormona. Genes relacionados a la síntesis de giberelinas como *GAI* (GA insensitive), y *GA 20-oxidase 3* mostraron una fuerte expresión en respuesta a baja R/FR (Casal, 2012).

### 2.3.2 LncRNAs en estímulos de luz y fotomorfogénesis

En ensayos de microarreglos de plantas etioladas de *A. thaliana* Col-0, plántulas *wt* y plántulas desetioladas sometidas a continua luz blanca por 1 y 6 horas, en donde se disectaron en cotiledones, hipocótilos y raíz para tener perfiles órgano-específicos (H. Wang et al., 2014). Un total de 13, 907 lncRNAs con respuesta a luz blanca se identificaron (H. Wang et al., 2014).

Por otro lado, se ha caracterizado un lncRNA de 236 nt que es un regulador positivo de la fotomorfogénesis (Y. Wang et al., 2014). *HID1* (Hidden Treasure 1) posee cuatro loops en su estructura secundaria, de los cuales dos horquillas principales son importantes para la modulación del crecimiento en hipocótilos en plántulas que crecen bajo cR (luz roja continua)(Y. Wang et al., 2014). Se sugiere que *HID1* se asocia con la cromatina y regula el control transcripcional de *PIF3* uniéndose a su región promotora bajo cR (Y. Wang et al., 2014). También se reporta una región altamente conservada aproximadamente de 74 nt en cinco especies de plantas de taxas dispares (*A. thaliana*, *A. lyrata*, *O. sativa*, *P. patens*, *S. moellendoffii*) (Y. Wang et al., 2014).

### **III Hipótesis**

Existen lncRNAs que participan en la regulación de la respuesta a sombra por proximidad vecina incida por luz roja lejana en *A. thaliana*.

### **IV Objetivos**

#### **4.1 Objetivo general**

Identificar y anotar lncRNAs expresados diferencialmente durante el fenómeno de proximidad vecina en *Arabidopsis thaliana*.

#### **4.2 Objetivos específicos**

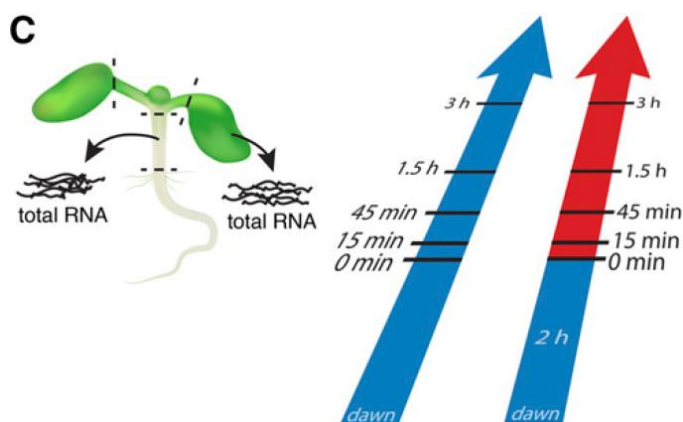
- Ensamblar transcriptomas espacio-temporales de *A. thaliana* durante la exposición baja y alta a R/FR.
- Identificar lncRNAs y genes codificantes en el transcriptoma de *A. thaliana*.
- Clasificar en base a su posición genómica los lncRNAs identificados.
- Determinar los lncRNAs y genes codificantes expresados diferencialmente durante SAS, así como su co-expresión con genes codificantes.
- Determinar enriquecimientos de términos GO en los genes expresados diferencialmente.
- Construir redes de co-expresión combinado con enriquecimiento de términos GO para inferirle una posible función a los lncRNAs diferencialmente expresados.

## V Metodología

### 5.1 Diseño Experimental

Las librerías se obtuvieron del trabajo Neighbor detection induces organ-specific transcriptomes, revealing patterns underlying hypocotyl-specific growth (Kohnen et al., 2016), donde se crecieron plántulas *A. thaliana* ecotipo Col-0 y se recolectaron hasta 5 días de edad, 16 h: 8 h en un fotoperiodo luz/obscuridad a 21° C. En el día sexto, las plántulas se mantuvieron en alta R/FR (control) o baja R/FR (tratamiento de sombra) según el caso del tratamiento. Los tiempos de exposición para cada tratamiento de luz fueron 0, 15, 45, 90 y 190 min donde posteriormente las plántulas se fijaron en acetona enfriada para llevar a cabo la disección de cotiledones e hipocótilos. Se muestrearon aproximadamente 50 plántulas y por cada punto de tiempo y condición de luz se prepararon dos duplicados. Después se procedió a extraer el RNA total empleando el RNAeasy kit y el protocolo de TruSeq. Por último, las librerías de RNA-seq se secuenciaron usando un secuenciador HISEQ 2500 (Illumina). Se generaron un total de 57 librerías direccionadas (*stranded*) con tres réplicas cada tratamiento (Figura 9).

**Figura 9.** Representación esquemática de la extracción de ARN para el análisis mediante RNA-seq. Se muestrearon plántulas en el día sexto en ambas condiciones por separado después del tiempo indicado en alta (azul) o baja F/FR. Las líneas punteadas representan los cortes realizados para delimitar los tejidos.





## 5.2 Preparación de librerías para ensamblado

Las 57 librerías se descargaron de GEO (Gene Expression Omnibus) con el ID de acceso GSE81202. Se examinó la calidad inicial de las secuencias crudas con el software visualizador FastQC (v 0.11.5) (Andrews, 2014). Posteriormente se utilizó Trimmomatic (v 0.32) (Bolger, Lohse, & Usadel, 2014), una herramienta para el corte y remoción de adaptadores para datos de Illumina (Fastq) empleando la siguiente línea de comando:

```
$ ILLUMINACLIP TrueSeq3-SE.fa:2:30:10 HEADCROP:10 SLIDINGWINDOW:4:15 MINLEN:80
```

Finalmente se examinaron nuevamente las secuencias con FastQC para corroborar la calidad de estas.

## 5.3 Ensamble de transcriptomas

RNA-seq es una potente herramienta para la identificación de lncRNAs tanto para su identificación y cuantificación, ya que a diferencia de los enfoques basados en microarreglos, RNA-seq trabaja a escalas completas del genoma, a resoluciones de nucleótidos únicos y no se limita a secuencias ya conocidas (Lee & Kikyo, 2012).

Para la identificación y localización de lncRNAs *bona fide* en cualquier genoma, es necesario realizar un ensamble de las lecturas de RNA-seq.

Para el ensamble de transcriptomas se optó por usar dos tipos de ensambladores: uno *de novo* y otro utilizando un genoma de referencia.

### 5.3.1 Ensamble *de novo* (Trinity)

Para el ensamble completo de transcriptomas se utilizó Trinity (Grabherr et al., 2011), que ensambla secuencias de transcritos a partir de datos Illumina de RNA-seq. Trinity divide los datos de secuencias en grafos de Bruijn, donde cada grafo representa la complejidad transcripcional en un gen, después se procesa cada gráfico de forma independiente para extraer isoformas enteras de splicing y separar transcritos derivados de genes parálogos.

En nuestro caso se optó por usar la opción de ensamble guiado ya que se cuenta con el genoma de referencia de TAIR10. Este parámetro primero alinea las lecturas al genoma, seguido por el ensamble *de novo* en cada locus.

Para iniciar el ensamble guiado, se concatenaron todas las librerías en una sola, partiendo con Bowtie2 (Langmead, Trapnell, Pop, & Salzberg, 2009), un alineador que indexa el genoma con un índice FM (Full-text-minute size) para mantener una huella de memoria pequeña y posteriormente realizar el mapeo de lecturas al genoma de referencia. Una vez creado el índice con Bowtie2, para realizar el mapeo al genoma se utilizó Tophat (Trapnell, Pachter, & Salzberg, 2009), donde se alineó la librería concatenada al genoma de TAIR 10 de *A. thaliana* con los parámetros siguientes:

```
$ tophat -i 20 -I 1000 --read-edit-dist 3 --read-realign-edit-dist 0 --library-type firststrand -g 1
```

Finalmente se ejecutó el comando de Trinity con la opción de ensamble guiado tomando como input el archivo bam generado del mapeo anterior. Para esto, se ordenaron los alineamientos por las coordenadas más a la izquierda utilizando samtools y la opción sort. Ya ordenado el archivo bam, se procedió a ensamblar el transcriptoma utilizando Trinity con las opciones siguientes:

```
$ Trinity -genome_guided_bam -jaccard_clip -genome_guided_max_intron 1000
```

### 5.3.2 Ensamble utilizando un genoma de referencia (Cufflinks)

Para este tipo de ensamble se ejecutó el pipeline *tuxedo suit*, el cual se resume en tres pasos principales: Bowtie (generación de índice FM), Tophat (alinear secuencias) y Cufflinks (ensamblador) (Trapnell et al., 2010). A diferencia del método anterior, en este tipo de ensamble se alineó con Tophat cada librería de manera individual, utilizando los mismos parámetros ya mencionados anteriormente. Para ensamblar el siguiente transcriptoma se utilizó Cufflinks (v 2.2.1) con la salida *accepted\_hits* (archivo bam) resultante de la alineación con los

siguientes parámetros: `–overlap-radius 1 –library-type fr-firststrand`. Finalmente, las coordenadas gtf del archivo de salida del ensamblado con cufflinks, se reunieron en un solo archivo de texto para realizar la unificación de todas las librerías con el programa cuffmerge.

## 5.4 Análisis de la calidad de los transcriptomas

Posteriormente, se realizaron los análisis para examinar la integridad de los ensamblados empleando herramientas como BUSCO (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015), RNAquast (Bushmanova et al., 2016) y una evaluación visual a través del programa IGV.

## 5.5 Identificación de lncRNAs

Para la identificación de lncRNAs a través de nuestro transcriptoma se optó por utilizar dos métodos independientes: Coding Potential Calculator (Kong et al., 2007) y el segundo es pipeline más estricto y robusto propuesto por (Cabili et al., 2011).

### 5.5.1 CPC (Coding Potential Calculator)

CPC es un método clasificador basado en una máquina de soporte de vectores, el cual como sus siglas lo indican es una herramienta para calcular el potencial que tiene un transcrito para codificar una proteína, que reúne seis características biológicamente importantes de secuencias que se mencionan enseguida: **log-odd score, cobertura del ORF predicho, la integridad del ORF, número de hits hacia una base de datos de secuencias proteínicas, hit score y frame-score**. Todo lo anterior se integra a un clasificador de aprendizaje de máquinas (machine learning). Mencionado lo anterior, el transcriptoma en formato fasta se empleó como entrada para la predicción de lncRNAs ejecutando la siguiente línea de comando de CPC (v 0.9 r2):

```
$ run_predict.sh cufflinks.fa result_in_table cpc result_evidences
```

Del archivo de salida *result\_evidences*, se filtraron los transcritos por número de nucleótidos y tipo de transcrito (codificante o no codificante).

### 5.5.2 Pipeline estricto (Cabili)

Esta tubería está basada en la propuesta de Cabili et al 2011. No es automatizado como el clasificador anterior, por lo tanto, para discriminar secuencias codificantes de nuestro transcriptoma se realizaron los siguientes filtros: 1) se descartaron todos los transcritos menores a 200 nucleótidos, 2) se extrajeron los marcos de lectura abiertos (ORFs) de cada transcrito y 3) Se mantuvieron los ORF menores a 75 nucleótidos, empleando una herramienta de Biopython llamada *extract\_orf* (Cock et al., 2009).

Se realizaron alineamientos con BLAST (Altschul et al., 1990), para secuencias nucleotídicas y secuencias peptídicas con los siguientes parámetros:

```
$ blastx -query cufflinks.fa -db nr -num_threads 8 -max_target_seqs 1 /  
-outfmt 6 > blastx_cufflinks.outmt6  
$ blastp -query cufflinks.fa -db nr -num_threads 8 -max_target_seqs 1 /  
-outfmt 6 > blastp_cufflinks.outmt6
```

También se calculó a partir de las secuencias peptídicas, que transcritos podrían llegar a tener péptido señal mediante el programa *signal P* (Nielsen & Krogh, 1998)

```
$ signal -f short -n signalp.out all_orf_protein.fa > signal.log
```

Mientras que con *tmhmm* calculamos si las secuencias peptídicas poseían hélices transmembranales (Krogh et al., 2001).

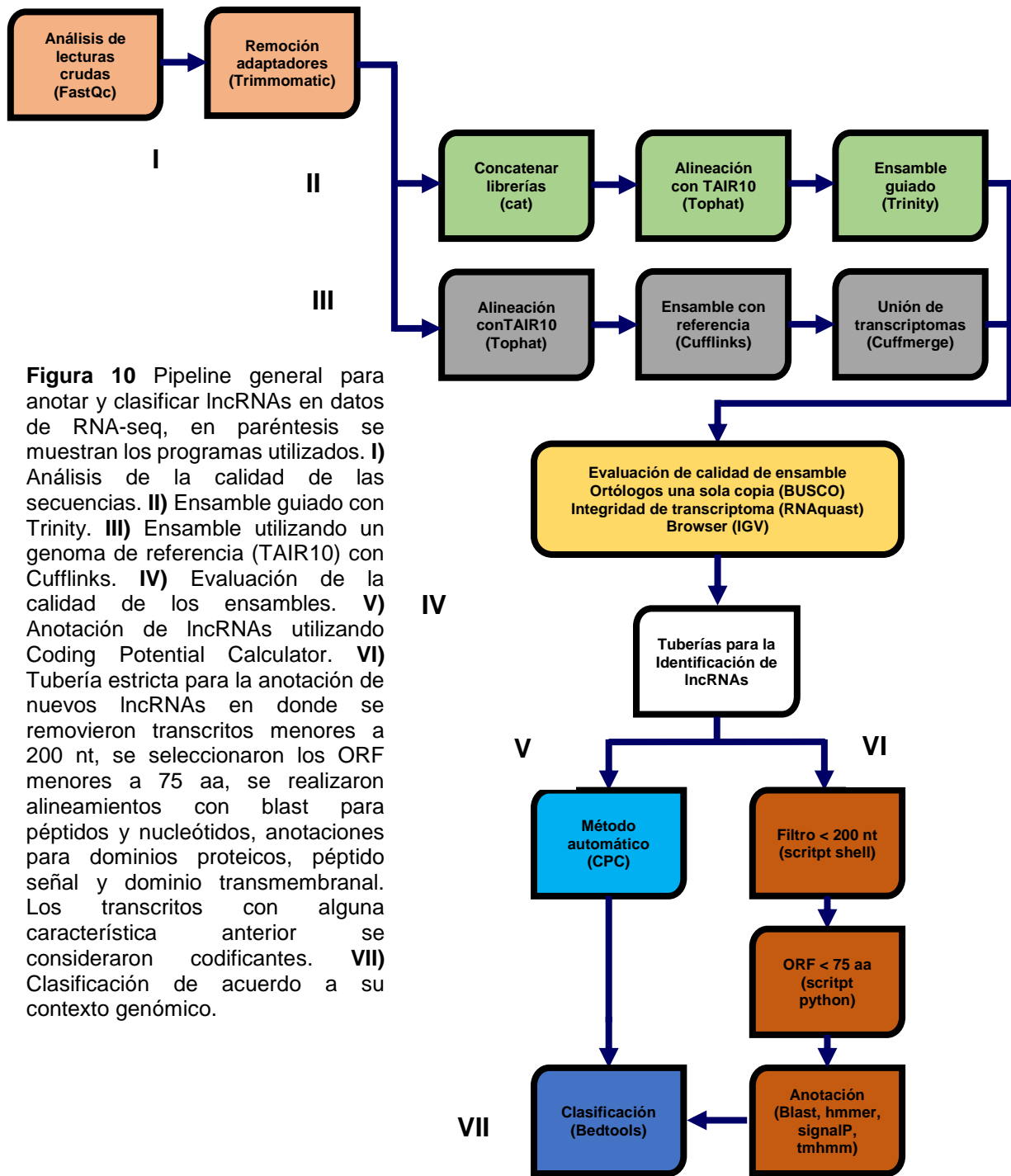
```
$ tmhmm -short < all_orf_protein,fa > tmhmm.out
```

Por otro lado, se determinaron los dominios proteicos con la herramienta *hmmscan* (Finn et al., 2011), utilizando una base de datos de Pfam mediante la siguiente línea de comando :

```
$ hmmscan -cpu 8 -domtblout TrinotatePFAM_aorf.out Pfam-A.hmm /
```

```
all_orf_protein.fa > pfam_aorf.log
```

Finalmente se creó una tabla del tipo binario donde las filas corresponden a los transcritos y las columnas son cada anotación característica de un gen codificante en donde 1 indica presencia y 0 indica ausencia. Seleccionamos aquellos transcritos sin ningún tipo de anotación. Por último, se unieron las listas finales de lncRNAs tanto de CPC y el pipeline de Cabili con el fin de obtener una lista absoluta con todos los lncRNAs identificados. En la Figura 10 se muestra un breve diagrama de flujo de la tubería empleada.



## 5.6 Clasificación de lncRNAs

Para la clasificación de lncRNAs con respecto a su posición en el genoma, se ejecutó el programa bedtools (Quinlan & Hall, 2010), el cual es una herramienta que nos permite realizar intersecciones genómicas, donde el archivo de entrada es una tabla de coordenadas con 12 columnas (bed12), para esto, es necesario dar formato a nuestro archivo gtf. Los transcritos se clasificaron en: superpuestos con genes codificantes, proximales intergénicos a genes codificantes (lincn500), intergénicos alejados 500 nt de un gen codificantes (lincw500) e intrónicos.

Para clasificar los lncRNAs que están superpuestos con genes codificantes, se intersectó la lista total de lncRNAs contra la lista de genes que codifican proteínas extraída de:

```
$ bedtools intersect -a lncrnas.bed -b igv_tair10_prot_bed12.bed > overlap.bed
```

Para clasificar los lncRNAs intergénicos primero se separaron los genes no codificantes que no intersectaron con los genes codificantes

```
$ bedtools intersect -a lncrnas.bed -b igv_tair10_prot_bed12.bed -v > linc.bed
```

Una vez teniendo la lista total de genes no codificantes en regiones intergénicas, se optó por utilizar una ventana de 500 bases para separar los lncRNAs en dos clases de transcritos intergénicos: los proximales que son los que cayeron dentro de la ventana de 500 bases ya sea en el extremo 5' o 3' de una región codificante, mientras que los transcritos que no caen en esa ventana se denominaron intergénicos:

```
$ bedtools window -a linc.bed -b TAIR10_GFF3_bed12_prot.bed -w 500 -v /  
> linc_w500_bed12.bed  
$ bedtools window -a linc.bed -b TAIR10_GFF3_bed12_prot.bed -w 500 /  
> linc_near500_bed12.bed
```

Para identificar los transcritos intrónicos se utilizó un script en perl para transformar las regiones intrónicas del genoma en exones, de este modo podremos

calcular las intersecciones dentro del exón (intrón) y clasificarlo como lncRNA intrónico:

```
$ perl get_intron_gtf_from_bed.pl input.bed output.gtf
$ bedtools intersect -a lncrnas.bed -b TAIR10_GFF3_introns.bed -f 1 -r /
> intronic.bed
```

## 5.7 Análisis de expresión diferencial de genes codificantes y lncRNAs

Para calcular la abundancia relativa de la expresión génica durante el fenómeno de sombra, primero se contaron las lecturas de cada muestra empleando cada archivo bam generados por nuestra alineación con tophat. Esto se realizó con la herramienta htseq, la cual realiza un conteo de las lecturas que alinean a cada gen anotado utilizando nuestro transcriptoma en formato gtf como referencia (Anders, Pyl, & Huber, 2015).

```
$ htseq-count -f bam -s no -r name sort_sample.bam cufflinks.gtf /
> sample.counts
```

El resultado fue una columna de cuentas para cada gen en cada muestra. Las columnas de cuentas se unieron para generar una matriz de cuentas.

```
$ paste $( ls *counts | grep cufflinks -v) | cut -f 1,2,4,6,8,10,12,14,16,18 /
20,22,24,26,28,30,32,34,36,38,40,42,44,46,48,50,52,54,56,58,60,62,64,66,68,70/
72,74,76,78,80,82,84,86,88,90,92,94,96,98,100,102,104,108,110,112 /
| grep '__' -v > cufflinks_counts_table.txt
```

Los nombres de las filas corresponden a todos los genes alineados a nuestro transcriptoma (incluyendo los alineados a TAIR10 y los alineados a los nuevos genes ensamblados), mientras que las columnas son las lecturas alineadas para cada uno de los tratamientos.



