



FACULTAD DE CIENCIAS AGRARIAS
UNIVERSIDAD NACIONAL DE ROSARIO

Estudio de la expresión génica mediada por
Brasinoesteroides en plantas de *Arabidopsis thaliana*

Mauricio Javier Grisolia

TRABAJO FINAL PARA OPTAR AL TÍTULO DE ESPECIALISTA EN
BIOINFORMÁTICA

Director: Lucas Daurelio

Co-Director: Luis Esteban

2018

Estudio de la expresión génica mediada por Brasinoesteroides en plantas de *Arabidopsis thaliana*

Mauricio Javier Grisolia

Licenciado en Biotecnología – Universidad Nacional de Rosario

Este Trabajo Final es presentado como parte de los requisitos para optar al grado académico de Especialista en Bioinformática, de la Universidad Nacional de Rosario y no ha sido previamente presentada para la obtención de otro título en ésta u otra Universidad. El mismo contiene los resultados obtenidos en investigaciones llevadas a cabo en el Centro de Estudios Fotosintéticos y Bioquímicos, durante el período comprendido entre el año 2013 y el año 2017, bajo la tutoría del Dr. Lucas Daurelio.

Mauricio Javier Grisolia

Lucas Daurelio

Luis Esteban

Defendida: 27 de Marzo de 2018

1 CONTENIDO

2	Listado de Ilustraciones y tablas.....	8
3	Introducción:.....	11
4	Objetivo General.....	13
4.1	Objetivos Específicos:	13
5	Materiales y métodos:.....	13
5.1	Datos crudos	13
5.2	Análisis de datos:	15
5.2.1	Uso de R:	15
5.3	Descripción de los paquetes de R:	16
5.3.1	Affy.....	16
5.3.2	arrayQualityMetrics.....	16
5.3.3	Limma.....	16
5.3.4	GOstats	16
5.3.5	Biobase	17
5.3.6	GEOquery.....	17
5.3.7	ath1121501probe	17
5.3.8	ath1121501cdf	17
5.3.9	ath1121501.db	17
5.4	Descripción de herramientas adicionales.....	17
5.4.1	CisGenome:	17
5.4.2	Galaxy:	18
5.4.3	Cytoscape:	18
5.4.4	Genemania:.....	18

5.5	Análisis de archivos de Microarreglos con Bioconductor	18
5.5.1	Curva BL vs tiempo.....	18
5.5.2	Col0 "Mock" vs Col0 "BL"	22
5.5.3	Depurado de lista de genes de expresión diferencial.....	24
5.5.4	Análisis de Microarreglos CATMA de Arabidopsis thaliana WS vs the1-.....	26
5.5.5	Análisis ChIP-chip con CisGenome.....	32
5.5.6	Análisis ChIP-Seq	33
5.6	Gráficos de Venn con el paquete Limma.....	34
5.6.1	Venn: comparación plantas tratadas y no tratadas con brasinoesteroides	35
5.6.2	Venn: comparación Goda 1h, Goda 30m y "consensoBL"	36
5.6.3	Venn: comparación genes consenso a tiempos menores	36
5.6.4	Venn: genes regulados por BR mediados y no mediados por los reguladores transcripcional BZR1 BES1 y PIF.	37
5.6.5	Venn: vía THE1/HERK.....	38
5.6.6	Venn: THE1/HERK1 / consenso BL / FT:.....	39
5.6.7	Venn: vía THE1/HERK1 vs genes de respuesta temprana y tardía.	40
5.7	Consulta de términos GO.....	41
5.7.1	Consulta GO: lista de genes BL CONSENSO	41
5.7.2	Consulta GO: lista de genes BL CONSENSO sin regulador	42
5.7.3	Consulta GO: lista de genes THE1 CONSENSO.....	42
6	Resultados y discusión:	43
6.1	Selección de los data set y obtención de datos crudos.....	43
6.2	Determinación del grupo de genes de respuesta consenso a Brasinoesteroides.	43
6.2.1	Selección de experimentos que muestren la respuesta de plantas silvestres al tratamiento con Brasinoesteroides:.....	43
6.2.2	Procesamiento de datos y Análisis expresión diferencial por microarreglo: Curva BL vs tiempo	43

6.2.3	Análisis microarreglo Affymetrix Col0 "Mock" vs Col0 "BL"	46
6.2.4	Determinación de la respuesta consenso al tratamiento con BR.	47
6.2.5	Consulta de términos GO para cada lista de identificadores.	51
6.2.6	Construcción de redes funcionales por "Genemania".	53
6.3	Determinación de genes regulados por los factores de transcripción BZR1, BES1 y PIF. 55	
6.3.1	Selección de experimentos de inmunoprecipitación de la cromatina, orientada a factores de transcripción (FT) asociados a la respuesta transcripcional a la acción de Brasinoesteroides:.....	55
6.3.2	Análisis ChIP-chip del regulador BES1	55
6.3.3	Análisis ChIP-chip del regulador BZR1.....	56
6.3.4	Análisis ChIP-Seq del regulador PIF1	57
6.3.5	Discriminación de grupos de genes regulados por BR mediados y no mediados por los reguladores transcripcional BZR1 BES1 y PIF.	57
6.3.6	Consulta de términos GO para cada lista de identificadores.	59
6.3.7	Construcción de redes funcionales por "Genemania".	61
6.4	Identificación de los genes específicos para la vía THE1/HERK y contrastación con los genes regulados por los factores de transcripción evaluados.	62
6.4.1	Selección de experimentos que muestren la respuesta transcripcional de plantas mediada por la vía THE1/HERK1:.....	62
6.4.2	Análisis del perfil de expresión de Arabidopsis thaliana WS vs the1-3.....	62
6.4.3	Análisis de los perfiles de expresión de plantas mutantes en the1 herk1:.....	63
6.4.4	Identificación genes específicos para la vía THE1/HERK.....	63
6.4.5	Correlacionar la vía de respuesta THE1/HERK1 con los asociados a respuesta temprana y tardía.	67
6.4.6	Consulta de términos GO para cada lista de identificadores.	68
6.4.7	Construcción de redes funcionales por "Genemania".	71
7	Conclusiones.....	73
8	Bibliografía:.....	74

2 LISTADO DE ILUSTRACIONES Y TABLAS

<i>Ilustración 1</i> Gráfico caja y bigotes mostrando la intensidad de fondo para el canal verde.....	27
<i>Ilustración 2</i> Gráfico caja y bigotes mostrando la intensidad de fondo para el canal rojo.....	27
<i>Ilustración 3</i> Heterogeneidad espacial de la intensidad de fondo para el primer microarreglo	28
<i>Ilustración 4</i> Heterogeneidad espacial de la intensidad de fondo para el primer microarreglo	28
<i>Ilustración 5</i> Gráfico de densidad de intensidad en los canales rojo y verde, para todo el set de datos sin normalizar	29
<i>Ilustración 6</i> Gráfico de densidad de intensidad en los canales rojo y verde, para todo el set de datos, con corrección de fondo	29
<i>Ilustración 7</i> Gráfico de densidad de intensidad en los canales rojo y verde, para todo el set de datos con normalización inter data set	30
<i>Ilustración 8</i> Esquema del Flujo de trabajo utilizado en galaxy.org	34
<i>Ilustración 9</i> Gráfico de clústeres representando todas las comparaciones ensayadas en manera grupal. Solo se muestran las primeras 40 sondas. Los valores varían des represión (magenta) a expresión (turquesa).	45
<i>Ilustración 10</i> Gráfico de clústeres representando las comparaciones de todos los miembros del set de datos. Solo se muestran las primeras 40 sondas. Los valores varían des represión (magenta) a expresión (turquesa).	46
<i>Ilustración 11</i> Diagrama de Venn entre tres experimentos de transcriptómica de plantas salvajes tratadas con BR. La intersección muestra 20 genes de respuesta consenso.....	48
<i>Ilustración 12</i> Diagrama de Venn entre la respuesta consenso a BR, y set de datos evaluados a 2 tiempos anteriores.	50
<i>Ilustración 13</i> Gráfico de frecuencias de términos GO asociados a la respuesta consenso a Brasinoesteroides	53
<i>Ilustración 14</i> Red funcional de respuesta consenso a Brasinoesteroides generada por Genemania. Los genes consulta se encuentran identificados con rayas. Las referencias se encuentran dentro del gráfico.	54
<i>Ilustración 15</i> Diagrama de Venn comparando la respuesta consenso a BR, y los genes controlados por tres FT distintos.	58
<i>Ilustración 16</i> Gráfico de frecuencias de términos GO asociados a la respuesta consenso a Brasinoesteroides sin regulador asociado.	60
<i>Ilustración 17</i> Red funcional de respuesta consenso a Brasinoesteroides sin regulador asignado, generada por Genemania. Los genes consulta se encuentran identificados con rayas. Las referencias se encuentran dentro del gráfico.	61
<i>Ilustración 18</i> Diagrama de Venn identificando la respuesta consenso a BR mediada por la vía THE1/HERK1.	64
<i>Ilustración 19</i> Diagrama de Venn comparando la respuesta consenso a BR, la respuesta mediada por la vía THE1 y la respuesta consenso sin regulador asignado.....	66
<i>Ilustración 20</i> Diagrama de Venn comparando la respuesta consenso a BR, la respuesta mediada por THE1 y la respuesta obtenida a 30 minutos y 1 hora post tratamiento con BR.	68

<i>Ilustración 21 Gráfico de frecuencias de términos GO asociados a la respuesta mediada por la vía THE1/HERK1.</i>	70
<i>Ilustración 22 Red funcional de respuesta consenso a Brasinoesteroides por la vía THE1/HERK1, generada por Genemania. Los genes consulta se encuentran identificados con rayas. Las referencias se encuentran dentro del gráfico.</i>	71
<i>Tabla 1 Resumen de los datos experimentales de estudios de transcriptómica a utilizar. Crecimiento: (O), Oscuridad continua; (L) Luz continua, (C) Ciclo luz oscuridad 16/8. (BL) Brassinolido, fitohormona natural que se clasifica dentro del grupo de los Brasinoesteroides. TPG: tiempo post germinación. NA: No aplicable.</i>	13
<i>Tabla 2 Grupos de comparación de datos de microarreglos para la curva BL-Tiempo de respuesta.</i>	44
<i>Tabla 3 Comparaciones realizadas en el análisis de expresión diferencial como respuesta a aplicación de BL a lo largo del tiempo.</i>	45
<i>Tabla 4 Grupos de comparación de datos de microarreglos para la comparación Col0 Mock vs Col0 BL.</i>	46
<i>Tabla 5 Grupo de datos, nombre de archivos y set de datos originales utilizados para la comparación de la respuesta consenso a Brasinoesteroides.</i>	47
<i>Tabla 6 Lista de genes que componen el repuesto consenso a Brasinoesteroides. La consulta fue realizada a partir de la herramienta Gene Description Search and Download de TAIR. NA: no aplica.</i>	49
<i>Tabla 7 Lista de genes que componen la repuesta consenso a tiempos medios a la acción de Brasinoesteroides. La consulta fue realizada a partir de la herramienta "Gene Description Search and Download" de TAIR. NA: no aplica.</i>	50
<i>Tabla 8 Reporte GOstats para la respuesta consenso a BR. ID de los términos GO (GOMFID); los p-value asociado al enriquecimiento del mismo en la selección de genes (Pvalue); la cantidad de veces que se encontró el término (Count); la cantidad esperada para cada término (ExpCount); la "razón de exceso" o "razón de probabilidades" (OddsRatio), parámetro que indica cuan fuertemente se presenta el termino en relación a lo esperado; y el número de genes asociado a este término en la base de datos utilizada (Size).</i>	52
<i>Tabla 9 Grupo de datos, nombre de archivos y set de datos originales utilizados para la clasificación de la respuesta a BL con y sin regulador identificado.</i>	57
<i>Tabla 10 Lista de genes que componen la repuesta consenso a Brasinoesteroides, con información sobre el regulador transcripcional asociado a la respuesta. En los genes cuyos reguladores no han sido estudiados o identificados se supone una vía de regulación alternativa. La consulta fue realizada a partir de la herramienta Gene Description Search and Download de TAIR. NA: no aplica.</i>	58
<i>Tabla 11 Salida de GOStats para la respuesta consenso a BR sin regulador identificado. ID de los términos GO (GOMFID); los p-value asociado al enriquecimiento del mismo en la selección de genes (Pvalue); la cantidad de veces que se encontró el término (Count); la cantidad esperada para cada término (ExpCount); la "razón de exceso" o "razón de probabilidades" (OddsRatio), parámetro que indica cuan fuertemente se presenta el termino en relación a lo esperado; y el número de genes asociado a este término en la base de datos utilizada (Size).</i>	60
<i>Tabla 12 Diseño del Intercambio de fluoróforos utilizados en la comparación WS the1-3.</i>	62

<i>Tabla 13 Grupo de datos, nombre de archivos y set de datos originales utilizados para la comparación de la respuesta consenso a Brasinoesteroides por la vía THE1/HERK1.....</i>	<i>63</i>
<i>Tabla 14 Lista de genes que componen la repuesta mediada por la vía THE1/HERK1. La consulta fue realizada a partir de la herramienta Gene Description Search and Download de TAIR. NA: no aplica.</i>	<i>64</i>
<i>Tabla 15 Lista de genes consenso involucrados en la respuesta mediada por THE1/HERK1, con información sobre el regulador transcripcional asociado a la respuesta. En los genes cuyos reguladores no han sido estudiados o identificados se supone una vía de regulación alternativa. La consulta fue realizada a partir de la herramienta Gene Description Search and Download de TAIR. NA: no aplica.</i>	<i>67</i>
<i>Tabla 16 Descripción de los archivos utilizados para asociar la respuesta THE1 con la curva de respuesta a lo largo del tiempo</i>	<i>67</i>
<i>Tabla 17 Reporte GOstats para la respuesta consenso mediada por THE1/HERK1. ID de los términos GO (GOMFID); los p-value asociado al enriquecimiento del mismo en la selección de genes (Pvalue); la cantidad de veces que se encontró el término (Count); la cantidad esperada para cada término (ExpCount); la “razón de exceso” o “razón de probabilidades” (OddsRatio), parámetro que indica cuan fuertemente se presenta el termino en relación a lo esperado; y el número de genes asociado a este término en la base de datos utilizada (Size).....</i>	<i>68</i>

3 INTRODUCCIÓN:

El crecimiento de los órganos en *Arabidopsis thaliana* puede dividirse en dos etapas: una etapa temprana de proliferación celular, generando las particiones del tejido de los órganos; y una etapa mediada por expansión celular donde ocurre el alargamiento celular, estableciendo el tamaño que tendrá el órgano maduro (Donnelly, Bonetta et al. 1999). En estas etapas la regulación hormonal juega un papel fundamental, desencadenando efectos pleiotrópicos según el requerimiento de la planta entre los que se encuentra la deposición y modificación del material de la pared celular (Reiter 2002). Existen tres hormonas principales involucradas en la regulación del crecimiento celular, Giberelinas, Auxinas y Brasinoesteroides, siendo las últimas dos las mayormente estudiadas (Sun, Fan et al. 2010).

Los Brasinoesteroides (BR) son hormonas esenciales para un amplio número de procesos fisiológicos a través de todo el ciclo de la vida de las plantas (Clouse and Sasse 1998). Estas hormonas, se clasifican como un grupo de hormonas esteroides de plantas, las cuales juegan un rol fundamental en orquestar las respuestas a estímulos ambientales y de desarrollo. Se encuentran ampliamente distribuidos entre las embriofitas, regulando múltiples procesos biológicos, tales como la elongación de la pared celular, la diferenciación vascular, fotomorfogénesis, senescencia y respuesta a estrés. Mutantes deficientes o insensibles a los BR presentan una gran variedad de defectos en el crecimiento, entre los que se encuentran enanismo, menor tasa de germinación de las semillas, hojas curvadas, floración tardía, macho esterilidad y fotomorfogénesis alterada (Li, Nagpal et al. 1996, Szekeres, Nemeth et al. 1996). Estudios Fisiológicos en *Arabidopsis thaliana* demostraron que los BR promueven la elongación celular, aumentan la tolerancia al estrés ambiental y aumentan la resistencia al ataque de patógenos. Los BR actúan a través del receptor del tipo quinasa BRI1, y utiliza una vía de transducción de señales para activar por desfosforilación miembros de los factores de transcripción de la familia BZR (Kim, Guan et al. 2009) y BES (Singh, Fridman et al. 2014). Actualmente se sugiere que la fosforilación controla a los FT BES1/BZR1 de diversas maneras, incluyendo el marcado para degradación mediante proteosoma, la retención de estos factores en el citoplasma, y la disminución en la actividad de unión a ADN (HEMATY). Estos reguladores son únicos en plantas y presentan un elevado porcentaje de identidad y similitud. Ambos presentan una región de unión a ADN del tipo hélice-giro-hélice básico en el extremo N terminal, y se unen a la región promotora E-Box (CANNTG) y BRRE (CGTGT/CG), respectivamente (He et al., 2005). De la totalidad de genes inducidos por BR, se cree que un 80% de ellos se encuentran regulados por la acción de BRI1 (Sun et al., 2010), mientras que los restantes podrían ser regulados por vías alternativas. Numerosos genes que responden al tratamiento por BR han sido identificados por diversos estudios de microarreglos (Mussig, Fischer et al. 2002, Goda, Sawa et al. 2004, Nemhauser, Mockler et al. 2004, Goda, Sasaki et al. 2008, Guo, Li et al. 2009). Sin embargo, el grupo de datos correspondiente a estos estudios se superpone solo parcialmente, presuntamente debido a diferencias en las condiciones ensayadas y a los efectos tejidos específicos de los BR. Esto sugiere que posiblemente no todos los genes de respuesta a BR hayan sido identificados todavía.

Entre los genes regulados por los BR se encuentran los que codifican enzimas con actividades asociadas al metabolismo de la pared celular como celulosas sintasas, pectinesterasas, xiloglucosiltransferasas y expansinas. Se cree que estas últimas direccionan la expansión celular en respuesta a BR, favoreciendo una desestructuración de la pared celular, seguida de una elongación mediada por las fuerzas asociadas al turgor celular (Cosgrove 1993). Estudios previos realizados en nuestro laboratorio indicaron que la expresión de módulos de unión a carbohidratos (CBM) en la pared celular (plantas EXPA8-SBD123) pueden favorecer la expansión celular de manera similar a la acción de las expansinas (Grisolia, Peralta et al. 2016). Es más, diversas expansinas presentan en su secuencia dominios de unión a carbohidratos sospechados de jugar un papel clave en el mecanismo de acción de las mismas (Marga, Grandbois et al. 2005).

La proteína THE1 pertenece a un grupo de receptores de membrana del tipo quinasa, con la capacidad de sensar la integridad de la pared celular y de desencadenar una señal al interior de la célula, generalmente asociada a la promoción del crecimiento y/o a la defensa contra patógenos. Estudios previos han caracterizado parcialmente la vía de señalización THE1/HERK y se ha propuesto que son capaces de regular la expresión de muchos genes diana por medio de un regulador aún no conocido. Muchos de estos genes se encuentran asociado al crecimiento y un subgrupo de ellos (incluyendo al mismo receptor THE1) son también regulados por BR (Guo, Ye et al. 2009). Por ejemplo, se conoce que la activación de THE1, desencadena la regulación positiva de cierto grupo de genes de defensa, inducidos por la activación de PRRs (Receptores de reconocimiento de patrones (Williamson, Burn et al. 2001). De esta manera, es concebible pensar que THE1 juegue un rol adicional en la defensa contra patógenos que dañen la pared celular. De hecho, la mayoría de los patógenos de plantas secretan enzimas modificadoras de la pared, y algunos incluso secretan inhibidores de la síntesis de celulosa, tal como la thaxtomina (Hematy, Sado et al. 2007). En las plantas EXPA8-SBD123 hemos encontrado un aumento de 13 veces en la expresión del receptor THE1 en comparación con las plantas salvajes (Col0) y una regulación luz dependiente de la tasa de elongación del hipocótilo, característica de una regulación hormonal (resultados no publicados). Por otro lado, Sun y colaboradores (Sun, Fan et al. 2010) mediante experimentos ChIP, determinaron una batería de genes cuyos promotores son blanco de unión al regulador BZR1. De esta forma, es posible discriminar entre los genes afectados por la acción de BR cuales lo hacen por acción de BZR1. Por otro lado, mediante la misma técnica, el mismo grupo ha determinado el conjunto de genes asociados a los reguladores transcripcionales BES1 y PIF. Es conveniente aclarar que este último regulador, se encuentra íntimamente asociado a la regulación transcripcional de genes en ausencia de luz y es capaz de interactuar con BZR1 y BES1 (Oh, Zhu et al. 2012). Este mecanismo de regulación es muy importante debido a que en las plantas EXPA8-SBD123, de la misma forma que en plantas silvestres tratadas con BR (Schroder, Lisso et al. 2014), presentan una atenuación en los fenotipos de incremento de crecimiento. Estos estudios dejan muchas preguntas abiertas, como, por ejemplo: ¿Qué grupos de genes asociados a la respuesta a BR son comunes en todas las condiciones? ¿Es posible asignar una función específica a estos genes? ¿Es posible que genes asignados a la regulación directa BR, en realidad estén asociados a la vía THE1/HERK? ¿Es posible que la vía THE1/HERK sea una vía de respuesta secundaria, activando genes de manera tardía? De esta forma, utilizando información

transcriptómica pública y estudios moleculares (ChIP-chip) y fisiológicos, propongo los siguientes objetivos:

4 OBJETIVO GENERAL

Contrastar la variación en la expresión de genes mediada por BR por la vía THE1/HERK en *Arabidopsis thaliana*.

4.1 OBJETIVOS ESPECÍFICOS:

- Determinar el grupo de genes que se expresan en común en las distintas condiciones de tratamiento con BR y clasificarlos según su función fisiológica.
- Determinar los genes regulados por acción de los BR, utilizando un entrecruzamiento con bases de datos ChIP-chip de los sitios de unión del regulador transcripcional BZR1, BES1 y PIF.
- Discriminar grupos de genes regulados por BR mediados y no mediados por BZR1, y cruzar estas bases de datos con ensayos de microarreglos en mutantes en la vía THE1/HERK.
- Identificar genes específicos para la vía THE1/HERK.
- Clasificar estos genes (GO) e interpretar los resultados.
- Utilizar esta información para describir una red de regulación con las herramientas apropiadas.

5 MATERIALES Y MÉTODOS:

5.1 DATOS CRUDOS

Los ensayos de transcriptómica fueron identificados de publicaciones específicas, según la *Tabla 1*.

Tabla 1 Resumen de los datos experimentales de estudios de transcriptómica a utilizar. Crecimiento: (O), Oscuridad continua; (L) Luz continua, (C) Ciclo luz oscuridad 16/8. (BL) Brassinolido, fitohormona natural que se clasifica dentro del grupo de los Brasinoesteroides. TPG: tiempo post germinación. NA: No aplicable.

