

# banuls\_marc\_ADO\_PEC1

Marc Bañuls Tornero

17/4/2020

## Contents

<b>Abstract</b>	<b>1</b>
<b>Objetivos</b>	<b>1</b>
<b>Materiales y Métodos</b>	<b>2</b>
Obtención de datos previos al análisis y materiales . . . . .	2
Métodos . . . . .	2
Comparaciones múltiples . . . . .	4
<b>Resultados</b>	<b>4</b>
Control de calidad . . . . .	4
Genes diferencialmente expresados . . . . .	8
Análisis de significación biológica . . . . .	11
Resumen de archivos obtenidos . . . . .	14
<b>Discusión</b>	<b>15</b>
<b>Conclusión</b>	<b>15</b>
<b>Apéndice</b>	<b>15</b>
<b>References</b>	<b>15</b>

El proyecto reproducible de este análisis se encuentra en un repositorio github. Para acceder a él se puede seguir este enlace: [https://github.com/MarcBanuls/banuls\\_marc\\_PEC1\\_OMICAS.git](https://github.com/MarcBanuls/banuls_marc_PEC1_OMICAS.git)

## Abstract

El factor de crecimiento transformante beta (TGF- $\beta$ ) es una citocina involucrada en varios procesos celulares relacionados principalmente con el desarrollo del organismo y la embriogénesis. Esta proteína (entre otras funciones) promueve la producción de fibras en el hígado, pero recientemente se ha descubierto que la proteína de membrana Caveolina-1 (CAV1) bloquea la señalización de las TGF- $\beta$  y promueve su apoptosis. Por ello, se quiere observar en qué genes y rutas metabólicas afecta concretamente CAV1 a la proteína TGF- $\beta$  para producir este efecto. Gracias al análisis de los arrays se han obtenido varias rutas metabólicas afectadas por el silenciamiento de CAV1 en un medio con TGF- $\beta$ .

## Objetivos

En esta investigación se pretende estudiar de manera concreta el efecto que tiene la presencia o ausencia de la proteína de membrana CAV1 en las funciones de la citocina TGF- $\beta$ . Para analizar esto, se realizan análisis de

microarrays procedentes del estudio que posteriormente se analizarán para comprobar qué rutas metabólicas y genes se ven alterados en presencia o ausencia de CAV1 cuando son tratados (o no) con TGF-B.

## Materiales y Métodos

### Obtención de datos previos al análisis y materiales

Las muestras originales utilizadas para realizar microarrays son células hepatocíticas procedentes de ratones. Éstas células son transfectadas con siCon o siCAV1 (control e inhibición de la proteína CAV1 respectivamente). Después de obtener las células transfectadas deseadas, se tratan las distintas células transfectadas con una solución de TGF-beta durante 48 horas (o no, debido a no realizar el tratamiento para obtener muestras control). Finalmente se obtiene el RNA de las distintas muestras para realizar los distintos microarrays.

El diseño experimental de este estudio consiste en realizar el análisis de 12 muestras. Se comparan los niveles de expresión génica entre hepatocitos control y hepatocitos con el gen CAV1 silenciado cuando están tratadas durante 2 días con TGF-B o no se tratan con éstos. Se realizan triplicados para cada uno de las cuatro combinaciones de tratamientos.

Los microarrays utilizados en este estudio son los utilizados en el paper de Han et al. (2019) de Affymetrix para el organismo de *Mus musculus*. Concretamente el microarray se llama [MoGene-1\_0-st] Affymetrix Mouse Gene 1.0 ST Array [CDF: MoGene10stv1\_Mm\_ENTREZG\_16.0.0].

### Métodos

El pipeline que hemos seguido para obtener los resultados del análisis, junto con los pasos realizados en cada parte del pipeline es:

#### Preparación de los datos

Descargamos los archivos CEL del experimento y los introducimos en el directorio Data. Creamos además el archivo “targets” en formato csv para determinar en que grupo se encuentra cada muestra y añadir más información. Hemos introducido 5 columnas en el archivo “targets.csv”:

- Filename: Nombre de la muestra CEL completa
- Group: Resumen de las condiciones en las que se encuentra la muestra
- Genotype: Genotipo de la muestra
- Treatment: Tipo de tratamiento realizado
- ShortName: Nombre creado para identificar fácilmente cada muestra específicamente

Podríamos haber introducido también la edad y especie de cada muestra, pero al ser todas las muestras iguales en ese aspecto no lo hemos considerado necesario.

#### Lectura de los archivos CEL

Leemos los archivos CEL teniendo en cuenta que están comprimidos (formato zip) y sustituimos los nombres de las muestras por sus nombres atribuidos en “Shortnames”.

#### Control de calidad y normalización

Realizamos un control de calidad de las muestras CEL crudas leídas para obtener un resumen de la calidad de estas muestras sin normalizar. Para obtener los resultados del control de calidad utilizamos el paquete *arrayQualityMetrics* (Kauffmann, Gentleman, and Huber (2009)).

Posteriormente utilizamos el método de normalización mediante RMA (Robust Multichip Analysis) (Irizarry et al. (2003)), para así obtener las muestras normalizadas a partir de los datos crudos. Para ello utilizamos el paquete *affy* (Gautier et al. (2004))

Realizamos de nuevo un control de calidad con los datos ya normalizados, obteniendo un resumen de la calidad de las muestras normalizadas.