El análisis de expresión diferencial se realizó con DESeq2 (Love, Anders, & Huber, 2014), utilizando nuestra matriz de cuentas para generar un objeto DESeq2. Los tejidos se analizaron por separado utilizando un modelo de interacción tratamiento:tiempo. Se generaron cuatro contrastes para o con esta interacción de tiempos continuos a lo largo del tiempo del experimento, es decir: 0 vs 15, 15 vs 45, 45 vs 90 y 90 vs 180 min. Se consideraron genes diferencialmente expresados los que tenían un FDR < 0.01 en cada contraste.

## **5.8 TopGO**

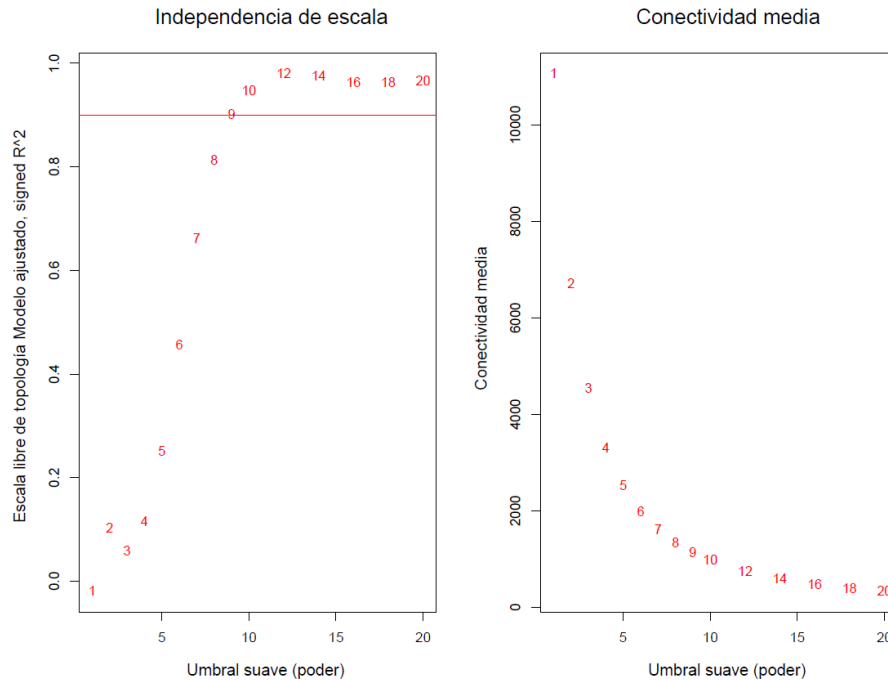
Los análisis de enriquecimiento funcional se efectuaron con topGO (Alexa & Rahnenfuhrer, 2016), utilizando la anotación de org.At.tair.db y los resultados del análisis de expresión diferencial, tomando en cuenta los términos anotados con un p value < 0.01 mediante una prueba de clásica de Fisher.

## **5.9 Construcción de la red de co-expresión (WGCNA) *A. thaliana***

Debido a que no existen categorías funcionales anotadas para lncRNAs, se optó por usar un enfoque llamado guilty by association (culpable por asociación), el cual consiste en agrupar lncRNAs de acuerdo con cómo su presencia (co-expresión) correlaciona con la expresión de genes codificantes, lo cual nos puede dar pistas sobre los procesos celulares en que se involucran estos transcritos y de este modo poder inferir una posible función.

Para lo anterior fue necesario la construcción una red de co-expresión génica utilizando WGCNA (Langfelder & Horvath, 2008). Se utilizó la tabla de cuentas generada con htseq para expresión diferencial de las 57 librerías de RNA-seq. La transformación de nuestra tabla se llevó a cabo con VST (Variance Stabilizing Transformation) de DESeq2 y posteriormente se realizó un filtrado por expresión media de los genes normalizados utilizando una métrica denominada MAD (Median Absolute Deviation), para eliminar los genes con poca variación en cada una de las muestras.

La función de adyacencia se ponderó por el poder de la correlación entre los diferentes genes. El parámetro ponderado  $\beta$  fue determinado por la ley de redes de escala libre. Para asegurar la conectividad promedio de la red sea continua, escogimos el valor de  $\beta$  de acuerdo con el gráfico de diagnóstico. Se escogió un valor  $\beta = 12$ , el cual es el valor más bajo para el cual la curva de índice de topología sin escala se mantiene estacionaria al alcanzar un valor de 0.9 (Figura 11).

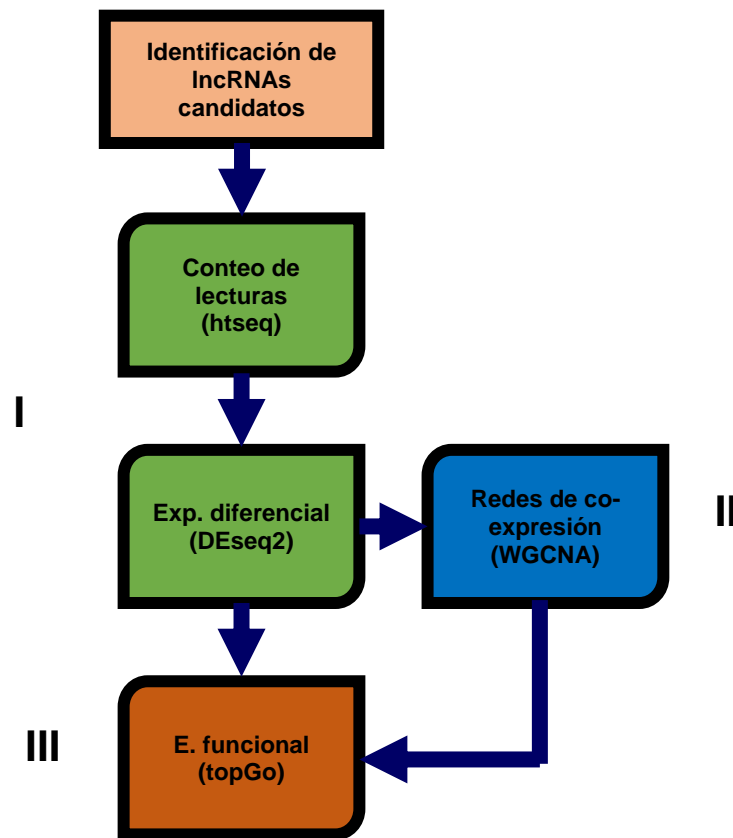


**Figura 11.** Análisis de la topología de red para varias potencias de umbral suave.

Una red ponderada del tipo *signed* se construyó utilizando el modo automático, creando una matriz mediante una correlación bicor (biweight midcorrelation) con la opción de módulos separados (unmerged). Genes con alta similitud se agruparon en el sobrepuesto topológico. Los perfiles de expresión de cada módulo se representaron por su componente principal (module eigengene), el cual nos indica la variación de los niveles de expresión en cada módulo.

Se realizó un análisis de enriquecimiento de categorías GO con topGO para cada módulo co-expresado con funciones de tipo celular, proceso biológico y función molecular tomando en cuenta los términos GO significativamente anotados si el *padj* fue  $< 0.01$ .

En la figura se describe de manera general la tubería utilizada para la identificación de lncRNAs candidatos para su futura validación (Figura 12).



**Figura 12.** Diagrama de flujo para encontrar lncRNAs candidatos. **I)** En verde se representan los análisis de expresión diferencial. **II)** En azul el análisis de redes de co-expresión. **III)** En naranja el Enriquecimiento funcional de genes diferencialmente expresados y de módulos co-expresados. En paréntesis se muestra la herramienta utilizada.

## 5.10 Generación gráfica de redes para cada módulo

Para cada gen de cada módulo se calculó el kme, la cual es una medida de conectividad basada en el eigengene, definida como la correlación entre el perfil de expresión del gen estudiado y del eigengene.

$$K_{cor,i}^{(q)} := \text{COR}(x_i, E^{(q)}).$$

Donde  $x_i$  es el perfil del nodo  $i$  y  $E^{(q)}$  es el módulo eigengene del módulo  $q$ . La métrica de miembros de módulo  $K_{cor,i}^{(q)}$  se encuentra en  $[-1,1]$ , y especifica que tan cercano está el nodo  $i$  al módulo  $q$ .  $q = 1, \dots, Q$ .

Se tomó como corte un kme menor al tercer cuartil y un umbral de adyacencia de 0.1 para cada módulo (debido a que existen módulos más grandes que otros, es decir que contienen más genes), usando la función `wgcna2igraph` para convertir la red en objeto `igraph` y poder visualizar los nodos representativos.

## VI Resultados

### 6.1 Ensamble de transcriptomas

Un total de 1,078,879,782 lecturas crudas en 57 librerías se filtraron por calidad y posteriormente alinearon al genoma de TAIR10. Estos alineamientos se utilizaron para realizar ensambles de transcriptoma *de novo* guiado (Trinity) y guiado por genoma (Cufflinks).

Para evaluar que tan completo estaba cada transcriptoma ensamblado, se utilizó BUSCO (Benchmarking Universal Single-Copy Orthologs), con la base de datos de embriofitas con 1,440 ortólogos de una sola copia para ambos ensambles. En el caso del transcriptoma de Trinity se encontró un 97.1% de BUSCOs completos, mientras que el transcriptoma de Cufflinks demostró tener un mayor porcentaje (99.5%) de BUSCOs completos en comparación al transcriptoma de Trinity. Adicionalmente, el ensamble de Trinity presentó una cantidad mayor de BUSCOs fragmentados y perdidos, 1.7% y 1.2% respectivamente en comparación con el ensamble de Cufflinks, dónde fueron 0.2% y 0.3% (Tabla 1).

Métricas BUSCO	Trinity	Cufflinks
BUSCOs completos	1,398 (97.1%)	1,433 (99,5%)
BUSCOs completos y una sola copia	945 (65.6%)	599 (41.6%)
BUSCOs completos y duplicados	453 (31.5%)	834 (57.9%)
BUSCOs fragmentados	24 (1.7%)	3 (0.2%)
BUSCOs ausentes	18 (1.2%)	4(0.3%)
Grupos BUSCO examinados	1440 (100%)	1440 (100%)

**Tabla 1** Métricas de BUSCO para Trinity y Cufflinks

En nuestro análisis por RNAquast para cada transcriptoma, se calcularon algunas métricas que se resumen a continuación (Tabla 2):

Un total de 54,286 transcritos se ensamblaron por medio de Trinity, mientras que el ensamble de Cufflinks ensambló 71,920 transcritos. Cufflinks posee más transcritos mayores 500 pares bases (66,579 transcritos) en comparación con Trinity (27,675 transcritos), lo cual resulta favorable para la anotación de RNAs largos no

codificantes. En transcritos alineados por medio de blat, de los 71,920 transcritos ensamblados, en Cufflinks se alinearon al genoma 71,086 mientras que Trinity se alienaron solo 53,868 transcritos de los 54,286 ensamblados. Cufflinks presentó un menor promedio de mismatches en contraste con Trinity, mientras que Trinity presentó un mayor número de transcritos quiméricos (Cufflinks tuvo ausencia de estos). La cobertura (sensibilidad) al genoma de TAIR10 fue de 0.99 en Cufflinks, mientras que en Trinity fue de 0.58. El genoma de *A. thaliana* de TAIR10 posee un total de 28,775 genes, en donde un total de 28,527 se ensamblaron por arriba del 95%, mientras que en Trinity solo se lograron ensamblar 14,636. La cobertura de genes al 50 y 95% fue similar a los genes ensamblados en Cufflinks. Por otro lado, los genes cubiertos al 95% en Trinity fueron 15,576 transcritos. Cufflinks presentó una mayor cobertura y mayor cantidad de isoformas ensambladas a diferencia de Trinity. La última métrica fue la especificidad de transcritos, en donde la tendencia fue favorable para Cufflinks presentando un emparejamiento positivo de 52,000 transcritos aproximadamente, mientras que Trinity presentó solo 9,811 de transcritos con un 95% de emparejamiento. Cufflinks también demostró tener un mayor número de genes sin anotar a diferencia de Trinity, lo cual no puede dar un pequeño indicio de la expresión o presencia de nuevos transcritos no codificantes durante la condición de sombra. En función a todo lo descrito anteriormente, se decidió trabajar con el transcriptoma ensamblado con Cufflinks para realizar la identificación y anotaciones de lncRNAs expresados durante la condición de sombra.

	<b>Trinity</b>	<b>Cufflinks</b>
Genes	28,775	28,775
Número promedio de exones por isoforma	5.704	5.704
<b>Métricas básicas de transcritos</b>		
Transcritos	54,286	71,920
Transcritos > 500 pb	35,040	66,579
Transcritos > 1000 pb	27,675	54,427
<b>Métricas de alineamientos</b>		
Alineados	54,286	71,834

Únicamente alineados	53,868	71,086
Alineamientos múltiples	361	748
Sin alinear	0	86
<b>Métricas de alineamiento para transcritos non-missambled</b>		
Fracción promedio alineada	0.999	1.0
Longitud promedio alineada	1,525.564	1,810.281
Desajustes promedio por transcrito	0.16	0.005
<b>Métricas de alineamiento para transcritos quiméricos</b>		
Errores de montaje	57	0
<b>Integridad de ensamble (sensibilidad)</b>		
Cobertura de la base de datos	0.588	0.998
Genes ensamblados 50 %	17,638	28,535
Genes ensamblados 95 %	14,636	28,527
Cobertura de genes 50 %	18,190	28,535
Cobertura de genes 95 %	15,576	28,527
Isoformas ensambladas 50 %	18,350	36,562
Isoformas ensambladas 95 %	15,159	36,549
Cobertura de isoformas 50%	18,918	36,562
Cobertura de isoformas 95 %	16,111	36,549
Cobertura media de isoformas	0.806	1.0
Ensamble promedio de isoformas	0.781	1.0
<b>Especificidad de ensamble</b>		
50% emparejados	31,763	64,170
95% emparejados	9,811	52,643
Sin anotar	4,927	6,456
Fracción media de transcrito emparejado	0.553	0.874

**Tabla 2.** Métricas de ensamble por RNAQuast para Trinity y Cufflinks

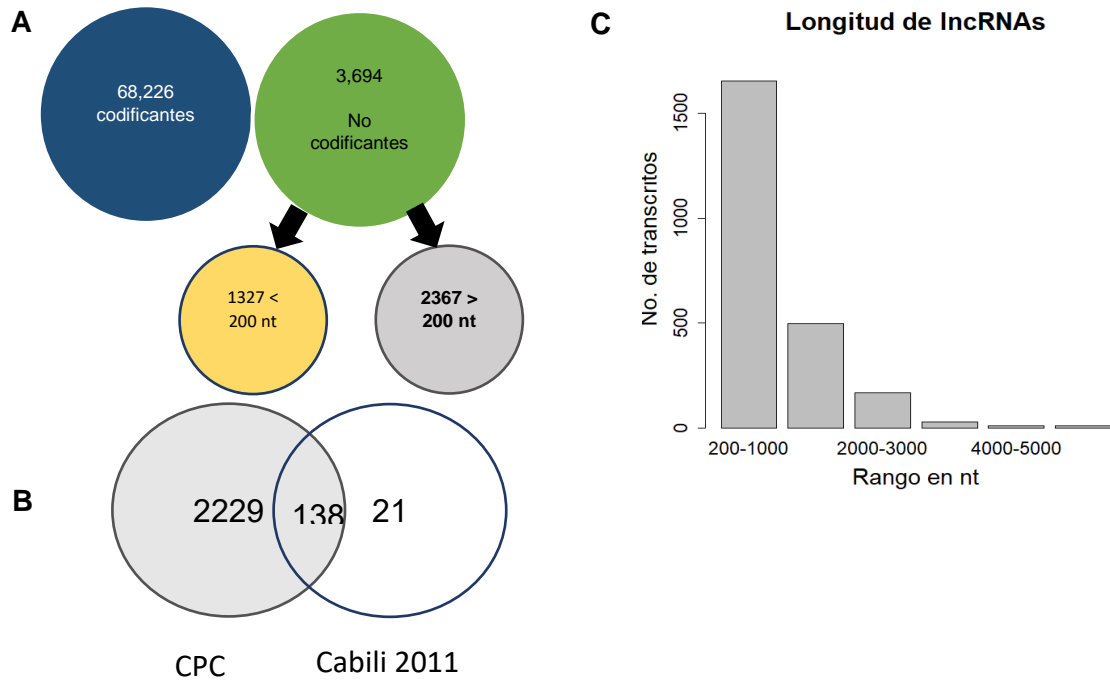
## 6.2 Identificación y clasificación de lncRNAs

Para la identificación de lncRNAs durante el fenómeno de sombra se emplearon dos pipelines independientes (CPC y un pipeline más riguroso al que denominamos Cabili – ver métodos) pero con el mismo fin: identificar y anotar genes no codificantes. Con el primer método, de una cifra total de 71,920 transcritos ensamblados, 68,226 transcritos resultaron ser codificantes, mientras que el resto (3,694 transcritos) resultaron ser transcritos no codificantes. Sin embargo, esa pequeña porción aún contenía transcritos menores a 200 nucleótidos, por lo tanto, al realizar el corte obtuvimos 2,367 lncRNAs putativos. La segunda tubería es un método con filtros más estrictos por lo cual nos arrojó una cifra de 159 transcritos largos no codificantes (Figura 13A). Una vez que se obtuvieron las dos listas, se realizó una intersección para ver que transcritos se compartían en ambas listas. De la intersección 138 genes se compartían en ambas listas (Figura 13B).

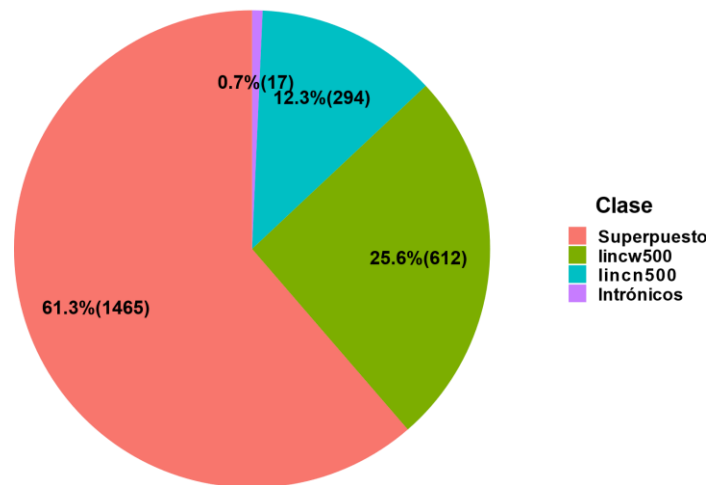
Se fusionaron ambas listas para obtener una sola lista de lncRNAs: un total de 2,388 lncRNAs se anotaron del transcriptoma de Cufflinks. De acuerdo con la Figura 13C, una gran fracción de nuestros genes no codificantes se encuentran en un rango de 200 a 1,000 nucleótidos, disminuyendo conforme va aumentando el rango de tamaño.

Para la clasificación de lncRNAs de acuerdo a su posición genómica se ejecutó el programa bedtools, en donde la mayor parte de nuestros transcritos (**61.3%**) se clasificaron como superpuestos con regiones codificantes, seguidos de un **25.6%** de lncRNAs intergénicos distales 500 pb a regiones codificantes (lincRNAs – long intergenic non coding RNAs en inglés), seguido por un **12.3%** de los lincRNAs proximales 500 pb a regiones codificantes, mientras que en una fracción muy pequeña de **0.7%** se encontraron los lncRNAs intrónicos (Figura 14).