DB	Acc. ID	Técnica	Entorno	Trat.	Crec.	TPG	ID muestra	Referencia
GEO	GSE14788	Microarreglo	Col0	mock	O	7-days	GSM351326	(Song, Zhou et al. 2009)
GEO	GSE14788	Microarreglo	Col0	mock	O	7-days	GSM351327	(Song, Zhou et al. 2009)
GEO	GSE14788	Microarreglo	Col0	200 nM BL	O	7-days	GSM351328	(Song, Zhou et al. 2009)
GEO	GSE14788	Microarreglo	Col0	200 nM BL	0	7-days	GSM351329	(Song, Zhou et al. 2009)

TAIR	1007966053	Microarreglo	Col0	Mock (30 min)	L	7-days	RIKEN-Goda Sample1	(Goda, Sasaki et al. 2008)
TAIR	1007966053	Microarreglo	Col0	Mock (30 min)	L	7-days	RIKEN-Goda Sample1	(Goda, Sasaki et al. 2008)
TAIR	1007966053	Microarreglo	Col0	10 nM BL (30 min)	L	7-days	RIKEN-Goda Sample8	(Goda, Sasaki et al. 2008)
TAIR	1007966053	Microarreglo	Col0	10 nM BL (30 min)	L	7-days	RIKEN-Goda Sample8	(Goda, Sasaki et al. 2008)
TAIR	1007966053	Microarreglo	Col0	Mock (1h)	L	7-days	RIKEN-Goda Sample9	(Goda, Sasaki et al. 2008)
TAIR	1007966053	Microarreglo	Col0	Mock (1h)	L	7-days	RIKEN-Goda Sample9	(Goda, Sasaki et al. 2008)
TAIR	1007966053	Microarreglo	Col0	10 nM BL (1h)	L	7-days	RIKEN-Goda Sample16	(Goda, Sasaki et al. 2008)
TAIR	1007966053	Microarreglo	Col0	10 nM BL (1h)	L	7-days	RIKEN-Goda Sample16	(Goda, Sasaki et al. 2008)
TAIR	1007966053	Microarreglo	Col0	Mock (3h)	L	7-days	RIKEN-Goda Sample17	(Goda, Sasaki et al. 2008)
TAIR	1007966053	Microarreglo	Col0	Mock (3h)	L	7-days	RIKEN-Goda Sample17	(Goda, Sasaki et al. 2008)
TAIR	1007966053	Microarreglo	Col0	10 nM BL (3h)	L	7-days	RIKEN-Goda Sample24	(Goda, Sasaki et al. 2008)
TAIR	1007966053	Microarreglo	Col0	10 nM BL (3h)	L	7-days	RIKEN-Goda Sample24	(Goda, Sasaki et al. 2008)
GEO	GSE7937	2 color Microarreglo	WS vs the1	0	Dark grown	5-days	GSM194873	(Hematy, Sado et al. 2007)
GEO	GSE7937	2 color Microarreglo	The1 vs WS	0	O	5-days	GSM194874	(Hematy, Sado et al. 2007)
NA	Tabla S1 y S2	NA	Col0 BR+ vs BR-	1 uM BL (2,5h)	C	20-days	NA	(Guo, Li et al. 2009)
NA	Tablas S3 y S4	NA	herk the1 BR+ vs BR-	1 uM BL (2,5h)	C	20-days	NA	(Guo, Li et al. 2009)
GEO	GSE24684	ChIP-chip BES1	BES1	NA	C	14-days	GSM608139	(Yu, Li et al. 2011)
GEO	GSE24684	ChIP-chip BES1	BES1	NA	C	14-days	GSM608140	(Yu, Li et al. 2011)
GEO	GSE24684	ChIP-chip BES1	BES1	NA	C	14-days	GSM608141	(Yu, Li et al. 2011)
GEO	GSE24684	ChIP-chip BES1	GFP	NA	C	14-days	GSM608142	(Yu, Li et al. 2011)
GEO	GSE24684	ChIP-chip BES1	GFP	NA	C	14-days	GSM608143	(Yu, Li et al. 2011)
GEO	GSE24684	ChIP-chip BES1	GFP	NA	C	14-days	GSM608144	(Yu, Li et al. 2011)
GEO	GSE23774	ChIP-chip BZR1 (2C)	pBZR1::BZR1-CFP vs. Col0	-	C	28-days	GSM586804	(Sun, Fan et al. 2010)
GEO	GSE43286	ChIP-Seq PIF	PIF1	-	O	3-days	GSM1059957	(Pfeiffer, Shi et al. 2014)
GEO	GSE43286	ChIP-Seq PIF	PIF1	-	O	3-days	GSM1059958	(Pfeiffer, Shi et al. 2014)

GEO	GSE43286	ChIP-Seq PIF	PIF1	-	0	3-days	GSM1059959	(Pfeiffer, Shi et al. 2014)
GEO	GSE43286	ChIP-Seq PIF	WT	-	0	3-days	GSM1059960	(Pfeiffer, Shi et al. 2014)
GEO	GSE43286	ChIP-Seq PIF	WT	-	0	3-days	GSM1059961	(Pfeiffer, Shi et al. 2014)
GEO	GSE43286	ChIP-Seq PIF	WT	-	0	3-days	GSM1059962	(Pfeiffer, Shi et al. 2014)

En el caso que los datos crudos se encontraran disponible en la base de datos GEO (Barrett, Wilhite et al. 2013), se procedió a la descarga de la misma, o a su consulta directa a través de paquetes de R específicos. Cuando los datos crudos no se encontraban disponibles, se procedió a la identificación de los genes diferencialmente expresados en la condición consultada. Este fue el caso de los datos utilizados en el trabajo de Guo y col, 2009. En este caso, se realizó una modificación del formato de la lista de genes, generando un archivo de texto utilizable para consultas y construcción de base de datos.

5.2 ANÁLISIS DE DATOS:

5.2.1 Uso de R:

R es un entorno y lenguaje de programación con un enfoque al análisis estadístico. R es una implementación de software libre del lenguaje S pero con soporte de alcance estático. Se trata de uno de los lenguajes más utilizados en investigación por la comunidad estadística. A esto contribuye la posibilidad de cargar diferentes bibliotecas o paquetes con funcionalidades de cálculo y gráficas. R puede ser utilizado mediante RStudio. Este es un entorno de desarrollo integrado (IDE) para R (lenguaje de programación). Incluye una consola, editor de sintaxis que apoya la ejecución de código, así como herramientas para el trazado, la depuración y la gestión del espacio de trabajo, proporcionando un entorno informático que permite el análisis y desarrollo de análisis de datos con R. Adicionalmente, este trabajo fue escrito en este entorno, mediante la ayuda de "rmarkdown" y "knitr". El primero es un paquete de R que permite insertar código en un documento de texto, y generar un documento final, el cual incluye el resultado del código mediante la implementación de un paquete adicional denominado "knitr". Para el uso de R en el análisis de experimentos de transcriptómica, es necesaria la instalación de paquetes a partir del repositorio Bioconductor, según el siguiente script.

```
#instalación del repositorio bioconductor
Fuente("https://bioconductor.org/biocLite.R")
biocLite()
#instalación de paquetes de bioconductor
biocLite(c("affy", "ath1121501probe", "arrayQualityMetrics", "limma", "ath1121501.db", "GOstats", "ath1121501cdf", "Biobase", "GEOquery", "plyr"))
```

5.3 DESCRIPCIÓN DE LOS PAQUETES DE R:

Bioconductor: Los paquetes de R utilizados se obtuvieron a partir del repositorio Bioconductor. Este es un programa libre y de código abierto, diseñado para el análisis y la comprensión de datos genómicos generados en experimentos de biología molecular. Este programa se basa en el lenguaje R, aunque permite la contribución de lenguajes alternativos. Presenta una gran cantidad de paquetes de anotación genómica predefinidos, lo que facilita el análisis de los datos crudos. Las librerías incluyen flujos de trabajo enfocados en el pre-procesamiento, evaluación de calidad, expresión diferencial, agrupado y clasificación, entre otros análisis importantes a nivel genómico. Adicionalmente, Bioconductor ofrece una interfase extensiva con recursos comunitarios, incluyendo GEO, ArrayExpress, Biomart, genome browsers, GO, KEGG, y otras fuentes de anotación (por ejemplo, TAIR).

5.3.1 Affy

Bioconductor permite el análisis de microarreglos de diversas fuentes comerciales, incluyendo Affymetrix, Illumina, Nimblegen, Agilent, y otras tecnologías de 1 o 2 colores (Bolstad, Collin et al. 2005). El paquete *affy* presenta las herramientas para el análisis de arreglo de oligonucleótidos manufacturados por Affymetrix, permitiendo el acceso y la manipulación de datos de intensidad (Gautier, Cope et al. 2004).

5.3.2 arrayQualityMetrics

El paquete *arrayQualityMetrics* produce un informe de calidad completo (HTML), a partir del llamado de una sola función. Estos reportes permiten evaluar la calidad de cada microarreglo, y facilitar la obtención de diagnósticos, tales como los efectos de lote, y estimar la calidad global del set de datos. Adicionalmente, puede ser utilizado para determinar la calidad de grupo de datos luego de un paso de normalización. La misma función puede ser utilizada para diferentes tipos de microarreglos (1 color, 2 colores, Affymetrix, Illumina) (Kauffmann, Gentleman et al. 2008).

5.3.3 Limma

Limma es un paquete de R para el análisis de datos expresión génica obtenido especialmente a partir de microarreglos. Utiliza modelos lineales para el análisis de los experimentos designados y para la posterior determinación de expresión diferencial. Además, utiliza métodos Bayesianos empíricos para proveer resultados estables, aun cuando se utiliza un pequeño número de réplicas. Ambas técnicas pueden ser utilizadas para todas las tecnologías de microarreglos (Ritchie, Phipson et al. 2015).

5.3.4 GOstats

GOstats es un paquete de Bioconductor escrito en R que permite a los usuarios determinar los términos GO a partir de una lista de genes, utilizando un test hipergeométrico clásico, o un método

hipergeométrico condicional para evaluar la relación entre términos GO (Falcon and Gentleman 2007).

5.3.5 Biobase

Biobase es parte del proyecto Bioconductor, y es utilizado como dependencia de múltiples paquetes. Contiene estructuras de datos organizados y estandarizados para representar datos genómicos. La clase *ExpressionSet* se designa para combinar diferentes fuentes de información en una estructura simple. Esta puede ser manipulada convenientemente como entrada o salida de muchas funciones de Bioconductor. Además, esta clase coordina toda la metadata del grupo de datos, entre las que se encuentra; la descripción de muestras experimentales, anotaciones, características de las plataformas, información relacionada al protocolo y los autores.

5.3.6 GEOquery

GEOquery es una herramienta que permite el acceso directo a GEO a partir de Bioconductor, eliminando la necesidad de dar formato a las entradas de cada análisis. Además, permite la manipulación simple de la metadata requerida para ciertas funciones como las de Biobase (Davis and Meltzer 2007).

5.3.7 ath1121501probe

Secuencia de sondas para microarreglos del tipo ath1121501 de Affymetrix.

5.3.8 ath1121501cdf

Es un archivo .CDF que describe el diseño de un arreglo GeneChip de Affymetrix, incluyendo el nombre de las sondas, y las coordenadas.

5.3.9 ath1121501.db

ath1121501ACCNUM es un objeto de R que provee información del mapeo entre los códigos identificatorios (IDs) del fabricante, y los AGI locus IDs. Adicionalmente presenta un objeto denominado *ath1121501GO*, el cual mapea las relaciones entre los identificadores del fabricante y términos GO directamente asociados con ellos.

5.4 DESCRIPCIÓN DE HERRAMIENTAS ADICIONALES

5.4.1 CisGenome:

CisGenome es un programa para el análisis de datos de inmunoprecipitación de la cromatina (ChIP-data). Permite la visualización, normalización, detección de picos, determinación de FDR, asociación pico-gen, y análisis de secuencias y motivos. Presenta una interfase gráfica y un navegador interno para la visualización de imágenes de arreglos, señales y estructura de genes (Ji, Jiang et al. 2008).

5.4.2 Galaxy:

Galaxy es una plataforma de integración de datos biológicos. Soporta la carga de datos locales, y remotos, incluyendo recursos públicos como UCSC Genome Browser, BioMart o InterMine. También soporta una gran cantidad de formatos de datos, y permite la interconversión de los mismos. Adicionalmente, incluye una gran cantidad de herramientas de manipulación y procesamiento de datos. Por último, permite la ejecución de estas herramientas en la nube, independizando al usuario de las limitaciones de hardware (Afgan, Baker et al. 2016).

5.4.3 Cytoscape:

Cytoscape es una herramienta de código abierto para la visualización de redes de interacción molecular, vías biológicas, y para la integración de las mismas con múltiples datos disponibles públicamente. Permite el manejo de la misma a través de una GUI y la incorporación de aplicaciones adicionales (Apps), las cuales permiten expandir las funcionalidades (Cline, Smoot et al. 2007).

5.4.4 Genemania:

Genemania es una aplicación para la visualización y análisis de relaciones biomoleculares. A partir de una lista de identificadores génicos, se produce la evaluación de los mismos a tres niveles. El primer nivel incluye la consulta y el almacenamiento de información genómica, de interacción proteína-proteína, GO y de vías metabólicas. El segundo nivel genera una red global utilizando el conjunto de los datos. El tercer nivel, permite la interacción de las redes con Cytoscape como una aplicación del mismo. Genemania utiliza información pública de co-expresión obtenidos de GEO, Interacción proteína obtenida de BioGRID y PathwayCommons, interacción genética proveniente de BioGRID, dominios proteicos comunes proveniente de bases de datos como InterPro, SMART y Pfam. También se incluyen información sobre Co-localización y de vías metabólica (Reactome y BioCyc) entre otros parámetros que no encajan con ninguna de estas categorías. Toda esta información es utilizada para generar redes funcionales a partir de una lista de identificadores de genes.

5.5 ANÁLISIS DE ARCHIVOS DE MICROARREGLOS CON BIOCONDUCTOR

5.5.1 Curva BL vs tiempo

Los archivos de intensidad obtenidos (.CEL), fueron analizados en Bioconductor con los siguientes scripts.

En primer lugar, se deben cargar las librerías necesarias

```
library(affy)
library(limma)
library(arrayQualityMetrics)
```

```
library(ath1121501.db)
library(ath1121501probe)
library(ath1121501cdf)
```

Luego se define el directorio de trabajo y se procede a leer los archivos .CEL en el directorio de trabajo por medio del paquete affy.

```
setwd("C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Curva_BL_Tiempo")
dir(pattern="*.cel")
## imprime secuencias y sus posiciones mapeadas para las primeras IDs
print.data.frame(ath1121501probe[1:6,1:4])
## Lee todos los archivos .cel en el directorio de trabajo
mydata <- ReadAffy()
```

Los archivos de intensidades se normalizan con RMA y llevo a cabo el control de calidad de los datos con el paquete arrayQualityMetrics

```
eset_rma <- rma(mydata)
# Imprime las primeras 4 lineas de la estructura de datos
exprs(eset_rma)[1:4,1:12]
# analisis exploratorio con array quality metrics
# El argumento force obliga a reemplazar el directorio si el mismo ya existe
arrayQualityMetrics(eset_rma, force = TRUE)
```

Se lleva a cabo la anotación de las sondas usando el objeto ath1121501ACCNUM

```
Annot <- data.frame(ACCNUM=sapply(contents(ath1121501ACCNUM), paste, collapse=", "),
  SYMBOL=sapply(contents(ath1121501SYMBOL), paste, collapse=", "),
  DESC=sapply(contents(ath1121501GENENAME), paste, collapse=", "))
Annot[3:4,]
```

La comparación de los distintos grupos se lleva a cabo con limma. En primer lugar, se importa información experimental desde el archivo "matriz.txt".

```
# Vuelvo a setear el directorio de trabajo
setwd("C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Curva_BL_Tiempo")
# se lee el archivo "matriz.txt" en donde se relacionan los nombres de archivos con los
# grupos de comparación
targets <- readTargets("matriz.txt")
# Control CEL files with targets files
data <- ReadAffy(filenamees=targets$FileName)
# Normalización con RMA
eset <- rma(data)
```

```

exprs(eset) <- log2(exprs(eset))
#Lista de los archivos analizados
pData(eset)
# Exporta los valores de expresion
write.exprs(eset, file="affy_all.xls")

#Creacion de un diseño de matriz
design <- model.matrix(~ -1+factor(c("G4","G4","G5","G5","G1","G1","G6","G6","G2","G2",
"G3","G3")))
colnames(design) <- c("group1", "group2", "group3", "group4", "group5", "group6") # Assigns c
olumn names.
design[1:12,]
# ajuste de un modelo lineal para cada gen.
fit <- lmFit(eset, design)
# creacion de una matriz de contraste apropiada
contrast.matrix <- makeContrasts(group2-group1, group4-group3, group6-group5, levels=desig
n)
contrast.matrix
#contraste segun la matriz de expresion
fit2 <- contrasts.fit(fit, contrast.matrix)
fit2 <- eBayes(fit2)

```

La función topTable es una función que extrae una tabla de los genes mejor clasificados en un ajuste de modelo lineal, en este caso "fit2". La tabla presenta las siguientes columnas: *logFC*: Logaritmo base 2 del "fold change" entre dos condiciones. Se interpreta como el log2 del nivel de expresión. *AveExpr*: Expresión promedio (logFC) de este gen a lo largo de todo el set de datos; *t*: Estadística t-student moderada; *P.value*: valor p asociado a la estadística t; *adj.P.Value*: es el valor p ajustado para testeos múltiples.

A partir de los datos de comparación, se imprimen los mismos según el coeficiente dado. Luego se reordenan los genes por p-value y finalmente se imprimen en archivos .txt necesario para comparaciones posteriores.

```

#TABLA CON LA COMPARACION 1-2 con nombre de línea
write.table(topTable(fit2, coef=1, adjust="fdr", sort.by="B", number=Inf), file="C:/Users/Mauri
notebook/Desktop/Mauricio_Grisolia/Curva_BL_Tiempo/limma_complete1.xls", row.names=T,
sep="\t")
#TABLA CON LA COMPARACION 3-4 con nombre de línea
write.table(topTable(fit2, coef=2, adjust="fdr", sort.by="B", number=Inf), file="C:/Users/Mauri
notebook/Desktop/Mauricio_Grisolia/Curva_BL_Tiempo/limma_complete2.xls", row.names=T,
sep="\t")
#TABLA CON LA COMPARACION 5-6 con nombre de línea

```

```

write.table(topTable(fit2, coef=3, adjust="fdr", sort.by="B", number=Inf), file="C:/Users/Mauri
notebook/Desktop/Mauricio_Grisolia/Curva_BL_Tiempo/limma_complete3.xls", row.names=T,
sep="\t")

#Muestra una tabla de los genes, ordenados por p-value.
todo1 <- merge(Annot, topTable(fit2, coef=1, adjust="fdr", sort.by="B", number=Inf), by.x=0, b
y.y=0, all=T)
# reestablece el orden correcto por p-value ajustado
todo1 <- todo1[order(todo1$adj.P.Val), ]
#se imprime el archivo en la dirección y con el nombre brindado
write.table(todo1, file="C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Curva_BL_Tiemp
o/30m.txt", sep="\t", col.names = NA)

#se realiza lo mismo para las siguientes 2 comparaciones
todo2 <- merge(Annot, topTable(fit2, coef=2, adjust="fdr", sort.by="B", number=Inf), by.x=0, b
y.y=0, all=T)
todo2 <- todo2[order(todo2$adj.P.Val), ]
write.table(todo2, file="C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Curva_BL_Tiemp
o/1h.txt", sep="\t", col.names = NA)

todo3 <- merge(Annot, topTable(fit2, coef=3, adjust="fdr", sort.by="B", number=Inf), by.x=0, b
y.y=0, all=T)
todo3 <- todo3[order(todo3$adj.P.Val), ]
write.table(todo3, file="C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Curva_BL_Tiemp
o/3h.txt", sep="\t", col.names = NA)

```

Los archivos fueron impresos en otra carpeta cuyo camino ("path") está definido en el argumento "file". Finalmente se procede a construir un gráfico de clústeres para cada comparación

```

#Gráfico de Clúster
my_fct <- function(tT) hclust(tT, method="complete")
heatmap(as.matrix(2^exprs(eset)[1:40,]), col = cm.colors(256), hclustfun=my_fct)

```

Para obtener la lista de identificadores que será utilizada en comparaciones posteriores, realizamos lo siguiente:

```

#a partir del TT, solo me quedo con los transcriptos cuya expresión sea significativa
#para comparación Grupo2-1 30 minutos
aS <- topTable(fit2, coef=1, adjust="fdr", sort.by="B", p.value=0.05, number=Inf)##$ID
affySample <- row.names(aS)

geneSample <- na.omit(as.vector(unlist(mget(affySample, ath1121501ACCNUM, ifnotfound=

```

```

NA))))
#imprime el archivo de la lista de genes para usar el diagrama de Venn
write.table(geneSample, file="C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/30mlista.txt", sep="\t", col.names = F, row.names = F, quote = F)

#para comparación Grupo4-3 1hora
aS <- topTable(fit2, coef=2, adjust="fdr", sort.by="B", p.value=0.05, number=Inf)##$ID
affySample <- row.names(aS)
geneSample <- na.omit(as.vector(unlist(mget(affySample, ath1121501ACCNUM, ifnotfound=NA))))
#imprime el archivo de la lista de genes para usar el diagrama de Venn
write.table(geneSample, file="C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/1hlista.txt", sep="\t", col.names = F, row.names = F, quote = F)

#para comparación Grupo6-5 3horas
aS <- topTable(fit2, coef=3, adjust="fdr", sort.by="B", p.value=0.05, number=Inf)##$ID
affySample <- row.names(aS)
geneSample <- na.omit(as.vector(unlist(mget(affySample, ath1121501ACCNUM, ifnotfound=NA))))
#imprime el archivo de la lista de genes para usar el diagrama de Venn
write.table(geneSample, file="C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/3hlista.txt", sep="\t", col.names=F, row.names=F, quote=F)

```

5.5.2 Col0 "Mock" vs Col0 "BL"

Como primer paso se procede a cargar las librerías necesarias para este análisis:

```

library(Biobase)
library(GEOquery)
library(limma)
library(affy)
library(ath1121501probe)
library(ath1121501.db)
library(ath1121501cdf)

```

A continuación, se procede a cargar el directorio de trabajo en donde se encuentran los datos crudos de este experimento y a conectar directamente Bioconductor con GEO a través de la interfase creada por "GEOquery". Debido a que el ID brindado incluye todas las comparaciones llevadas a cabo en ese trabajo, se seleccionan los sets de datos requeridos mediante un vector apropiado, y una posterior eliminación del set de datos no utilizados.

```

setwd("C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Song")
# cargar la serie de datos y la matriz directamente desde GEO

```

```

gset <- getGEO("GSE14788", GSEMatrix =TRUE)
if (length(gset) > 1) idx <- grep("GPL198", attr(gset, "names")) else idx <- 1
gset <- gset[[idx]]

# Crear los nombres de columna
fvarLabels(gset) <- make.names(fvarLabels(gset))

# Elegir los nombres de los grupos a comprar según las condiciones en las que fue generada cada
a archivo.
sml <- c("X","X","X","X","X","X","G0","G0","G1","G1","X","X","X","X");

# eliminar muestras "X"
sel <- which(sml != "X")
sml <- sml[sel]
gset <- gset[,sel]

# transformar a log2
ex <- exprs(gset)
qx <- as.numeric(quantile(ex, c(0., 0.25, 0.5, 0.75, 0.99, 1.0), na.rm=T))
LogC <- (qx[5] > 100) ||
(qx[6]-qx[1] > 50 && qx[2] > 0) ||
(qx[2] > 0 && qx[2] < 1 && qx[4] > 1 && qx[4] < 2)
if (LogC) { ex[which(ex <= 0)] <- NaN
exprs(gset) <- log2(ex) }

# configuración de los datos y lectura del diseño matriz
fl <- as.factor(sml)
gset$description <- fl
design <- model.matrix(~ description + 0, gset)
colnames(design) <- levels(fl)

```

Seguidamente, se utiliza el paquete "limma" para realizar el análisis de expresión diferencial de la misma manera que en el paso anterior, generando un modelo lineal para cada gen, y un cómputo estadístico para calcular las estadísticas asociadas a la expresión diferencial.

```

fit <- lmFit(gset, design)
cont.matrix <- makeContrasts(G1-G0, levels=design)
fit2 <- contrasts.fit(fit, cont.matrix)
fit2 <- eBayes(fit2, 0.01)
tT <- topTable(fit2, adjust="fdr", sort.by="B", number=Inf)