### Filtraje no específico

Realizamos un filtraje de los genes mediante el paquete *genefilter* (Gentleman et al. (2019)) que menos variación tengan o que no estén anotados adecuadamente a partir de las muestras normalizadas, para así facilitar los posteriores análisis. De esta manera obtenemos las muestras normalizadas y filtradas (eliminando IDs no anotados correctamente).

### Identificación de genes diferencialmente expresados

Para encontrar los genes diferencialmente expresados debemos observar los genes que presentan diferencias entre los distintos grupos. Para ello, podemos crear una gráfica de variabilidad de cada gen entre los distintos grupos utilizando la desviación estándar de estos genes como referencia de variabilidad. Para encontrar los genes más diferencialmente expresados, podemos filtrar por aquellos que tienen una desviación estándar mayor al 90-95% de las desviaciones estándar.

### Selección de genes diferencialmente expresados

Para realizar la selección de los genes que se encuentran diferencialmente expresados nos basamos en la utilización del paquete *limma* (Ritchie et al. (2015)) para construir un modelo lineal.

**Matriz de diseño** Primero hemos de construir una matriz de diseño. Como se sugiere en el manual de *limma*, interpretaremos los grupos del estudio como un modelo de un factor con 4 niveles: + CAV1.TGF (células con el gen CAV1 silenciado y tratadas durante 48 horas con TGF-B) + CAV1.CTRL (células con el gen CAV1 silenciado sin tratamiento) + WT.TGF (células control y tratadas durante 48 horas con TGF-B) + WT.CTRL (células control sin tratamiento)

##	CAV1.TGF	CAV1.CTRL	WT.TGF	WT.CTRL
## CAV1.TGF.1	1	0	0	0
## CAV1.TGF.2	1	0	0	0
## CAV1.TGF.3	1	0	0	0
## CAV1.CTRL.1	0	1	0	0
## CAV1.CTRL.2	0	1	0	0
## CAV1.CTRL.3	0	1	0	0
## WT.TGF.1	0	0	1	0
## WT.TGF.2	0	0	1	0
## WT.TGF.3	0	0	1	0
## WT.CTRL.1	0	0	0	1
## WT.CTRL.2	0	0	0	1
## WT.CTRL.3	0	0	0	1

**Contrastes** A partir de esta matriz de diseño ya podemos formular las comparaciones entre grupos convenientes. Para ello, utilizaremos matrices de contrastes. Las comparaciones que queremos hacer es: - ¿Hay diferencias entre las células con el gen CAV1 silenciado y las que no lo tienen cuando son tratadas con TGF-B? - ¿Hay diferencias entre las células con el gen CAV1 silenciado y las que no lo tienen cuando no son tratadas con TGF-B? - ¿Existe una relación entre el tipo de célula (con el gen CAV1 silenciado o no) y la realización (o no) del tratamiento en estas células?

Para ello, utilizaremos de nuevo el paquete *limma*.

##	Contrasts
## Levels	CAV1.TGFvsWT.TGF CAV1.CTRLvsWT.CTRL INT
## CAV1.TGF	1 0 1
## CAV1.CTRL	0 1 -1

##	WT.TGF	-1	0	-1
##	WT.CTRL	0	-1	1

De esta manera observaremos el efecto de tratar las células con TGF-B durante 48 horas o no realizar este tratamiento en las células que tienen o no silenciado el gen CAV1, y la interacción entre las células con el gen CAV1 silenciado y el haberla tratado o no.

**Estimación del modelo y selección de genes** Pasamos a estimar los modelos y contrastes junto con la realización de pruebas de significación, utilizando el método implementado en el paquete *limma*, el cual utiliza el análisis mediante Bayes empírico.

Ahora que tenemos el modelo de Bayes empírico en nuestro modelo, deseamos realizar tres comparaciones:

- Comparación 1: Genes con cambios en su expresión entre cepas cuando se encuentran en ayunas durante 6 horas:
- Comparación 2: Genes con cambios en su expresión entre cepas cuando no se encuentran en ayunas:
- Comparación 3: Genes con cambios en su expresión entre las comparaciones 1 y 2

## Anotación de los resultados

Teniendo todos los genes cuya expresión ha sido más significativa en los contrastes realizados, debemos identificarlos correctamente. Para ello, utilizamos la función creada y suministrada en el material del curso (Pla, Ricardo, and Gonzalo (2019)), y guardamos los genes con sus valores en nuevas tablas.

## Comparaciones múltiples

Para realizar comparaciones múltiples utilizamos también el paquete *limma*. De esta manera, podemos observar el número de genes que se han visto “down-regulated” o “up-regulated” (o sin diferencias significativas) en cada comparación.

## Análisis de significación biológica

con las comparaciones de genes podemos interpretar como afectan los genes con significación dentro del contexto biológico. Para ello, utilizamos el paquete ReactomePA de Bioconductor. En este análisis reslizamos una selección poco restrictiva de los genes diferencialmente expresados para poder tener una cantidad de genes aceptable para analizar. Para ello preparamos la lista de genes para analizar y utilizaremos como universo todos los genes que tengan almenos alguna anotación en Gene Ontology. Para realizar este proceso utilizamos el paquete *AnnotationDbi* (Pagès et al. (2019)).

Utilizando el paquete *reactomePA* (Yu and He (2016)) podemos realizar un análisis de enriquecimiento utilizando el universo y la selección de genes diferencialmente expresados anteriormente. Con este análisis obtenemos un esquema de las rutas metabólicas y gráficos de barras que contienen los genes diferencialmente expresados.