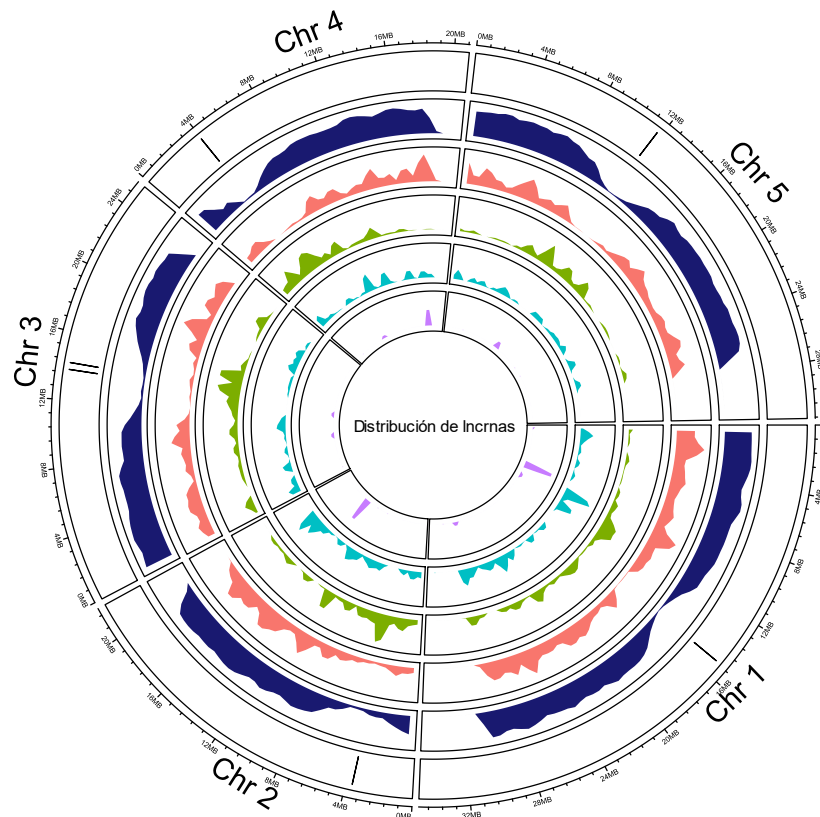
**Figura 13.** lncRNAs identificados en el transcriptoma. **A)** Identificación de transcritos en Cufflinks por medio de CPC, en color azul los transcritos codificantes, en verde los no codificantes, en oro los < a 200 nt y gris los > a 200 nt. **B)** Diagrama de Venn de la intersección de lncRNAs entre el archivo de salida de CPC y el método de Cabili. **C)** Histograma para agrupar los lncRNAs de acuerdo a su longitud para ambas clasificaciones.



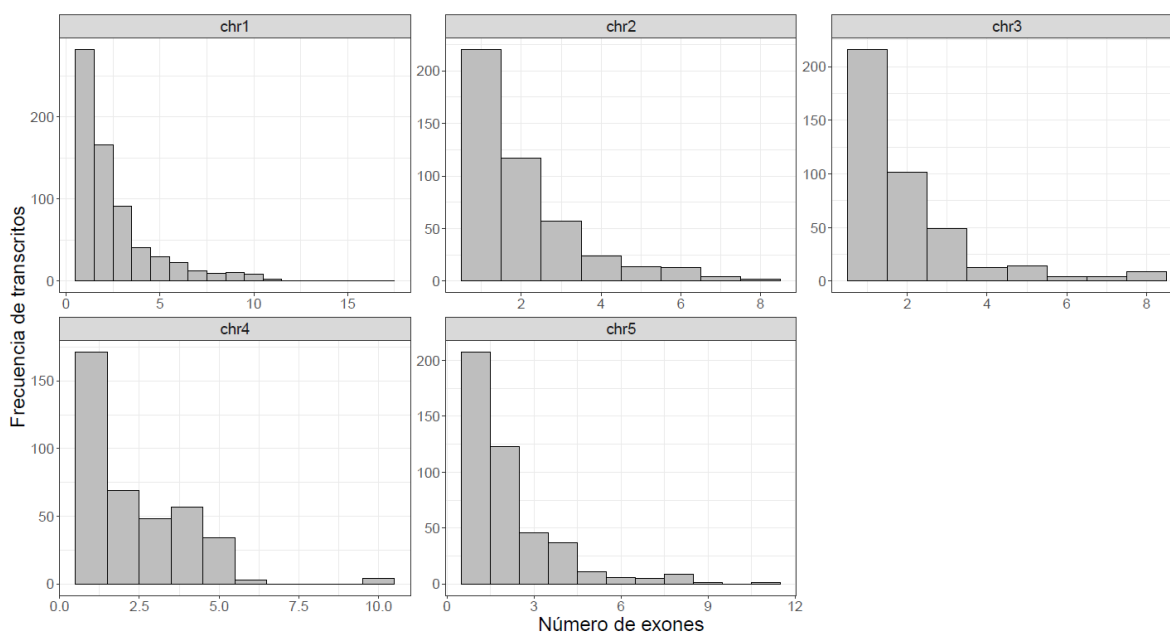
**Figura 14.** Clasificación de lncRNAs de acuerdo a su contexto genómico

Los lncRNAs identificados se distribuyeron a lo largo de todo el genoma de *A. thaliana*, con algunos mostrando acumulaciones puntuales particularmente los intrónicos (Figura 15).

También se clasificaron los lncRNAs de acuerdo con el número de exones que contenía cada transcrito (Figura 16). Se graficó el número de exones de cada transcrito para cada cromosoma de *A. thaliana*, en donde encontramos que los lncRNAs con un solo exón son los más abundantes en todos los cromosomas. Sin embargo, pudimos observar en la distribución que también se ensamblaron transcritos con más de un exón o que sufren se splicing alternativo. Por mencionar un ejemplo, en el cromosoma 1 podemos observar que existe una minoría de transcritos con un número > a 10 exones.



**Figura 15.** Distribución de los lncRNAs a lo largo del genoma clasificados de acuerdo a su posición genómica. El primer carril representa la posición de los centrómeros. El eje externo representa cada cromosoma de *A. thaliana* excluyendo los organelos. En azul marino se representan los genes codificantes, en los lncRNAs superpuestos, en verde los intergénicos alejados 500 nt de genes codificantes, en azul turquesa los intergénicos cercanos genes codificantes y en morado los intrónicos.



**Figura 16.** Histogramas del número de exones en lncRNAs para cada cromosoma de *A. thaliana*, donde chr es cromosoma. El eje horizontal representa cuantos exones tiene cada transcrito no codificante mientras que en el vertical se encuentra la frecuencia acumulada de transcritos.

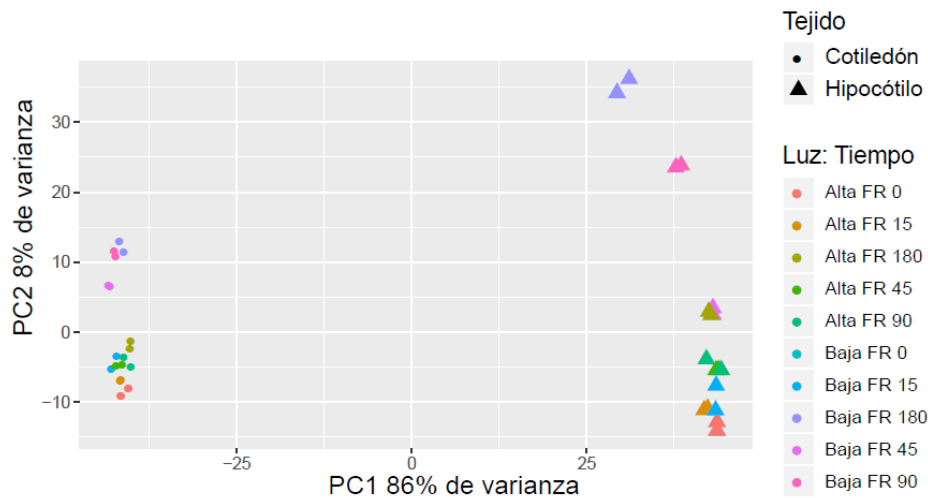
### 6.3 Expresión diferencial de genes codificantes y no codificantes

El análisis de expresión diferencial se realizó con el paquete de R Bioconductor DESeq2, tomando en cuenta los genes anotados en el genoma de TAIR10 y los identificados de novo por medio del transcriptoma.

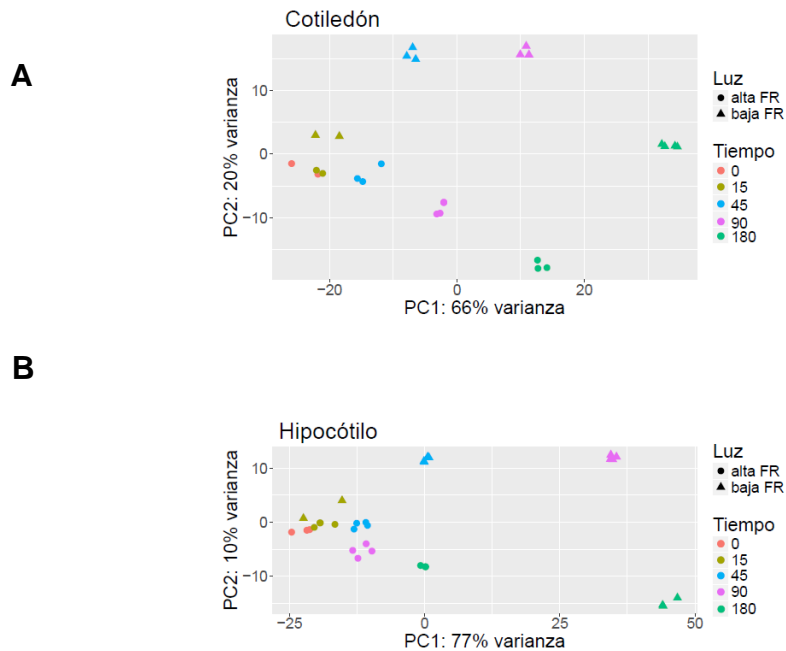
Como un primer acercamiento a nuestros datos, se graficó un análisis de componentes principales (PCA) del conjunto total de muestras, en donde, de acuerdo a lo esperado, sobre el componente principal 1 (PC1) se percibió un 86% de varianza debido a los tejidos (cotiledón e hipocótilos), mientras que en PC2, la varianza fue menor (8%), lo que corresponde a la interacción del tiempo y la exposición a la sombra simulada (Figura 17).

Para percibir mejor la varianza entre la interacción del tratamiento dependiente del tiempo, se realizó un análisis de componentes principales para cada tejido (Figura 18). Para el tejido de Cotiledón (Figura 18A), el PC1 presentó un 66% de la varianza debido principalmente a la variable tiempo, mientras que en el PC2, la varianza fue

del 20% y se vio más influenciada por el tratamiento de sombra simulada. Una tendencia muy similar se vio en tejido de Hipocótilo (Figura 18B), en donde la varianza de las muestras principalmente se dio por el tiempo (PC1 77%), mientras solo un 10 % de la varianza en el PC2, estuvo influenciada más por los tratamientos de sombra.



**Figura 17.** PCA para la visualización de la distancia entre muestras Tejidos de Cotiledón e Hipocótilo. High FR es el tratamiento de alta proporción luz roja lejana y Low FR es el tratamiento de baja proporción luz roja lejana.

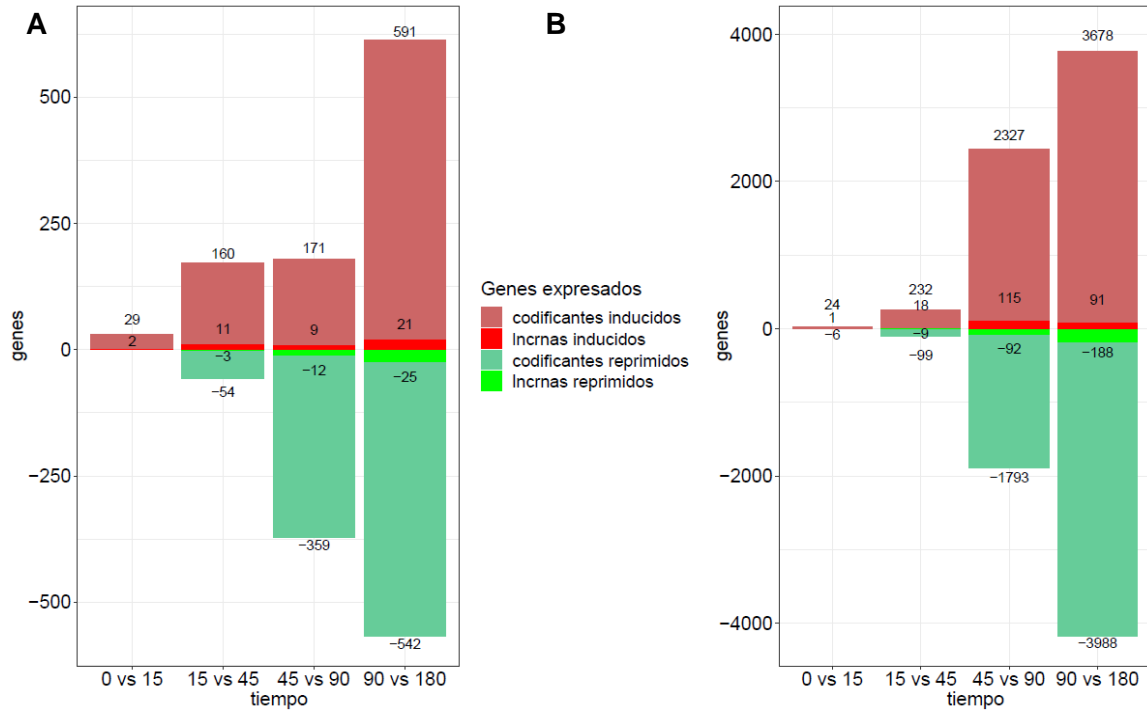


**Figura 18.** PCA de tejidos independientes. **A)** PCA de muestras de Cotiledón. **B)** PCA para el tejido de Hipocótilo.

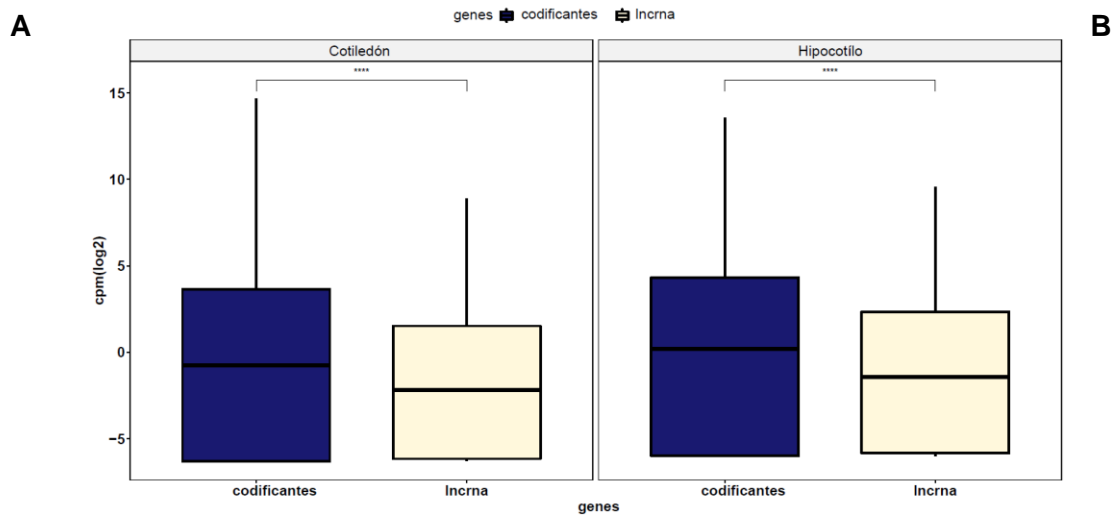
Posteriormente se realizaron los contrastes para detectar los genes expresados diferencialmente en cada tejido, tomando en cuenta la interacción entre el tratamiento de sombra y el tiempo (Figura 19).

Durante el contraste de los tiempos 0 vs 15, en el tejido de Cotiledón un total de 31 genes expresaron diferencialmente (Figura 19A) (29 codificantes y 2 lncRNAs sobre expresados) al igual que en Hipocótilo (Figura 19B), solo que 25 se sobre expresaron (24 codificantes y 1 lncRNA) y 6 genes codificantes se reprimieron. En el contraste de 15 vs 45, para el caso de Cotiledón (Figura 19A), un cambio en la expresión génica de 228 genes, donde ~70% de genes se indujeron, donde solo 11 genes eran lncRNAs y 3 eran lncRNAs reprimidos, mientras que en Hipocótilo (Figura 19B) los genes inducidos en respuesta a la sombra fueron 358, de los cuales ~69% eran genes inducidos, donde 18 eran lncRNAs y en el caso de transcritos reprimidos había 9 lncRNAs. En cotiledón durante el contraste de 45 vs 90 (Figura 19A), se indujo un total de 551 genes, donde ~67% de los genes se reprimieron durante este tiempo, induciéndose solo 9 lncRNAs y reprimiéndose 12. En hipocótilo (Figura 19B) se expresaron diferencialmente un total de 4,327 genes (~56% sobre expresados) de los cuales 115 genes inducidos eran lncRNAs y 92 eran lncRNAs reprimidos. En el último contraste, en cotiledón (Figura 19A) se expresaron 1,179 genes (~51% inducidos), donde 21 lncRNAs se sobre expresaron y 25 lncRNAs se reprimieron. En hipocótilo (Figura 19B) se encontraron 7,945 genes expresados diferencialmente, un poco más del 52% de los genes se reprimieron (188 lncRNAs) mientras que 91 genes no codificantes se sobre expresaron en este último tiempo.

Por otro lado, se compararon los niveles de expresión en cpm (conteos por millón) entre los lncRNAs y los genes codificantes (Figura 20). En ambos tejidos se puede observar que los niveles de expresión de los lncRNAs son ligeramente más bajos significativamente comparados con los genes codificantes (Wilcox,  $p_{adj} < 0.01$ ).



**Figura 19.** Total de genes inducidos y reprimidos en cada contraste de tiempo en **a)** Cotiledón e **b)** Hipocótulo durante el tratamiento de sombra. En rojo claro se representa el número genes codificantes sobre expresados, mientras que en rojo intenso se muestran los lncRNAs sobre expresados. En verde claro se representa el número genes codificantes reprimidos, mientras que en verde intenso se muestran los lncRNAs reprimidos.  $P_{adj} (<0.01)$ .



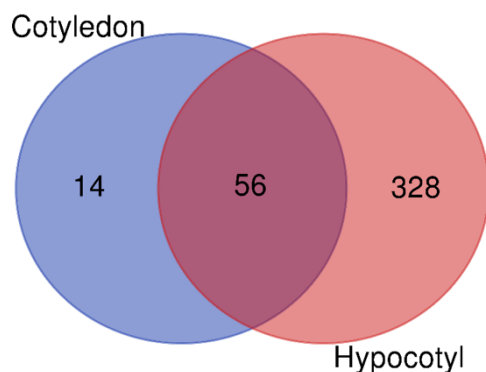
**Figura 20.** Boxplot de conteos por millón (cpmlog<sub>2</sub>) entre transcritos codificantes (cajas azules) y no codificantes (cajas grises) en **a)** cotiledón e **b)** hipocótulo. Comparación prueba Wilcoxon, ( $p_{adj} < 0.01$ ).

Para observar si existe una relación tejido-específica, se graficó un diagrama de Venn para identificar los lncRNAs diferencialmente expresados compartidos en

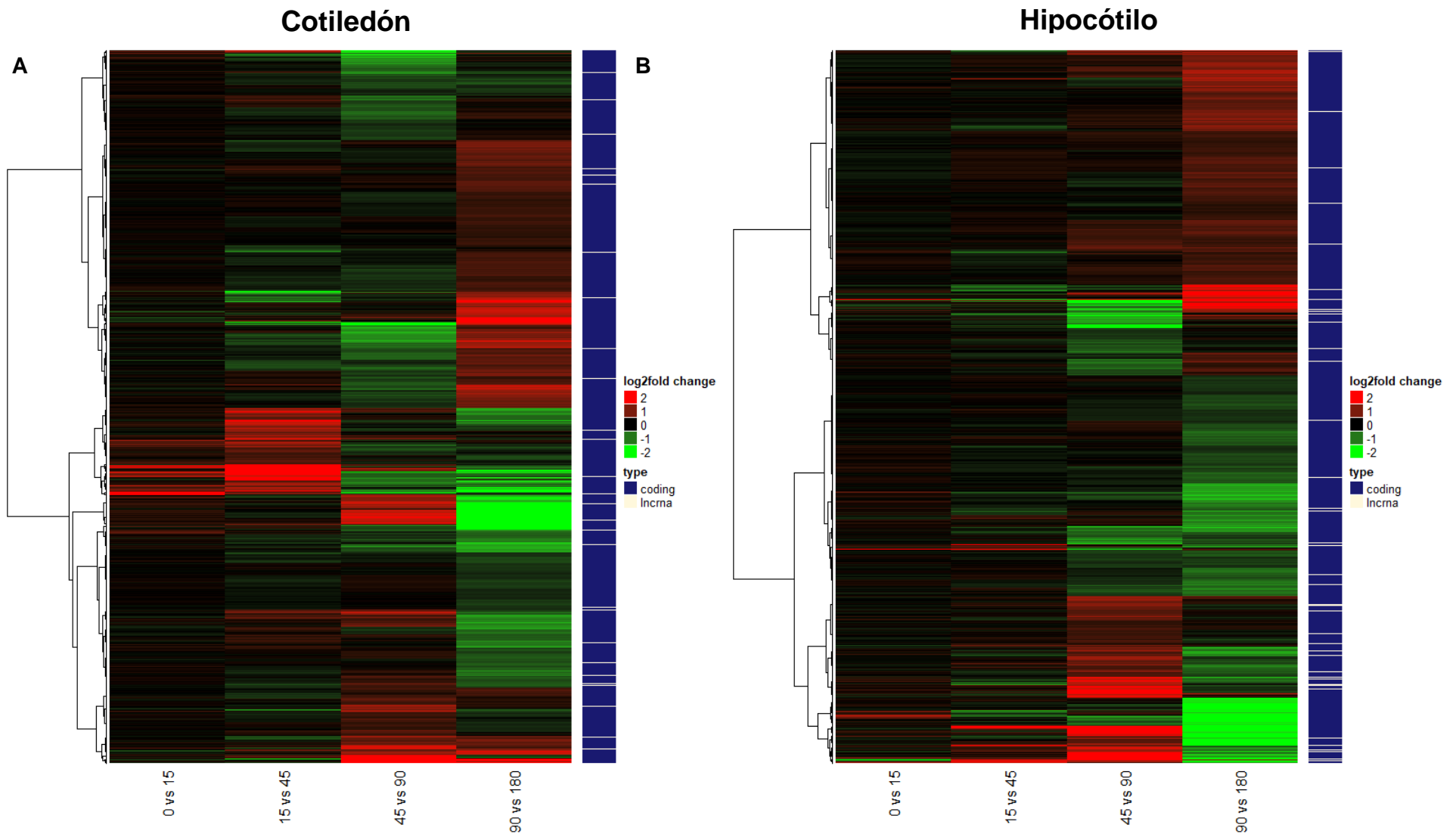
ambos tejidos (Figura 21). Se encontraron 56 lncRNAs compartidos en ambos tejidos, 14 que eran únicos de cotiledón y 328 únicos de hipocótilos.