```

Nuevamente se vincula a GEO para cargar la anotación de genes, se reemplaza la misma en la salida de la comparación de expresión diferencial y por último se restablece el orden correcto por "P.Value"

```
# cargar la anotación de la plataforma desde NCBI
gpl <- annotation(gset)
platf <- getGEO(gpl, AnnotGPL=TRUE)
ncbifd <- data.frame(attr(dataTable(platf, "table"))

# Reemplazo de la anotación original
tT <- tT[setdiff(colnames(tT), setdiff(fvarLabels(gset), "ID"))]
tT <- merge(tT, ncbifd, by="ID")
#restablecer el orden correcto
tT <- tT[order(tT$P.Value), ]
tT <- subset(tT, select=c("Gene.symbol", "ID", "adj.P.Val", "P.Value", "t", "B", "logFC", "Gene.title"))
write.table(tT, file="tT.xls", row.names=F, sep="\t")
```

Se imprime la lista de genes para ser utilizada en pasos posteriores

```
#Filtrar tT con p.Value < 0.05
tT2 <- tT[tT$P.Value < 0.05, "ID"]
write.table(tT2, file="tT2.txt", sep="\t", col.names = F, row.names = F, quote = F)
affySample <- readLines("tT2.txt")
geneSample <- na.omit(as.vector(unlist(mget(affySample, ath1121501ACCNUM, ifnotfound=NA))))
#imprime el archivo de la lista de genes para usar el diagrama de Venn
write.table(geneSample, file="C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/BL+BL-.txt", sep="\t", col.names = F, row.names=F, quote = F)
```

Como último paso, se realiza el gráfico de clústeres

```
#Gráfico de clústeres
my_fct <- function(tT) hclust(tT, method="complete")
heatmap(as.matrix(2^exprs(gset)[1:40,]), col = cm.colors(256), hclustfun=my_fct)
```

5.5.3 Depurado de lista de genes de expresión diferencial

Se obtuvo del trabajo de Guo y colaboradores (Guo, 2009) una lista de genes la cual puede ser utilizadas en la construcción de diagramas de Venn, y para la consulta de los términos GO de cada gen. Como primer paso, cargo las librerías necesarias:

```
#carga las librerías necesarias
library("stringr")
```

```
library(ath1121501.db)
library(GOstats)
library(ath1121501cdf)
```

Luego, se leen los archivos .txt provenientes del material suplementario de (Guo,2009) ST1_PDF.txt: contiene todos los genes significativamente sobre-expresados en plantas control tratadas con BL. ST2_PDF.txt: contiene todos los genes significativamente infra-expresados en plantas control tratadas con BL. ST3_PDF.txt: contiene todos los genes significativamente sobre-expresados en plantas the1 herk1 (mutantes) tratadas con BL. ST4_PDF.txt: contiene todos los genes significativamente infra-expresados en plantas the1 herk1 (mutantes) tratadas con BL.

```
setwd("C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Guo")
#lee la el archivo txt e imprime el encabezado
wtup <- readLines("ST1_PDF.txt")
wtdo <- readLines("ST2_PDF.txt")
mutup <- readLines("ST3_PDF.txt")
mutdo <- readLines("ST4_PDF.txt")

#detecta las líneas que presenta la siguiente expresion regular en all_data
wtup.clean <- grep("AT[012345].....", wtup)
wtdo.clean <- grep("AT[012345].....", wtdo)
mutup.clean <- grep("AT[012345].....", mutup)
mutdo.clean <- grep("AT[012345].....", mutdo)

#crea e imprime el encabezado del vector solo con las líneas que presentan la expresión regular
wtup.l <- wtup[wtup.clean]
wtdo.l <- wtdo[wtdo.clean]
mutup.l <- mutup[mutup.clean]
mutdo.l <- mutdo[mutdo.clean]

#obtengo los primeros 10 caracteres de cada línea
wtup.s <- substr(wtup.l, 1, 10)
wtdo.s <- substr(wtdo.l, 1, 10)
mutup.s <- substr(mutup.l, 1, 10)
mutdo.s <- substr(mutdo.l, 1, 10)

#se eliminan los espacios en blanco
wtup.ids <- str_trim(wtup.s, side = "both")
wtdo.ids <- str_trim(wtdo.s, side = "both")
mutup.ids <- str_trim(mutup.s, side = "both")
mutdo.ids <- str_trim(mutdo.s, side = "both")
```

```
#unir las listas de genes regulados positiva y negativamente
```

```
wt.t <- c(wtup.ids, wtdo.ids)
```

```
mut.t <- c(mutup.ids, mutdo.ids)
```

Una vez generadas las listas de genes, imprimimos las mismas de la siguiente manera.

```
#imprimir archivos de cada condición en carpeta específica para Venn. Son "character strings"
```

```
write.table(wt.t, file = "C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/col0B  
L.txt", quote = F, sep = "\t", row.names = FALSE, col.names = FALSE)
```

```
write.table(mut.t, file = "C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/mut  
BL.txt", quote = F, sep = "\t", row.names = FALSE, col.names = FALSE)
```

5.5.4 Análisis de Microarreglos CATMA de *Arabidopsis thaliana* WS vs the1-

De la misma manera que antes, se inicia cargando las librerías necesarias:

```
library(ath1121501.db)
```

```
library(GOstats)
```

```
library(ath1121501cdf)
```

```
library(limma)
```

```
library(plyr)#librería para realizar el Join
```

Utilizando la función de *limma* "readTargets()", se lee por defecto el archivo "Targets.txt" que se encuentre en el mismo directorio de trabajo. Posteriormente leo los archivos de intensidad del arreglo de dos colores (plataforma GenePix)

```
#definir directorio de trabajo
```

```
setwd("C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Hematy")
```

```
#Leer archivo "Targets.txt"
```

```
targets <- readTargets()
```

```
#leo los archivos de intensidad .gpr pertenecen a genepix
```

```
RG <- read.maimages(targets$FileName, Fuente="genepix")
```

En este caso, realizaremos una evaluación manual de la calidad de los arreglos. Para los arreglos de 2 colores, es un paso esencial ver los gráficos no normalizados de cada grupo de datos. La intensidad de fondo también es útil como guía. Normalmente se realizan gráficos del tipo "box" para cada intensidad de fondo.

```
#Genera gráfico .png
```

```
plotMA3by2(RG)
```

```
#Grafico de intensidad de fondo para cada láser
```

```
boxplot(data.frame(log2(RG$Gb)),main="Green background")
```

Green background

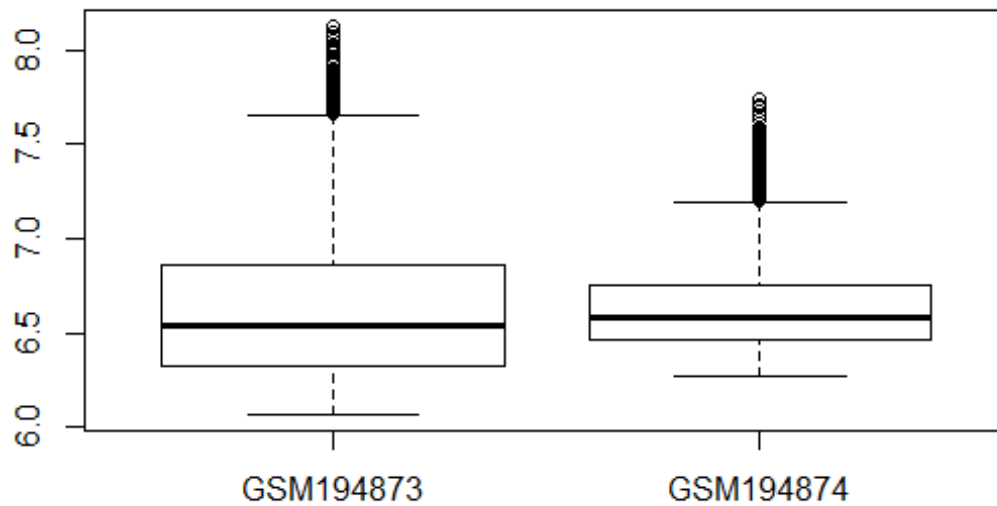


Ilustración 1 Gráfico caja y bigotes mostrando la intensidad de fondo para el canal verde.

```
boxplot(data.frame(log2(RG$Rb)),main="Red background")
```

Red background

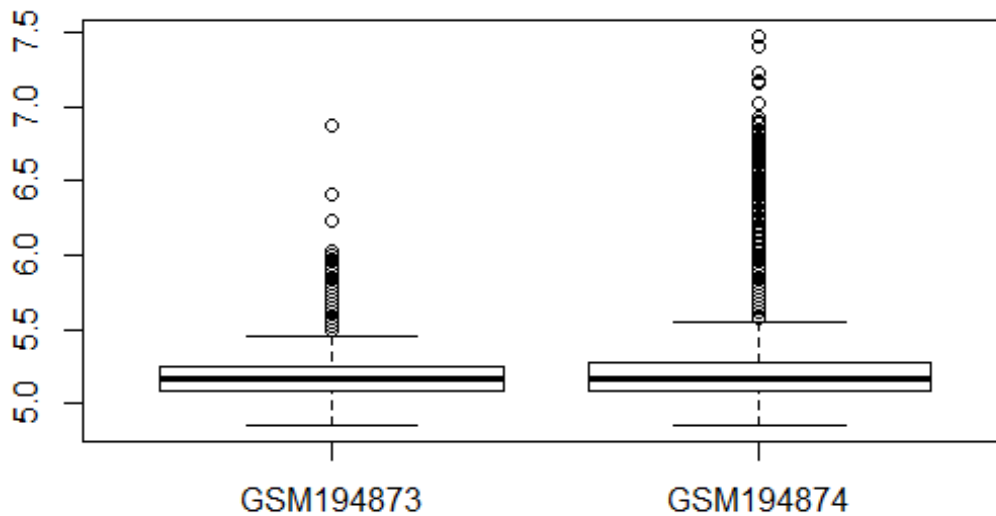


Ilustración 2 Gráfico caja y bigotes mostrando la intensidad de fondo para el canal rojo.

Finalmente, evaluamos la heterogeneidad espacial de los microarreglos individuales, generando una imagen de las intensidades de fondo con respecto al área del microarreglo.

```
#Heterogeneidad espacial  
imageplot(log2(RG$Gb[,1]),RG$printer)
```

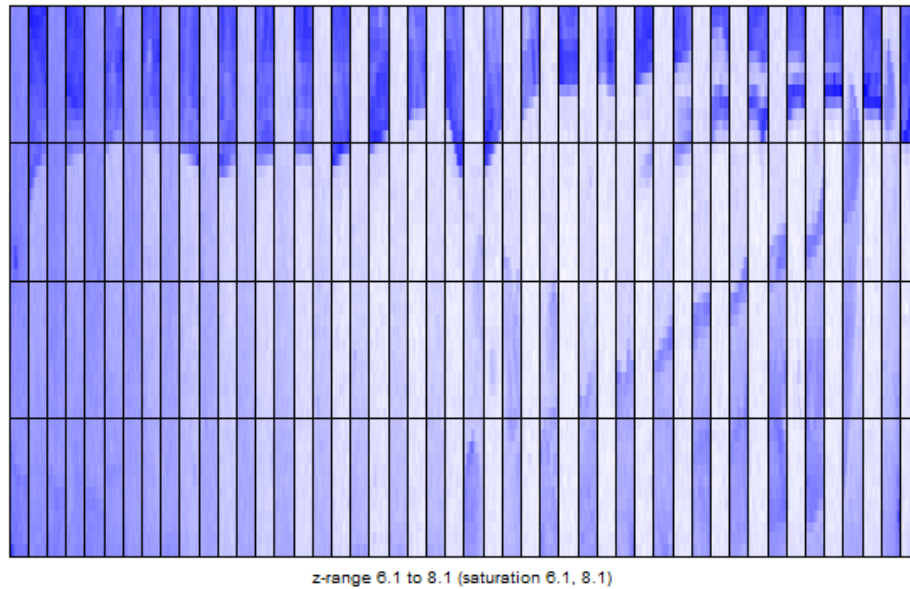


Ilustración 3 Heterogeneidad espacial de la intensidad de fondo para el primer microarreglo

```
imageplot(log2(RG$Gb[,2]),RG$printer)
```

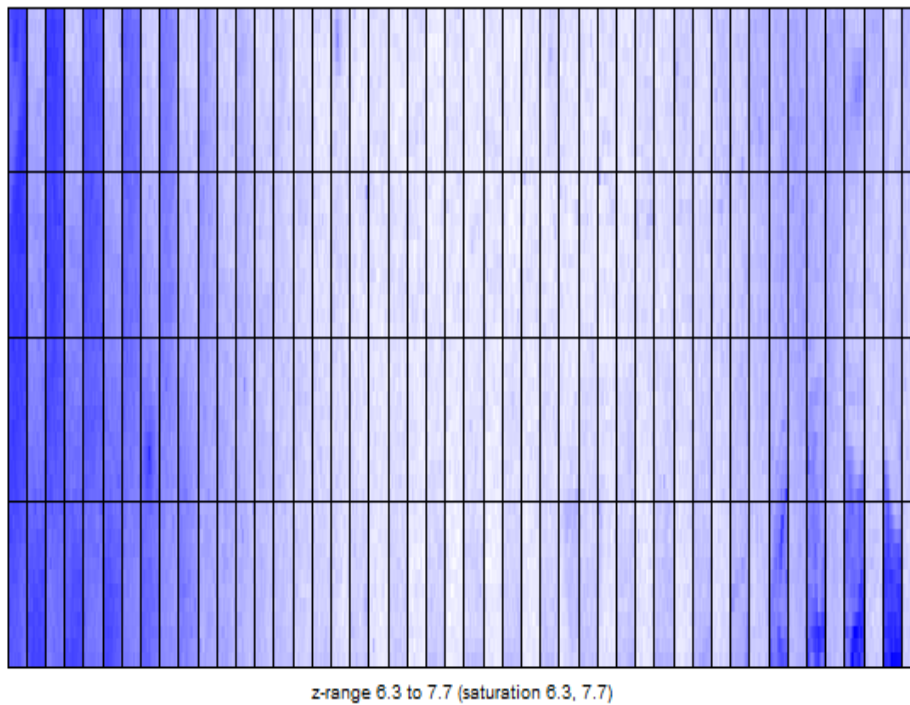


Ilustración 4 Heterogeneidad espacial de la intensidad de fondo para el primer microarreglo

Finalmente se grafican las densidades sin normalización para evaluar las variaciones entre canales.

```
plotDensities(RG)
```

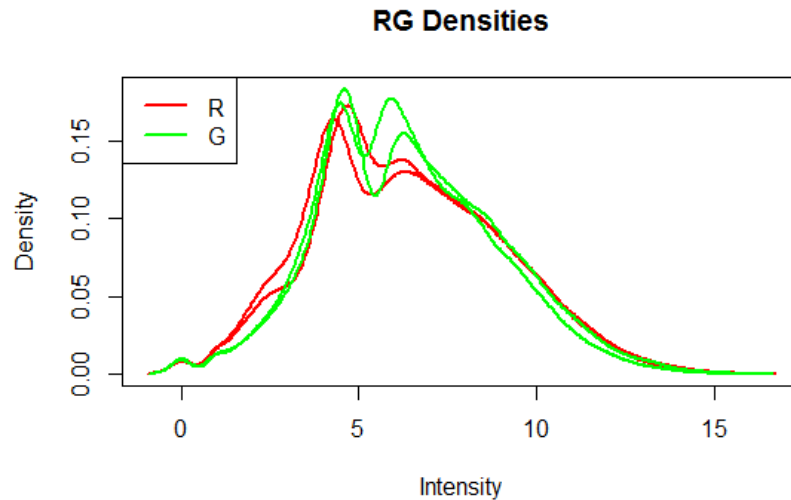


Ilustración 5 Gráfico de densidad de intensidad en los canales rojo y verde, para todo el set de datos sin normalizar

Es necesario realizar una corrección del fondo. Existen varios métodos para llevar a cabo este proceso. En nuestro caso preferimos el método "normexp" ya que es el recomendado para evaluación de expresión diferencial de genes.

```
RG <- backgroundCorrect(RG, method="normexp", offset=50)
```

```
#Graficar densidades corregidas
```

```
plotDensities(RG)
```

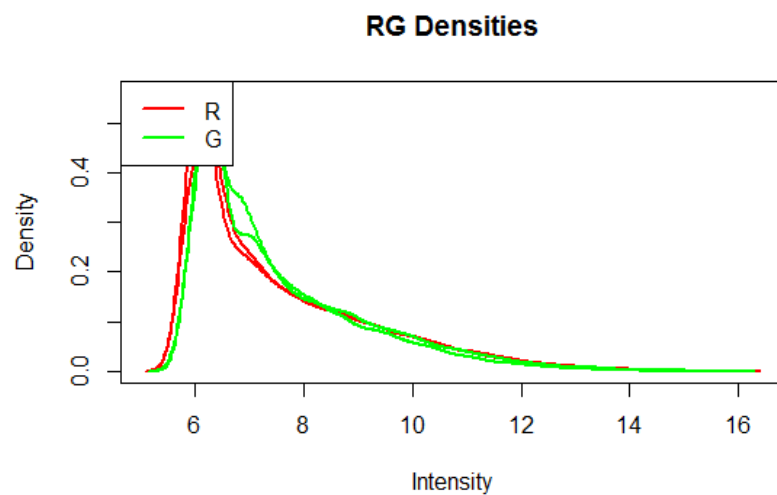


Ilustración 6 Gráfico de densidad de intensidad en los canales rojo y verde, para todo el set de datos, con corrección de fondo

Como último paso de corrección, se lleva a cabo la normalización entre los diferentes arreglos

```
#Normalización entre microarreglos  
MA.p <- normalizeWithinArrays(RG)  
#Aplicar la normalización por cuantiles a la normalización anterior, genera la misma distribución entre todos los arreglos y todos los canales  
MA.pAq <- normalizeBetweenArrays(MA.p, method="Aquantile")  
#Se grafica las densidades  
plotDensities(MA.pAq)
```

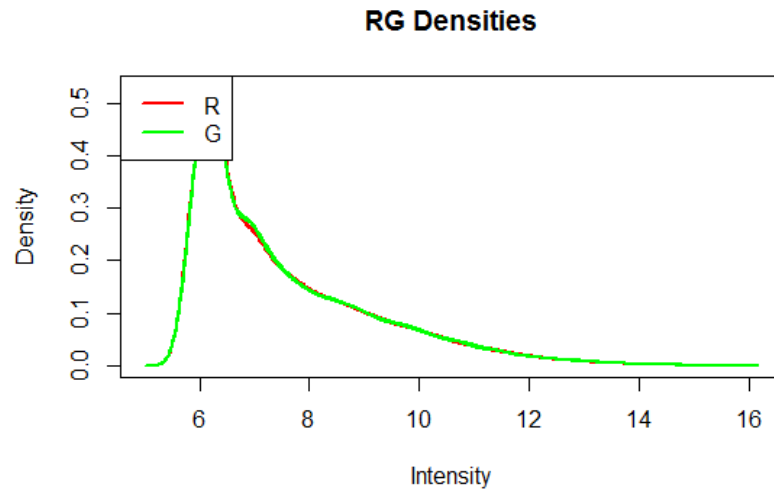


Ilustración 7 Gráfico de densidad de intensidad en los canales rojo y verde, para todo el set de datos con normalización inter data set.

Una vez corregido el grupo de datos, se define el diseño de comparación con referencia a la variable "targets". El diseño de "Dye Swap" requiere que siempre se compare "Control vs mutante". De esa manera, la variable diseño será un vector con los caracteres c(1,-1). *Limma* luego procede a generar los modelos lineales y a evaluar la expresión diferencial.

```
design <- c(1,-1)  
fit <- lmFit(MA.pAq, design)  
fit <- eBayes(fit)  
#valor Inf del parámetro "number", muestra todos los valores con p<0.05. En caso contrario muestra 10  
tt <- topTable(fit, sort.by = "p", p.value = 0.05, number = Inf)  
#el data frame tt, presenta los ID de genes (Coordenadas CATMA) con algunas letras en minúscula. Para los próximos pasos es necesario que las mismas estén todas en mayúscula  
tt$ID <- toupper(tt$ID)  
#se escribe el archivo "toptable.txt" dentro de WD  
write.table(tt, file = "toptable.txt", quote = F, sep = "\t", row.names = F)
```

Para asociar los ID CATMA con los ID de los genes, es necesario utilizar un archivo en donde se indican la relación entre ambas variables. El mismo se encuentra en el directorio de trabajo y se denomina "CATMA_2.2_07122011.txt" y es suministrado por el grupo que construyó el arreglo.

```
##definir directorio de trabajo
setwd("C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Hematy")
#Leer archivo
ids <- read.delim("CATMA_2.2_07122011.txt", header = F, sep = "\t", dec = ";")
#renombrar las columnas
colnames(ids) <- c("ID", "PROBE_TYPE", "GENE_ID", "GENE_TYPE", "DESCRIPTION")
#modificar la columna ids$ID a la clase y con el formato apropiado
ids$ID <- as.character(ids$ID)
ids$ID <- toupper(ids$ID)
#unir ambas tablas como "left join", ya que solo requiero los ID de genes identificados en tt
tids <- join(tt, ids, type="left", match="all")
# Filtro las filas con NA
row.has.na <- apply(tids, 1, function(x){any(is.na(x))})
tids.wna <- tids[!row.has.na,]
#Escribo archivo output
write.table(tids.wna, file = "tids.txt", quote = F, sep = "\t")
```

Finalmente debo generar una lista de genes para ser utilizados en evaluaciones posteriores

```
#quedarse con la columna de GENE_ID
geneSample <- tids.wna$GENE_ID

# Me quedo solo con los que empiezan con AT
geneSample <- grep("^AT*", geneSample, value = TRUE)

# Declaro variables auxiliares
geneSample.usar <- list()
geneSample.iterar <- unlist(geneSample)

# Me fijo en cada valor de geneSample, para ver cuales tienen ";" ya que esto afecta los procesos
posteriores
for(value in geneSample.iterar)
{
  if(grepl(";",value)) # SI tiene ;
  {
    # Lo separamos y solo guardamos el primer ID
    valor.auxiliar = unlist(strsplit(value, ";"))
    geneSample.usar <- c(geneSample.usar, valor.auxiliar[1])
  }
}
```

```

} else { # Si no tiene ; lo guardamos
geneSample.usar <- c(geneSample.usar, value)
}

}

#Reasignamos la variable geneSample
geneSample <- unlist(geneSample.usar)
#Escribimos el archivo lista de ID para Venn
write.table(geneSample, file = "C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/wsthe1.txt", quote = F, sep = "\t", row.names = FALSE, col.names = FALSE)

```

5.5.5 Análisis ChIP-chip con CisGenome

5.5.5.1 *BES1*

Para el análisis de los ficheros descargados se utilizó el programa CisGenome v2.0 (Ji and Wong 2005, Ji, Vokes et al. 2006, Ji, Jiang et al. 2008). Este programa fue seleccionado para el análisis, debido a que los paquetes de R analizados ("Ringo", "Starr") (Toedling, Sklyar et al. 2007, Zacher, Kuan et al. 2010) dependían de archivos no accesibles. Antes de analizar los datos, es necesario crear el set de datos, a partir de los archivos de Affymetrix. Los archivos .CEL guardan información de intensidad y calidad de los pixeles. Los archivos .bpmmap, son archivos binarios que contienen información del diseño de los arrays. Los mismos incluyen la posición y secuencia de cada sonda dentro del microarreglo, inclusive la posición de la sonda "mismatch". Como primer paso, se deben importar los archivos .CEL y .bpmmap mediante "File > Load Data > Tiling Array Dataset > Import from Affymetrix CEL+BPMAP". Es necesario que cada archivo .CEL tenga asociado un archivo .bpmmap. Los archivos CEL deben estar asociados a un tipo de muestra. En este caso, fueron creados dos grupos de tres muestras. El grupo "1" se nombró "IP", e incluyó los ficheros "GSM608139.CEL", "GSM608140.CEL", "GSM608141.CEL". El grupo "2" se nombró "CONTROL", e incluyó los ficheros "GSM608142.CEL", "GSM608143.CEL", "GSM608144.CEL". El próximo paso consistió en la normalización de los datos. Para esto, solo es necesario hacer click en el menú "Tiling Array > Normalization > Quantile (CEL+BPMAP)". A continuación, se realiza la detección de picos, mediante el menú "Tiling Array > Peak Detection (TileMap)". Es importante indicar al programa el tipo de comparación que se desea hacer. En este caso, 1>2, "IP" > "CONTROL". Esto genera un archivo de texto .COD, el cual presenta la lista de sondas que presentan interacción en el grupo de "IP", pero no en el grupo "CONTROL". Adicionalmente se presenta el parámetro FDR para evaluar la significancia de cada sonda. Luego, se procede a la anotación de genes. Para esto es necesario descargar el genoma de *Arabidopsis thaliana*. Existen varias versiones disponibles para descargar en la página de Cisgenome. En este caso utilizamos la versión del genoma TAIR 9 ya que es la versión más actual dentro de las incluidas en la página. Una vez descargada la base de datos genómica, se procede a cargar la misma en el programa mediante el menú "File > Load Data > Genome Database", y se selecciona el archivo .CGW. Finalmente, se procede a anotar las regiones genómicas con los genes vecinos. Para

esto, se seleccionaron solo las coordenadas con un FDR < 0.05 y se accede al menú correspondiente, de la siguiente manera "Genome > Annotate with > Closest Gene". Se define la máxima distancia de la región al gen y se determina el método de determinación de distancia. En este caso se utilizaron los parámetros por defecto. La salida de este procedimiento es una lista de genes, indicando el ID, las coordenadas genómicas, la hebra codificante y el cromosoma. Por último, se procede a la selección de la lista de genes para ser utilizada en la sección de diagrama de Venn y análisis de GO.

```
#Defino directorio de trabajo  
setwd("C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/BES1")  
#defino variable cod a partir de los datos obtenidos en CisGenome  
cod <- read.table("AnotacionFDR")  
atg <- cod[,6, drop=FALSE]  
write.table(atg, file = "C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/bes1.txt", quote = FALSE, row.names = FALSE, col.names = FALSE)
```

5.5.5.2 BZR1

De esta manera, se utilizó la misma metodología que en la sección anterior, y la salida del programa CisGenome, fue utilizada para obtener la lista de genes regulados por BZR1, para ser utilizados en análisis posterior.

```
setwd("C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/BZR1")  
#defino variable cod a partir de los datos obtenidos en CisGenome  
cod <- read.table("AnotacionFDR")  
atg <- cod[,6, drop=FALSE]  
write.table(atg, file = "C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/BZR1.txt", quote = FALSE, row.names = FALSE, col.names = FALSE)
```

5.5.6 Análisis ChIP-Seq

El análisis en Galaxy fue llevado a cabo mediante el siguiente flujo de trabajo: "ChIP-seq analysis (create BAM, bigWig and PeakCalling)".