# Resultados

## Control de calidad

### Tablas resumen del análisis de outliers

Durante el control de calidad de los datos crudos hemos observado que existen varios arrays que presentan outliers pero sólo en uno de los tres métodos de análisis utilizados por el paquete. Concretamente los outliers aparecen en los gráficos “MA”. Al no corroborarse la detección de outliers en los distintos arrays por más de un método, podemos proseguir con la normalización.

	array	sampleNames	*1	*2	*3	Group	Genotype	Treatment	ShortName
<input type="checkbox"/>	1	CAV1.TGF.1				TGF	CAV1	48-Hours with TGF-B	CAV1.TGF.1
<input type="checkbox"/>	2	CAV1.TGF.2				TGF	CAV1	48-Hours with TGF-B	CAV1.TGF.2
<input type="checkbox"/>	3	CAV1.TGF.3			x	TGF	CAV1	48-Hours with TGF-B	CAV1.TGF.3
<input type="checkbox"/>	4	CAV1.CTRL.1				CTRL	CAV1	No treatment	CAV1.CTRL.1
<input type="checkbox"/>	5	CAV1.CTRL.2				CTRL	CAV1	No treatment	CAV1.CTRL.2
<input type="checkbox"/>	6	CAV1.CTRL.3			x	CTRL	CAV1	No treatment	CAV1.CTRL.3
<input type="checkbox"/>	7	WT.TGF.1				TGF	WT	48-Hours with TGF-B	WT.TGF.1
<input type="checkbox"/>	8	WT.TGF.2			x	TGF	WT	48-Hours with TGF-B	WT.TGF.2
<input type="checkbox"/>	9	WT.TGF.3				TGF	WT	48-Hours with TGF-B	WT.TGF.3
<input type="checkbox"/>	10	WT.CTRL.1			x	CTRL	WT	No treatment	WT.CTRL.1
<input type="checkbox"/>	11	WT.CTRL.2			x	CTRL	WT	No treatment	WT.CTRL.2
<input type="checkbox"/>	12	WT.CTRL.3				CTRL	WT	No treatment	WT.CTRL.3

Figure 1: Detección de outliers en los microarrays del estudio con datos crudos

En el control de calidad de los datos normalizados no se observa ningún tipo de outlier dentro de los métodos de detección de estos, confirmando que los posibles outliers del array con los datos crudos eran debidos a una falta de normalización.

	array	sampleNames	*1	*2	*3	Group	Genotype	Treatment	ShortName
<input type="checkbox"/>	1	CAV1.TGF.1				TGF	CAV1	48-Hours with TGF-B	CAV1.TGF.1
<input type="checkbox"/>	2	CAV1.TGF.2				TGF	CAV1	48-Hours with TGF-B	CAV1.TGF.2
<input type="checkbox"/>	3	CAV1.TGF.3				TGF	CAV1	48-Hours with TGF-B	CAV1.TGF.3
<input type="checkbox"/>	4	CAV1.CTRL.1				CTRL	CAV1	No treatment	CAV1.CTRL.1
<input type="checkbox"/>	5	CAV1.CTRL.2				CTRL	CAV1	No treatment	CAV1.CTRL.2
<input type="checkbox"/>	6	CAV1.CTRL.3				CTRL	CAV1	No treatment	CAV1.CTRL.3
<input type="checkbox"/>	7	WT.TGF.1				TGF	WT	48-Hours with TGF-B	WT.TGF.1
<input type="checkbox"/>	8	WT.TGF.2				TGF	WT	48-Hours with TGF-B	WT.TGF.2
<input type="checkbox"/>	9	WT.TGF.3				TGF	WT	48-Hours with TGF-B	WT.TGF.3
<input type="checkbox"/>	10	WT.CTRL.1				CTRL	WT	No treatment	WT.CTRL.1
<input type="checkbox"/>	11	WT.CTRL.2				CTRL	WT	No treatment	WT.CTRL.2
<input type="checkbox"/>	12	WT.CTRL.3				CTRL	WT	No treatment	WT.CTRL.3

Figure 2: Detección de outliers en los microarrays del estudio con datos normalizados

## Gráficos PCA

También podemos observar en los gráficos PCA generados en cada análisis de detección de outliers de los datos crudos y normalizados:

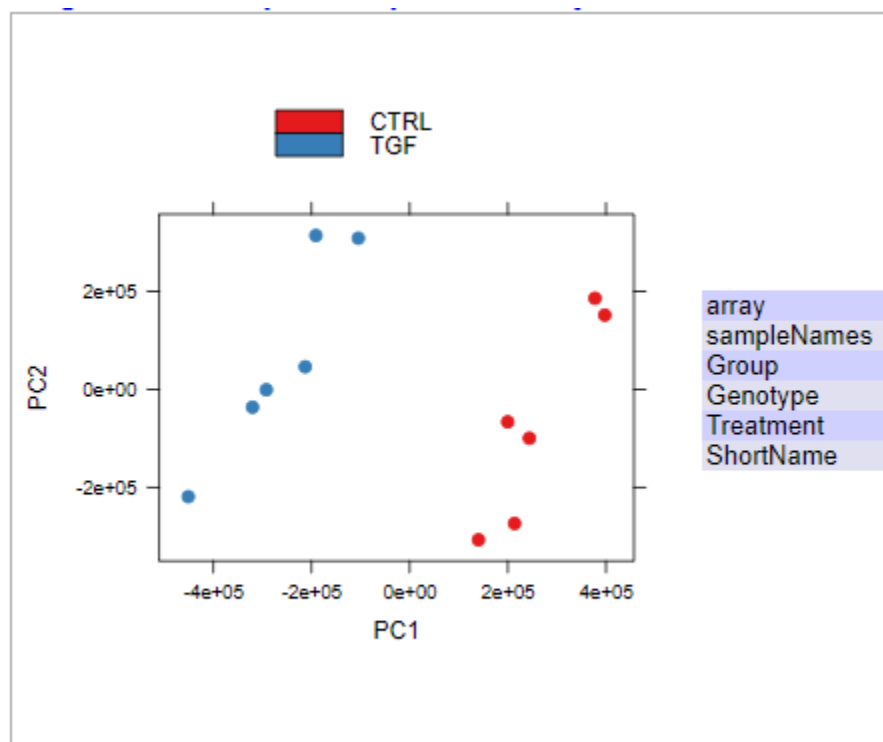


Figure 3: Detección de outliers mediante PCA

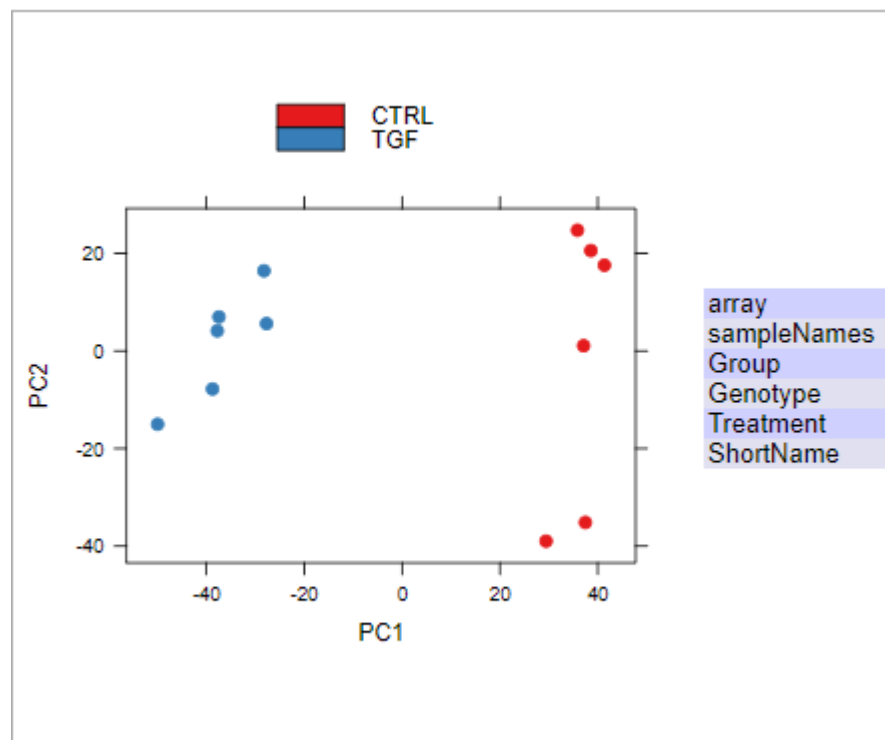


Figure 4: Detección de outliers mediante PCA

En ambos gráficos PCA no observamos ninguna anomalía que nos indique alguna variación de outliers.

### Boxplots

Por último observamos mediante gráficos de boxplot en ambos datos (crudos y normalizados) la presencia de outliers:

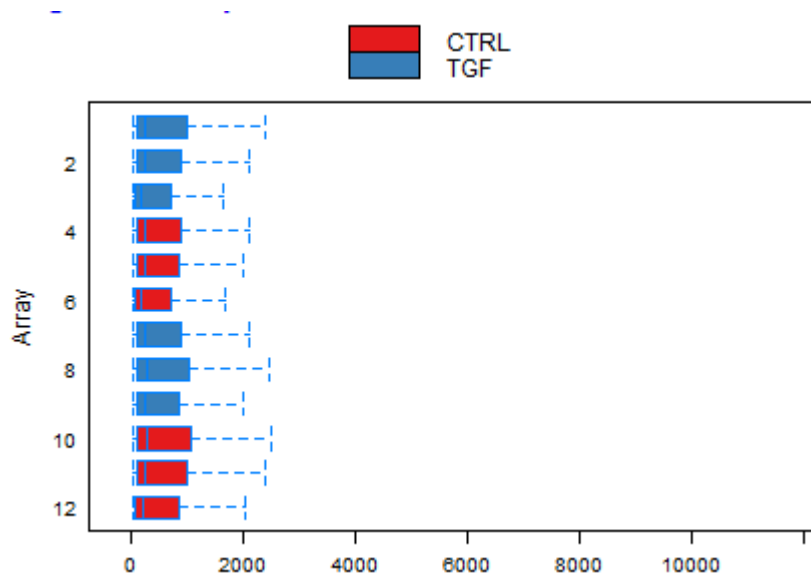


Figure 5: Detección de outliers en datos crudos mediante boxplots

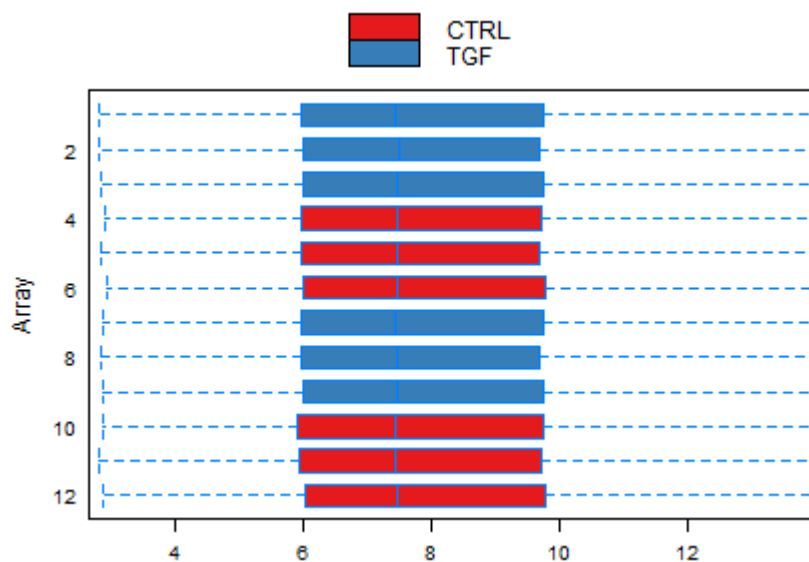


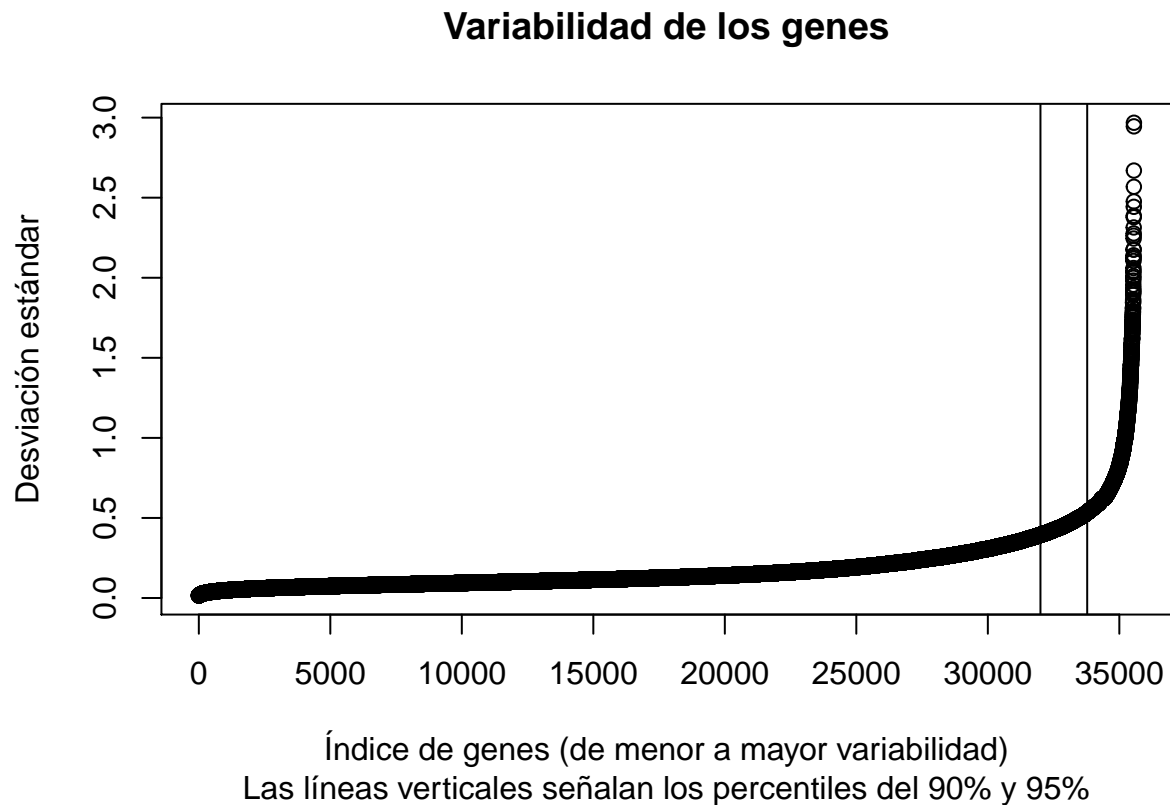
Figure 6: Detección de outliers en datos normalizados mediante boxplots

Tampoco observamos ningún outlier significativo en los gráficos de cajas, tanto para los datos crudos como para los normalizados, por lo que podemos proseguir con seguridad con el análisis del estudio.

## Genes diferencialmente expresados

### Genes con mayor variabilidad

En este gráfico podemos hacernos una idea de los genes con mayor variabilidad:



Observamos que de los 35000 genes del estudio, tan solo unos 2000 tienen una desviación estándar entre los percentiles al 90-95%.