En el tejido de cotiledón, 1,569 genes se expresaron diferencialmente en alguna parte del tiempo durante la exposición a luz roja lejana (Figura 22A). En este tejido se puede observar que hay una minoría de genes que se encienden al principio (15 min) una vez iniciada la exposición, se incrementa el número de genes expresados a través del tiempo transcurrido. Respecto al tejido de hipocótilo, se encontraron 9,621 genes expresados diferencialmente durante alguna parte del tiempo (Figura 22B). En contraste con cotiledón, en hipocótilo los genes de respuesta a SAS se indujeron en los tiempos finales (90 y 180 min), sin embargo, los genes de respuesta en hipocótilo son muchos más que en cotiledón. Aproximadamente un 4% de los transcritos no codificantes se expresaron diferencialmente en hipocótilo (Figura 23A), de ese porcentaje la mayoría de transcritos que se expresaron se derivan de genes codificantes (318 unión), seguido de los intergénicos vecinos a codificantes (32 linc cercanos 500 pb de una región codificante) y por último los menos expresados fueron los intergénicos posicionados en una ventana mayor a los 500 nucleótidos. En cotiledón (Figura 23B), un 4.5% resultaron ser lncRNAs diferencialmente expresados, de los cuales al igual que en hipocótilo la mayoría eran transcritos superpuestos (63 superpuestos), seguido de los lincn500 (6 lincn500) y por último solo un lincw500.

Sin embargo, en cotiledón e hipocótilo se encontraron 2 y 18 lncRNAs nuevos respectivamente expresados diferencialmente expresados.

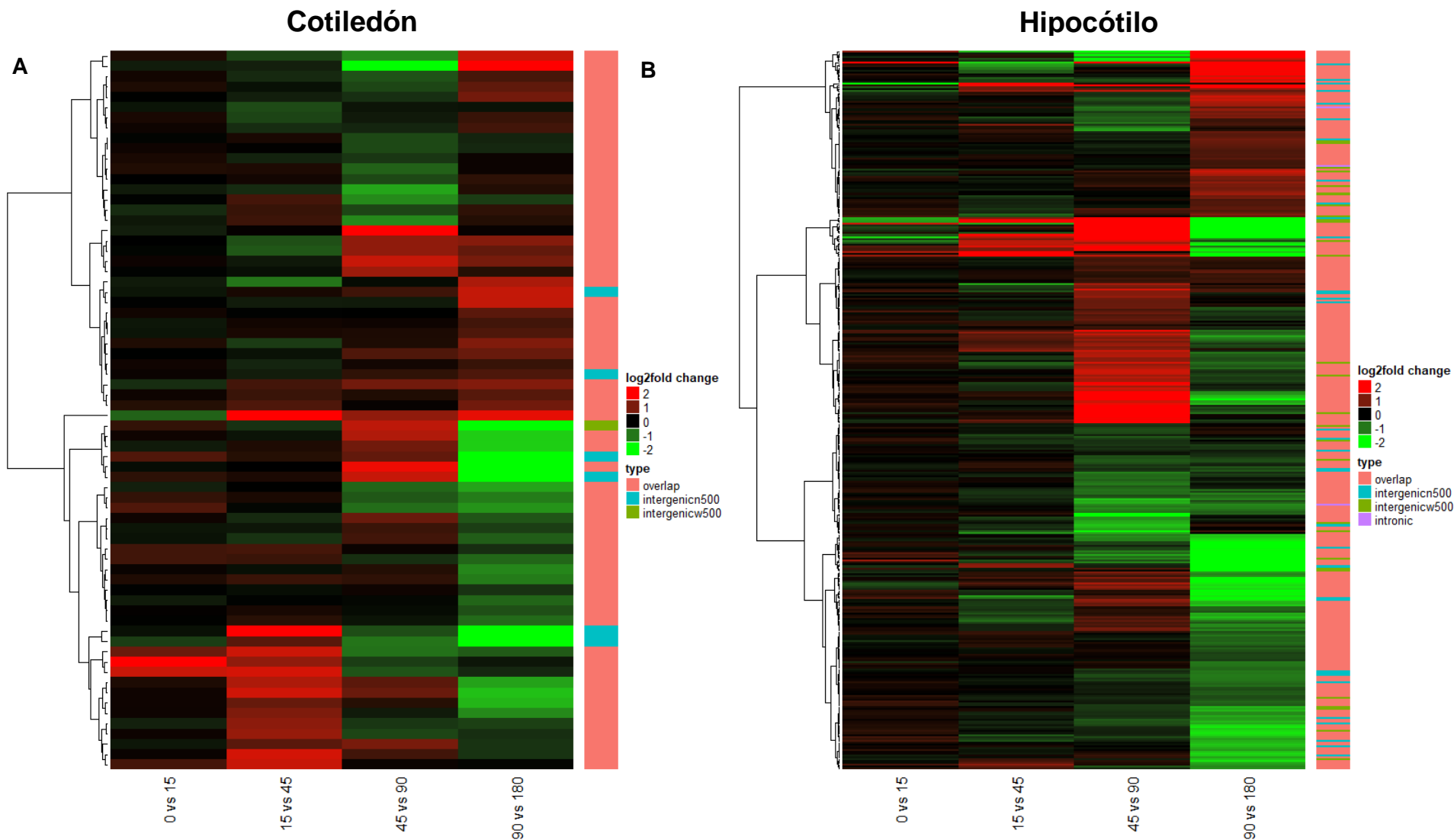


**Figura 21.** Diagrama de Venn de lncRNAs que se comparten en ambos tejidos.



**Figura 22.** Heatmap de todos los genes expresados diferencialmente para **A)** Cotiledón e **B)** Hipocótilo durante el fenómeno de luz roja lejana. Cada bloque muestra el log2fold change de cada gen a través de los contrastes entre tiempos. La anotación del lado derecho indica si pertenece a un gen codificante o no codificante. Los genes están clusterizados en tres grupos principales representados a la izquierda. Los parámetros tomados en cuenta para representar los genes en el heatmap, fueron que al menos un gen cambiara significativamente durante algún contraste entre tiempos. Para determinar si el gen está diferencialmente expresado se tomó un pvalue (padj) > 0.01.





**Figura 23.** Heatmap de todos los lncRNAs contrastados expresados diferencialmente para **A)** Cotiledón e **B)** Hipocótilo durante el fenómeno de luz roja lejana. Cada bloque muestra el log2fold change de cada gen (lncRNA) a través de los contrastes entre tiempos. La anotación al lado derecho indica la posición en donde se encuentra el lncRNA en el genoma. Los genes están agrupados en tres grupos principales representados a la izquierda. Los parámetros tomados en cuenta para representar los genes en el heatmap, fueron que al menos un gen cambiara significativamente durante algún contraste entre tiempos. Para determinar si el gen está diferencialmente expresado se tomó un pvalue ( $p_{adj}$ ) > 0.01.

## **6.4 Análisis de enriquecimiento funcional de genes diferencialmente expresados**

Para asignarle un posible proceso biológico en el que participan el conjunto de genes diferencialmente expresados, se realizó un análisis de enriquecimiento funcional con topGO de los genes expresados diferencialmente en ambos tejidos.

En el caso de los genes inducidos (up-regulated) en el tejido de cotiledón, se encontraron funciones enriquecidas significativamente relacionadas a la respuesta a sombra (Figura 24A), por ejemplo, términos que resaltan en el primer tiempo (0 a 15 min), fueron los relacionados a respuesta a hormonas, prevención a la sombra, biosíntesis de auxinas, respuesta a luz roja, la ruta de señalización por luz roja lejana y la ruta de señalización a brasinoesteroides, entre otros. Durante el segundo tiempo sobresaltaron términos GO relacionados a la organización celular, gravitropismo, regulación del tamaño celular, crecimiento de células de punta, transporte polar de auxinas etc. En el tiempo de exposición entre los tiempos de 45 y 90 minutos se acentuaron términos como respuesta UV-B, a ácido abscísico y a estímulos de luz. Por último, en el transcurso entre 90 a 180 minutos, términos como biogénesis de pared celular, proceso de biosíntesis de polisacáridos, respuesta a sacarosa, a fructosa y karrikina fueron procesos muy enriquecidos. Términos de respuesta a auxinas, respuesta a brasinoesteroides y crecimiento celular multidimensional, fueron términos enriquecidos en más de dos tiempos.

Para los genes reprimidos en cotiledón (down-regulated) durante el fenómeno de sombra el panorama fue distinto (Figura 24B). En el transcurso del primer tiempo (0-15 min) no presentó categorías enriquecidas. Por otra parte, en el segundo tiempo que sucede entre los 15 y 45 minutos, términos como regulación de la germinación de semillas, del proceso biosintético de clorofilas, desarrollo del óvulo de la planta, gravitropismo y la vía de señalización del ácido giberélico fueron enriquecidos significativamente. Por adición, algunos términos enriquecidos en el tiempo transcurrido entre los 45 y 90 minutos fueron los relacionados a la morfogénesis de la hoja, crecimiento multidimensional celular, vía de señalización del ácido jasmónico y proceso biosintético de brasinoesteroides. Finalmente,

durante el tiempo entre 90 y 180 minutos, términos de respuesta a giberelinas, procesos relacionados a la fotosíntesis, aflojamiento de la pared celular, estímulo a respuesta a luz, procesos de biosíntesis para oligosacáridos y transcripción de ADN templado fueron enriquecidos para genes reprimidos. Algunos términos como respuesta a auxinas, brasinoesteroides y prevención a la sombra se acentuaron en los tiempos finales.

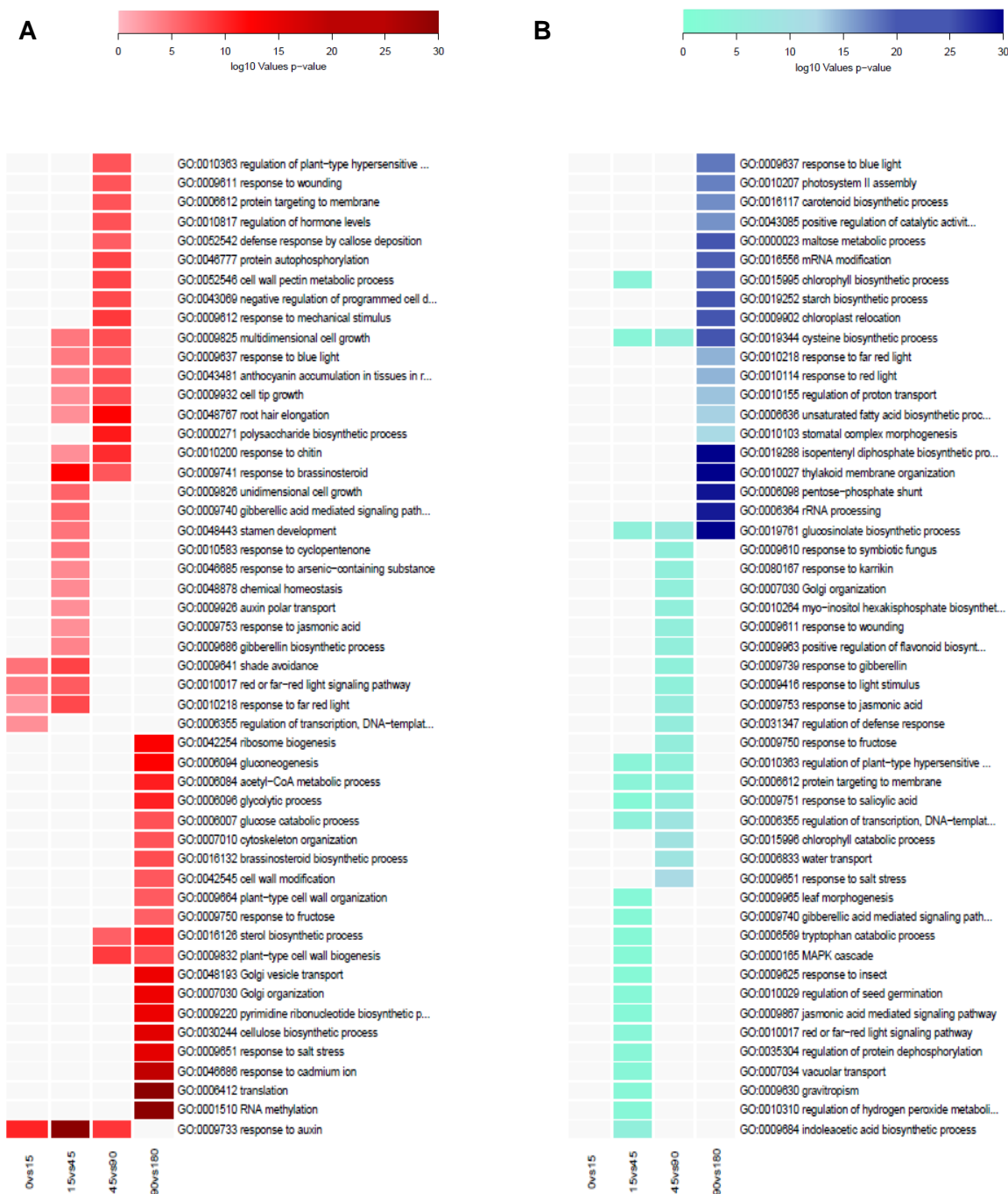
Para el tejido de hipocótilo (Figura 25A), siendo el caso de los genes inducidos (up-regulated) durante el tiempo inicial de 0 a 15 minutos, algunos términos enriquecidos significativamente fueron los relacionados a respuesta a auxinas, prevención a la sombra (shade avoidance), regulación de la transcripción, vía señalización de luz roja lejana y respuesta a luz roja lejana. Los siguientes términos enriquecidos durante el segundo tiempo (15 a 45 min) fueron los relacionados a crecimiento celular multidimensional y unidimensional, respuesta a luz azul, acumulación de antocianinas, crecimiento de células de punta, transporte polar de auxinas y respuesta a brasinoesteroides, entre otros. El tercer tiempo, que va de los 45 a 90 minutos resaltaron términos GO referentes a los niveles de hormonas, procesos metabólicos para la biosíntesis de pectina de pared y procesos biosintéticos de pared celular. Por último, en los tiempos finales del tratamiento, se representaron términos relacionados a la glucólisis, por mencionar algunos tenemos los relacionados a la gluconeogénesis, procesos metabólicos de acetil CoA, procesos glicolíticos y proceso catabólico de la glucosa, por otra parte términos relacionados con la pared celular tales como la modificación de la pared celular, organización de la pared celular, proceso biosintéticos de celulosa, biogénesis de la pared fueron significativamente enriquecidos. En el caso de hipocótilo, términos muy sobrerrepresentados durante el tiempo final fueron la metilación de ARN y traducción.

Para los genes reprimidos en hipocótilo (Figura 25B), al igual que en cotiledón no se anotaron significativamente genes en términos durante el primer tiempo de exposición a luz roja. Sin embargo, a partir del segundo tiempo (15 a 45 min) términos GO de genes reprimidos incluyen respuesta al ácido salicílico, vías de

señalización mediadas por ácido giberélico y por luz roja lejana, proceso biosintético de ácido indolacético y gravitropismo se acentuaron significativamente. Durante el lapso de tiempo que va de los 45 a 90 min de exposición, términos de respuesta a giberelinas, estímulo de luz, ácido jasmónico, procesos catabólicos y transporte de agua se representaron significativamente. Finalmente, términos de respuesta a luz azul, ensamble del fotosistema II, organización de la membrana tilakoide, proceso biosintético de carotenoides y clorofilas, así como relocalización de clorofilas presentaron genes anotados significativamente. También procesos relacionados al metabolismo de azúcares como procesos metabólicos de maltosa, biosíntesis de almidón se inhibieron severamente en el último tiempo. Términos relacionados al proceso biosintético de cisteína y glucosinolatos se sobre representaron durante los tres tiempos.



**Figura 24.** Enriquecimiento funcional de términos GO de los genes diferencialmente expresados en **Cotiledón** durante el tratamiento de sombra, donde los bloques representan el p-value < 0.01 (Classic Fisher) en una escala del -log<sub>10</sub>. A) Bloques rojos: Genes inducidos. B) Bloques azules: Genes reprimidos.



**Figura 25.** Enriquecimiento funcional de términos GO de los genes diferencialmente expresados en **Hipocótilo** durante el tratamiento de sombra, donde los bloques representan el p-value < 0.01 (Classic Fisher) en una escala del -log<sub>10</sub>. A) Bloques rojos: Genes inducidos. B) Bloques azules: Genes reprimidos

## **6.5 Construcción de redes de co-expresión y anotación funcional para la identificación de lncRNAs candidatos**

Para poder inferir una posible función a la lista de lncRNAs expresados diferencialmente en ambos tejidos, se construyeron redes de co-expresión a partir de una población de 21,112 genes (genes codificantes y no codificantes) para agruparlos en módulos de acuerdo a sus perfiles de expresión, utilizando un enfoque de análisis de redes de co-expresión génicas ponderadas (WGCNA) para describir patrones de correlación entre genes a través de muestras de RNA-seq. Este método nos permite agrupar genes altamente correlacionados (formar grupos o módulos con genes codificantes) e identificar genes blanco para su futura caracterización (Langfelder & Horvath, 2008). Un total de 40 módulos altamente co-expresados se detectaron en nuestra red. El módulo gris se descartó, debido a que en este módulo se agrupan los genes que no co-expresan con ningún gen. Módulos como el turquesa y el azul presentaron el mayor número de genes co-expresados con 3,710 y 3,262 genes respectivamente. Sin embargo, dentro de esos dos grandes clados se presentaron pequeños módulos más específicos mayores a mil genes como el café, verde, gris, el rojo y el amarillo. El módulo más pequeño fue el rojo naranja 4 con 55 genes.

Para anotar las funciones de cada módulo, se identificaron términos GO enriquecidos en cada módulo usando topGo, en donde se anotaron los genes para los parámetros de proceso biológico, componente celular y función molecular. Considerando que se anotaran los módulos en al menos un parámetro de los mencionados anteriormente, se identificaron 29 módulos con enriquecimientos funcionales específicos.

Posteriormente se realizó un escrutinio de los lncRNAs diferencialmente expresados y co-expresados en cada módulo para crear una lista de posibles lncRNAs candidatos para cada tejido. De los 29 módulos con enriquecimiento en términos GO, en 19 módulos se hallaron lncRNAs diferencialmente expresados de cotiledón (Figura 26A). Enfocándonos en los módulos que resultaron interesantes (términos que participan en el fenómeno de proximidad vecina), el magenta, el rojo

naranja 4, azul cielo 3 y amarillo presentaron 4, 2, 2 y 1 lncRNAs expresados diferencialmente respectivamente de los cuales todos resultaron ser lncRNAs que superponen con genes codificantes.