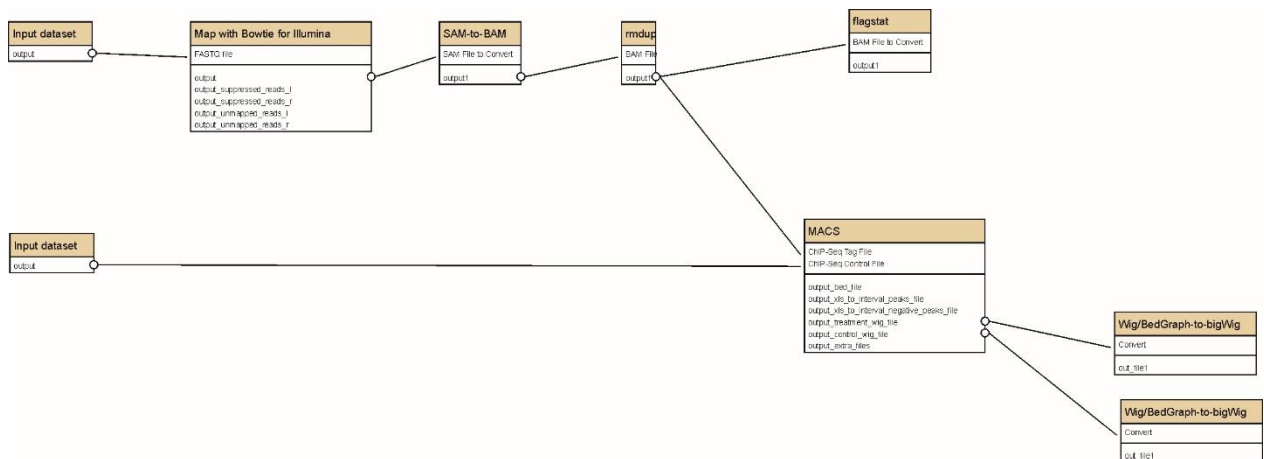


Ilustración 8 Esquema del Flujo de trabajo utilizado en galaxy.org

Bowtie: Alineador de secuencias cortas a una secuencia de referencia, generando un archivo SAM de coordenadas.

SAM to BAM: Conversión del archivo de coordenadas SAM a un archivo binario BAM.

rmDup: Remueve duplicados de PCR. Si la herramienta encuentra dos lecturas “PE” que alinean en la misma posición, se suponen que son duplicados por PCR de los pasos de enriquecimiento y se eliminan del data set.

MACS: permite hacer el "peak calling", generando un archivo de coordenadas “.BED”.

El archivo. BED, fue descargado y utilizado en CisGenome para convertirlo a .COD siguiendo el siguiente menú: "File > File format conversion > BED > COD". Una vez obtenido el archivo .COD, se utilizó el genoma TAIR9 para hacer la anotación de genes. El archivo generado fue formateado de la siguiente manera para consultas posterior.

```
setwd("C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/PIF1")
#defino variable cod a partir de los datos obtenidos en CisGenome
cod <- read.table("AnotatePIF")
atg <- cod[,6, drop=FALSE]
write.table(atg, file = "C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/PIF1.txt", quote = FALSE, row.names = FALSE, col.names = FALSE)
```

5.6 GRÁFICOS DE VENN CON EL PAQUETE LIMMA

Para realizar estas comparaciones en R, es necesario utilizar el paquete limma:

```
library(limma)
```

5.6.1 Venn: comparación plantas tratadas y no tratadas con brasionesteroides

```
#Definir directorio de trabajo
setwd("C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn")
#La función read.table lee los archivos indicados como un data frame
set1 <- read.table('3hlista.txt')
set2 <- read.table('BL+BL-.txt')
set3 <- read.table('col0BL.txt')

#quedarse con los miembros únicos de los sets (en el caso que existan ID repetidos)
set1.u <- unique(unlist(set1))
set2.u <- unique(unlist(set2))
set3.u <- unique(unlist(set3))

#"union" une todos los datos (de a dos) generando una variable nueva que contenga todos los IDs. La función Sort ordena los IDs por orden creciente.
universe <- sort(union(set1.u, union(set2.u, set3.u)) )

#se genera estructura de datos llamados "Counts" para el uso en la función VennCounts
Counts <- matrix(0, nrow=length(universe), ncol=3)
#se nombra la columna perteneciente a cada grupo de datos
colnames(Counts) <- c("Goda 3h", "Song", "Guo")
#se cuenta cada uno de los integrantes de "universe" las veces que aparece en cada set de datos únicos
for (i in 1:length(universe))
{
  Counts[i,1] <- universe[i] %in% set1.u
  Counts[i,2] <- universe[i] %in% set2.u
  Counts[i,3] <- universe[i] %in% set3.u
}

#Ahora se puede usar la función "vennDiagram" y "vennCounts" para generar el diagrama:
vennDiagram( vennCounts(Counts) )
```

Para realizar la intersección de los sets de datos

```
#intersección de los data sets
int.set <- sort(intersect(set1.u, intersect(set2.u, set3.u)) )
write.table(int.set, file = "C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/consensoBL.txt", quote = FALSE, row.names = FALSE, col.names = FALSE)
```

5.6.2 Venn: comparación Goda 1h, Goda 30m y "consensoBL"

```
#Definir directorio de trabajo
setwd("C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn")
#AHORA HAY QUE HACER UNA INTERSECCIÓN ENTRE ESTE GRUPO DE GENES Y LOS DE 1 H
set4 <- read.table('1hlista.txt')
set5 <- read.table('30mlista.txt')
#quedarse con los miembros únicos de los sets (en el caso que existan ID repetidos)
set4.u <- unique(unlist(set4))
set5.u <- unique(unlist(set5))

#esta función genera y ordena los ID de los tres data set. La función Sort ordena los IDs por ord en creciente.
universe <- sort(union(int.set, union(set5.u, set5.u)) )
#se genera estructura de datos llamados "Counts" para el uso en la función VennCounts

Counts <- matrix(0, nrow=length(universe), ncol=3)
#se nombra la columna perteneciente a cada grupo de datos
colnames(Counts) <- c("Goda 1h", "Goda 30m", "ConsensoBL")
#se cuenta cada uno de los integrantes de "universe" las veces que aparece en cada set de datos únicos
for (i in 1:length(universe))
{
  Counts[i,1] <- universe[i] %in% set4.u
  Counts[i,2] <- universe[i] %in% set5.u
  Counts[i,3] <- universe[i] %in% int.set
}
#Ahora se puede usar la función "vennDiagram" y "vennCounts" para generar el diagrama:
vennDiagram( vennCounts(Counts) )
```

5.6.3 Venn: comparación genes consenso a tiempos menores

La lista de estos genes se obtiene de la siguiente manera:

```
#intersección de los data sets tiempo final-tiempo inicial
int.set.total <- sort(intersect(int.set, intersect(set4.u, set5.u)) )
write.table(int.set.total, file = "C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/int_tiempo_total.txt", quote = FALSE, row.names = FALSE, col.names = FALSE)
#intersección tiempo final - 1 h
int.set.medio <- sort(intersect(int.set, set4.u) )
```

```
write.table(int.set.medio, file ="C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/int_tiempo_medio.txt", quote = FALSE, row.names = FALSE, col.names = FALSE)
```

5.6.4 Venn: genes regulados por BR mediados y no mediados por los reguladores transcripcional BZR1 BES1 y PIF.

De la misma manera que en los puntos anteriores, se define la librería y el directorio de trabajo.

```
#se necesita la librería Limma  
library(limma)
```

Finalmente se realiza el gráfico de Venn apropiado

```
#Definir directorio de trabajo  
setwd("C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn")  
#leer tablas  
set1 <- read.table('bes1.txt')  
set2 <- read.table('BZR1.txt')  
set3 <- read.table('PIF1.txt')  
set4 <- read.table('consensoBL.txt')  
  
#quedarse con los miembros únicos de los sets (en el caso que existan ID repetidos)  
set1.u <- unique(unlist(set1))  
set2.u <- unique(unlist(set2))  
set3.u <- unique(unlist(set3))  
set4.u <- unique(unlist(set4))  
#la intersección de los archivos me generará un grupo de genes consenso:  
#esta función genera y ordena los ID de los tres data set. La función Sort ordena los IDs por ord  
en creciente.  
universe <- sort(union(set1.u, union(set2.u, union(set3.u, set4.u))))  
#se genera estructura de datos llamados "Counts" para el uso en la función VennCounts  
  
Counts <- matrix(0, nrow=length(universe), ncol=4)  
#se nombra la columna perteneciente a cada grupo de datos  
colnames(Counts) <- c("BES1", "BZR1", "PIF1", "CONS")  
#se cuenta cada uno de los integrantes de "universe" las veces que aparece en cada set de datos  
únicos  
for (i in 1:length(universe))  
{  
  Counts[i,1] <- universe[i] %in% set1.u  
  Counts[i,2] <- universe[i] %in% set2.u  
  Counts[i,3] <- universe[i] %in% set3.u  
  Counts[i,4] <- universe[i] %in% set4.u
```

```
}
```

#Ahora se puede usar la función "vennDiagram" y "vennCounts" para generar el diagrama:
vennDiagram(vennCounts(Counts))

Para identificar estos genes, debemos realizar lo siguiente:

```
#Genes consenso regulados por PIF1  
int.set <- sort(intersect(set3.u, set4.u) )  
write.table(int.set, file = "C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/PIF-  
CONS.txt", quote = FALSE, row.names = FALSE, col.names = FALSE)  
#Genes consenso regulados por BZR1  
int.set1 <- sort(intersect(set2.u, set4.u) )  
write.table(int.set, file = "C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/BZ  
R1-CONS.txt", quote = FALSE, row.names = FALSE, col.names = FALSE)  
#Genes consenso regulados por BES1  
int.set2 <- sort(intersect(set1.u, set4.u) )  
write.table(int.set, file = "BES1-CONS.txt", quote = FALSE, row.names = FALSE, col.names =  
FALSE)  
  
#Genes Consenso sin regulador asignado  
#Primero se define el grupo de genes con regulador  
genes.reg <- sort(union(int.set2, union(int.set, int.set1)))  
#setdiff  
genes.sin.reg <- sort(setdiff(set4.u, genes.reg))  
write.table(genes.sin.reg, file = "C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Ve  
nn/SR-CONS.txt", quote = FALSE, row.names = FALSE, col.names = FALSE)
```

5.6.5 Venn: vía THE1/HERK

```
#se necesita la librería Limma  
library(limma)  
#Definir directorio de trabajo  
setwd("C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn")  
  
set1 <- read.table('Col0BL.txt')  
set2 <- read.table('mutBL.txt')  
set3 <- read.table('wsthe1.txt')  
  
#quedarse con los miembros únicos de los sets (en el caso que existan ID repetidos)  
set1.u <- unique(unlist(set1))  
set2.u <- unique(unlist(set2))
```

```

set3.u <- unique(unlist(set3))

#Debo sustraer Col0BL-MutBL: Resta para ver los genes diferencialmente expresados en la mut
ante the1herk1 tratados con BL
genes.mut <- sort(setdiff(set1.u, set2.u))
#Ws VS the1-3: muestra los genes que se modifican en ausencia del receptor THE1- Estos genes
se modifican debido a la ausencia de the1
#La intersección entre la resta anterior y WS vs the1.3 daría un consenso de genes regulados por
THE1.
universe <- sort(union(genes.mut, set3.u))
#se genera estructura de datos llamados "Counts" para el uso en la función VennCounts
Counts <- matrix(0, nrow=length(universe), ncol=2)
#se nombra la columna perteneciente a cada grupo de datos
colnames(Counts) <- c("resp_the1_BL", "afect_the1-")
#se cuenta cada uno de los integrantes de "universe" las veces que aparece en cada set de datos
únicos
for (i in 1:length(universe))
{
  Counts[i,1] <- universe[i] %in% genes.mut
  Counts[i,2] <- universe[i] %in% set3.u
}
#Ahora se puede usar la función "vennDiagram" y "vennCounts" para generar el diagrama:
vennDiagram( vennCounts(Counts) )

```

Para identificar estos genes, realizamos lo siguiente:

```

int.set <- sort(intersect(set3.u, genes.mut) )
write.table(int.set, file = "C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/the1
_reg.txt", quote = FALSE, row.names = FALSE, col.names = FALSE)

```

5.6.6 Venn: THE1/HERK1 / consenso BL / FT:

```

#Definir directorio de trabajo
setwd("C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn")

set4 <- read.table('consensoBL.txt')
set5 <- read.table('SR-CONS.txt')
#únicos
set4.u <- unique(unlist(set4))
set5.u <- unique(unlist(set5))

universe <- sort(union(int.set, union(set5.u, set4.u)))

```

#se genera estructura de datos llamados "Counts" para el uso en la función VennCounts

```
Counts <- matrix(0, nrow=length(universe), ncol=3)
```

#se nombra la columna perteneciente a cada grupo de datos

```
colnames(Counts) <- c("the1", "SR", "CONSBL")
```

#se cuenta cada uno de los integrantes de "universe" las veces que aparece en cada set de datos únicos

```
for (i in 1:length(universe))
```

```
{
```

```
Counts[i,1] <- universe[i] %in% int.set
```

```
Counts[i,2] <- universe[i] %in% set5.u
```

```
Counts[i,3] <- universe[i] %in% set4.u
```

```
}
```

#Ahora se puede usar la función "vennDiagram" y "vennCounts" para generar el diagrama:

```
vennDiagram( vennCounts(Counts) )
```

Para obtener la lista de genes, realizamos los siguientes pasos:

```
int.set1 <- sort(intersect(int.set, intersect(set5.u, set4.u) ))
```

```
write.table(int.set1, file ="C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/the1-SR-CONS.txt", quote = FALSE, row.names = FALSE, col.names = FALSE)
```

5.6.7 Venn: vía THE1/HERK1 vs genes de respuesta temprana y tardía.

#se necesita la librería Limma

```
library(limma)
```

#Definir directorio de trabajo

```
setwd("C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn")
```

```
set1 <- read.table('the1_reg.txt')
```

```
set2 <- read.table('int_tiempo_total.txt')
```

```
set3 <- read.table('int_tiempo_medio.txt')
```

```
set4 <- read.table('consensoBL.txt')
```

#quedarse con los miembros únicos de los sets (en el caso que existan ID repetidos)

```
set1.u <- unique(unlist(set1))
```

```
set2.u <- unique(unlist(set2))
```

```
set3.u <- unique(unlist(set3))
```

```
set4.u <- unique(unlist(set4))
```

```
universe <- sort(union(set1.u, union(set2.u, union(set3.u, set4.u))))
```

#se genera estructura de datos llamados "Counts" para el uso en la función VennCounts

```

Counts <- matrix(0, nrow=length(universe), ncol=4)
#se nombra la columna perteneciente a cada grupo de datos
colnames(Counts) <- c("the1", "30m", "1h", "consenso")
#se cuenta cada uno de los integrantes de "universe" las veces que aparece en cada set de datos
únicos
for (i in 1:length(universe))
{
Counts[i,1] <- universe[i] %in% set1.u
Counts[i,2] <- universe[i] %in% set2.u
Counts[i,3] <- universe[i] %in% set3.u
Counts[i,4] <- universe[i] %in% set4.u
}

#Ahora se puede usar la función "vennDiagram" y "vennCounts" para generar el diagrama:
vennDiagram( vennCounts(Counts) )

```

```

int.set <- sort(intersect(set4.u, set1.u) )
write.table(int.set, file ="C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/the1
_CONS.txt", quote = FALSE, row.names = FALSE, col.names = FALSE)

#1 de estos genes se comparte entre the1 y 1 h
int.set <- sort(intersect(set3.u, set1.u) )
write.table(int.set, file ="C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/Salida/Venn/the1
_1h.txt", quote = FALSE, row.names = FALSE, col.names = FALSE)

```

5.7 CONSULTA DE TÉRMINOS GO

5.7.1 Consulta GO: lista de genes BL CONSENSO

Para realizar la consulta:

```

#GOstats permite consultar los términos GO a partir de una lista de genes
library(ath1121501.db)
library(GOstats)
library(ath1121501cdf)

#Definir directorio de trabajo
setwd("C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/GO")

#Es necesario incluir el código para hacer consultas GO de cada uno

```

```

bl <- readLines("consensoBL.txt")
blsr <- readLines("SR-CONS.txt")
the1 <- readLines("the1_reg.txt")

#consultar los términos GO para una lista de genes BL CONSENSO
affyUniverse <- ls(ath1121501cdf)
geneUniverse <- as.vector(unlist(mget(affyUniverse, ath1121501ACCNUM, ifnotfound=NA)))
params <- new("GOHyperGParams", geneIds=bl, universeGeneIds=geneUniverse,
  annotation="ath1121501", ontology="MF", pvalueCutoff=0.05, conditional=FALSE,
  testDirection = "over")
hgOver <- hyperGTest(params)
tabla1 <- summary(hgOver)
write.table(tabla1, file = "C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/GO/consensoGO
.txt", quote = FALSE, row.names = FALSE, col.names = FALSE, sep = "\t")

```

5.7.2 Consulta GO: lista de genes BL CONSENSO sin regulador

```

affyUniverse <- ls(ath1121501cdf)
geneUniverse <- as.vector(unlist(mget(affyUniverse, ath1121501ACCNUM, ifnotfound=NA)))
params <- new("GOHyperGParams", geneIds=blsr, universeGeneIds=geneUniverse,
  annotation="ath1121501", ontology="MF", pvalueCutoff=0.05, conditional=FALSE,
  testDirection = "over")
hgOver <- hyperGTest(params)
tabla2 <- summary(hgOver)
write.table(tabla2, file = "C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/GO/SRGO.txt",
  quote = FALSE, row.names = FALSE, col.names = FALSE, sep = "\t")

```

5.7.3 Consulta GO: lista de genes THE1 CONSENSO

```

affyUniverse <- ls(ath1121501cdf)
geneUniverse <- as.vector(unlist(mget(affyUniverse, ath1121501ACCNUM, ifnotfound=NA)))
params <- new("GOHyperGParams", geneIds=the1, universeGeneIds=geneUniverse,
  annotation="ath1121501", ontology="MF", pvalueCutoff=0.05, conditional=FALSE,
  testDirection = "over")
hgOver <- hyperGTest(params)
tabla3 <- summary(hgOver)
write.table(tabla3, file = "C:/Users/Mauri notebook/Desktop/Mauricio_Grisolia/GO/THE1GO.txt
", quote = FALSE, row.names = FALSE, col.names = FALSE, sep = "\t")

```

6 RESULTADOS Y DISCUSIÓN:

6.1 SELECCIÓN DE LOS DATA SET Y OBTENCIÓN DE DATOS CRUDOS

El conjunto de datos necesario para este trabajo, se encuentra publicado en los repositorios de datos, tal como GEO (*Gene expression omnibus*, <http://www.ncbi.nlm.nih.gov/geo/>). Existe una diversidad importante en los datos de transcriptómica recopilados ya que los ensayos han sido llevados a cabo en diversas condiciones experimentales. Los datos obtenidos deben organizarse teniendo en cuenta las condiciones de crecimiento, ya sea oscuridad o luz continua, o en ciclos de día largo (16h luz/8h oscuridad), las dosis de BR utilizadas, el “entorno” genético (Col0, WS) y los días de crecimiento (Tabla 1). La estandarización de estas condiciones es fundamental para realizar una correcta comparación experimental.

6.2 DETERMINACIÓN DEL GRUPO DE GENES DE RESPUESTA CONSENSO A BRASINOESTEROIDES.

6.2.1 Selección de experimentos que muestren la respuesta de plantas silvestres al tratamiento con Brasinoesteroides:

GSE14788: Comparación de Microarreglos de 1 canal, de plantas *A. thaliana* Col0 tratadas y no tratadas (Mock) con brasinolido 200 nM (Song, Zhou et al. 2009). Las plantas fueron crecidas por 7 días en ausencia total de luz, debido a que en esas condiciones el crecimiento del hipocótilo se basa exclusivamente en la expansión celular. Plataforma: Affymetrix.

TAIR 1007966053: Experimento a gran escala que compara la respuesta a distintas hormonas en tres puntos del tiempo: 30 minutos, 1 hora y 3 horas post tratamiento, mediante microarreglos de 1 canal de Affymetrix (Goda, Sasaki et al. 2008). Los mismos fueron conducidos en plantas de entorno Col0 crecidas en luz continua a lo largo de 7 días post germinación.

Por último, contamos con la lista de genes diferencialmente expresados en un experimento de microarreglo de 1 canal de Affymetrix (Guo, Li et al. 2009) de plantas Col0 adultas (20 días post germinación) crecidas en presencia de 1 μ M de BL en un ciclo de luz/oscuridad 15h/9h. Lamentablemente los datos crudos de este experimento no se encuentran en la base de datos GEO.