### Gráficos de volcán

Después de normalizar los datos pasamos y encontrar los genes diferencialmente expresados pasamos a observar los genes diferencialmente expresados en cada comparación. Podemos realizar un gráfico de volcán para observar qué genes son los que están más (o menos) diferencialmente expresados:



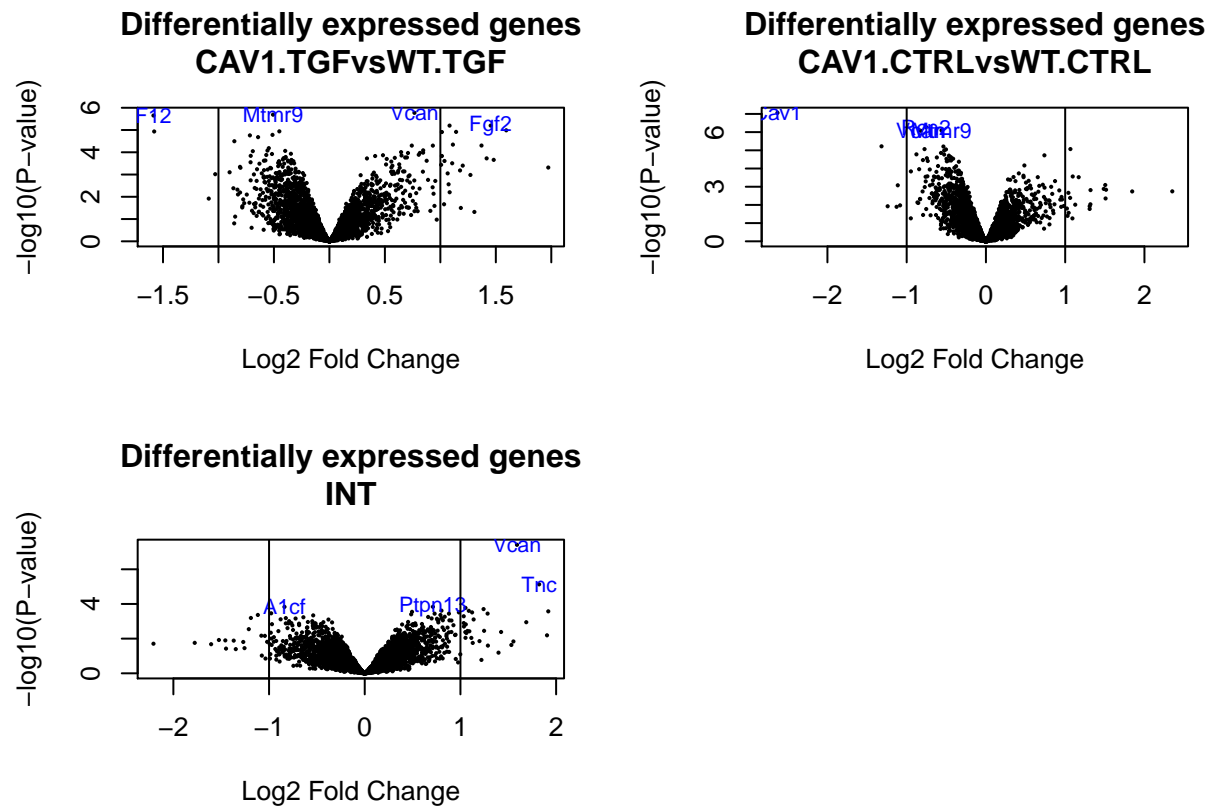
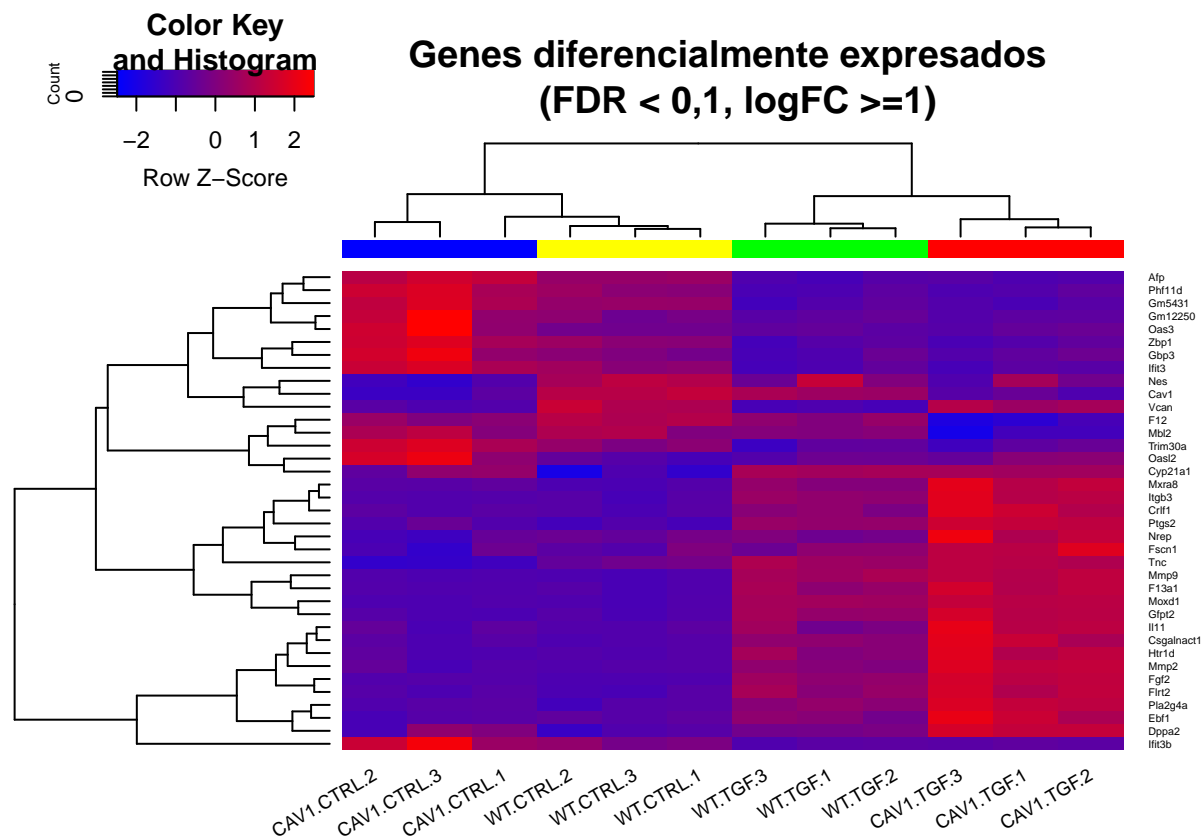