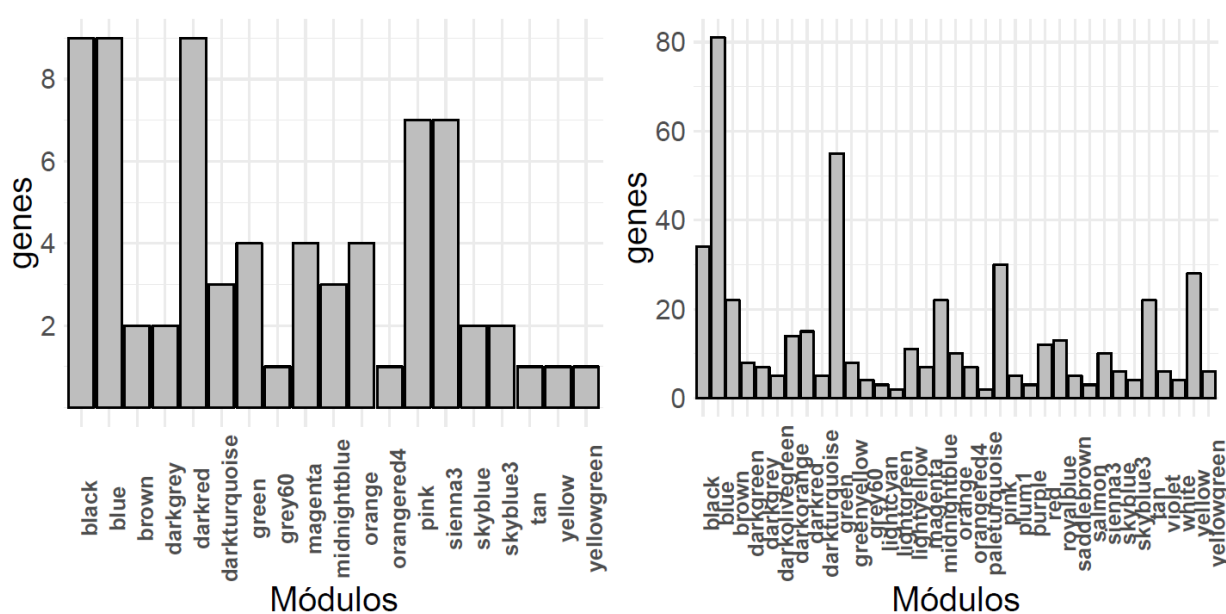
Por otra parte, el hipocótilo presentó mayor abundancia de lncRNAs co-expresados en 35 módulos distintos (Figura 26B). Enfocándonos en los lncRNAs de los módulos interesantes en hipocótilo, en el módulo magenta, rojo naranja4, púrpura, rojo, azul cielo3, violeta y amarillo se encontraron 5, 4, 2, 8, 4, 3 y 20 transcritos no codificantes expresados diferencialmente. De los lncRNAs diferencialmente co-expresados, en el módulo magenta, dos lncRNAs se encuentran en regiones intergénicas, mientras que los tres restantes eran superpuestos con genes codificantes. De los cuatro co-expresados en el módulo rojo naranja4, uno se clasificó como intergénico, mientras que los tres faltantes también se superpusieron con genes codificantes. Los genes no codificantes de los módulos púrpura, rojo y azul cielo2 resultaron ser superpuestos. En el módulo violeta se identificó un lncRNA intergénico y un superpuesto. En el módulo amarillo en el que se encontraron la mayoría de lncRNAs diferencialmente co-expresados, 18 resultaron ser superpuestos y 2 intergénicos.

La manera en que se agrupan los genes en los módulos es un indicador que pueden tener las mismas funciones o que pueden participar en procesos biológicos similares lo cual causa que se co-regulen. En caso de cotiledón, los lncRNAs expresados diferencialmente se agruparon principalmente en módulos como el negro (con términos GO relacionados a la reproducción de organismos multicelulares, transporte de proteínas intracelulares, etc.) y el azul (transcripción del promotor de plástido, proceso metabólico de almidón etc.) con 9 lncRNAs ambos módulos, mientras que en tejido de hipocótilo al igual que en cotiledón la mayor abundancia de lncRNAs diferencialmente expresados se encontró en el módulo azul y el módulo verde (con enriquecimiento para componente celular en vacuola) con 81 y 55 transcritos no codificantes respectivamente.

Sin embargo, nos enfocamos en los términos interesantes implicados en la respuesta a sombra (Figura 27). Módulos como el **rojo naranja 4**, **azul cielo 3**



presentaron términos similares como respuesta a auxinas, respuesta a hormonas y respuesta a compuestos orgánicos. Por otro lado, módulos como el **violeta** y **amarillo** presentaron similitud en términos relacionados a procesos biosintéticos pared celular, por mencionar algunos términos GO enfocados al metabolismo de pared celular, procesos metabólicos de galacturanos, pectinas y hemicelulosa, biogénesis de pared celular y metabolismo de polisacáridos de pared celular. El módulo **rojo** y el **magenta** presentaron términos relacionados a respuesta a proceso fotosintéticos como procesos catabólicos relacionados a pigmentos, proceso catabólico de clorofilas y respuesta a luz UV.



Para cada uno de los módulos de interés se calculó la expresión representativa del módulo o eigengene. Esta es una métrica indicada por un gen representativo en el cual se resumen los perfiles de expresión de cada módulo. Los módulos donde incrementó el módulo eigengene en tratamientos de sombra fueron el **rojo naranja4** y **azul cielo3** en ambos tejidos en tiempos finales principalmente (Figura 28). En módulos **magenta**, **purpura** y **rojo**, la expresión incrementó ligeramente en

cotiledón al final del tratamiento de sombra mientras que en el módulo **rojo** y **purpura** aumentó en hipocótilo en tiempos finales de tratamiento mucho más que en hipocótilo (Figura 28). En módulos como el **amarillo** y el **violeta** la expresión disminuyó en los últimos tiempos en hipocótilo (Figura 28).

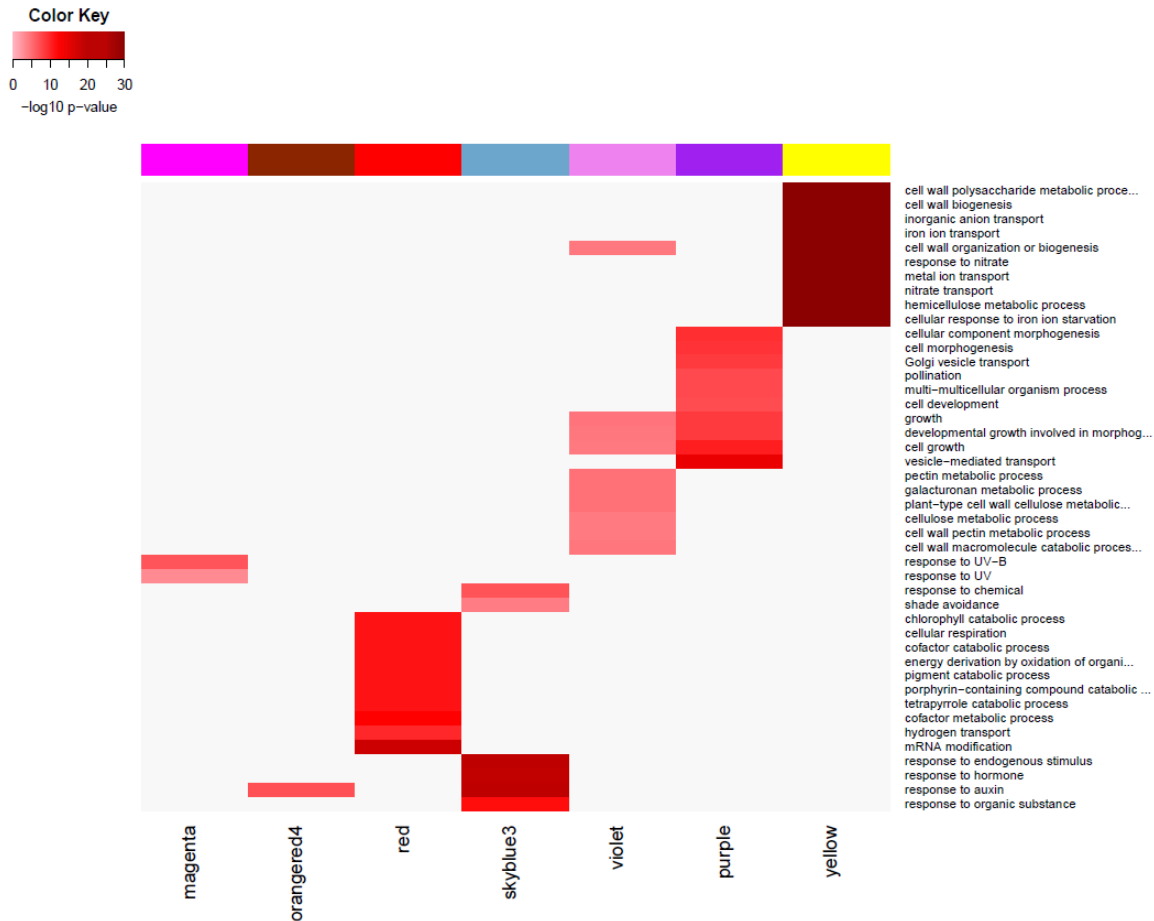
Para ver el comportamiento de los genes en los módulos interesantes durante la condición de sombra se graficaron los niveles de expresión, tanto de genes codificantes como de lncRNAs encontrados cada uno de los módulos (Figura 29). En cotiledón (Figura 29A), se puede observar que genes en los módulos **azul cielo3** y **rojo naranja 4** (los cuales comparten términos similares como respuesta a auxinas y respuesta a hormonas), se expresaron durante tiempos iniciales de 0 a 45 min, reduciendo su expresión en tiempos finales. La mayoría de los genes en el módulo como el **amarillo** presentaron niveles de expresión de represión principalmente en el tiempo de 90 a 180 min lo cual concuerda con el módulo eigengene. Principalmente genes en el módulo **magenta** (con términos GO de respuesta a luz UV) se indujeron durante la respuesta a sombra en el tiempo de 90 a 180 min.

En el caso de hipocótilo (Figura 29B) al igual que en cotiledón, la mayor parte de genes en el módulo **amarillo** se reprimieron durante la exposición a sombra en concordancia con el comportamiento del módulo eigengene. Módulos como el **rojo**, **púrpura**, **violeta** y **magenta** se agruparon, induciendo su expresión principalmente en los tiempos finales de tratamiento (180 min). Algunos genes del módulo **rojo naranja 4** se indujeron en el penúltimo tiempo (entre 45 y 90 min).

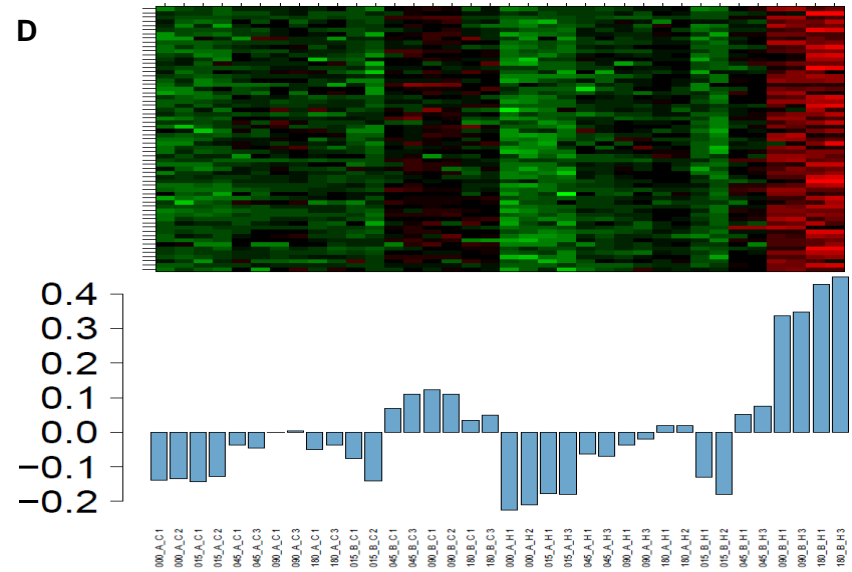
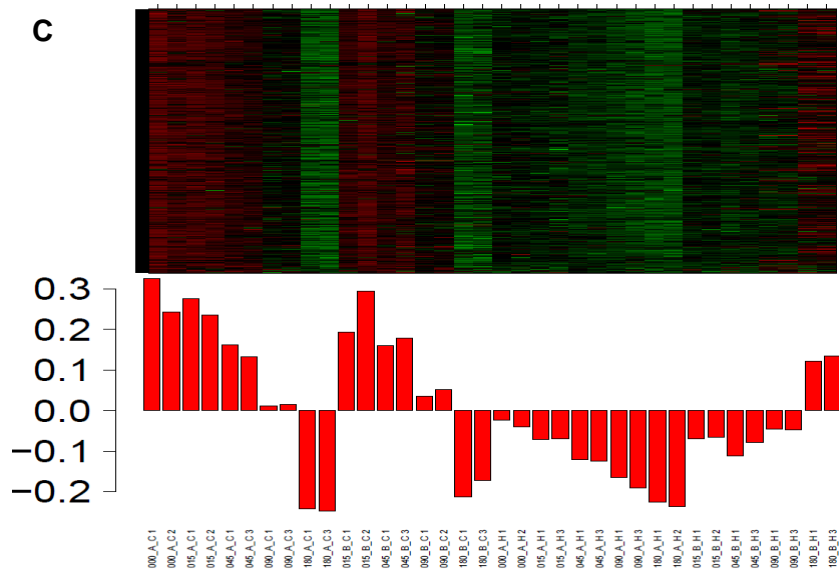
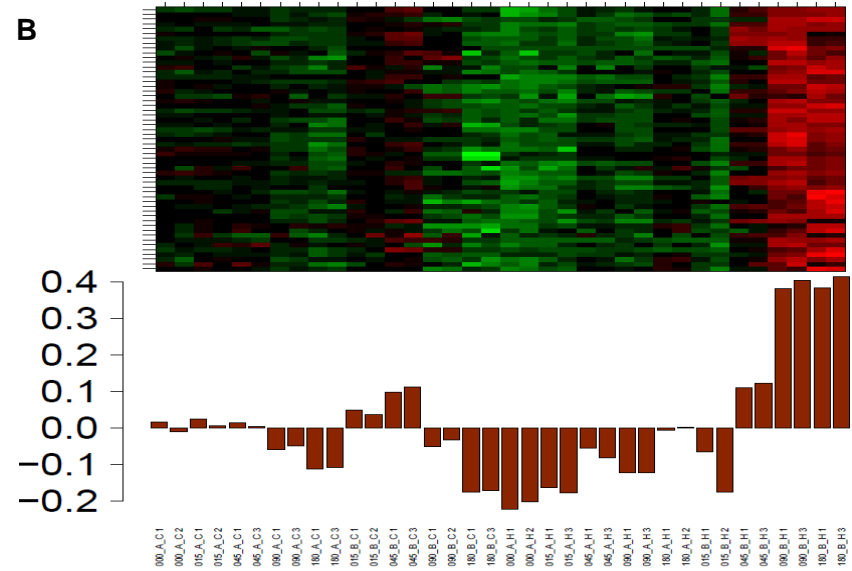
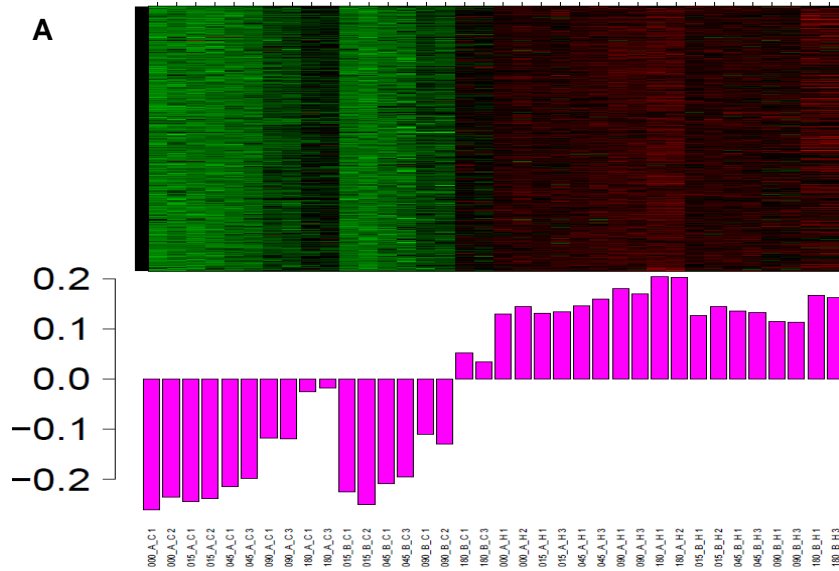
En redes de expresión génica, es importante medir la conectividad entre nodos (genes) para representar posibles interacciones. Se graficó cada módulo de interés para poder observar lncRNAs que pudieran servir como genes centrales, es decir nodos maestros que posiblemente estén interconectados con otros genes de interés.

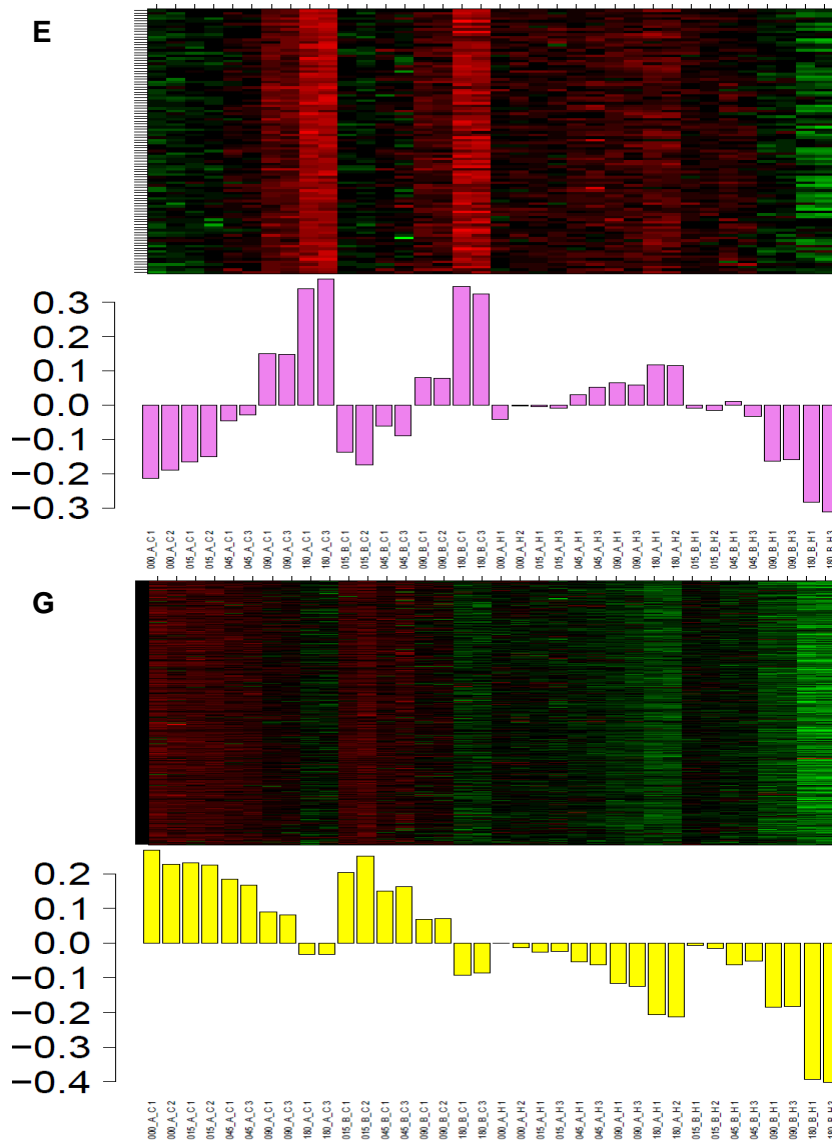
En al menos tres módulos que contenían genes centrales no codificantes (lncRNAs) fuertemente interconectados con genes no codificantes. En el caso de la red **rojo** encontramos dos lncRNAs superpuestos, los genes *AT1G28395* y *AT1G56045*. En

la red **amarilla** identificamos también dos genes los cuales también resultaron ser lncRNAs superpuestos: el gen *AT1G06730* y un transcrito que se encuentra en las coordenadas del gen *CYCD6.6*. Finalmente, el en módulo **azul cielo3** encontramos un lncRNA superpuesto derivado del gen *PSK4*.