El objetivo es comparar la respuesta de plantas crecidas en distintas condiciones, por diferentes períodos de tiempo y tratadas bajo concentraciones distintas de Brasinoesteroides, para poder encontrar una respuesta en común.

6.2.2 Procesamiento de datos y Análisis expresión diferencial por microarreglo: Curva BL vs tiempo

El trabajo de Goda y col. (2008) implicó el análisis a gran escala de líneas de *Arabidopsis thaliana* Col0 frente a la acción de una variedad de hormonas, entre las que se encuentran Auxinas,

Giberelinas, etileno y Brasinoesteroides, a 3 puntos diferentes en el tiempo (30 min, 1 hora y 3 horas). En el caso de BR las plántulas presentan un tiempo post germinación de 7 días, y las mismas fueron tratadas con Brasinolido BL 10 nM y un tratamiento control ("mock").

El análisis inicial constó en determinar tres grupos de datos a partir de las réplicas biológicas de cada condición (Tabla 2).

Tabla 2 Grupos de comparación de datos de microarreglos para la curva BL-Tiempo de respuesta.

Archivo	Grupo	Descripción
RIKENGODA1AC.cel	1	Grupo 1 Mock 30 m
RIKENGODA1BC.cel	1	Grupo 1 Mock 30 m
RIKENGODA8A.cel	2	Grupo 2 BL 30 m
RIKENGODA8B.cel	2	Grupo 2 BL 30 m
RIKENGODA9AC.cel	3	Grupo 3 Mock 1 h
RIKENGODA9BC.cel	3	Grupo 3 Mock 1 h
RIKENGODA16A.cel	4	Grupo 3 bl 1 h
RIKENGODA16B.cel	4	Grupo 3 bl 1 h
RIKENGODA17AC.cel	5	Grupo 5 Mock 3 h
RIKENGODA17BC.cel	5	Grupo 5 Mock 3 h
RIKENGODA24A.cel	6	Grupo 6 bl 3 h
RIKENGODA24B.cel	6	Grupo 6 bl 3 h

Los niveles de expresión fueron determinados por la plataforma de microarreglo GeneChip de Affymetrix (ver sección 5.5.1) para cada comparación (1-2; 3-4;5-6). El análisis de expresión diferencial fue llevado a cabo con el software Limma. A partir de la Tabla 2, se diseñó la matriz de comparación la cual se aplicó para el posterior análisis de comparación. Los archivos de intensidad fueron normalizados, y se utilizó un modelo lineal para cada gen, el cual fue utilizado para ajustar según la matriz de comparación. Posteriormente se utilizó la función *topTable* para extraer una tabla de los genes mejor clasificados con el ajuste utilizado. La tabla consta de las siguientes columnas

logFC: Logaritmo base 2 del "fold change" entre dos condiciones. Se interpreta como el log₂ del nivel de expresión.

AveExpr: Expresión promedio (logFC) de este gen a lo largo de todo el set de datos.

t: Estadística t-student moderada

P.value: valor p asociado a la estadística t

adj.P.Value: es el valor p ajustado para tests múltiples

6.2.3 Análisis microarreglo Affymetrix Col0 "Mock" vs Col0 "BL"

Uno de los factores críticos en la modulación de la respuesta transcripcional a las hormonas, es la presencia o ausencia de estímulos de luz. De esta forma, se seleccionó un grupo de datos de plantas crecidas en el mismo entorno y estadio y ausencia de luz. Estos fueron utilizados para determinar los niveles de expresión de genes en ausencia (mock) y presencia de BL (200 nM) (Tabla 1). En el trabajo de Song y col, 2009 (Song, 2009), se evaluó la respuesta transcripcional en estas condiciones mediante un microarreglo de la plataforma Affymetrix (Tabla 4). Utilizando el ID de acceso GSE14788, se determinaron los genes diferencialmente expresados (FDR < 0,05). A diferencia del análisis anterior, se descargaron los datos crudos directamente desde R con la librería "GEOquery" (ver sección 5.5.2).

Tabla 4 Grupos de comparación de datos de microarreglos para la comparación Col0 Mock vs Col0 BL.

Archivo	Grupo	Descripción
GSM351326	1	Grupo 1 Mock
GSM351327	1	Grupo 1 Mock
GSM351328	2	Grupo 2 200 nM BL
GSM351329	2	Grupo 2 200 nM BL

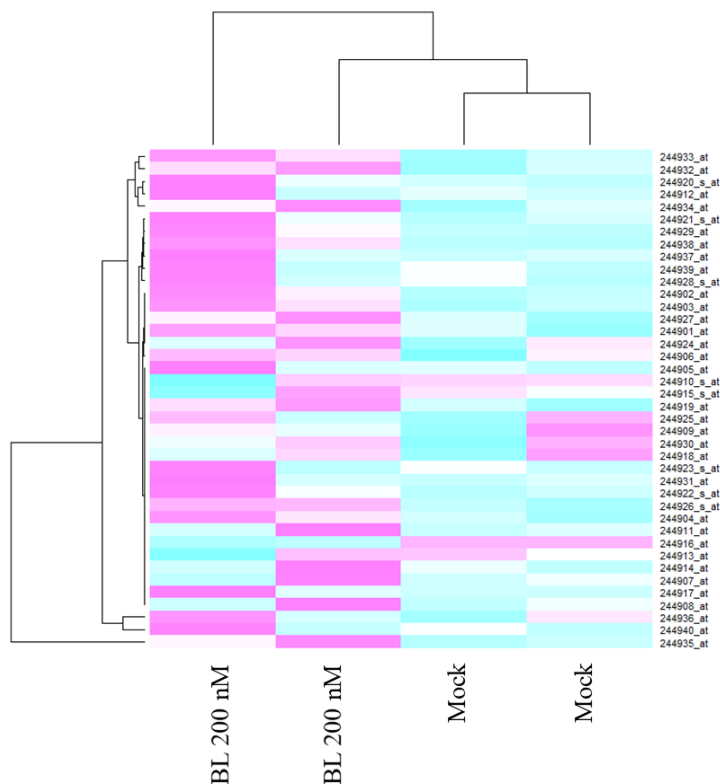


Ilustración 10 Gráfico de clústeres representando las comparaciones de todos los miembros del set de datos. Solo se muestran las primeras 40 sondas. Los valores varían des expresión (magenta) a expresión (turquesa).

Nuevamente, los niveles de expresión fueron determinados por la plataforma de microarreglo Affymetrix para la única comparación (1-2) y el análisis fue llevado a cabo con el software Limma. A partir de la Tabla 4, se diseñó la matriz de comparación. Los archivos de intensidad fueron normalizados y modelados (ver sección 5.5.2). Nuevamente se utilizó la función *topTable* para extraer una tabla de los genes mejor clasificados con el ajuste utilizado. La Ilustración 10 muestra los niveles de expresión para la comparación “Mock” vs “BL”: La primera inspección de los valores de intensidad por sonda, permite evaluar un agrupamiento evidente de las réplicas a nivel de tratamiento.

De la misma forma que en la sección anterior, los valores de $FDR < 0.05$ fueron utilizados para filtrar las tablas y generar las listas de genes diferencial para evaluar respuestas consenso en secciones posteriores (6.2.4).

6.2.4 Determinación de la respuesta consenso al tratamiento con BR.

El uso de gráficos de Venn para la comparación de genes expresados diferencialmente en distintos experimentos de microarreglo, es un recurso gráfico muy utilizado, ya que permite evaluar a primera vista el grado de solapamiento entre grupos de respuesta. Con la intención de determinar los genes asociados a la respuesta a BR son comunes en todas las condiciones, y generar un grupo de genes consenso, se realizó la primera comparación de set de datos, agrupando los experimentos de plantas tratadas y no tratadas con Brasinoesteroides (ver 5.6.1).

Adicionalmente a las listas de genes generadas en las dos secciones previas (6.2.2 y 6.2.3), fue necesario realizar la depuración de las tablas incluidas en los materiales suplementarios del trabajo realizado por Guo y colaboradores (Guo, Li et al. 2009) (ver sección 5.5.3).

De esta manera, los grupos de datos a comparar se resumen en la Tabla 5.

Tabla 5 Grupo de datos, nombre de archivos y set de datos originales utilizados para la comparación de la respuesta consenso a Brasinoesteroides

DB	Acc. ID	Técnica	Entorno	Tratamiento	Crec.	Edad	ID muestra	Nombre Archivo
GEO	GSE14788	Microarreglo	Col0	mock	Dark grown	7-days	GSM351326	BL+BL-.txt
GEO	GSE14788	Microarreglo	Col0	mock	Dark grown	7-days	GSM351327	
GEO	GSE14788	Microarreglo	Col0	200 nM BL	Dark grown	7-days	GSM351328	
GEO	GSE14788	Microarreglo	Col0	200 nM BL	Dark grown	7-days	GSM351329	
-	Tabla S1 y S2	-	Col0 BR+ vs BR-	1 uM BL (2,5h)	15/9h	20-days	-	Col0BL.txt
TAIR	1007966053	Microarreglo	Col0	Mock (30 min)	cont. light	7-days	RIKEN-Goda Sample1	30mlista.txt

TAIR	1007966053	Microarreglo	Col0	Mock (30 min)	cont. light	7- days	RIKEN-Goda Sample1	
TAIR	1007966053	Microarreglo	Col0	10 nM BL (30 min)	cont. light	7- days	RIKEN-Goda Sample8	
TAIR	1007966053	Microarreglo	Col0	10 nM BL (30 min)	cont. light	7- days	RIKEN-Goda Sample8	
TAIR	1007966053	Microarreglo	Col0	Mock (1h)	cont. light	7- days	RIKEN-Goda Sample9	1hlista.txt
TAIR	1007966053	Microarreglo	Col0	Mock (1h)	cont. light	7- days	RIKEN-Goda Sample9	
TAIR	1007966053	Microarreglo	Col0	10 nM BL (1h)	cont. light	7- days	RIKEN-Goda Sample16	
TAIR	1007966053	Microarreglo	Col0	10 nM BL (1h)	cont. light	7- days	RIKEN-Goda Sample16	
TAIR	1007966053	Microarreglo	Col0	Mock (3h)	cont. light	7- days	RIKEN-Goda Sample17	3hlista.txt
TAIR	1007966053	Microarreglo	Col0	Mock (3h)	cont. light	7- days	RIKEN-Goda Sample17	
TAIR	1007966053	Microarreglo	Col0	10 nM BL (3h)	cont. light	7- days	RIKEN-Goda Sample24	
TAIR	1007966053	Microarreglo	Col0	10 nM BL (3h)	cont. light	7- days	RIKEN-Goda Sample24	

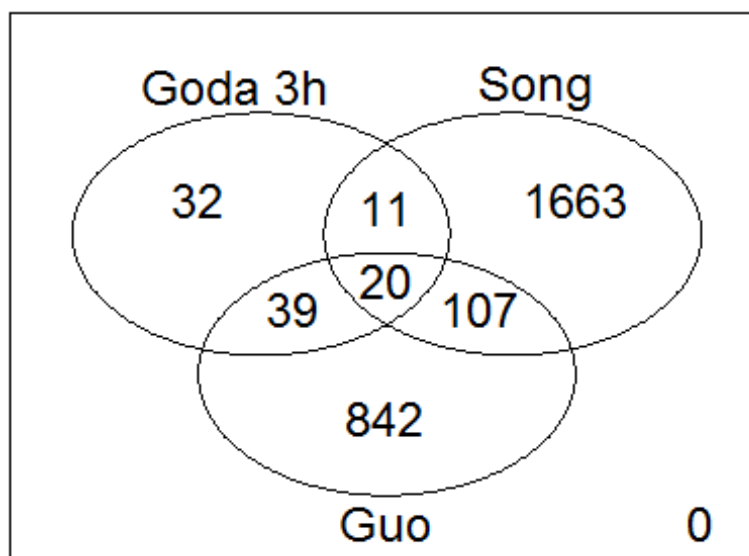


Ilustración 11 Diagrama de Venn entre tres experimentos de transcriptómica de plantas salvajes tratadas con BR. La intersección muestra 20 genes de respuesta consenso.

Así, a partir de los archivos referidos en la Tabla 5, se pudo generar un Diagrama de Venn, el cual permite encontrar fácilmente patrones de respuesta en común utilizando los 3 sub grupos de respuesta tardía (sección 5.6.1).

Una vez obtenido el diagrama de Venn, se realiza la intersección de los sets de datos con la función "intersect", generando el archivo "consensoBL.txt", que muestra la respuesta consenso por acción de BR a t= 3h o más, formado por la intersección de los sets de datos de "Goda 3h", "Song", "Guo". Los genes identificados como consenso fueron identificados en la base de datos de TAIR, generando la Tabla 6.

Tabla 6 Lista de genes que componen el repuesto consenso a Brasinoesteroides. La consulta fue realizada a partir de la herramienta Gene Description Search and Download de TAIR. NA: no aplica.

Locus ID	Descripción	Símbolo primario
AT2G34300	Metiltransferasa dependiente de S-adenosil-L-metionina (Fuente: Araport11)	NA
AT2G34510	Proteína putativa de función desconocida DUF642 (Fuente: Araport11)	NA
AT5G57785	Proteína hipotética (Fuente: Araport11)	NA
AT3G13980	Proteína con dominio SKI/DACH (Fuente: Araport11)	NA
AT5G01740	Pertenece a la familia del factor de transporte nuclear 2 (NTF2) (Fuente: Araport11)	NA
AT5G24570	Proteína hipotética (Fuente: Araport11)	NA
AT4G25780	CAP (Proteína secretora rica en Cisteína) (Fuente: Araport11)	ATCAPE2
AT4G02330	Pectina metilesterasa. Sensible a frío y a Brasinoesteroides.	ATPMEPCRB
AT4G17460	Proteína HD-ZIP clase II - Regula actividad meristemática (gineceo).	HAT1
AT1G30690	Familia fosfatidilinositol transferasa Sec14p-like (Fuente: Araport11)	PATL4
AT3G30180	Citocromo p450 cataliza la última reacción en la producción de Brasinolido.	BR6OX2
AT3G50660	22 α hidroxilasa limitante en la producción de brasinoesteroides. Miembro familia CYP90B.	DWF4
AT5G25190	Miembro de la subfamilia ERF B-6 (Factor de respuesta a etileno).	ESE3
AT2G33570	Glicosiltransferasa (DUF23) (Fuente: Araport11)	GALS1
AT3G28340	Galacturonosiltransferasa putativa.	GATL10
AT2G19800	Pertenece a la familia Mio-inositol oxigenasa.	MIOX2
AT1G76160	SKU5 similar 5 (Fuente: Araport11)	SKS5
AT4G36110	Miembro de familia de proteína de respuesta a auxina del tipo SAUR; (Fuente: Araport11)	SAUR9
AT4G30290	Xiloglucano endotransglucosilasa/hidrolasa.	XTH19
AT1G10550	Proteína de membrana con función predicha en la modificación de la pared celular.	XTH33

Para corroborar que genes de esta respuesta consenso se encuentran alterados a tiempos menores, se realiza el mismo análisis anterior, pero entre el set de datos de Goda 1h, Goda 30m y el "consensoBL".

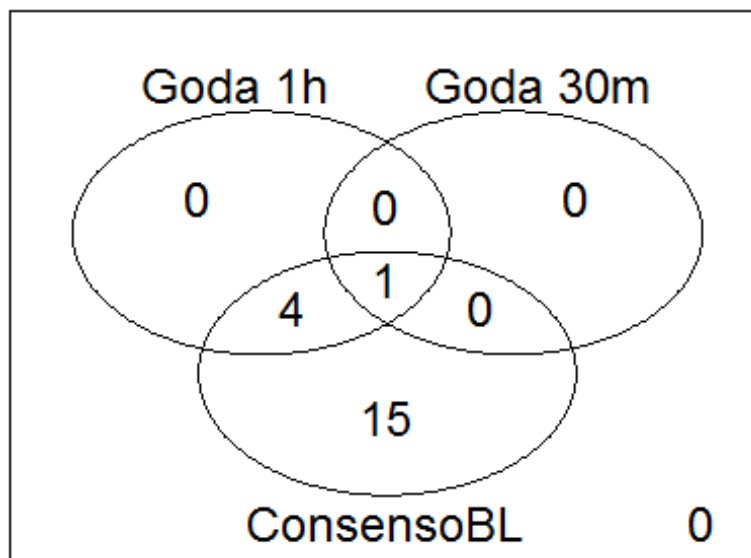


Ilustración 12 Diagrama de Venn entre la respuesta consenso a BR, y set de datos evaluados a 2 tiempos anteriores.

Cuando se realiza la identificación de los genes de respuesta media (intersección Goda 1h vs Consenso BL), se obtienen 4 genes consenso alterados (Tabla 7). Uno de estos genes codifica a un FT que modula la respuesta a infecciones (HAT1) (Zou, Deng et al. 2016). Otro gen codifica a una oxidasa de brasinoesteroides (BR6OX2) que es una enzima citocromo p450 que cataliza el último paso de la producción de Brasinolido. Probablemente esto genere un bucle de retroalimentación positiva involucrado en la transmisión de la señal gatillada por BL en el resto del tejido (Katsumata, Hasegawa et al. 2008). Otro gen codifica por un factor de transporte nuclear (NTF2) (AT5G01740) y el último gen codifica a una proteína rica en cisteína asociada a patogénesis (CAP). Entre estos genes, se encuentra el gen denominado SAUR9 (SMALL AUXIN UPREGULATED RNA 9) y se conoce que es expresado en respuesta a Auxinas.

Tabla 7 Lista de genes que componen la repuesta consenso a tiempos medios a la acción de Brasinoesteroides. La consulta fue realizada a partir de la herramienta "Gene Description Search and Download" de TAIR. NA: no aplica.

Locus ID	Descripción	Símbolo primario
AT5G01740	Pertenece a la familia del factor de transporte nuclear 2 (NTF2) (Fuente: Araport11)	NA
AT4G25780	CAP (Proteína secretora rica en Cisteína) (Fuente: Araport11)	ATCAPE2
AT4G17460	Proteína HD-ZIP clase II - Regula actividad meristemática (gineceo).	HAT1
AT3G30180	Citocromo p450 cataliza la última reacción en la producción de Brasinolido.	BR6OX2
AT4G36110	Miembro de familia de proteína de respuesta a auxina del tipo SAUR; (Fuente: Araport11)	SAUR9

La respuesta a este tiempo parece estar asociada a la retroalimentación positiva en la síntesis de Brasinoesteroides, y a la inducción de una respuesta tardía mediada por factores de transcripción.

Cuando se evalúa la respuesta temprana (consenso a 30 minutos), solo se observa la expresión diferencial del gen SAUR9.

Así, es importante observar cómo se comporta el grupo de genes consenso expresados en presencia de BL. A 30 minutos post tratamiento, solo se encuentra expresado significativamente el gen AT4G36110. Luego de 1 hora post tratamiento, el mismo gen (SAUR9) se encuentra significativamente alterado, aunque se incluye 4 genes adicionales. A partir de las 3 horas, la respuesta estaría mediada por un grupo de 20 genes.

Se ha validado la inducción del gen SAUR9 (y SAUR10) en presencia de IAA y BR, demostrando una respuesta sinérgica extremadamente fuerte hasta 100 veces el nivel de las condiciones control. Adicionalmente, se demostró que SAUR9 se reprime en presencia de Giberelinas, responde en ausencia de luz promueve la elongación celular a través de la acidificación de la pared celular (van Mourik, 2017).

De esta forma, la proteína codificada por gen, parece tener un rol fundamental en esta respuesta. Sería necesario el estudio de este gen para validar su importancia a nivel biológico.

6.2.5 Consulta de términos GO para cada lista de identificadores.

Si bien la especie en estudio es *Arabidopsis thaliana*, cuyos genes se encuentran anotados en la mayoría de los casos, la simple identificación de los mismos no permite relacionar todos los aspectos que caracterizan los productos de cada gen. La integración de la función biológica, la función molecular o la localización subcelular de producto, puede llevar a conclusiones biológicas más precisas.

Una opción fácilmente aplicable a partir de los datos obtenidos, es realizar una consulta de términos GO (“Gene Ontology”) en R por medio del software “GOstats”, el cual puede ser utilizado de manera directa a partir de los datos obtenidos de microarreglos. Esta herramienta utiliza un modelo Hipergeométrico, en donde cada término se trata como una clasificación independiente. Cada gen seleccionado se clasifica en forma cruzada de acuerdo a la anotación GO presente en la base de datos utilizada: *ath1121501GO* dentro de la base de datos *ath1121501* de Affimetrix. Luego, se genera una probabilidad Hipergeométrica (P-value) que determina si el número de genes seleccionados para un término GO (*Count*) es mayor al esperado (ExpCount) (Tabla 8). Este valor esperado depende de la tecnología y plataforma utilizada en el análisis. Así, al utilizar la misma base de datos local que en el análisis de expresión diferencial, la comparación es más rápida, y requiere menor poder de cómputo, y puede realizarse directamente desde R. El inconveniente de esta herramienta es que los términos pueden estar desactualizado. Una opción es usar el software BLAST2GO ya que consulta directamente con la base de datos del NCBI, pfam y KEGG, y, por lo tanto, los términos están actualizados. Sin embargo, este software es privativo y requiere suscripción.

Es necesario aclarar que al observar solo los *p-values* de expresión diferencial para identificar los genes a consultar, se pierde la direccionalidad de la expresión. Esto conlleva a asociar categorías GO de genes que no se encuentran expresados en la misma dirección. Una forma de abordar este problema, es separando los genes seleccionados en dos listas de acuerdo a la dirección de expresión, y realizando dos análisis paralelos. Este es un enfoque más elegante, pero cae por fuera de los objetivos de este trabajo final.

Con este paquete es posible consultar la asociación con vías KEGG “*KEGGHyperGParams*” y dominios proteicos PFAM “*PFAMHyperGParams*”, aunque en este trabajo se restringió la consulta utilizando el argumento “*GOHyperGParams*”, el cual solo ensaya la comparación de la lista de genes con los términos GO de la DB.

Además, en la consulta realizada se puede especificar un tipo de ontología por vez: BP “*Biological Process*”, CC (“*Cellular Compartment*”), and MF (“*Molecular Function*”). En este caso, solo utilizamos el parámetro MF (ver sección 5.7.1). Adicionalmente, utilizamos un *p-value* de corte igual a 0.05 (“*pvalueCutoff*”) para mostrar en la salida solo los términos significativamente enriquecidos, un test “no condicional”, y la direccionalidad apropiada para determinar los términos sobre-representados (“*testDirection = over*”).

La salida de la primera consulta, respuesta consenso a Brasinoesteroides: "consensoBL.txt" se representa en la Tabla 8. La estructura de la misma consta de los ID de los términos GO (GOMFID); los *p-value* asociados al enriquecimiento del mismo en la selección de genes (P-value); la cantidad de veces que se encontró el término (Count); la cantidad esperada para cada término (ExpCount); la “razón de exceso” o “razón de probabilidades” (OddsRatio), parámetro que indica cuan fuertemente se presenta el termino en relación a lo esperado; y el número de genes asociado a este término en la base de datos utilizada (Size).

Tabla 8 Reporte GOstats para la respuesta consenso a BR. ID de los términos GO (GOMFID); los *p-value* asociado al enriquecimiento del mismo en la selección de genes (Pvalue); la cantidad de veces que se encontró el término (Count); la cantidad esperada para cada término (ExpCount); la “razón de exceso” o “razón de probabilidades” (OddsRatio), parámetro que indica cuan fuertemente se presenta el termino en relación a lo esperado; y el número de genes asociado a este término en la base de datos utilizada (Size).