Figure 7: Detección de genes diferencialmente expresados en las distintas comparaciones

### Mapa de calor

Ahora realizamos un mapa de calor para observar qué genes son los más o menos expresados y su comparación con los otros grupos (escogemos los genes que han estado diferencialmente expresados en al menos una comparación):



De esta manera observamos una elevada cantidad de genes que están inhibidos en las muestras no tratadas y expresados en las muestras tratadas (donde se encuentra una mayor diferenciación en las células con el gen CAV1 inhibido) gracias a la observación de la expresión positiva o negativa mediante código de colores (azul para expresión baja y rojo para sobre-expresión).

### Diagrama de Venn

Construimos una matriz con los genes significativos entre las comparaciones para observar la cantidad de genes sobre-expresados o inhibidos en cada comparación.

##	CAV1.TGFvsWT.TGF	CAV1.CTRLvsWT.CTRL	INT
## Down	3	3	0
## NotSig	15	22	35
## Up	19	12	2

Con una significancia del 1%, observamos que en la comparación entre células con el gen CAV1 silenciado o no con tratamiento de TGF-B, hay 19 genes sobre-expresados y 3 genes inhibidos. en la comparación entre células con el gen CAV1 silenciado o no sin tratamiento, hay 12 genes sobre-expresados y 3 genes inhibidos. Finalmente en la comparación entre estas dos comparaciones tan solo observamos la sobre-expresión de 2 genes.

Además, se puede observar de manera gráfica mediante un diagrama de Venn el número de genes en común diferencialmente expresados entre las distintas comparaciones.

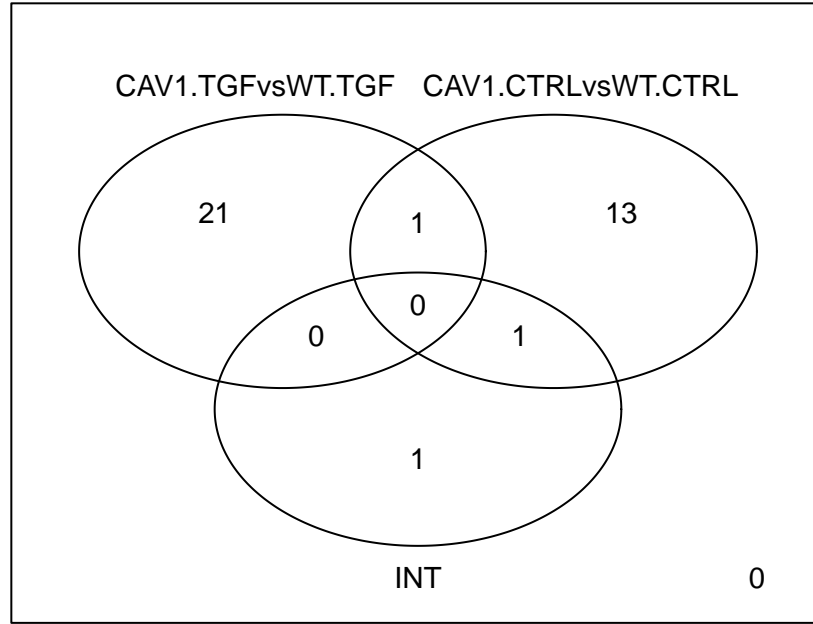


Figure 8: Diagrama de Venn con las relaciones de genes diferencialmente expresados entre comparaciones

### Análisis de significación biológica

Finalmente, para obtener resultados de significación biológica a partir de los arrays analizados, hemos realizado un análisis de enriquecimiento con un nivel de significación del 5%. Como la comparación entre las dos primeras comparaciones (llamada “INT”) no tiene valores por encima de este nivel de significación, no podemos realizar un análisis de enriquecimiento para ésta. Por ello, mostramos las tablas con las rutas metabólicas en las dos primeras comparaciones, mostrando sólo las cinco primeras rutas metabólicas significativas:

Table 1: Rutas biológicas significativamente expresadas en CAV1.CTRLvsWT.CTRL

	Description	GeneRatio	BgRatio	pvalue	p.adjust
R-MMU-453279	Mitotic G1-G1/S phases	13/255	126/8772	7.14323534787607e-05	0.0463595974077157
NA	NA	NA	NA	NA	NA
NA.1	NA	NA	NA	NA	NA
NA.2	NA	NA	NA	NA	NA

Table 2: Rutas biológicas significativamente expresadas en CAV1.TGFvsWT.TGF

	Description	GeneRatio	BgRatio	pvalue	p.adjust
R-MMU-2142753	Arachidonic acid metabolism	8/249	57/8772	0.00018678486627873	0.046696679721
R-MMU-8957322	Metabolism of steroids	12/249	125/8772	0.000215376662369126	0.046696679721
R-MMU-975634	Retinoid metabolism and transport	7/249	46/8772	0.000283566978918989	0.046696679721
R-MMU-6806667	Metabolism of fat-soluble vitamins	7/249	50/8772	0.000480499222650633	0.046696679721