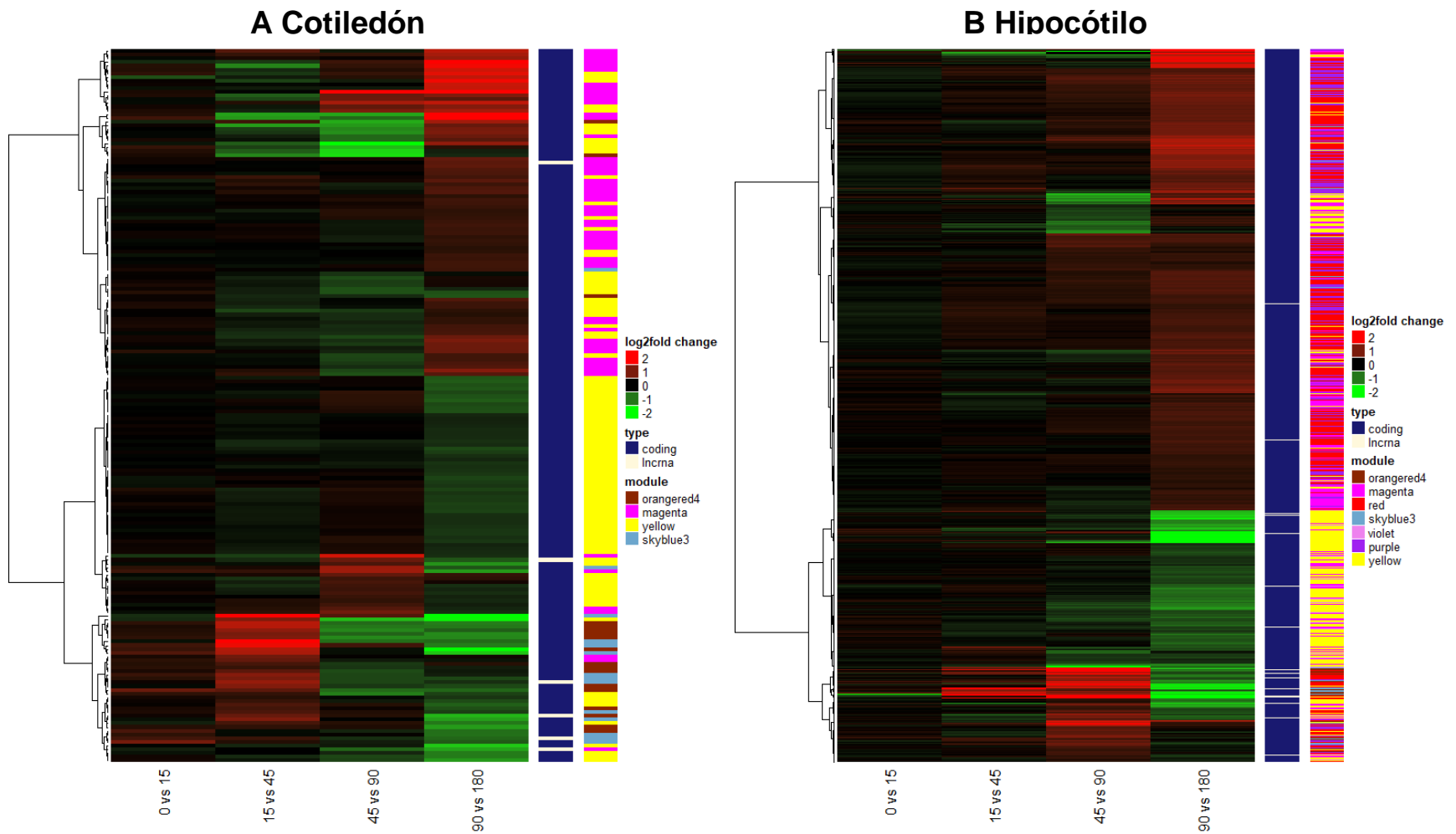


**Figura 27.** Heatmap de los módulos co-expresados de los términos GO con funciones biológicas de interés implicados en el fenómeno de sombra. Módulos interesantes implicados en la respuesta a hormonas como el magenta, el rojonaranja4, azul cielo3. Por otro lado, los implicados en la biogénesis de compuestos de pared celular como el violeta y el amarillo. El módulo rojo implicado a procesos implicados a fotorreacciones.





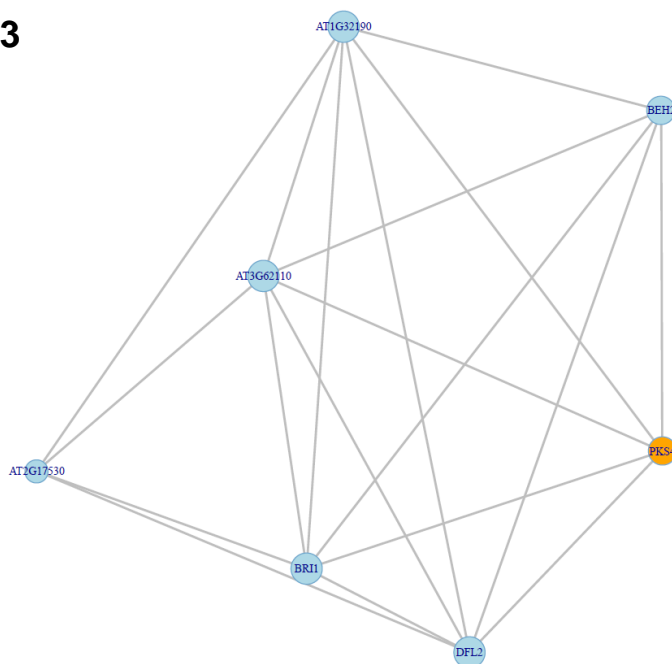
**Figura 28.** Módulos de co-expresión relacionados al fenómeno de sombra. Se graficó el valor eigengene de cada gen para cada módulo en cotiledón ordenados de la siguiente manera: tratamiento control (high R:FR) en tiempo de exposición ascendente seguido del tratamiento de sombra (low R:FR) en tiempo ascendente de exposición. El mismo orden para hipocótilo. Los módulos se presentan en el siguiente orden: **a)** magenta, **b)** rojo naranja4, **c)** rojo, **d)** azul cielo3, **e)** violeta, **f)** púrpura y **g)** amarillo.



**Figura 29.** Heatmap de los genes co-expresados diferencialmente en módulos interesantes durante el fenómeno de luz roja lejana que contienen al menos un lncRNA para **A)** Cotiledón e **B)** Hipocótilo. La primera anotación al lado derecho indica el tipo de transcrito (codificante o no codificantes). La segunda anotación indica el módulo en el que se co-expresados encuentran. Los parámetros tomados en cuenta para representar los genes en el heatmap, fueron que al menos un gen cambiara significativamente durante algún contraste entre tiempos. Para determinar si el gen está diferencialmente expresado se tomó un pvalue (padj) < 0.01



### C Azul cielo 3



**Figura 30.** Redes de co-expresión de los módulos con funciones relevantes en la condición de sombra. **A)** Para el módulo rojo, **B)** para el amarillo y **C)** el azul cielo3. Nodos en azul representan a los genes codificantes, mientras que los nodos naranjas representan a los lncRNAs. El tamaño del nodo indica el hubscore (centralidad de nodo). Nodos en la red son representados si el kme es > al umbral del kme (tercer cuartil). Solo se representaron los genes diferencialmente expresados.

De acuerdo con la anotación de factores de transcripción, se anotaron los lncRNAs que superpusieron con estos, en donde encontramos 29 lncRNAs que se derivan de locus de factores de transcripción, de los cuales 17 están diferencialmente expresados.

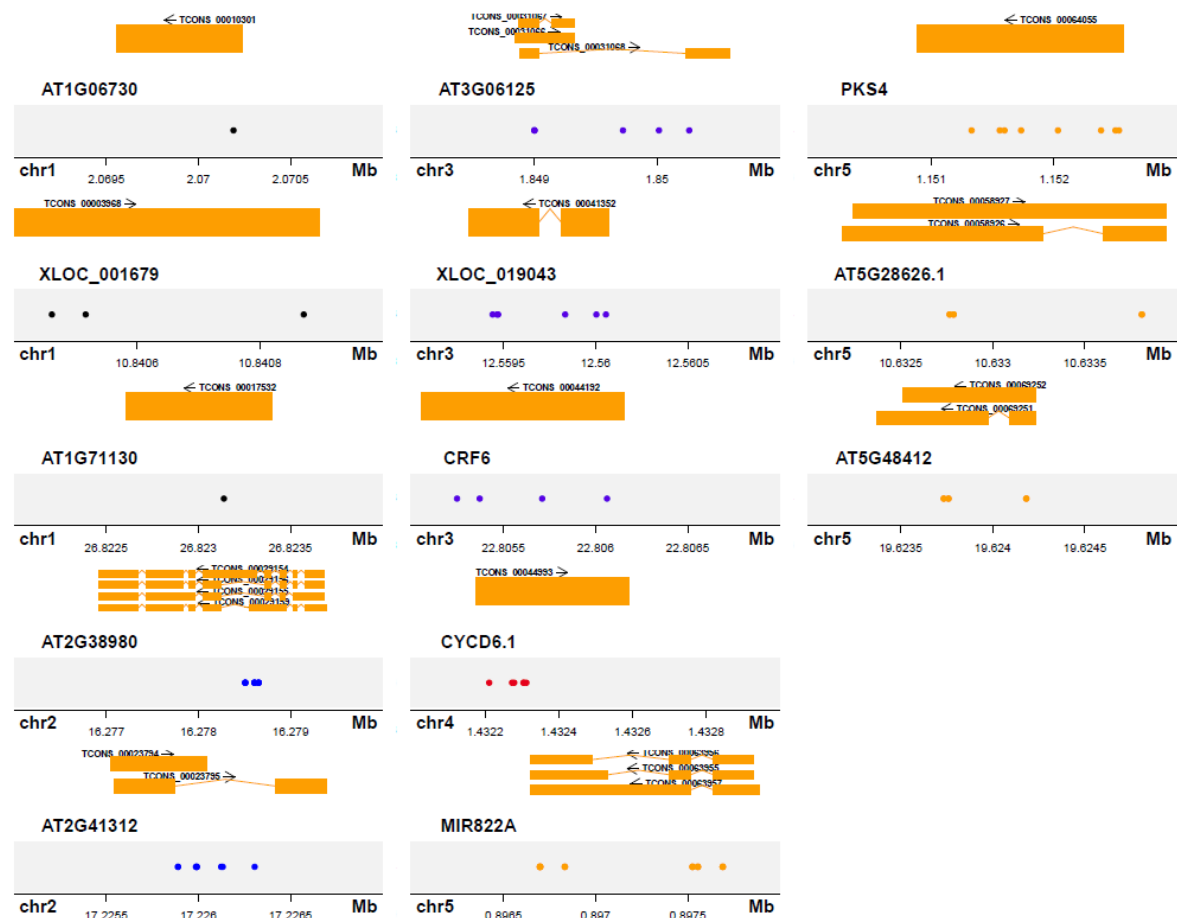
En total generamos una lista de 26 transcritos anotados como lncRNAs, los cuales se derivaron de 15 locus distintos y se hallaron expresados diferencialmente en hipocótilo. Para esta selección nos enfocamos principalmente en lncRNAs que co-expresaran en módulos de interés (SAS o similar) de acuerdo con la clasificación funcional en términos GO y que se encontraran en regiones intergénicas o estuvieran superpuestos parcialmente en un locus. También seleccionamos algunos transcritos que superpusieron con factores de transcripción, tales como *CRF6* y AT1G71130, así como otros ya identificados anteriormente como AT4G34881 y AT1G10682. En cotiledón si se encontraron lncRNAs, pero la mayoría se encontraban en regiones codificantes por lo que se descartaron (Tabla 3).



Para finalizar se descargaron las coordenadas de mutantes insercionales (T-DNA, Transposones) en *A. thaliana* para buscar mutantes en regiones no codificantes y poder validar los lncRNAs anotados. Se realizó una intersección con bedtools, en donde se encontró que 13 locus de los 15 fueron afectados por una inserción, de los cuales los transcritos *XLOC\_019043* y *XLOC\_001679* son un nuevos lncRNAs no anotados.

Símbolo Tair	Posición	Módulos	Tcons	chr	chrSt	chrE	Exones	tamaño (nt)
XLOC_001679	intern500	red	TCONS_00003968	1	10840250	10840897	1	647
AT1G71130	overlap	orange	TCONS_00017532	1	26822605	26823402	1	797
AT1G10682	interw500	blue	TCONS_00010894	1	3544990	3545471	1	481
AT1G06730	overlap	yellow	TCONS_00010301	1	2069555	2070236	1	681
AT2G38980	intern500	violet	TCONS_00029159	2	16276915	16279396	7	2481
AT2G38980	intern500	violet	TCONS_00029154	2	16276915	16279367	8	2452
AT2G38980	intern500	violet	TCONS_00029155	2	16276915	16279371	6	2456
AT2G38980	intern500	violet	TCONS_00029156	2	16276915	16279371	8	2456
AT2G41312	intern500	yellow	TCONS_00023794	2	17225522	17226045	1	523
AT2G41312	interw500	yellow	TCONS_00023795	2	17225541	17226693	2	1152
XLOC_019043	intercw500	purple	TCONS_00041352	3	12559313	12560069	2	756
AT3G06125	interw500	yellow	TCONS_00031066	3	1848847	1849332	1	485
AT3G06125	interw500	yellow	TCONS_00031068	3	1848882	1850596	2	1714
AT3G06125	interw500	yellow	TCONS_00031067	3	1848875	1849332	2	457
CRF6	overlap	pink	TCONS_00044192	3	22805057	22806154	1	1097
AT4G34881	overlap	pink	TCONS_00049365	4	16617622	16618506	2	884
CYCD6.1	overlap	yellow	TCONS_00044993	4	1432176	1432590	1	414
AT5G28626.1	interw500	magenta	TCONS_00058927	5	10632240	10633944	1	1704
AT5G28626.1	interw500	magenta	TCONS_00058926	5	10632185	10633944	2	1759
AT5G48412	intercn500	magenta	TCONS_00069252	5	19623515	19624232	1	717
AT5G48412	overlap	magenta	TCONS_00069251	5	19623372	19624232	2	860
MIR822A	intern500	orangered4	TCONS_00063958	5	897019	897356	1	337
MIR822A	intern500	orangered4	TCONS_00063956	5	896649	897855	3	1206
MIR822A	interw500	orangered4	TCONS_00063955	5	896649	897855	3	1206
MIR822A	interw500	orangered4	TCONS_00063957	5	896649	897889	2	1240
PKS4	overlap	skyblue3	TCONS_00064055	5	1150882	1152562	1	1680

**Tabla 3.** Lista de genes candidatos (lncRNAs) expresados diferencialmente en módulos de interés (respuesta a SAS).



**Figura 31.** Locus de transcritos no codificantes con inserción de T-DNA en alguna parte del cuerpo del gen.

## VII Discusión de resultados

En comparación con los microarreglos, RNA-seq es una herramienta mucho más poderosa para la detección y anotación de nuevos transcritos no codificantes, gracias a su sensibilidad, alto rendimiento y sin información de secuenciación previa (Lv et al., 2016).

En este trabajo se analizaron exhaustivamente transcriptomas de *A. thaliana* durante cuatro tiempos de exposición a luz roja lejana (condición de sombra, baja R:FR) en dos tejidos (cotiledones e hipocótilos) para la identificación y anotación de transcritos no codificantes expresados durante este fenómeno.

Existen diferentes criterios muy establecidos para la identificación de lncRNAs mediante enfoques bioinformáticos. El primer criterio es tomar en cuenta la longitud de la molécula. El corte por lo general es de 200 nucleótidos basados en las longitudes ya establecidas para ARNs cortos tales como mirRNAs, piRNAs, ARNs asociados a promotores (PASRs) y ARNs asociados a terminadores.

El presentar un ORF no es indicador de que se traduzca un transcrito. Es de esperarse que los lncRNAs puedan contener casualmente ORF parciales o completos. Por lo anterior se tiene que analizar la presencia de una proteína predicha antes de discriminar un ORF funcional de uno no funcional, ya que se asume que un ORF muy largo puede traducirse a una proteína. En nuestro pipeline utilizamos un ORF con un corte en 75 aa, un corte muy estricto comparado a otros pipelines donde utilizan cortes  $\leq$  a 100 aa ( $>$  100 aa se puede considerar codificante), porque, aunque existen lncRNAs con codones mayores a 100 (*H19*, *Xist*, *Mirg*, *Gtl2* y *KcnqOT1* por mencionar algunos), el utilizar un corte mayor puede dar lugar a falsos positivos, es decir que podemos tomar algunas proteínas menores de 100 aa (Dinger et al., 2008).

Para el tercer criterio se contempló que las secuencias no presentaran homología con proteínas, por lo tanto, se alinearon las secuencias con bases de datos de proteínas para descartar el potencial codificante.

Tomando en cuenta la reanotación del genoma de *A. thaliana* de Araport11 en donde encontraron un total de 3,559 lncRNAs (2,444 lincRNAs y 1,115 NATs) (Cheng et al., 2017), para nuestra lista de genes no codificantes clasificamos 906 lincRNAs (~63% menos que en Araport11) y genes que podrían resultar como NATs (superpuestos en nuestra clasificación) se determinaron 1,465 transcritos. Por lo que se puede decir que se encontraron más transcritos superpuestos en nuestra anotación comparada con Araport11, sin descartar las isoformas ensambladas.

En cuestión del tamaño, en el rango de 200 a 1,000 nucleótidos se encontró que la mayor parte de nuestros lncRNAs ensamblados, muy concordante con el trabajo en *A. thaliana*, en donde se anotaron 156 nuevos lncRNAs, los cuales se encontraban en un rango de 200 a 1400 nucleótidos con una longitud media de 982 nt .36 % menos de los genes codificantes (Lu et al., 2017). Comparando con un trabajo en algodón (*Gossypium hirsutum*), donde se anotaron un total de 1,117 lncRNAs únicos bajo estrés salino, se encontró que la mediana en longitud fue de 1,200 nt y que la mayoría también se encontró dentro de un rango menor a los 2,000 nt (Deng, Zhang, Wang, Yuan, & Shen, 2018).

Los lncRNAs al igual que los mRNAs pueden tener múltiples isoformas, lo cual nos indica que pueden poseer más de un exón de acuerdo a la anotación de lncRNAs en 17 especies (animales) (Hezroni et al., 2015). Aunque en nuestra anotación de lncRNAs, encontramos que la mayor parte de nuestros transcritos son monoexónicos, similar a un estudio donde anotaron lncRNAs (la mayoría monoexónicos) en respuesta a deficiencia de fósforo en *A. thaliana*, lo cual sugiere que los eventos de splicing son más abundantes en genes codificantes que en los no codificantes (Yuan et al., 2016). Sin embargo se encontró existencia de lncRNAs con más de un exón en nuestros datos, lo que indica que el splicing en lncRNAs es más común de lo que parece, ya que entre el 50-60% de estos contienen intrones (J. Liu, Wang, & Chua, 2015).

Por otra parte, en concordancia con el estudio de Soya, los lncRNAs tienen una longitud promedio comparados a los genes codificantes (Golicz, Singh, & Bhalla, 2017). También se cuantificaron los niveles de expresión, comparando lncRNAs

(intergénicos) con genes codificantes y de acuerdo con lo antes reportado, los genes no codificantes tienen una menor expresión que los codificantes (Golicz et al., 2017; Ulitsky, 2016).

De acuerdo con el análisis de expresión tomando en cuenta todos los genes (codificantes y no codificantes), encontramos varios genes reportados en respuesta a sombra. En cotiledones se encontraron genes codificantes involucrados en la respuesta de SAS como el gen *ATHB-4* (AT2G44910.1), un factor de transcripción homeodominio cierre de leucina clase II, el cual puede modular respuestas de SAS, controlando las respuestas moleculares y fisiológicas de hormonas (Sorin, Salla-Martret, Bou-Torrent, Roig-Villanova, & Martínez-García, 2009), genes *YUCC8* (AT4G28720) involucrados en la biosíntesis de auxinas (Y. Zhao et al., 2001), *ATHB-2* (AT4G16780) el cual es un regulador negativo inducido bajo las fluctuaciones de luz roja a luz roja lejana impidiendo la expansión de cotiledones (Steindler et al., 1999), así como proteínas de la superfamilia *bHLH* (AT3G21330 y AT5G66580) que participan en la transducción de señales del fitocromo (Duek & Fankhauser, 2005), *FLF1* (AT4G31380) encargado de la transición floral en periodos de días cortos en *A. thaliana* (Melzer, Kampmann, Chandler, & Appel, 1999), genes *IAA1* (AT4G14560) que codifican proteínas nucleares de vida corta que involucrados en la regulación de genes reguladores de respuesta a auxinas (Abel, Oeller, & Theologis, 1994), genes *ATKAT1* (AT5G46240) de canales de potasio que regulan la ingesta de potasio (Hirsch, R. E. Lewis B.D. Spalding E.P. Sussman, 1998), el factor de transcripción *BIM1*. (AT5G08130) involucrado en la señalización de brasinoesteroides en *A. thaliana* (Yin et al., 2005), *ESE3* (AT5G25190) involucrado en la respuesta a tolerancia salina inducido por etileno (L. Zhang et al., 2011) y fitocromos de la clase *phyA* (AT1G09570) (Reed, Nagatani, Elich, Fagan, & Chory, 1994) y el gen *EFE* (AT1G05010) catalizador de un precursor del etileno (Babula et al., 2006).