GOMFID	Pvalue	OddsRatio	ExpCount	Count	Size	Término
GO:0010012	0,002	922,6	0,00	1	2	esteroide 22- α hidroxilasa
GO:0048531	0,002	922,6	0,00	1	2	β -1,3-galactosiltransferasa
GO:0050113	0,002	922,6	0,00	1	2	inositol oxigenasa
GO:0016758	0,003	11,8	0,30	3	264	hexosil transferasa
GO:0004497	0,004	23,8	0,10	2	84	monooxigenasa
GO:0008395	0,005	307,5	0,00	1	4	esteroide hidroxilasa
GO:0080039	0,007	184,5	0,01	1	6	xiloglucano endotransglucosilasa
GO:0008378	0,016	70,9	0,02	1	14	galactosiltransferasa
GO:0016757	0,016	6,4	0,55	3	476	glucosil transferasa

GO:0043621	0,025	43,9	0,03	1	22	auto-asociación de proteínas
GO:0016701	0,026	41,9	0,03	1	23	oxidoreductasa
GO:0047262	0,027	40,1	0,03	1	24	poligalacturonato 4- α -galacturonosiltransferasa

A partir de estos datos, se construyó el siguiente gráfico para agrupar las funciones moleculares obtenidas de una manera visual.

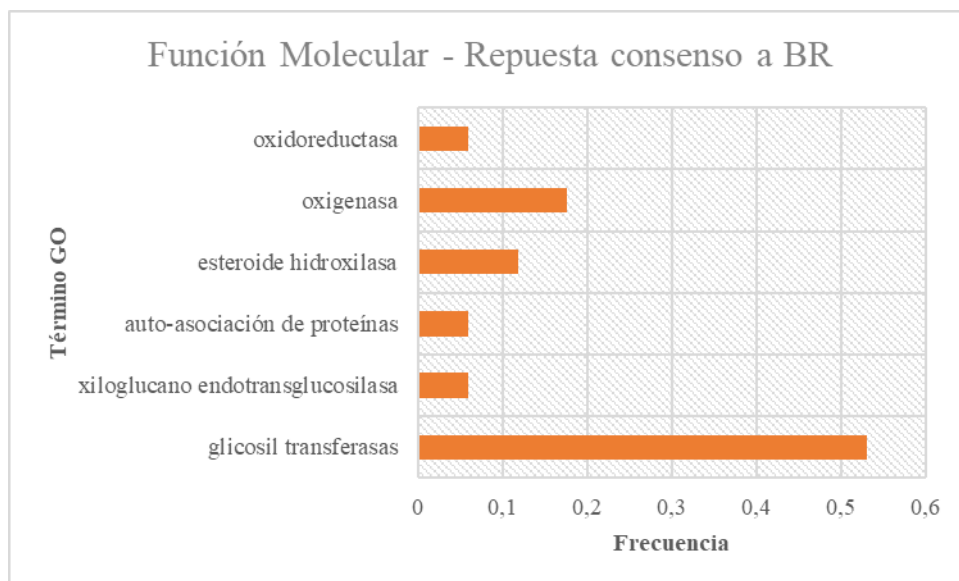


Ilustración 13 Gráfico de frecuencias de términos GO asociados a la respuesta consenso a Brasinoesteroides

De esta manera, se puede observar que los términos sobre-representados para esta respuesta están asociados a la biosíntesis de brasinoesteroides y a la síntesis/remodelación de la pared celular.

6.2.6 Construcción de redes funcionales por "Genemania".

En la actualidad, el crecimiento de la cantidad de datos biológicos a partir de las disciplinas "ómicas", y la presencia de diversas bases de datos curadas y no curadas, hace necesario la evaluación integrativa de cada resultado obtenido. Teniendo en cuenta que estos genes sólo fueron identificados a partir de análisis transcriptómicos, es útil la evaluación de los mismos a través de herramientas bioinformáticas actualmente disponibles. Específicamente, *Genemania* permite generar redes funcionales a partir de datos de co-expresión, Interacción física, Interacción genética, co-localización, vía metabólica, predicción de relación funcional entre genes y similitud en dominios proteicos.

Las redes incluyen los genes consulta (rayas) y genes estrechamente asociados con los mismos.

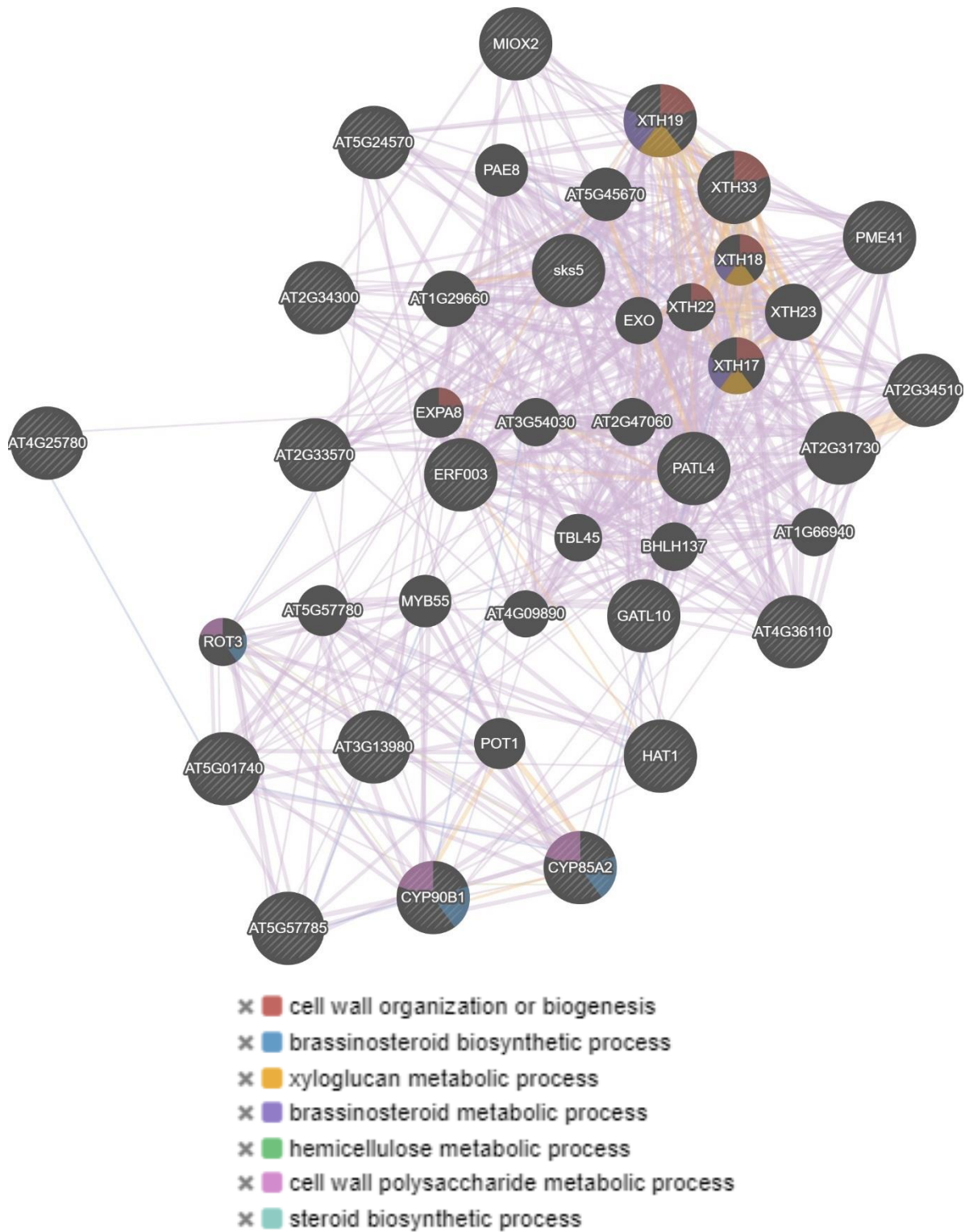


Ilustración 14 Red funcional de respuesta consenso a Brasinoesteroides generada por Genemania. Los genes consulta se encuentran identificados con rayas. Las referencias se encuentran dentro del gráfico.

A partir de los resultados obtenidos, es posible observar una lista de genes consenso que se encuentran regulados por la acción de Brasinoesteroides. Comparando los genes alterados en 3 set de datos distintos, se observó un conjunto de 20 genes, los cuales se agrupan en dos nodos de una red funcional. El primero y más denso, incluye genes asociados con la biogénesis y la organización de la pared celular de *A. thaliana*. El segundo nodo incluye a los genes asociados con la biosíntesis y los procesos metabólicos asociados con los brasinoesteroides. Esto se correlaciona con los grupos de términos GO enriquecidos de la sección 6.2.5.

Adicionalmente, la salida de esta red, permite identificar nuevos genes probablemente asociados a la respuesta. Esta información sería de extremada utilidad para la construcción de un modelo de respuesta a Brasinoesteroides.

6.3 DETERMINACIÓN DE GENES REGULADOS POR LOS FACTORES DE TRANSCRIPCIÓN BZR1, BES1 y PIF.

6.3.1 Selección de experimentos de inmunoprecipitación de la cromatina, orientada a factores de transcripción (FT) asociados a la respuesta transcripcional a la acción de Brasinoesteroides:

GSE24684: Experimento de inmunoprecipitación de la cromatina acoplada a microarreglos (ChIP-chip) del factor de transcripción BES1 (Yu, Li et al. 2011). Se utilizó un anticuerpo anti-BES1 en plantas de 14 días post germinación crecidas en plantas Col0 con una mutación de ganancia de función (*bes1-D*). El anticuerpo anti-GFP fue utilizado como control. Plataforma Affymetrix.

GSE23774: ChIP-chip del factor de transcripción BZR1 (Sun, Fan et al. 2010). Se utilizaron plantas Col0 transformadas con un vector que expresa el factor BZR1 fusionado al dominio CFP (*pBZR1::BZR1-CFP*). Se precipito BZR1 con el anticuerpo anti-CFP, mientras que el control se llevó a cabo en plantas silvestres con el mismo anticuerpo. Las plantas fueron crecidas por 28 días en un ciclo luz oscuridad de 16/8h. Plataforma Affumetrix.

GSE43286: Experimento de inmunoprecipitación de la cromatina acoplada a secuenciación paralela masiva (ChIP-Seq). Se utilizaron plantas Col0 transgénicas, sobreexpresantes del FT PIF1 fusionado al epítipo Myc (P1M), inmunoprecipitadas con el anticuerpo anti-Myc. La comparación se realizó contra plantas salvajes, usando el mismo anticuerpo. Plataforma Illumina.

De esta forma, podremos identificar genes que no presentarían regulación asociada a estas vías.

6.3.2 Análisis ChIP-chip del regulador BES1

Los Brasinoesteroides son reguladores importantes para el crecimiento y el desarrollo de las plantas. Estas hormonas gatillan una vía de señalización, partiendo del receptor de pared BRI1, y concluyendo con el control de la actividad de dos familias de factores de transcripción: BES1 y BZR1. La forma más directa de evaluar los genes que son regulados por un grupo de factores de transcripción, es por ensayos de inmunoprecipitación de la cromatina acoplado con microarreglos

en baldosas ("tiling array"): "ChIP-chip". La diferencia entre los *tiling arrays* y los microarreglos tradicionales es la naturaleza de las sondas. En los microarreglos tradicionales, las mismas cubren todos los genes con función determinada a predicha, sin embargo, los *tiling arrays* se utiliza una gran cantidad de sondas para alguna región del genoma, preferentemente continua, en la que las funciones locales son mayormente desconocidas. Adicionalmente, se puede combinar la inmunoprecipitación de la cromatina (ChIP) con NGS, generando un nuevo enfoque denominado ChIP-Seq. El objetivo de estas técnicas, es localizar los sitios de unión de proteínas los cuales pueden ayudar a identificar elementos funcionales en el genoma. En cualquiera de los dos casos, al mapear las regiones a las que se unen los factores de transcripción, es posible determinar los genes cercanos y asociar una posible relación FT-promotor de gen diana.

La evaluación de los genes directamente regulados por estos factores de transcripción (FT) permite identificar cuáles de estos, se regulan directamente por la acción del receptor BRI1, y en consecuencia por la acción de los Brasinoesteroides. Para esto, el primer paso implicó la determinación de los grupos de datos públicos que cumplen con los requerimientos. Luego, los genes identificados en estos tratamientos serán comparados con los genes que interactúan directamente con los tres reguladores mencionados (BZR1, BES1 y PIF). De esta forma, podremos identificar genes que no presentarían regulación asociada a estas vías.

En la actualidad, he podido determinar dos ensayos de ChIP-chip, uno para el regulador BES1 y otro para el regulador BZR1 (Sun, Fan et al. 2010, Yu, Li et al. 2011) y un ensayo de ChIP-Seq para el regulador PIF (Pfeiffer, Shi et al. 2014).

En esta sección nos centraremos en el ensayo ChIP-chip de BES1, realizado en la plataforma de Affymetrix - GPL10977 [At35b_MR] Arabidopsis Tiling 1.0R Array. De esta forma, se procede al análisis de los sets de datos GSE24684 obtenidos por inmunoprecipitación de la cromatina (ChIP) de *Arabidopsis thaliana* por un anticuerpo anti-BES1, utilizando el anticuerpo anti-GFP como control por triplicado, a través del software CisGenome (ver sección 5.5.5.1).

La salida de este programa es una lista de genes seleccionados ($FDR < 0.05$), indicando el ID, las coordenadas genómicas, la hebra codificante y el cromosoma en el que se encuentra. A partir de esta, se obtuvo los identificadores de locus de los genes regulados directamente por BES1 (899 genes, ver material suplementario).

6.3.3 Análisis ChIP-chip del regulador BZR1

En el trabajo llevado a cabo por Sun y colaboradores (Sun, 2010), se realizó un ChIP-chip a partir de plantas transformadas con el vector pBZR1::BZR1-CFP, en comparación con plantas Col-0 salvaje. Ambas líneas fueron inmunoprecipitadas por triplicado utilizando un anticuerpo anti-CFP, y el DNA obtenido fue hibridado en la misma plataforma de Affymetrix que el caso anterior (GPL10977 [At35b_MR] Arabidopsis Tiling 1.0R Array). El análisis de este set de datos fue llevado a cabo de manera similar a la sección anterior (ver sección 5.5.5.2), obteniéndose una lista de genes regulados directamente por BZR1 (2395 genes, ver material suplementario).

6.3.4 Análisis ChIP-Seq del regulador PIF1

En el trabajo realizado por Pfeiffer y colaboradores (Pfeiffer, Shi et al. 2014), se realizó un experimento de ChIP-Seq, en donde se compararon plantas Col0 transgénicas, y plantas sobreexpresantes del FT PIF1 (ver 6.3.1). De esta forma, se generaron 6 bibliotecas, las cuales se secuenciaron en una línea de la plataforma Illumina HiSeq 2000, con una configuración 1 x 50 bp. El llamado de bases fue realizado con el “*pipeline*” CASAVA-1.8.0. Finalmente, los FASTQ disponibles de la base de datos, fueron utilizados para su reanálisis por Galaxy (ver 5.5.6). Los archivos “.BED” fueron utilizados en CisGenome para la anotación de genes. Esto generó una lista de genes que fue utilizada en análisis posteriores (2016 genes, ver material suplementario).

6.3.5 Discriminación de grupos de genes regulados por BR mediados y no mediados por los reguladores transcripcional BZR1 BES1 y PIF.

Con la intención de encontrar cuales de los genes consenso se encuentran regulados por FT conocidos, decidimos realizar la intersección de los sets de datos obtenidos en los análisis de inmunoprecipitación de la cromatina dirigida a los reguladores PIF1, BZR1 y BES1, con los genes consenso encontrados en la sección anterior (ver sección 5.6.4). En la siguiente tabla, se encuentra una descripción asociada a los archivos utilizados.

Tabla 9 Grupo de datos, nombre de archivos y set de datos originales utilizados para la clasificación de la respuesta a BL con y sin regulador identificado

DB	Acc. ID	Técnica	Entorno	Crec.	Edad	ID muestra	Nombre Archivo
GEO	GSE24684	ChIP-chip BES1	BES1	16/8h	14-days	GSM608139	bes1.txt
GEO	GSE24684	ChIP-chip BES1	BES1	16/8h	14-days	GSM608140	
GEO	GSE24684	ChIP-chip BES1	BES1	16/8h	14-days	GSM608141	
GEO	GSE24684	ChIP-chip BES1	GFP	16/8h	14-days	GSM608142	
GEO	GSE24684	ChIP-chip BES1	GFP	16/8h	14-days	GSM608143	
GEO	GSE24684	ChIP-chip BES1	GFP	16/8h	14-days	GSM608144	
GEO	GSE23774	ChIP-chip BZR1 (2C)	pBZR1::BZR1-CFP vs. Col0	16/8h	28-days	GSM586804	BZR1.txt
GEO	GSE43286	ChIP-Seq PIF	PIF1	dark-grown	3-days	GSM1059957	PIF1.txt
GEO	GSE43286	ChIP-Seq PIF	PIF1	dark-grown	3-days	GSM1059958	
GEO	GSE43286	ChIP-Seq PIF	PIF1	dark-grown	3-days	GSM1059959	
GEO	GSE43286	ChIP-Seq PIF	WT	dark-grown	3-days	GSM1059960	
GEO	GSE43286	ChIP-Seq PIF	WT	dark-grown	3-days	GSM1059961	
GEO	GSE43286	ChIP-Seq PIF	WT	dark-grown	3-days	GSM1059962	

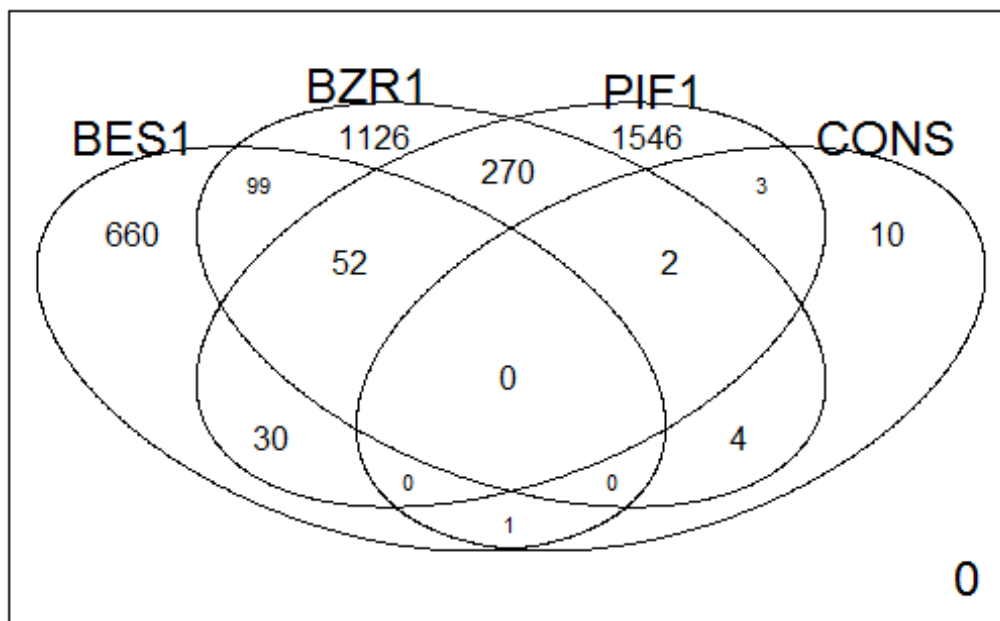


Ilustración 15 Diagrama de Venn comparando la respuesta consenso a BR, y los genes controlados por tres FT distintos.

De los genes consenso, solo 3 se regulan solo PIF1, 4 solo BZR1 y 1 solo BES1. Además, 2 genes se regulan por PIF1 y BZR1. En consecuencia, 10 genes de respuesta consenso se estarían regulando por otra vía (ver Tabla 10).

Tabla 10 Lista de genes que componen la repuesta consenso a Brasinoesteroides, con información sobre el regulador transcripcional asociado a la respuesta. En los genes cuyos reguladores no han sido estudiados o identificados se supone una vía de regulación alternativa. La consulta fue realizada a partir de la herramienta Gene Description Search and Download de TAIR. NA: no aplica.

Locus ID	Descripción	Regulador asociado	Símbolo primario
AT2G34300	Metiltransferasa dependiente de S-adenosil-L-metionina (Fuente: Araport11)	Alternativa	NA
AT5G57785	Proteína hipotética (Fuente: Araport11)	Alternativa	NA
AT3G13980	Proteína con dominio SKI/DACH (Fuente: Araport11)	Alternativa	NA
AT5G24570	Proteína hipotética (Fuente: Araport11)	Alternativa	NA
AT4G17460	Proteína HD-ZIP clase II - Regula actividad meristemática (gineceo).	Alternativa	HAT1
AT2G33570	Glicosiltransferasa (DUF23) (Fuente: Araport11)	Alternativa	GALS1
AT3G28340	Galacturonosiltransferasa putativa.	Alternativa	GATL10

AT4G36110	Miembro de familia de proteína de respuesta a auxina del tipo SAUR; (Fuente: Araport11)	Alternativa	SAUR9
AT4G30290	Xiloglucano endotransglucosilasa/hidrolasa.	Alternativa	XTH19
AT1G10550	Proteína de membrana con función predicha en la modificación de la pared celular.	Alternativa	XTH33
AT2G34510	Proteína putativa de función desconocida DUF642 (Fuente: Araport11)	BES1	NA
AT5G01740	Pertenece a la familia del factor de transporte nuclear 2 (NTF2) (Fuente: Araport11)	BZR1	NA
AT4G25780	CAP (Proteína secretora rica en Cisteína) (Fuente: Araport11)	BZR1	ATCAPE2
AT3G50660	22 α hidroxilasa limitante en la producción de brasinoesteroides. Miembro familia CYP90B.	BZR1	DWF4
AT2G19800	Pertenece a la familia Mio-inositol oxigenasa.	BZR1	MIOX2
AT4G02330	Pectina metilesterasa. Sensible a frío y a Brasinoesteroides.	PIF	ATPMEPCRB
AT1G30690	Familia fosfatidilinositol transferasa Sec14p-like (Fuente: Araport11)	PIF	PATL4
AT1G76160	SKU5 similar 5 (Fuente: Araport11)	PIF	SKS5
AT3G30180	Citocromo p450 cataliza la última reacción en la producción de Brasinolido.	PIF/BZR1	BR6OX2
AT5G25190	Miembro de la subfamilia ERF B-6 (Factor de respuesta a etileno).	PIF/BZR1	ESE3

De esta manera, estos tres reguladores estudiados parecerían estar involucrados en la expresión de genes asociados a la amplificación de la respuesta (hacia una respuesta secundaria) y a la producción de Brasinoesteroides (retroalimentación positiva). Por otro lado, los genes de respuesta mayoritariamente asociados a la remodelación de la pared celular, parecen estar asociados a vías alternativas.

En referencias bibliográficas (van Mourik, 2017), se ha identificado a SAUR9 como blanco de los reguladores ARF6/ARF8 and BZR1, aunque no se ha identificado el motivo de unión a BZR en la región promotora de SAUR9. Estos resultados, juntos con los observados en este análisis, sugieren que la inducción de SAUR9 puede estar mediada por otro regulador.

6.3.6 Consulta de términos GO para cada lista de identificadores.

Para avanzar en el estudio de los genes consenso sin regulador asociado, se realizó la identificación de los términos GO enriquecidos en este grupo de genes (ver sección 5.7.2). De esta manera, en la Tabla 11 se puede observar la salida de “GOSTats” para la consulta de términos MF.

Tabla 11 Salida de GOSTats para la respuesta consenso a BR sin regulador identificado. ID de los términos GO (GOMFID); los p-value asociado al enriquecimiento del mismo en la selección de genes (Pvalue); la cantidad de veces que se encontró el término (Count); la cantidad esperada para cada término (ExpCount); la “razón de exceso” o “razón de probabilidades” (OddsRatio), parámetro que indica cuan fuertemente se presenta el termino en relación a lo esperado; y el número de genes asociado a este término en la base de datos utilizada (Size).