Algunos genes reprimidos en cotiledones durante los últimos tiempos de exposición a luz roja fueron relacionados a la superfamilia de Citocromos P450 afectando reacciones del metabolismo secundario (AT3G50660) (Bolwell, Holloway, & Index-

cytochrome, 1994), *JAIBA* (AT4G17460) que participan en el desarrollo de distintos órganos y regulación meristemática (Zúñiga-Mayo, Marsch-Martínez, & De Folter, 2012).

Al igual que en cotiledón, en hipocótilo encontramos algunos genes inducidos bajo sombra de manera compartida, como lo fue el gen *ATHB-2* (AT4G16780), *ATHB-4* (AT2G44910), factores que interactúan con el fitocromo como *PIF3*(AT2G46970), los genes *SAUR22* y *SAUR28* (AT5G18050, AT3G03830) que promueve la expansión celular (Spartz et al., 2012), *ATXYL1* (AT1G68560) que participa en la biosíntesis de xiloglucanos de hemicelulosa destinado a pared celular (Pauly et al., 2013), el fotorreceptor *phyB* (AT2G18790) (Reed et al., 1994), genes de la familia *GH3* (AT2G23170) de respuesta temprana a auxinas involucrados en la elongación de hipocótilos durante la exposición a luz roja lejana (Takase, Nakazawa, Ishikawa, Manabe, & Matsui, 2003), *NPH3* (AT1G67900) involucrado en la respuesta fototrópica (Motchoulski & Liscum, 1999).

Algunos genes reprimidos compartidos en hipocótilo durante la condición de sombra fueron los de la superfamilia del Citocromo P450 (AT4G36380) y *PBP1* (AT3G16420) que es una proteína de unión a *PYK10*.

La mayoría de nuestros lncRNAs presentaron perfiles de expresión espaciotemporales, ya sea que se expresen específicamente en tejidos, durante alguna etapa de desarrollo o bajo algún tipo de condición abiótica o biótica.

En nuestro análisis encontramos una mayor abundancia de genes codificantes en comparación a nuestros transcritos no codificantes tanto en cantidad afín al análisis en *Miscanthus lutarioriparius* donde encontraron la misma postura en cuanto a la proporción de genes codificantes y lncRNAs (Xu et al., 2017).

Simplemente con un enfoque bioinformático, es muy difícil encontrar directa y/específicamente una función biológica o mecanismo de un gen. Otro problema es que aún no existen bases de datos de lncRNAs estrictamente anotados funcionalmente en plantas (J. Wang, Meng, Dobrovolskaya, Orlov, & Chen, 2017). Sin embargo existe el enfoque (culpable por asociación), el cual consiste en

construir redes de coexpresión para agrupar tanto genes codificantes como no codificantes de acuerdo a sus perfiles de co-expresión, seguido de un análisis de enriquecimiento funcional (Feature, 2016; Signal, Gloss, & Dinger, 2016). La presencia de lncRNAs en clusters correlacionada con la expresión de ciertas proteínas nos puede ayudar a inferir el proceso biológico o celular en el cual puede estar involucrado (Feature, 2016). Por lo que podemos inferir que múltiples lncRNAs anotados en nuestro análisis están siendo regulados durante procesos de crecimiento celular y procesos de biosíntesis de hemicelulosa y polisacáridos de pared celular. Por otra parte, podrían existir diversos lncRNAs expresados durante la detección de luz roja lejana en tejidos fotosintéticos.

La luz es un factor ecológico muy importante en procesos regulatorios de crecimiento y desarrollo en las plantas, por lo que estudios en mecanismos como la foto morfogénesis en un punto clave para la biología de plantas (Jiao, Lau, & Deng, 2007).

Son muy pocos los transcritos no codificantes dependientes de una vía de luz, a los cuales se le han encontrado funciones particulares (X. Liu et al., 2015b). Tenemos el caso de *HID1*, un lncRNA que regula negativamente la fotomorfogénesis y el reverdecimiento en plántulas de *A. thaliana*, uniéndose a la región UTR y evitando la transcripción de *PIF3*. A diferencia de *APOLO*, *HID1* es un transcrito constitutivo, sin cambios en su expresión dependiente de luz roja. Por otro lado, también se han encontrado otros lncRNAs que interactúan con varios *PIFs* para reprimir y evitar que se transcriban genes de prevención a la sombra (genes SAS) durante condiciones de luz intensa. Se sugiere que esos lncRNAs están reprimidos en condición de sombra ya que solo se inducen bajo condiciones de luz alta.

Otro lncRNA dependiente de luz es *FLORE*, un transcrito anti sentido al gen *CDF5*, que regula el reloj circadiano del tiempo de floración, por lo que también se sugiere que este transcrito también se encuentra desregulado debido al estadio de las plántulas sometidas a sombra (Henriques et al., 2017).

En un reporte de Wang se identificaron 626 pares concordantes NAT y 766 pares discordantes NAT cuyos perfiles de expresión están relacionados con vías de

señalización por luz (condiciones normales de luz) (H. Wang et al., 2014). Debido a que no conocemos la dirección de nuestros lncRNAs, denominamos a los que probablemente resulten ser NAT como transcritos superpuestos, los cuales fueron en su mayoría regulados por el estado de sombra, tanto en cotiledones e hipocótilos. En el mismo estudio también se determinó si los efectos de luz tenían expresión específico temporal de desarrollo, en donde se representó que la mayor parte de transcritos (tanto codificantes como no codificantes) prefería expresarse en cotiledones (H. Wang et al., 2014), caso contrario al fenómeno de sombra donde los transcritos tanto codificantes y no codificantes tenían mayor preferencia por el tejido de hipocótilo.

En nuestro trabajo también surgió el interés de *APOLO*, un lncRNA implicado en la formación de loops de cromatina el cual se posiciona en el tercer cromosoma de *A. thaliana* (Ariel et al., 2014). Este lncRNA se expresa en presencia de auxinas en raíz. Interesantemente, en nuestro transcriptoma no pudimos encontrar este transcrito, lo cual nos habla de una especificidad de tejido ya que este gen es exclusivo de raíz. Mencionado esto, se cree que algunos de los lncRNAs expresados en módulos con función de respuesta a auxinas sean directamente dependientes de la hormona.

También existen diversos ncRNAs del tipo sRNAs, que actúan en procesos dependientes de luz. Por mencionar algunos, miR157d y miR319 que degradan a sus genes blancos, los cuales son los factores de transcripción *HY5* y *TCP* (*TEOSINTE BRANCHED1*, *CYCLOIDEA*, y *PCF*) que regulan positiva y negativamente la fotomorfogénesis (Sánchez-Retuerta, Suárez-López, & Henriques, 2018). miR171 que regula positivamente la producción de clorofilas uniéndose a su blanco *SCL*, permitiendo la sobre expresión de *POR* (protochlorophyllide oxidoreductase), la cual es clave para la síntesis de clorofilas (Sánchez-Retuerta et al., 2018).

Con todos los estudios mencionados, aún no hay lncRNAs anotados o identificados en condiciones de sombra (baja R:FR), sin embargo, en este trabajo se ha



incrementado el catálogo de transcritos, por lo que se puede especular un sinfín de funciones para nuestros candidatos.

Los lncRNAs intergénicos son funcionalmente diversos y se ha mantenido un gran interés en estudiarlos en plantas, ya que muy pocos han sido experimentalmente caracterizados (Marques & Ponting, 2014). De acuerdo con diversos estudios se habla de que los lncRNAs son tejido-específicos, y aunque faltan más ensayos para determinar esto, se demuestra esta tendencia en nuestros datos, ya que en hipocótilo encontramos más lncRNAs expresados diferencialmente los cuales no se expresaron en cotiledón. También los lncRNAs poseen pocos exones (un exón en promedio), se expresan relativamente bajos en comparación a los genes codificantes y son de tamaño pequeño. Tal y como lo observamos en nuestra lista final de genes, nuestros lncRNAs candidatos poseen las características de un lncRNA *bona fide*.

En cuanto a su manera de actuar sobre genes codificantes, se ha reportado que los lncRNAs pueden regular a los genes vecinos (*cis*), sin embargo los genes vecinos a los nuestros transcritos candidatos no se encuentran co-expresados, por lo cual podemos inferir que el lncRNAs está regulando de forma distal *trans* (Marques & Ponting, 2014).

Como se ha reportado antes, algunos lncRNAs pueden ser clasificados como pseudogenes o pueden derivarse de un elemento transponible (Marques & Ponting, 2014; Ponting et al., 2009). De acuerdo a la visualización de nuestro transcriptoma, dos de nuestros lncRNAs candidatos se derivan de estos elementos.

En un trabajo donde se identificaron 76 lncRNAs en librerías de cDNA en *A. thaliana* involucrados en respuesta a estrés abiótico (Amor et al., 2009b), donde nosotros pudimos identificar al menos 25 lncRNAs en nuestro transcriptoma, sin embargo, solo 10 lncRNAs que se derivaban de mRNAs estaban expresados en hipocótilo. En nuestra anotación, transcritos como el npc311 y npc072 anotados en sus librerías, los cuales se sometieron a diversos tipos de estrés abiótico para después medir su abundancia, se encontró que el primero se indujo en raíces, se reprimió en déficit de fósforo y en estrés hídrico mientras que el segundo se reprimió en estrés

salino (Amor et al., 2009b). Los transcritos mencionados anteriormente se encontraron diferencialmente expresados en hipocótilo, lo cual podría indicar que estos lncRNAs participan en la regulación mediada por luz roja lejana. Por otro lado, un lncRNA en particular, ncp375 (TCONS\_00019670) se expresó en el tejido floral (Amor et al., 2009b), por lo tanto no lo encontramos diferencialmente expresado debido a que nuestro transcriptoma era de plántulas.

Algunos lncRNAs pueden actuar directamente sobre genes vecinos en forma *cis* (Ariel, Romero-Barrios, Jégu, Benhamed, & Crespi, 2015). De acuerdo con nuestra anotación, identificamos 15 lncRNAs expresados que se derivaban de factores de transcripción, tal como el caso de *COLDAIR* y *COOLAIR*, cuyos transcritos se derivan del primer intrón y el segundo sobrelapa parcialmente la región 5' de *FLC*, para reprimir indirectamente el mismo locus (Csorba et al., 2014; Heo & Sung, 2011). Al menos 2 transcritos se derivaban de los factores de transcripción *CRF6* y *CRF8* (CITOKININ RESPONSE FACTOR 6 y 8), tal como un lncRNA lnc-225 identificado en estímulo de luz alta donde se expresaba simultáneamente junto con su gen vecino, el gen *CRF4* (Di et al., 2014). Hormonas como auxinas y citoquininas son importantes para controlar el crecimiento y desarrollo de una planta, en el caso de respuesta a sombra la planta debe ser capaz de alargarse para alcanzar la luz foto asimilable. Se ha reportado que algunos *CRFs* pueden regular la expresión de algunos genes *PIN* (transportadores de auxinas), específicamente de *PCRE1* y *PCRE7* gracias al reconocimiento de un motivo por factores de transcripción (Šimášková et al., 2015). En nuestro caso estos lncRNAs derivados de *CRFs* los encontramos reprimidos en hipocótilos durante sombra, lo cual era de esperarse, ya que, según lo reportado, estos lncRNAs se inducen en condición de luz intensa.

Una manera de actuar de los lncRNAs es reprimir la transcripción, secuestrando factores de transcripción o dirigiéndose directamente al ADN. En nuestra red amarilla y roja, encontramos diversos factores interconectados con lncRNAs hub, por lo que se puede inferir que estos pueden estar actuando de manera transcripcional o postranscripcional.

También se ha reportado que algunos lncRNAs pueden ser precursores de algunos sRNAs, tal como el caso de miR869, donde el lncRNA se sobreacumuló en mutantes de *dcl4* (Amor et al., 2009a). De acuerdo con nuestra anotación, se identificó un lncRNA expresado en el locus de miR822a, el cual se sugiere que podría ser un gen joven precursor de un miRNA que participe en el fenómeno de sombra.

Finalmente se encontraron 20 nuevos lncRNAs diferencialmente expresados (2 en cotiledón y 18 en hipocótilo, de los cuales dos están en nuestra lista de candidatos) no anotados en una base de datos. Estos y todos los transcritos identificados en este trabajo, no deberían considerarse lncRNAs hasta realizar ensayos rigurosos de su potencial codificante. Antes de su anotación se debe demostrar la restricción evolutiva que actúa sobre nuestras nuevas secuencias o quizás otra evidencia (Moghe et al., 2013).

## VIII Conclusiones

- Se anotaron 2,367 lncRNAs mediante el programa CPC y 159 con la anotación más estricta, de los cuales 138 intersecan en ambas anotaciones. 159 posibles nuevos lncRNAs no se encontraron en la anotación de Araport11.
- Según lo reportado, nuestros lncRNAs son principalmente monoexónicos (sin embargo, hay una gran cantidad de transcritos multiexónicos) y en general se expresan en niveles más bajos que los genes codificantes.
- Se encontraron 70 lncRNAs expresados en cotiledón y 384 expresados en hipocótilo, de los cuales 2 en cotiledón y 18 en hipocótilo no se encontraron anotados en ninguna base de datos.
- Los lncRNAs más abundantes son los que superponen con genes codificantes (~61%), seguido de los intergénicos (~37%) y por último los intrónicos (~0.7%).
- Se encontraron 40 módulos co-expresados de los cuales 29 arrojaron anotaciones con al menos un enriquecimiento en proceso biológico, función molecular o componente celular. 7 módulos presentaron enriquecimiento funcional interesante para participar en la prevención de sombra.
- Exclusivamente en hipocótilo, se identificaron 26 lncRNAs candidatos expresados con funciones relacionadas a la respuesta a sombra en 15 locus distintos.
- De los 15 locus con lncRNAs, 13 presentaron mutagénesis insercional (T-DNA), identificándose dos lncRNAs no anotados (*XLOC\_001679* y *XLOC\_019043*).

## IX Perspectivas

- Corroborar la transcripción en tejidos de *A. thaliana* durante el fenómeno de luz roja lejana en los transcritos candidatos.
- Estudiar mutantes insercionales de lncRNAs con pérdida función para evidenciar si hay o no un fenotipo.
- Determinar el mecanismo de acción de estos lncRNAs.
- Determinar la estructura secundaria y terciaria para esclarecer su posible mecanismo.
- Determinar la dirección de los transcritos que sobrelapan con genes codificantes.
- Estudiar tipos de conservación de los lncRNAs en diferentes especies cercanas.

## IX Bibliografía

- Abel, S., Oeller, P. W., & Theologis, A. (1994). Early auxin-induced genes encode short-lived nuclear proteins. *Biochemistry*, *91*(January), 326–330. <https://doi.org/10.1073/pnas.91.1.326>
- Alexa, A., & Rahnenfuhrer, J. (2016). topGO: Enrichment analysis for Gene Ontology. R package version 2.28.0. *Bioconductor*.
- Allen, T., Duek, P. D., Whitelam, G. C., Fankhauser, C., & Building, G. (2008). Phytochrome-mediated inhibition of shade avoidance involves degradation of growth-promoting bHLH transcription factors, *1*, 312–323. <https://doi.org/10.1111/j.1365-313X.2007.03341.x>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.
- Amor, B. Ben, Wirth, S., Merchan, F., Laporte, P., Hirsch, J., Maizel, A., ... Thermes, C. (2009a). Novel long non-protein coding RNAs involved in. *Genome Research*, (Vaucheret 2006), 57–69. <https://doi.org/10.1101/gr.080275.108.1>
- Amor, B. Ben, Wirth, S., Merchan, F., Laporte, P., Hirsch, J., Maizel, A., ... Thermes, C. (2009b). Title: Novel long non-protein coding RNAs involved in. *Sciences-New York*, (Vaucheret 2006), 57–69. <https://doi.org/10.1101/gr.080275.108.1>
- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. <https://doi.org/10.1093/bioinformatics/btu638>
- Andrews, S. (2014). FastQC A Quality Control tool for High Throughput Sequence Data. <http://www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/>.
- Ariel, F., Jegu, T., Latrasse, D., Romero-Barrios, N., Christ, A., Benhamed, M., & Crespi, M. (2014). Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. *Molecular Cell*, *55*(3), 383–396.
- Ariel, F., Romero-Barrios, N., Jégú, T., Benhamed, M., & Crespi, M. (2015). Battles and hijacks: Noncoding transcription in plants. *Trends in Plant Science*, *20*(6), 362–371. <https://doi.org/10.1016/j.tplants.2015.03.003>
- Azzalin, C. M., & Lingner, J. (2015). Telomere functions grounding on TERRA firma. *Trends in Cell Biology*, *25*(1), 29–36. <https://doi.org/10.1016/j.tcb.2014.08.007>
- Babula, D., Misztal, L. H., Jakubowicz, M., Kaczmarek, M., Nowak, W., & Sadowski, J. (2006). Genes involved in biosynthesis and signalisation of ethylene in Brassica oleracea and Arabidopsis thaliana: Identification and genome comparative mapping of specific gene homologues. *Theoretical and Applied Genetics*, *112*(3), 410–420. <https://doi.org/10.1007/s00122-005-0136-7>

- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bolwell, G. P., Holloway, R., & Index-cytochrome, K. W. (1994). Review article number 96 plant cytochrome p450. *Review Literature And Arts Of The Americas*, *37*(6).
- Bushmanova, E., Antipov, D., Lapidus, A., Suvorov, V., & Prjibelski, A. D. (2016). RnaQUAST: A quality assessment tool for de novo transcriptome assemblies. *Bioinformatics*, *32*(14), 2210–2212. <https://doi.org/10.1093/bioinformatics/btw218>
- Cabili, M., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., & Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes and Development*. <https://doi.org/10.1101/gad.17446611>
- Casal, J. J. (2012). Shade Avoidance. *The Arabidopsis Book*, *10*, e0157. <https://doi.org/10.1199/tab.0157>
- Casal, J. J. (2013). Photoreceptor Signaling Networks in Plant Responses to Shade. *Annual Review of Plant Biology*, *64*(1), 403–427. <https://doi.org/10.1146/annurev-arplant-050312-120221>
- Chalei, V., Sansom, S. N., Kong, L., Lee, S., Montiel, J. F., Vance, K. W., & Ponting, C. P. (2014). The long non-coding RNA Dali is an epigenetic regulator of neural differentiation. *ELife*, *3*(November), 1–24. <https://doi.org/10.7554/eLife.04530>
- Chekanova, J. A. (2015). Long non-coding RNAs and their functions in plants. *Current Opinion in Plant Biology*, *27*, 207–216. <https://doi.org/10.1016/j.pbi.2015.08.003>
- Chen, M., Chen, J., & Zhang, D. (2015). Exploring the secrets of long noncoding RNAs. *International Journal of Molecular Sciences*, *16*(3), 5467–5496. <https://doi.org/10.3390/ijms16035467>
- Cheng, C. Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., & Town, C. D. (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant Journal*, *89*(4), 789–804. <https://doi.org/10.1111/tpj.13415>
- Costa, F. F. (2005). Non-coding RNAs: New players in eukaryotic biology. *Gene*, *357*(2), 83–94. <https://doi.org/10.1016/j.gene.2005.06.019>
- Csorba, T., Questa, J. I., Sun, Q., & Dean, C. (2014). Antisense COOLAIR mediates the coordinated switching of chromatin states at *FLC* during vernalization. *Proceedings of the National Academy of Sciences*, *111*(45), 16160–16165. <https://doi.org/10.1073/pnas.1419030111>
- Delás, M. J., & Hannon, G. J. (2017). lncRNAs in development and disease: from functions to mechanisms. *Open Biology*, *7*(7), 170121. <https://doi.org/10.1098/rsob.170121>

- Deng, F., Zhang, X., Wang, W., Yuan, R., & Shen, F. (2018). Identification of *Gossypium hirsutum* long non-coding RNAs (lncRNAs) under salt stress. *BMC Plant Biology*, *18*(1), 1–14. <https://doi.org/10.1186/s12870-018-1238-0>
- Di, C., Yuan, J., Wu, Y., Li, J., Lin, H., Hu, L., ... Lu, Z. J. (2014). Characterization of stress-responsive lncRNAs in *Arabidopsis thaliana* by integrating expression, epigenetic and structural features. *Plant Journal*, *80*(5), 848–861. <https://doi.org/10.1111/tpj.12679>
- Ding, J., Lu, Q., Ouyang, Y., Mao, H., Zhang, P., Yao, J., ... Li, X. (2012). A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice, 317784(Nongken 58). <https://doi.org/10.1073/pnas.1121374109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1121374109>
- Dinger, M. E., Pang, K. C., Mercer, T. R., & Mattick, J. S. (2008). Differentiating Protein-Coding and Noncoding RNA: Challenges and Ambiguities. *PLOS Computational Biology*, *4*(11), e1000176. Retrieved from <https://doi.org/10.1371/journal.pcbi.1000176>
- Duek, P. D., & Fankhauser, C. (2005). bHLH class transcription factors take centre stage in phytochrome signalling, *10*(2), 11–14.
- Fang, Y., & Fullwood, M. J. (2016). Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer. *Genomics, Proteomics and Bioinformatics*, *14*(1), 42–54. <https://doi.org/10.1016/j.gpb.2015.09.006>
- Feature, T. (2016). Finding Function in novel, 6–8. Retrieved from [https://www.researchgate.net/profile/Michael\\_Hengartner/publication/7107661\\_Kaletta\\_T\\_Hengartner\\_MO\\_Finding\\_function\\_in\\_novel\\_targets\\_C\\_elegans\\_a\\_s\\_a\\_model\\_organism\\_Nat\\_Rev\\_Drug\\_Discov\\_5\\_387-398/links/09e4150a41cb5b578f000000.pdf](https://www.researchgate.net/profile/Michael_Hengartner/publication/7107661_Kaletta_T_Hengartner_MO_Finding_function_in_novel_targets_C_elegans_a_s_a_model_organism_Nat_Rev_Drug_Discov_5_387-398/links/09e4150a41cb5b578f000000.pdf)
- Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, *39*(suppl\_2), W29–W37.
- Fox, A. H., & Lamond, A. I. (2010). Paraspeckles. *Cold Spring Harbor Perspectives in Biology*, *2*(7), a000687.
- Franklin, K. A., & Whitelam, G. C. (2005). Phytochromes and shade-avoidance responses in plants. *Annals of Botany*, *96*(2), 169–175. <https://doi.org/10.1093/aob/mci165>
- Goff, L. A., & Rinn, J. L. (2015). Linking RNA biology to lncRNAs. *Genome Research*, *25*(10), 1456–1465. <https://doi.org/10.1101/gr.191122.115>
- Golicz, A., Singh, M. B., & Bhalla, P. L. (2017). The long intergenic non-coding RNA (lincRNA) landscape of the soybean genome. *Plant Physiology*, *176*(March), pp.01657.2017. <https://doi.org/10.1104/pp.17.01657>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652.



<https://doi.org/10.1038/nbt.1883>

- Henriques, R., Wang, H., Liu, J., Boix, M., Huang, L. F., & Chua, N. H. (2017). The antiphase regulatory module comprising CDF5 and its antisense RNA FLORE links the circadian clock to photoperiodic flowering. *New Phytologist*. <https://doi.org/10.1111/nph.14703>
- Heo, J. B., Lee, Y.-S., & Sung, S. (2013). Epigenetic regulation by long noncoding RNAs in plants. *Chromosome Research*, 21(6–7), 685–693. <https://doi.org/10.1007/s10577-013-9392-6>
- Heo, J. B., & Sung, S. (2011). Vernalization-Mediated Epigenetic Silencing by a Long Intronic Noncoding RNA Vernalization-Mediated Epigenetic Silencing by a Long Intronic Noncoding RNA. *Science*, 331(6013), 76–79.
- Hezroni, H., Koppstein, D., Schwartz, M. G., Avrutin, A., Bartel, D. P., & Ulitsky, I. (2015). Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Reports*, 11(7), 1110–1122.
- Hirsch, R. E. Lewis B.D. Spalding E.P. Sussman, M. . (1998). A role for the AKT1 potassium c in plant nutrition. *Science*, 280(5365), 918–921. <https://doi.org/10.1126/science.280.5365.918>
- Jiao, Y., Lau, O. S., & Deng, X. W. (2007). Light-regulated transcriptional networks in higher plants. *Nature Reviews Genetics*, 8(3), 217.
- Kim, E. D., & Sung, S. (2012). Long noncoding RNA: Unveiling hidden layer of gene regulatory networks. *Trends in Plant Science*, 17(1), 16–21. <https://doi.org/10.1016/j.tplants.2011.10.008>
- Kohnen, M. V., Schmid-Siebert, E., Trevisan, M., Petrolati, L. A., Sénéchal, F., Müller-Moulé, P., ... Fankhauser, C. (2016). Neighbor Detection Induces Organ-Specific Transcriptomes, Revealing Patterns Underlying Hypocotyl-Specific Growth. *The Plant Cell*, 28(12), 2889–2904. <https://doi.org/10.1105/tpc.16.00463>
- Kong, L., Zhang, Y., Ye, Z. Q., Liu, X. Q., Zhao, S. Q., Wei, L., & Gao, G. (2007). CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research*, 35(SUPPL.2), 345–349. <https://doi.org/10.1093/nar/gkm391>
- Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology*, 305(3), 567–580. <https://doi.org/10.1006/jmbi.2000.4315>
- Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9(1), 559.
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3). <https://doi.org/10.1186/gb-2009-10-3-r25>
- Lee, C., & Kikyo, N. (2012). Strategies to identify long noncoding RNAs involved in gene regulation. *Cell & Bioscience*, 2(1), 37. <https://doi.org/10.1186/2045->

- Leivar, P., Monte, E., Oka, Y., Liu, T., Carle, C., Castillon, A., ... Quail, P. H. (2008). Multiple Phytochrome-Interacting bHLH Transcription Factors Repress Premature Seedling Photomorphogenesis in Darkness. *Current Biology*, 18(23), 1815–1823. <https://doi.org/10.1016/j.cub.2008.10.058>
- Leivar, P., Tepperman, J. M., Cohn, M. M., Monte, E., Al-Sady, B., Erickson, E., & Quail, P. H. (2012). Dynamic Antagonism between Phytochromes and PIF Family Basic Helix-Loop-Helix Factors Induces Selective Reciprocal Responses to Light and Shade in a Rapidly Responsive Transcriptional Network in *Arabidopsis*. *The Plant Cell*, 24(4), 1398–1419. <https://doi.org/10.1105/tpc.112.095711>
- Li, X., Wu, Z., Fu, X., & Han, W. (2014). LncRNAs: Insights into their function and mechanics in underlying disorders. *Mutation Research - Reviews in Mutation Research*, 762, 1–21. <https://doi.org/10.1016/j.mrrev.2014.04.002>
- Lin, Y., Schmidt, B. F., Bruchez, M. P., & McManus, C. J. (2018). Structural analyses of NEAT1 lncRNAs suggest long-range RNA interactions that may contribute to paraspeckle architecture. *Nucleic Acids Research*, (March), 1–11. <https://doi.org/10.1093/nar/gky046>
- Liu, J., Wang, H., & Chua, N. H. (2015). Long noncoding RNA transcriptome of plants. *Plant Biotechnology Journal*, 13(3), 319–328. <https://doi.org/10.1111/pbi.12336>
- Liu, X., Hao, L., Li, D., Zhu, L., & Hu, S. (2015a). Long Non-coding RNAs and Their Biological Roles in Plants. *Genomics, Proteomics and Bioinformatics*, 13(3), 137–147. <https://doi.org/10.1016/j.gpb.2015.02.003>
- Liu, X., Hao, L., Li, D., Zhu, L., & Hu, S. (2015b). Long Non-coding RNAs and Their Biological Roles in Plants. *Genomics, Proteomics & Bioinformatics*, 13(3), 137–147. <https://doi.org/10.1016/j.gpb.2015.02.003>
- Love, M. I., Anders, S., & Huber, W. (2014). *Differential analysis of count data - the DESeq2 package*. *Genome Biology* (Vol. 15). <https://doi.org/10.1186/s13059-014-0550-8>
- Lu, Z., Xia, X., Jiang, B., Ma, K., Zhu, L., Wang, L., & Jin, B. (2017). Identification and characterization of novel lncRNAs in *Arabidopsis thaliana*. *Biochemical and Biophysical Research Communications*, 488(2), 348–354. <https://doi.org/10.1016/j.bbrc.2017.05.051>
- Lv, Y., Liang, Z., Ge, M., Qi, W., Zhang, T., Lin, F., ... Zhao, H. (2016). Genome-wide identification and functional prediction of nitrogen-responsive intergenic and intronic long non-coding RNAs in maize (*Zea mays* L.). *BMC Genomics*, 17(1), 350. <https://doi.org/10.1186/s12864-016-2650-1>
- Ma, L., Bajic, V. B., & Zhang, Z. (2013). On the classification of long non-coding RNAs. *RNA Biology*, 10(6), 924–933. <https://doi.org/10.4161/rna.24604>
- Marchese, F. P., Raimondi, I., & Huarte, M. (2017). The multidimensional mechanisms of long noncoding RNA function. *Genome Biology*, 18(1), 1–13.

<https://doi.org/10.1186/s13059-017-1348-2>

- Marques, A. C., & Ponting, C. P. (2014). Intergenic lncRNAs and the evolution of gene expression. *Current Opinion in Genetics and Development*, 27, 48–53. <https://doi.org/10.1016/j.gde.2014.03.009>
- Melzer, S., Kampmann, G., Chandler, J., & Appel, K. (1999). FPF1 modulates the competence to flowering in Arabidopsis. *Plant Journal*, 18(4), 395–405. <https://doi.org/10.1046/j.1365-313X.1999.00461.x>
- Mercer, T. R., Dinger, M. E., & Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nature Reviews Genetics*, 10(3), 155–159. <https://doi.org/10.1038/nrg2521>
- Metzker, M. L. (2010). Sequencing technologies the next generation. *Nature Reviews Genetics*, 11(1), 31–46. <https://doi.org/10.1038/nrg2626>
- Moghe, G. D., Lehti-Shiu, M. D., Seddon, A. E., Yin, S., Chen, Y., Juntawong, P., ... Shiu, S.-H. (2013). Characteristics and Significance of Intergenic Polyadenylated RNA Transcription in Arabidopsis. *Plant Physiology*, 161(1), 210–224. <https://doi.org/10.1104/pp.112.205245>
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, 5(7), 621–628. <https://doi.org/10.1038/nmeth.1226>
- Motchoulski, A., & Liscum, E. (1999). Arabidopsis NPH3: A NPH1 photoreceptor-interacting protein essential for phototropism. *Science*, 286(5441), 961–964. <https://doi.org/10.1126/science.286.5441.961>
- Nielsen, H., & Krogh, A. (1998). Prediction of signal peptides and signal anchors by a hidden Markov model. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology; ISMB. International Conference on Intelligent Systems for Molecular Biology*, 6, 122–130. <https://doi.org/10.1.1.47.4026>
- Pauly, M., Gille, S., Liu, L., Mansoori, N., de Souza, A., Schultink, A., & Xiong, G. (2013). Hemicellulose biosynthesis. *Planta*, 238(4), 627–642. <https://doi.org/10.1007/s00425-013-1921-1>
- Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S., & Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. *Nature*, 379(6561), 131–137. <https://doi.org/10.1038/379131a0>
- Pierik, R., Mommer, L., & Voeselek, L. A. (2013). Molecular mechanisms of plant competition: Neighbour detection and response strategies. *Functional Ecology*, 27(4), 841–853. <https://doi.org/10.1111/1365-2435.12010>
- Ponting, C. P., Oliver, P. L., & Reik, W. (2009). Evolution and Functions of Long Noncoding RNAs. *Cell*, 136(4), 629–641. <https://doi.org/10.1016/j.cell.2009.02.006>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>

- Reed, J. W., Nagatani, A., Elich, T. D., Fagan, M., & Chory, J. (1994). Phytochrome A and Phytochrome B Have Overlapping but Distinct Functions in Arabidopsis Development. *Plant Physiology*, 104(4), 1139–1149. <https://doi.org/10.1104/1139> [pii]
- Rinn, J. L., & Chang, H. Y. (2012). Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry*, 81(1), 145–166. <https://doi.org/10.1146/annurev-biochem-051410-092902>
- Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., ... Chang, H. Y. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell*, 129(7), 1311–1323. <https://doi.org/10.1016/j.cell.2007.05.022>
- Roig-Villanova, I., & Martínez-García, J. F. (2016). Plant Responses to Vegetation Proximity: A Whole Life Avoiding Shade. *Frontiers in Plant Science*, 7(February), 1–10. <https://doi.org/10.3389/fpls.2016.00236>
- Sánchez-Retuerta, C., Suárez-López, P., & Henriques, R. (2018). Under a New Light: Regulation of Light-Dependent Pathways by Non-coding RNAs. *Frontiers in Plant Science*, 9(July), 1–17. <https://doi.org/10.3389/fpls.2018.00962>
- Schmitt, A. M., & Chang, H. Y. (2016). Long Noncoding RNAs in Cancer Pathways. *Cancer Cell*, 29(4), 452–463. <https://doi.org/10.1016/j.ccell.2016.03.010>
- Seo, J. S., Sun, H., Park, B. S., Huang, C., Yeh, S., & Jung, C. (2017). ELF18-INDUCED LONG-NONCODING RNA Associates with Mediator to Enhance Expression of Innate Immune Response Genes in Arabidopsis, 29(May), 1024–1038. <https://doi.org/10.1105/tpc.16.00886>
- Signal, B., Gloss, B. S., & Dinger, M. E. (2016). Computational Approaches for Functional Prediction and Characterisation of Long Noncoding RNAs. *Trends in Genetics*, 32(10), 620–637. <https://doi.org/10.1016/j.tig.2016.08.004>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Šimášková, M., O'Brien, J. A., Khan, M., Van Noorden, G., Ötvös, K., Vieten, A., ... Benková, E. (2015). Cytokinin response factors regulate PIN-FORMED auxin transporters. *Nature Communications*, 6. <https://doi.org/10.1038/ncomms9717>
- Sorin, C., Salla-Martret, M., Bou-Torrent, J., Roig-Villanova, I., & Martínez-García, J. F. (2009). ATHB4, a regulator of shade avoidance, modulates hormone response in Arabidopsis seedlings. *Plant Journal*, 59(2), 266–277. <https://doi.org/10.1111/j.1365-313X.2009.03866.x>
- Spartz, A. K., Lee, S. H., Wenger, J. P., Gonzalez, N., Itoh, H., Inzé, D., ... Gray, W. M. (2012). The SAUR19 subfamily of SMALL AUXIN UP RNA genes promote cell expansion. *Plant Journal*, 70(6), 978–990. <https://doi.org/10.1111/j.1365-313X.2012.04946.x>

- Steindler, C., Matteucci, A., Sessa, G., Weimar, T., Ohgishi, M., Aoyama, T., ... Moro, P. A. (1999). Shade avoidance responses are mediated by the ATHB-2 HD-Zip protein, a negative regulator of gene expression, *4245*, 4235–4245.
- Takase, T., Nakazawa, M., Ishikawa, A., Manabe, K., & Matsui, M. (2003). DFL2, a New Member of the Arabidopsis GH3 Gene Family, is Involved in Red Light-Specific Hypocotyl Elongation. *Plant and Cell Physiology*, *44*(10), 1071–1080. <https://doi.org/10.1093/pcp/pcg130>
- Taylor, D. H., Chu, E. T.-J., Spektor, R., & Soloway, P. D. (2015). Long non-coding RNA regulation of reproduction and development. *Molecular Reproduction and Development*, *82*(12), 932–956. <https://doi.org/10.1002/mrd.22581>
- Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*, *25*(9), 1105–1111. <https://doi.org/10.1093/bioinformatics/btp120>
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., ... Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, *28*(5), 511–515. <https://doi.org/10.1038/nbt.1621>
- Tuck, A. C., & Tollervey, D. (2013). A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. *Cell*, *154*(5), 996–1009. <https://doi.org/10.1016/j.cell.2013.07.047>
- Ulitsky, I. (2016). Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nature Reviews Genetics*, *17*(10), 601–614. <https://doi.org/10.1038/nrg.2016.85>
- Wahlestedt, C. (2013). Targeting long non-coding RNA to therapeutically upregulate gene expression. *Nature Reviews Drug Discovery*, *12*(6), 433–446. <https://doi.org/10.1038/nrd4018>
- Wang, H., Chung, P. J., Liu, J., Jang, I. C., Kean, M. J., Xu, J., & Chua, N. H. (2014). Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. *Genome Research*, *24*, 3. <https://doi.org/10.1101/gr.165555.113.2006>
- Wang, J., Meng, X., Dobrovolskaya, O. B., Orlov, Y. L., & Chen, M. (2017). Non-coding RNAs and Their Roles in Stress Response in Plants Wang J et al / miRNA and lncRNA in Plant Stress Response. *Genomics, Proteomics and Bioinformatics*, *15*(5), 301–312. <https://doi.org/10.1016/j.gpb.2017.01.007>
- Wang, Y., Fan, X., Lin, F., He, G., Terzaghi, W., Zhu, D., & Deng, X. W. (2014). Arabidopsis noncoding RNA mediates control of photomorphogenesis by red light. *Proceedings of the National Academy of Sciences*, *111*(28), 10359–10364. <https://doi.org/10.1073/pnas.1409457111>
- Wilusz, J. E., Sunwoo, H., & Spector, D. L. (2009). Long noncoding RNAs : functional surprises from the RNA world Long noncoding RNAs : functional surprises from the RNA world, 1494–1504. <https://doi.org/10.1101/gad.1800909>

- Wright, M. W., & Bruford, E. a. (2011). Naming “junk”: human non-protein coding RNA (ncRNA) gene nomenclature. *Human Genomics*, 5(2), 90–98. <https://doi.org/10.1186/1479-7364-5-2-90>
- Xu, Q., Song, Z., Zhu, C., Tao, C., Kang, L., Liu, W., ... Sang, T. (2017). Systematic comparison of lncRNAs with protein coding mRNAs in population expression and their response to environmental change. *BMC Plant Biology*, 17(1), 1–15. <https://doi.org/10.1186/s12870-017-0984-8>
- Yin, Y., Vafeados, D., Tao, Y., Yoshida, S., Asami, T., & Chory, J. (2005). A new class of transcription factors mediates brassinosteroid-regulated gene expression in Arabidopsis. *Cell*, 120(2), 249–259. <https://doi.org/10.1016/j.cell.2004.11.044>
- Yoshida, M., Tsuiji, H., Furuno, M., Yoshimoto, R., Hasegawa, Y., & Nakagawa, S. (2011). Competition between a noncoding exon and introns: Gomafu contains tandem UACUAAC repeats and associates with splicing factor-1. *Genes to Cells*, 16(5), 479–490. <https://doi.org/10.1111/j.1365-2443.2011.01502.x>
- Yuan, J., Zhang, Y., Dong, J., Sun, Y., Lim, B. L., Liu, D., & Lu, Z. J. (2016). Systematic characterization of novel lncRNAs responding to phosphate starvation in Arabidopsis thaliana. *BMC Genomics*, 17(1), 655. <https://doi.org/10.1186/s12864-016-2929-2>
- Zhang, L., Li, Z., Quan, R., Li, G., Wang, R., & Huang, R. (2011). An AP2 Domain-Containing Gene, ESE1, Targeted by the Ethylene Signaling Component EIN3 Is Important for the Salt Response in Arabidopsis. *Plant Physiology*, 157(2), 854–865. <https://doi.org/10.1104/pp.111.179028>
- Zhang, Y. C., & Chen, Y. Q. (2013). Long noncoding RNAs: New regulators in plant development. *Biochemical and Biophysical Research Communications*, 436(2), 111–114. <https://doi.org/10.1016/j.bbrc.2013.05.086>
- Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J., & Lee, J. T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*, 322(5902), 750–756. <https://doi.org/10.1126/science.1163045>
- Zhao, Y., Christensen, S. K., Fankhauser, C., Cashman, J. R., Cohen, J. D., Weigel, D., & Chory, J. (2001). A Role for Flavin Monooxygenase – Like Enzymes in Auxin Biosynthesis. *Science*, 291(January), 306–310. <https://doi.org/10.1126/science.291.5502.306>
- Zúñiga-Mayo, V. M., Marsch-Martínez, N., & De Folter, S. (2012). JAIBA, a class-II HD-ZIP transcription factor involved in the regulation of meristematic activity, and important for correct gynoecium and fruit development in Arabidopsis. *Plant Journal*, 71(2), 314–326. <https://doi.org/10.1111/j.1365-313X.2012.04990.x>