GOMFID	Pvalue	Odds Ratio	ExpCount	Count	Size	Término
GO:0016758	0,000	29,6	0,15	3	264	hexosil transferasa
GO:0048531	0,001	1961,6	0,00	1	2	β -1,3-galactosiltransferasa
GO:0016757	0,002	16,1	0,27	3	476	glucosil transferasa
GO:0080039	0,003	392,2	0,00	1	6	xiloglucano endotransglucosilasa
GO:0008378	0,008	150,8	0,01	1	14	galactosiltransferasa
GO:0043621	0,013	93,3	0,01	1	22	auto-asociación de proteínas
GO:0047262	0,014	85,2	0,01	1	24	poligalacturonato 4- α -galacturonosiltransferasa
GO:0016798	0,017	12,4	0,20	2	355	glucosil hidrolasa

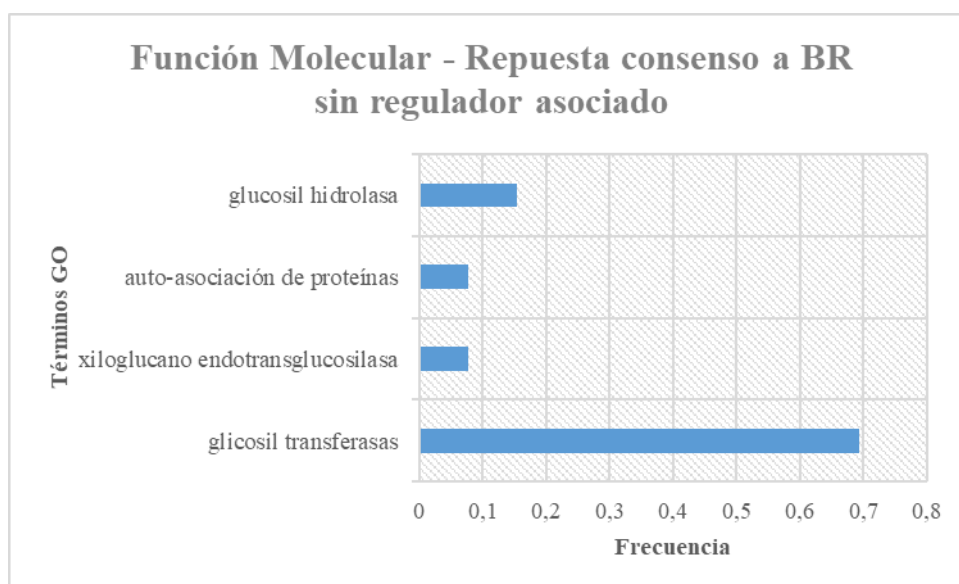


Ilustración 16 Gráfico de frecuencias de términos GO asociados a la respuesta consenso a Brasinoesteroides sin regulador asociado.

Es llamativo el enriquecimiento de términos relacionado con enzimas remodeladoras de la pared celular de plantas (Ilustración 16).

6.3.7 Construcción de redes funcionales por "Genemania".

Con el objetivo de representar la relación actualmente conocida entre los genes de respuesta a BR sin regulador asignados, y con la intención de expandir estas asociaciones a genes estrechamente relacionados a los mismos, se obtuvo la siguiente red funcional mediante Genemania (Ilustración 17).

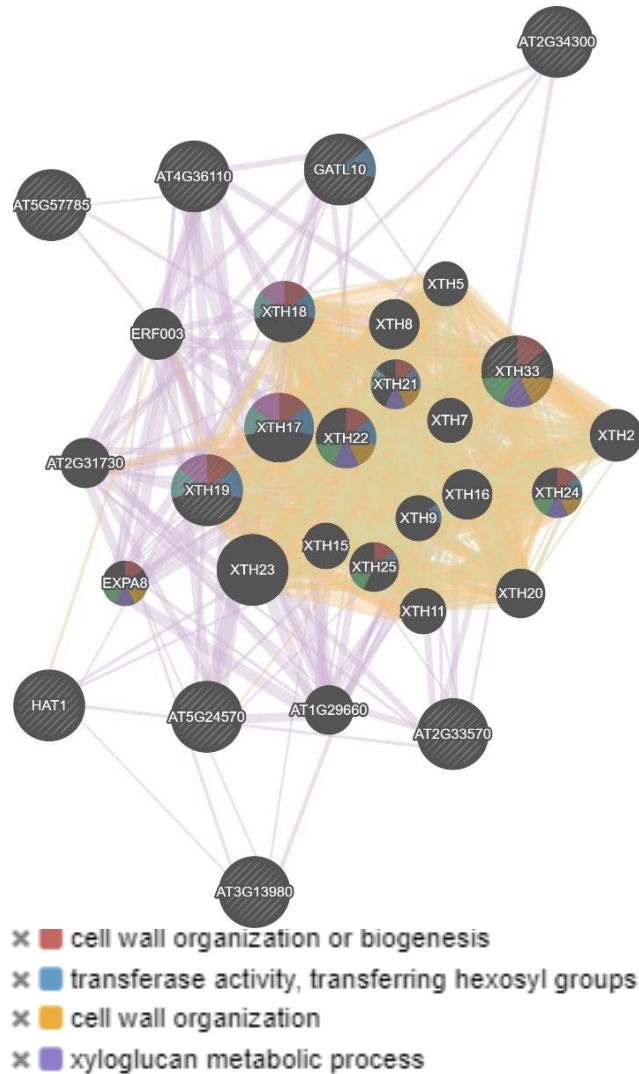


Ilustración 17 Red funcional de respuesta consenso a Brasinoesteroides sin regulador asignado, generada por Genemania. Los genes consulta se encuentran identificados con rayas. Las referencias se encuentran dentro del gráfico.

Esta red funcional, agrupa genes basándose en cuatro procesos funcionales: Biogénesis de la pared celular, Actividad hexosil transferasa, organización de la pared celular y metabolismo del xiloglucano.

6.4 IDENTIFICACIÓN DE LOS GENES ESPECÍFICOS PARA LA VÍA THE1/HERK Y CONTRASTACIÓN CON LOS GENES REGULADOS POR LOS FACTORES DE TRANSCRIPCIÓN EVALUADOS.

6.4.1 Selección de experimentos que muestren la respuesta transcripcional de plantas mediada por la vía THE1/HERK1:

GSE7937: Comparación de Microarreglos de dos canales, de plantas *A. thaliana* Ws (*Wassilewskija*) silvestre, y plantas Ws mutantes delecionales en el receptor *the1* (Hematy, Sado et al. 2007). Las mismas fueron crecidas durante 5 días en completa oscuridad. Estos microarreglos incluye un diseño de intercambio de fluoróforos ("Dye Swap") (ver sección 6.4). Plataforma CATMA. Lamentablemente, el entorno genético de las *Arabidopsis thaliana* utilizadas en esta comparación (*wassilewskija*) es distinta a las utilizadas en el resto (Col0).

Por último, el mismo trabajo utilizado en la sección 6.1.1, contamos con la lista de genes diferencialmente expresados en un experimento de microarreglo de 1 canal de Affymetrix (Guo, Li et al. 2009) de plantas Col0 adultas (20 días post germinación) *the1 herk1* crecidas en presencia de 1 μ M de BL en un ciclo de luz/oscuridad 15h/9h. Como se comentó más arriba, los datos crudos de este experimento no se encuentran en la base de datos GEO.

El entrecruzamiento de la salida del análisis de estos experimentos puede ayudar a determinar distintas consultas, con el objetivo de responder preguntas de interés biológico.

6.4.2 Análisis del perfil de expresión de *Arabidopsis thaliana* WS vs the1-3

La comparación de los perfiles de expresión de plantas de 5 días post germinación en oscuridad continua, fue llevado a cabo en la plataforma "CATMA microarray". El Proyecto Complete Arabidopsis Transcriptome MicroArray (CATMA) tuvo como objetivo diseñar y producir etiquetas de alta calidad, con la capacidad de cubrir la mayoría de los genes de *A. thaliana*. Los diseños incluyeron microarreglos de dos colores, los cuales se hibridan con cDNA provenientes de dos muestras a ser comparadas, marcadas con diferentes fluoróforos. La intensidad de cada fluoróforo es utilizada para determinar los genes que se encuentran expresados diferencialmente. El marcado recíproco con fluoróforos ("*dye swap*"), es un tipo de diseño en el cual se compara las intensidades en dos microarreglos independientes en los cuales las dos muestras intercambian los fluoróforos. Esto se realiza para evitar los sesgos inherentes a cada fluoróforo (Hematy, Sado et al. 2007). El diseño de las comparaciones silvestre (*A. thaliana wassilewskija*) vs the1-3, implicó una primera comparación WS (Cy3) vs the1-3 (Cy5), y una segunda comparación the1-3 (Cy3) vs WS (Cy5).

Tabla 12 Diseño del Intercambio de fluoróforos utilizados en la comparación WS the1-3

FileName	Cy3	Cy5
GSM194873.gpr	WS	the1-3
GSM194874.gpr	the1-3	WS

Este diseño fue incluido, junto con los nombres de archivos de los datos crudos, en un archivo de texto llamado "Targets.txt", el cual fue leído por defecto por el paquete limma. Los datos crudos de estos microarreglos fueron obtenidos en formato. GPR. Los mismos pertenecen a la salida típica del programa de análisis "GenePix". A continuación, se incluye el script utilizado para el análisis. Los pasos incluyen la lectura de los datos crudos, la normalización intra e inter arreglo, y la determinación de los transcritos diferencialmente expresados. Por último, se procedió a imprimir un archivo de lista de genes para la comparación mediante diagramas de Venn, y a la determinación de la ontología de este grupo de genes mediante el paquete *GOstats* (ver sección 5.5.3).

6.4.3 Análisis de los perfiles de expresión de plantas mutantes en *the1 herk1*:

Además del receptor quinasa THE1, otro receptor de tipo quinasa (RLKs) que se encuentra transcripcionalmente inducido por BR, es el HERCULES1. Estos dos receptores pertenecen a una familia de CrRLK en *A. thaliana* de 17 miembros (Guo, 2009). En trabajo de Guo y colaboradores, se llevó a cabo experimentos de microarreglos con plantas *A. thaliana* Col0 de 24 días post germinación. Un grupo de plantas silvestre fueron rociadas con BL 1M e incubadas por 2.5 horas, mientras que se realizó lo mismo con agua en un grupo control ("mock"). Adicionalmente se realizó el mismo ensayo en plantas mutantes *the1 herk1*. El RNA de las mismas fue extraído e hibridado en microarreglos comerciales de Affymetrix (*Arabidopsis* Genomic arrays). Se realizaron dos comparaciones. La primera incluye la comparación de plantas silvestres tratadas con BL vs plantas silvestres no tratadas (mock). La segunda comparación incluye plantas mutantes tratadas con BL vs plantas mutantes no tratadas (mock). Lamentablemente, los datos crudos de este experimento no se encuentran disponibles y no fueron compartidos conmigo ante la consulta a los autores de la publicación. Sin embargo, he obtenido las tablas que incluyen los genes que presentaron expresión diferencial significativa en ambas comparaciones, y he dado formato a las mismas para determinar la lista de genes apropiada a cada comparación (ver Sección 5.5.4). La lista de genes se encuentra en el archivo "wsthe1.txt" en el material suplementario.

6.4.4 Identificación genes específicos para la vía THE1/HERK

Para determinar que genes estarían controlados por la vía THE1/HERK1, contamos con tres grupos de identificadores de genes.

Tabla 13 Grupo de datos, nombre de archivos y set de datos originales utilizados para la comparación de la respuesta consenso a Brasinoesteroides por la vía THE1/HERK1

DB	Acc. ID	Técnica	Entorno	Tratamiento	Crec.	Edad	ID muestra	Nombre Archivo
-	Tabla S1 y S2	-	Col0 BR+ vs BR-	1 uM BL (2,5h)	15/9h	20-days	-	Col0BL.txt
-	Tablas S3 y S4	-	herk the1 BR+ vs BR-	1 uM BL (2,5h)	15/9h	20-days	-	mutBL.txt

GEO	GSE7937	2 color Microarreglo	WS vs the1	0	Dark grown	5- days	GSM194873	wsthe1.txt
GEO	GSE7937	2 color Microarreglo	The1 vs WS	0	Dark grown	5- days	GSM194874	

De esta manera, la primera comparación implica determinar la diferencia entre los grupos de datos Col0BL y MutBL. Esto permite evaluar los genes diferencialmente expresados en la mutante *the1 herk1* tratados con BL. Se podría pensar que el grupo de genes regulados en la planta silvestre tratada, pero no en la planta mutante tratada, estarían controlados por la vía ausente en las plantas mutantes. Para tratar de diferenciar estos genes, de otros regulados por pleiotropismos, realizamos la intersección de este grupo de genes, con los que se encuentran diferencialmente expresados en entre plantas *A. thaliana* WS y la mutante en *THE1* (ver sección 5.6.5).

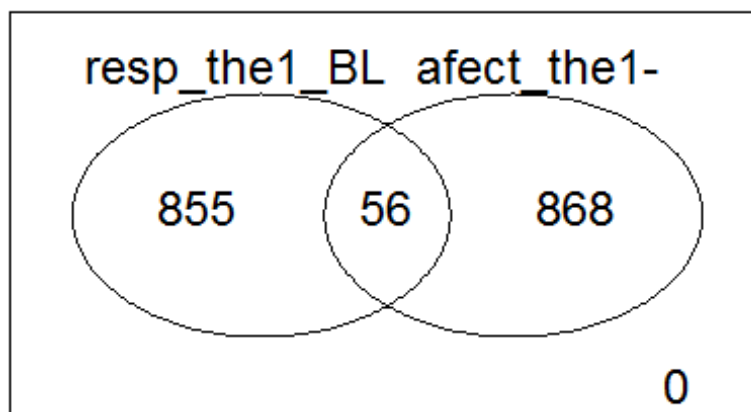


Ilustración 18 Diagrama de Venn identificando la respuesta consenso a BR mediada por la vía *THE1/HERK1*.

Podemos observar que existen 56 genes compartidos en estos dos sets de datos, formando lo que se denomina la "respuesta a BL por la vía *THE1/HERK1*".

Tabla 14 Lista de genes que componen la repuesta mediada por la vía *THE1/HERK1*. La consulta fue realizada a partir de la herramienta *Gene Description Search and Download* de TAIR. NA: no aplica.

Locus ID	Descripción	Símbolo primario
AT2G39980	Acil - Transferasa del tipo HXXXD-tipo	NA
AT5G45670	Acil - Transferasa motivo GDSL	NA
AT5G24490	Proteína de la subunidad menor ribosomal 30S	NA
AT2G03350	Proteína sin función conocida DUF538	NA
AT4G32870	Policétido ciclasa	NA
AT3G51750	Proteína hipotética	NA
AT3G56880	Proteína con motivo VQ	NA

AT4G28780	Acil - Transferasa motivo GDSL	NA
AT5G01740	Pertenece a la familia del factor de transporte nuclear 2 (NTF2) (Fuente: Araport11)	NA
AT5G14120	Fosfatidilinositol transferasa	NA
AT4G37240	Regulador transcripcional del tipo HTH	NA
AT3G19030	Proteína tipo factor de iniciación de transcripción TFIID	NA
AT2G44670	Proteína asociada a la senescencia (DUF581)	NA
AT5G57785	Proteína hipotética (Fuente: Araport11)	NA
AT3G15630	Galacturonosiltransferasa putativa	NA
AT5G03120	Proteína transmembrana	NA
AT1G48570	Proteína de unión a RAN - Dedos de Zn	NA
AT3G17380	Proteína del tipo TRAF	NA
AT5G64090	Xiloglucano endotransglucosilasa	NA
AT1G29660	Acil - Transferasa motivo GDSL	NA
AT3G62550	Proteína tipo Adenina nucleótido α hidrolasa	NA
AT2G29670	Proteína del tipo TPR	NA
AT1G07135	Proteína rica en glicina	NA
AT3G07010	Pectina liasa	NA
AT1G51090	Proteína de respuesta inmunológica	ATHMAD1
AT4G21960	Peroxidasa	PRXR1
AT1G64900	Citocromo P450	CYP89A2
AT3G18080	B-S glucosidasa 44	BGLU44
AT3G13750	β - galactosidasa	BGAL1
AT5G63810	Glicósido Hidrolasa	BGAL10
AT3G48360	Factor de transcripción relacionado a respuestas múltiples	BT2
AT4G04700	Quinasa dependiente de Calcio	CPK27
AT2G46830	Factor de transcripción asociado al ritmo circadiano	CCA1
AT5G15410	canal de catión activado por nucleótido cíclico	CNGC2
AT1G28330	Proteína asociada a dormancia	DYL1
AT4G08950	Proteína de respuesta a fosfato	EXO
AT5G09820	Fibrilina 5	FBN5
AT3G15540	Respuesta primaria a Auxinas	IAA19
AT2G35060	Transportador de grupos fosfato	KUP11
AT1G78600	Proteína de dedos Zn regulada por luz	LZF1
AT5G52310	Respuesta a frío	LTI78
AT1G68570	Transportador de Giberelinas	NPF3.1

AT1G22530	Proteína regulada por Auxinas	PATL2
AT3G49220	Inhibidor de Invertasa/Pectina metilesterasa	PME34
AT3G29670	Malonil transferasa	PMAT2
AT5G04190	Fitocromo quinasa	PKS4
AT4G02075	Proteína dedos Zn RING/FYVE/PHD	PIT1
AT5G37260	Proteína asociada a la regulación del ritmo circadiano	RVE2
AT2G30520	Proteína asociada a respuesta fototrópica	RPT2
AT3G48460	Acil - Transferasa motivo GDSL	SFAR4
AT3G18280	Inhibidor de transferencia lipídica en semillas	TED4
AT1G70560	Producción de precursor de Auxinas	TAA1
AT4G32770	Tocoferol ciclasa	VTE1
AT5G06690	Tioredoxina	WCRKC1
AT1G11545	Xiloglucano endotransglucosilasa	XTH8
AT3G62720	Xilosiltransferasa	XT1

Para contestar la pregunta sobre cuántos de estos genes se encuentran regulados en el consenso BL, y cuántos de estos lo hacen por intermedio de un FT desconocido (ver sección 5.6.6), comparamos:

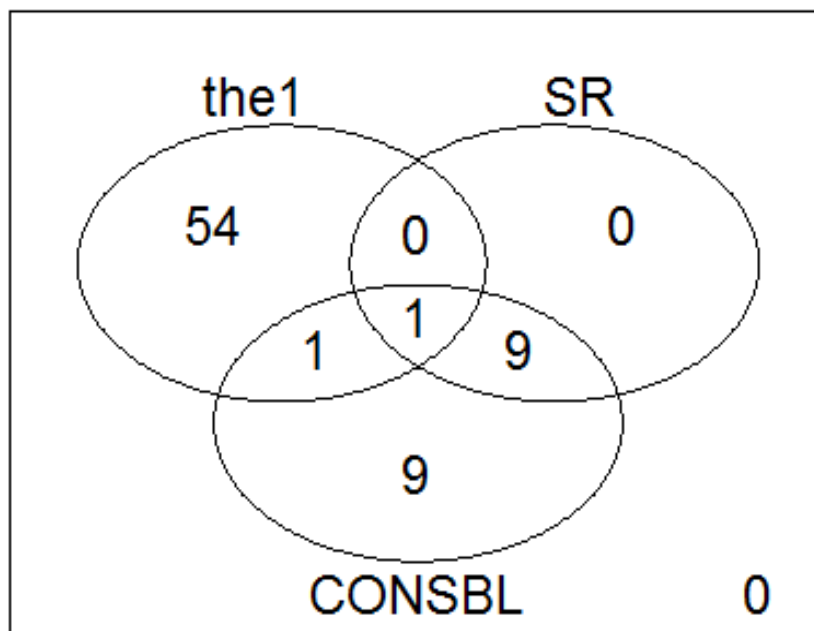


Ilustración 19 Diagrama de Venn comparando la respuesta consenso a BR, la respuesta mediada por la vía THE1 y la respuesta consenso sin regulador asignado.

De esta manera, observamos que dos genes se comparten con la respuesta consensoBL. Solo uno de estos genes se regula por medio de uno de los tres FT evaluados (BZR1) (Tabla 15).

Tabla 15 Lista de genes consenso involucrados en la respuesta mediada por THE1/HERK1, con información sobre el regulador transcripcional asociado a la respuesta. En los genes cuyos reguladores no han sido estudiados o identificados se supone una vía de regulación alternativa. La consulta fue realizada a partir de la herramienta Gene Description Search and Download de TAIR. NA: no aplica.

Locus ID	Descripción	Regulador asociado	Símbolo primario
AT5G01740	Pertenece a la familia del factor de transporte nuclear 2 (NTF2) (Fuente: Araport11)	BZR1	NA
AT5G57785	Proteína hipotética (Fuente: Araport11)	Alternativa	NA

De esta manera, observamos que existe un bajo solapamiento entre la respuesta consenso a BR. De los dos genes que se encontraron en común, solo uno es regulado por un FT conocido (BZR1). Así, la evidencia encontrada hasta el momento, sugiere que la respuesta consenso no mediada por los reguladores conocidos, no sería promovida por la vía THE1/HERK1.

Para contrastar la respuesta de esta vía con la mediada por brasinoesteroides a distintos tiempos post tratamiento, se procede de la siguiente manera.

6.4.5 Correlacionar la vía de respuesta THE1/HERK1 con los asociados a respuesta temprana y tardía.

Para identificar si alguno de los genes de la respuesta mediada por THE1, se encuentran también asociados a la respuesta temprana mediada por BR, utilizamos los siguientes grupos de datos (ver 5.6.7 y Tabla 16):

Tabla 16 Descripción de los archivos utilizados para asociar la respuesta THE1 con la curva de respuesta a lo largo del tiempo

Nombre archivo	N.º genes	Descripción
the1_reg.txt	56	Respuesta consenso por la vía THE1/HERK1
int_tiempo_total.txt	1	Respuesta consenso por acción de BR a t= 30m
int_tiempo_medio.txt	4	Respuesta consenso por acción de BR a t= 1h
consensoBL.txt	20	Respuesta consenso por acción de BR a t= 3h o más

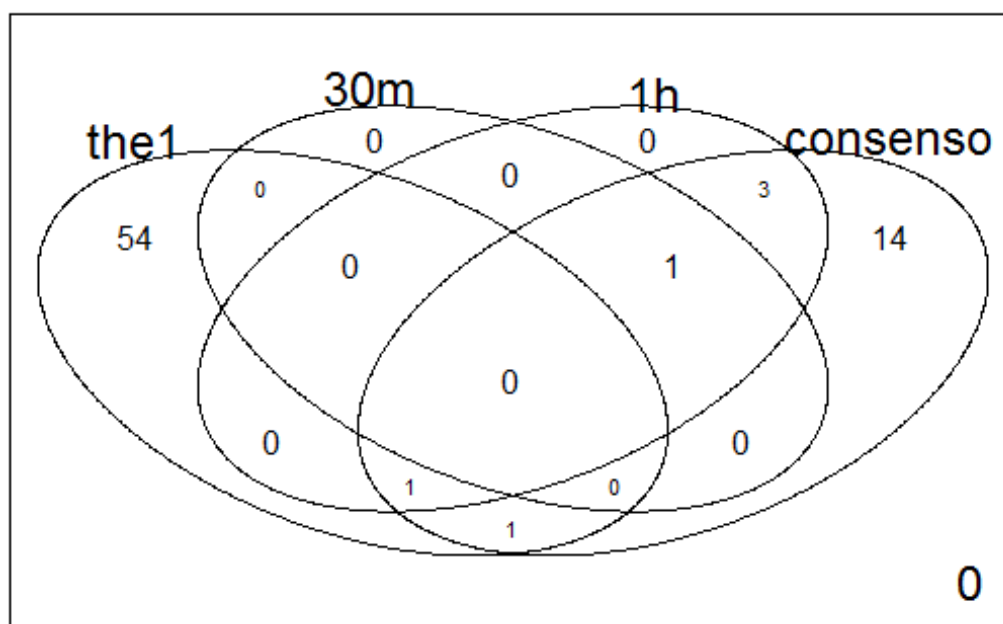


Ilustración 20 Diagrama de Venn comparando la respuesta consenso a BR, la respuesta mediada por THE1 y la respuesta obtenida a 30 minutos y 1 hora post tratamiento con BR.

En la Ilustración 20 se observa que de los dos genes identificados en la sección anterior (6.4.4), el regulado por BZR1, AT5G01740, también se encuentra en la respuesta temprana (1h). El gen AT5G57785, estaría respondiendo a tiempos más elevados. Esto podría indicar que la respuesta mediada por THE1/HERK1 podría iniciarse a tiempos post tratamiento más largos. Es necesaria mucha más evidencia para conectar ambas vías.

6.4.6 Consulta de términos GO para cada lista de identificadores.

Con el objetivo de evaluar los términos GO enriquecidos en la respuesta mediada por THE1/HERK1, consultamos a partir de “GOSTats” los términos de la lista de 56 genes identificados con anterioridad (5.7.3). La salida de esta consulta se puede encontrar en la Tabla 17.

Tabla 17 Reporte GOSTats para la respuesta consenso mediada por THE1/HERK1. ID de los términos GO (GOMFID); los p-value asociado al enriquecimiento del mismo en la selección de genes (Pvalue); la cantidad de veces que se encontró el término (Count); la cantidad esperada para cada término (ExpCount); la “razón de exceso” o “razón de probabilidades” (OddsRatio), parámetro que indica cuan fuertemente se presenta el termino en relación a lo esperado; y el número de genes asociado a este término en la base de datos utilizada (Size)

GOMFID	P-value	OddsRatio	ExpCount	Count	Size	Término
GO:0004565	0,001	50,9	0,05	2	17	β-galactosidasa
GO:0015925	0,002	38,1	0,06	2	22	galactosidasa
GO:0005222	0,003	Inf	0	1	1	canal de cationes intracelular activado cAMP
GO:0009976	0,003	Inf	0	1	1	tocoferol ciclasa

GO:0047312	0,003	Inf	0	1	1	L-fenilalanina:piruvato aminotransferasa
GO:0047668	0,003	Inf	0	1	1	amigdalina β -glucosidasa
GO:0050048	0,003	Inf	0	1	1	L-leucina:2-oxoglutarato aminotransferasa
GO:0070546	0,003	Inf	0	1	1	L-fenilalanina aminotransferasa
GO:0080081	0,003	Inf	0	1	1	4-metilumbelliferil- β -D-glucopiranosido β -glucosidasa
GO:0080082	0,003	Inf	0	1	1	esculin β -glucosidasa
GO:0080083	0,003	Inf	0	1	1	β -gentiobiosa β -glucosidasa
GO:0080098	0,003	Inf	0	1	1	L-tirosina:piruvato aminotransferasa
GO:0080099	0,003	Inf	0	1	1	L-metionina:2-oxoglutarato aminotransferasa
GO:0080100	0,003	Inf	0	1	1	L-glutamina:2-oxoglutarato aminotransferasa
GO:0080130	0,003	Inf	0	1	1	L-phenylalanina:2-oxoglutarato aminotransferasa
GO:0004567	0,005	372,8	0,01	1	2	β -manosidasa
GO:0070548	0,005	372,8	0,01	1	2	L-glutamina aminotransferasa
GO:0080079	0,005	372,8	0,01	1	2	celobiosa glucosidasa
GO:0004021	0,008	186,4	0,01	1	3	L-alanina:2-oxoglutarato aminotransferasa
GO:0010326	0,008	186,4	0,01	1	3	metionina-oxo-acido transaminasa
GO:0016420	0,008	186,4	0,01	1	3	maloniltransferasa
GO:0019904	0,008	186,4	0,01	1	3	unión a dominio proteico
GO:0030170	0,008	186,4	0,01	1	3	unión a piridoxal fosfato
GO:0033843	0,008	186,4	0,01	1	3	xiloglucano 6-xilosiltransferasa
GO:0047635	0,008	186,4	0,01	1	3	alanina-oxo-acido transaminasa
GO:0050362	0,008	186,4	0,01	1	3	L-triptofano:2-oxoglutarato aminotransferasa
GO:0050736	0,008	186,4	0,01	1	3	O-maloniltransferasa
GO:0070279	0,008	186,4	0,01	1	3	unión a vitamina B6
GO:0070529	0,008	186,4	0,01	1	3	L-triptófano aminotransferasa
GO:0080097	0,008	186,4	0,01	1	3	L-triptófano:piruvato aminotransferasa
GO:0015079	0,008	15,9	0,14	2	50	transportador de ion potasio
GO:0004838	0,011	124,3	0,01	1	4	L-tirosina:2-oxoglutarato aminotransferasa
GO:0005221	0,011	124,3	0,01	1	4	canal de catión intracelular activado por nucleótido cíclico
GO:0043855	0,011	124,3	0,01	1	4	canal de catión activado por nucleótido cíclico
GO:0070547	0,011	124,3	0,01	1	4	L-tirosina aminotransferasa
GO:0019842	0,014	93,2	0,01	1	5	unión a vitamina
GO:0016798	0,016	4,5	0,97	4	355	glucosil hidrolasa

GO:0035252	0,016	74,5	0,02	1	6	UDP-xilosiltransferasa
GO:0002020	0,019	62,1	0,02	1	7	unión a proteasa
GO:0005262	0,022	53,2	0,02	1	8	canal activado por calcio
GO:0005242	0,024	46,6	0,02	1	9	canal de potasio rectificador hacia adentro
GO:0008422	0,024	46,6	0,02	1	9	β -glucosidasa
GO:0099094	0,024	46,6	0,02	1	9	canal de catión activado por ligando
GO:0009975	0,027	41,4	0,03	1	10	ciclasa
GO:0015923	0,03	37,3	0,03	1	11	manosidasa
GO:0042285	0,032	33,9	0,03	1	12	xilosiltransferasa
GO:0015926	0,046	23,3	0,05	1	17	glucosidasa
GO:0004553	0,049	3,9	0,83	3	302	o-glucosil hidrolasa

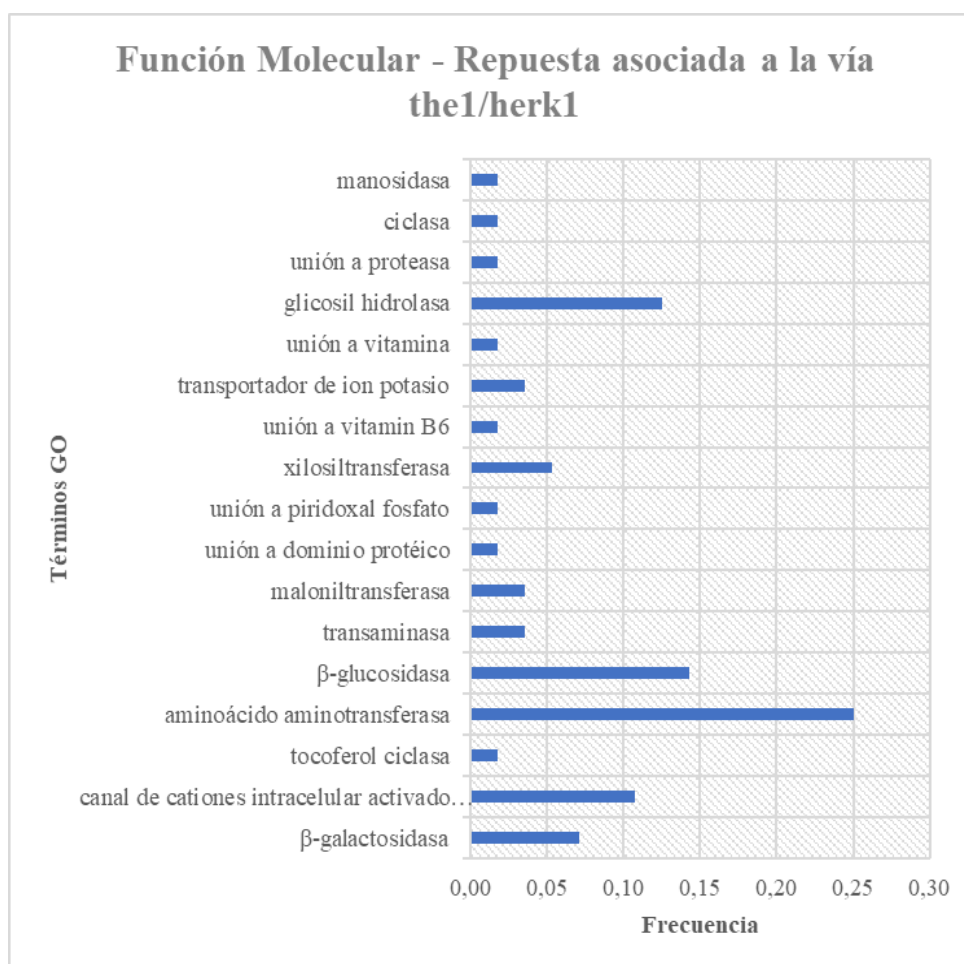


Ilustración 21 Gráfico de frecuencias de términos GO asociados a la respuesta mediada por la vía THE1/HERK1.

En este gráfico Ilustración 22, se puede encontrar un solo agrupamiento, relacionado a las siguientes respuestas: Fototropismo, Metabolismo de la pared celular, y respuesta a compuestos orgánicos cíclicos.

A partir de los datos de transcriptómica obtenido en mutantes de la vía THE1/HERK y la expresión diferencial obtenida en plantas mutantes en *THE1* tratadas con BL, podemos identificar 56 genes que estarían involucrados en una respuesta consenso mediada por esta vía. El análisis de GO y la red funcional construida, nos indican que la mayoría de estos genes se encuentran involucrados en la remodelación de la pared celular, en fototropismo y en la respuesta a compuestos orgánicos cíclicos. De estos 56 genes, solo 2 se comparten con los genes consenso en respuesta a BR. El primero es un factor de transporte nuclear (AT5G01740) que se encuentra regulado por la acción del FT BZR1. El segundo gen codifica para una proteína hipotética de respuesta a karrikina (AT5G57785). Las karrikinas son un compuesto activo del humo proveniente de la quema de material vegetal que se cree que estimula la germinación de semillas. Se sugiere que esto puede ser una adaptación evolutiva en respuesta a los incendios forestales (Halford 2010).

Debido a que 54 genes regulados por la vía THE1/HERK1 no se comparten con los genes consenso de respuesta a BR, sugerimos que los mismos pertenecen a una vía independiente de los BR, a diferencia de lo sugerido en diversos trabajos (Guo, Li et al. 2009, Guo, Ye et al. 2009) especialmente porque en estos trabajos se observan genes de respuesta en común y que la aplicación de BL induce la expresión tanto de THE1 como de HERK1. Sin embargo, es necesaria más evidencia en el caso que se requiera probar la conexión de estas vías. Por ejemplo, es necesario el estudio bioquímico de estos dos genes identificadas.

Teniendo en cuenta nuestros resultados, uno puede proponer que la respuesta a la acción de los brasinoesteroides puede dividirse en una etapa temprana y en una etapa tardía. La etapa temprana se centra en la regulación de FT y enzimas involucradas en la retroalimentación positiva de la señal, mientras que la etapa tardía incluye la remodelación de la pared celular asociada al crecimiento por expansión celular. La respuesta por la vía THE1/HERK1 puede gatillarse posteriormente, sensando la desestructuración de la membrana (la cual puede deberse a crecimiento o infección), gatillando una respuesta mayormente centrada a la remodelación de la pared celular. Esto concuerda con los datos experimentales obtenidos en mi tesis doctoral (Grisolia, Peralta et al. 2016), en la que la alteración de la estructura de la pared, gatilla la expresión de THE1 y de varios de sus genes bajo su control.

7 CONCLUSIONES

Se han utilizado y analizado una variedad de estudios de transcriptómica, y se ha podido procesar los resultados obtenidos mediante herramientas bioinformáticas diversas. No solo se incluyeron herramientas de análisis basadas en R y Bioconductor, sino que se incorporaron herramientas con GUI y sistemas de análisis integrado como Galaxy. Adicionalmente, se implementaron herramientas nativas de R para contestar consultas asociadas a preguntas biológicas definidas. A partir de los resultados encontrados, se ha podido sugerir un modelo de respuesta que ayuden a entender la respuesta mediada por Brasinoesteroides y el fenotipo de las plantas con pared celular alterada generada en un trabajo previo en nuestro laboratorio.

8 BIBLIOGRAFÍA:

- Afgan, E., D. Baker, M. van den Beek, D. Blankenberg, D. Bouvier, M. Cech, J. Chilton, D. Clements, N. Coraor, C. Eberhard, B. Gruning, A. Guerler, J. Hillman-Jackson, G. Von Kuster, E. Rasche, N. Soranzo, N. Turaga, J. Taylor, A. Nekrutenko and J. Goecks (2016). "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update." Nucleic Acids Res **44**(W1): W3-W10.
- Barrett, T., S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C. L. Robertson, N. Serova, S. Davis and A. Soboleva (2013). "NCBI GEO: archive for functional genomics data sets—update." Nucleic Acids Res **41**(Database issue): D991-995.
- Bolstad, B. M., F. Collin, J. Brettschneider, K. Simpson, L. Cope, R. A. Irizarry and T. P. Speed (2005). Quality Assessment of Affymetrix GeneChip Data. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. R. Gentleman, V. J. Carey, W. Huber, R. A. Irizarry and S. Dudoit. New York, NY, Springer New York: 33-47.
- Cline, M. S., M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, K. Hanspers, R. Isserlin, R. Kelley, S. Killcoyne, S. Lotia, S. Maere, J. Morris, K. Ono, V. Pavlovic, A. R. Pico, A. Vailaya, P.-L. Wang, A. Adler, B. R. Conklin, L. Hood, M. Kuiper, C. Sander, I. Schmulevich, B. Schwikowski, G. J. Warner, T. Ideker and G. D. Bader (2007). "Integration of biological networks and gene expression data using Cytoscape." Nat. Protocols **2**(10): 2366-2382.
- Clouse, S. D. and J. M. Sasse (1998). "BRASSINOSTEROIDS: Essential Regulators of Plant Growth and Development." Annu Rev Plant Physiol Plant Mol Biol **49**: 427-451.
- Cosgrove, D. J. (1993). "How Do Plant Cell Walls Extend?" How Do Plant Cell Walls Extend?
- Davis, S. and P. S. Meltzer (2007). "GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor." Bioinformatics **23**(14): 1846-1847.
- Donnelly, P. M., D. Bonetta, H. Tsukaya, R. E. Dengler and N. G. Dengler (1999). "Cell cycling and cell enlargement in developing leaves of Arabidopsis." Dev Biol **215**(2): 407-419.
- Falcon, S. and R. Gentleman (2007). "Using GStats to test gene lists for GO term association." Bioinformatics **23**(2): 257-258.
- Gautier, L., L. Cope, B. M. Bolstad and R. A. Irizarry (2004). "affy—analysis of Affymetrix GeneChip data at the probe level." Bioinformatics **20**(3): 307-315.
- Goda, H., E. Sasaki, K. Akiyama, A. Maruyama-Nakashita, K. Nakabayashi, W. Li, M. Ogawa, Y. Yamauchi, J. Preston, K. Aoki, T. Kiba, S. Takatsuto, S. Fujioka, T. Asami, T. Nakano, H. Kato, T. Mizuno, H. Sakakibara, S. Yamaguchi, E. Nambara, Y. Kamiya, H. Takahashi, M. Y. Hirai, T. Sakurai, K. Shinozaki, K. Saito, S. Yoshida and Y. Shimada (2008). "The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access." Plant J **55**(3): 526-542.

- Goda, H., S. Sawa, T. Asami, S. Fujioka, Y. Shimada and S. Yoshida (2004). "Comprehensive comparison of auxin-regulated and brassinosteroid-regulated genes in Arabidopsis." Plant Physiol **134**(4): 1555-1573.
- Grisolia, M. J., D. A. Peralta, H. A. Valdez, J. Barchiesi, D. F. Gomez-Casati and M. V. Busi (2016). "The targeting of starch binding domains from starch synthase III to the cell wall alters cell wall composition and properties." Plant Mol Biol **93**(1-2): 121-135.
- Guo, H., L. Li, H. Ye, X. Yu, A. Algreen and Y. Yin (2009). "Three related receptor-like kinases are required for optimal cell elongation in Arabidopsis thaliana." Proc Natl Acad Sci U S A **106**(18): 7648-7653.
- Guo, H., H. Ye, L. Li and Y. Yin (2009). "A family of receptor-like kinases are regulated by BES1 and involved in plant growth in Arabidopsis thaliana." Plant Signal Behav. **4**(8): 784-786.
- Halford, B. (2010). "Smoke Signals." Chemical & Engineering News **88**(15): 37-38.
- Hematy, K., P. E. Sado, T. A. Van, S. Rochange, T. Desnos, S. Balzergue, S. Pelletier, J. P. Renou and H. Hofte (2007). "A receptor-like kinase mediates the response of Arabidopsis cells to the inhibition of cellulose synthesis." Curr.Biol. **17**(11): 922-931.
- Ji, H., H. Jiang, W. Ma, D. S. Johnson, R. M. Myers and W. H. Wong (2008). "An integrated software system for analyzing ChIP-chip and ChIP-seq data." Nat Biotechnol **26**(11): 1293-1300.
- Ji, H., S. A. Vokes and W. H. Wong (2006). "A comparative analysis of genome-wide chromatin immunoprecipitation data for mammalian transcription factors." Nucleic Acids Research **34**(21): e146-e146.
- Ji, H. and W. H. Wong (2005). "TileMap: create chromosomal map of tiling array hybridizations." Bioinformatics **21**(18): 3629-3636.
- Katsumata, T., A. Hasegawa, T. Fujiwara, T. Komatsu, M. Notomi, H. Abe, M. Natsume and H. Kawaide (2008). "Arabidopsis CYP85A2 catalyzes lactonization reactions in the biosynthesis of 2-deoxy-7-oxalactone brassinosteroids." Biosci Biotechnol Biochem **72**(8): 2110-2117.
- Kauffmann, A., R. Gentleman and W. Huber (2008). "arrayQualityMetrics—a bioconductor package for quality assessment of microarray data." Bioinformatics **25**(3): 415-416.
- Kim, T. W., S. Guan, Y. Sun, Z. Deng, W. Tang, J. X. Shang, Y. Sun, A. L. Burlingame and Z. Y. Wang (2009). "Brassinosteroid signal transduction from cell-surface receptor kinases to nuclear transcription factors." Nat Cell Biol **11**(10): 1254-1260.
- Li, J., P. Nagpal, V. Vitart, T. C. McMorris and J. Chory (1996). "A role for brassinosteroids in light-dependent development of Arabidopsis." Science **272**(5260): 398-401.
- Marga, F., M. Grandbois, D. J. Cosgrove and T. I. Baskin (2005). "Cell wall extension results in the coordinate separation of parallel microfibrils: evidence from scanning electron microscopy and atomic force microscopy." Plant J **43**(2): 181-190.

- Mussig, C., S. Fischer and T. Altmann (2002). "Brassinosteroid-regulated gene expression." Plant Physiol **129**(3): 1241-1251.
- Nemhauser, J. L., T. C. Mockler and J. Chory (2004). "Interdependency of brassinosteroid and auxin signaling in Arabidopsis." PLoS Biol **2**(9): E258.
- Oh, E., J. Y. Zhu and Z. Y. Wang (2012). "Interaction between BZR1 and PIF4 integrates brassinosteroid and environmental responses." Nat Cell Biol **14**(8): 802-809.
- Pfeiffer, A., H. Shi, J. M. Tepperman, Y. Zhang and P. H. Quail (2014). "Combinatorial complexity in a transcriptionally centered signaling hub in Arabidopsis." Mol Plant **7**(11): 1598-1618.
- Reiter, W.-D. (2002). "Biosynthesis and properties of the plant cell wall." Current Opinion in Plant Biology **5**(6): 536-542.
- Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi and G. K. Smyth (2015). "limma powers differential expression analyses for RNA-sequencing and microarray studies." Nucleic Acids Res **43**(7): e47.
- Schroder, F., J. Lisso, T. Obata, A. Erban, E. Maximova, P. Giavalisco, J. Kopka, A. R. Fernie, L. Willmitzer and C. Mussig (2014). "Consequences of induced brassinosteroid deficiency in Arabidopsis leaves." BMC Plant Biol **14**(1): 309.
- Singh, A. P., Y. Fridman, L. Friedlander-Shani, D. Tarkowska, M. Strnad and S. Savaldi-Goldstein (2014). "Activity of the brassinosteroid transcription factors BRASSINAZOLE RESISTANT1 and BRASSINOSTEROID INSENSITIVE1-ETHYL METHANESULFONATE-SUPPRESSOR1/BRASSINAZOLE RESISTANT2 blocks developmental reprogramming in response to low phosphate availability." Plant Physiol **166**(2): 678-688.
- Song, L., X. Y. Zhou, L. Li, L. J. Xue, X. Yang and H. W. Xue (2009). "Genome-wide analysis revealed the complex regulatory network of brassinosteroid effects in photomorphogenesis." Mol Plant **2**(4): 755-772.
- Sun, Y., X. Y. Fan, D. M. Cao, W. Tang, K. He, J. Y. Zhu, J. X. He, M. Y. Bai, S. Zhu, E. Oh, S. Patil, T. W. Kim, H. Ji, W. H. Wong, S. Y. Rhee and Z. Y. Wang (2010). "Integration of brassinosteroid signal transduction with the transcription network for plant growth regulation in Arabidopsis." Dev Cell **19**(5): 765-777.
- Szekeres, M., K. Nemeth, Z. Koncz-Kalman, J. Mathur, A. Kauschmann, T. Altmann, G. P. Redei, F. Nagy, J. Schell and C. Koncz (1996). "Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450, controlling cell elongation and de-etiolation in Arabidopsis." Cell **85**(2): 171-182.
- Toedling, J., O. Sklyar and W. Huber (2007). "Ringo—an R/Bioconductor package for analyzing ChIP-chip readouts." BMC bioinformatics **8**(1): 221.
- van Mourik, H., van Dijk, A. D., Stortenbeker, N., Angenent, G. C., & Bemer, M. (2017). Divergent regulation of Arabidopsis SAUR genes: a focus on the SAUR10-clade. BMC plant biology, **17**(1), 245.

Williamson, R., J. E. Burn, R. Birch, T. I. Baskin, T. Arioli, A. Betzner and A. Cork (2001). "Morphology of *frsw1*, a cellulose-deficient mutant of *Arabidopsis thaliana*." Protoplasma **215**(1): 116-127.

Yu, X., L. Li, J. Zola, M. Aluru, H. Ye, A. Foudree, H. Guo, S. Anderson, S. Aluru, P. Liu, S. Rodermel and Y. Yin (2011). "A brassinosteroid transcriptional network revealed by genome-wide identification of BES1 target genes in *Arabidopsis thaliana*." Plant J **65**(4): 634-646.

Zacher, B., P. F. Kuan and A. Tresch (2010). "Starr: simple tiling ARRAY analysis of affymetrix ChIP-chip data." BMC bioinformatics **11**(1): 194.

Zou, L. J., X. G. Deng, X. Y. Han, W. R. Tan, L. J. Zhu, D. H. Xi, D. W. Zhang and H. H. Lin (2016). "Role of Transcription Factor HAT1 in Modulating *Arabidopsis thaliana* Response to Cucumber mosaic virus." Plant Cell Physiol **57**(9): 1879-1889.