Como se puede observar, la comparación CAV1.CTRLvsWT.CTRL tan solo tiene una ruta metabólica significativa, indicando que en principio la inhibición del gen CAV1 no parece tener elevadas consecuencias cuando no hay presencia de TGF-B, concretamente afectando únicamente a la ruta metabólica de la mitosis celular en las fases G1 y G1-S. Al contrario que en esta comparación, la comparación CAV1.TGFvsWT.TGF sí que indica varias rutas metabólicas afectadas. Esto implica que hay diferencias significativas en la expresión de varias rutas metabólicas entre las células con el gen CAV1 silenciado y las células control cuando ambas son tratadas con TGF-B. Como la comparación entre la interacción entre ambas comparaciones sólo tenía un gen con una significación al 1% se entiende que no hay rutas metabólicas afectadas entre las comparaciones mencionadas.

También podemos mostrar en un gráfico de redes los genes significativos que están implicados en cada ruta metabólica. Primero observamos la ruta metabólica de la comparación CAV1.CTRLvsWT.CTRL:

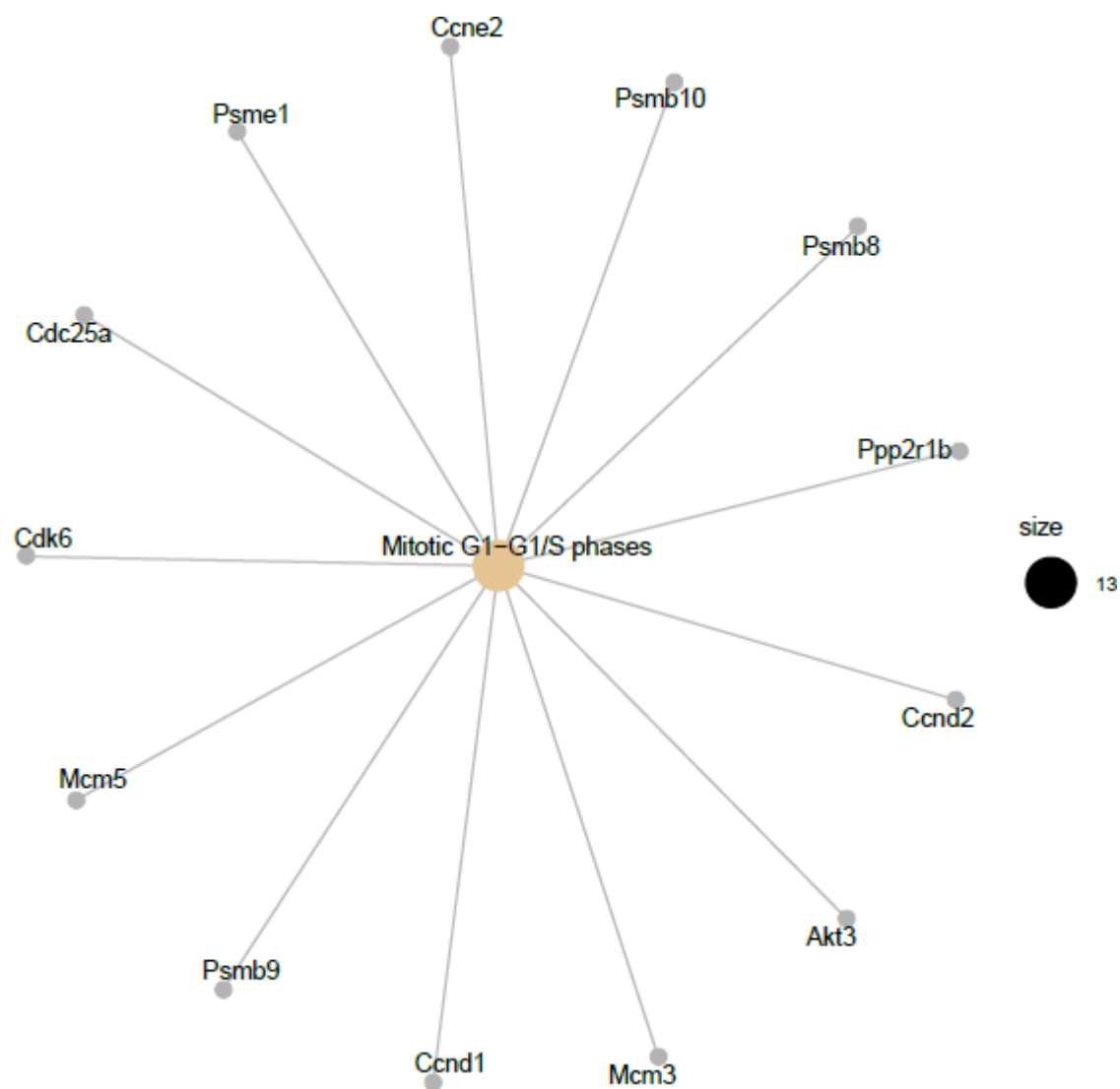


Figure 9: Gráfico de redes de la comparación CAV1.CTRLvsWT.CTRL

Y el gráfico de redes de cada ruta metabólica para la comparación CAV1.TGFvsWT.TGF:

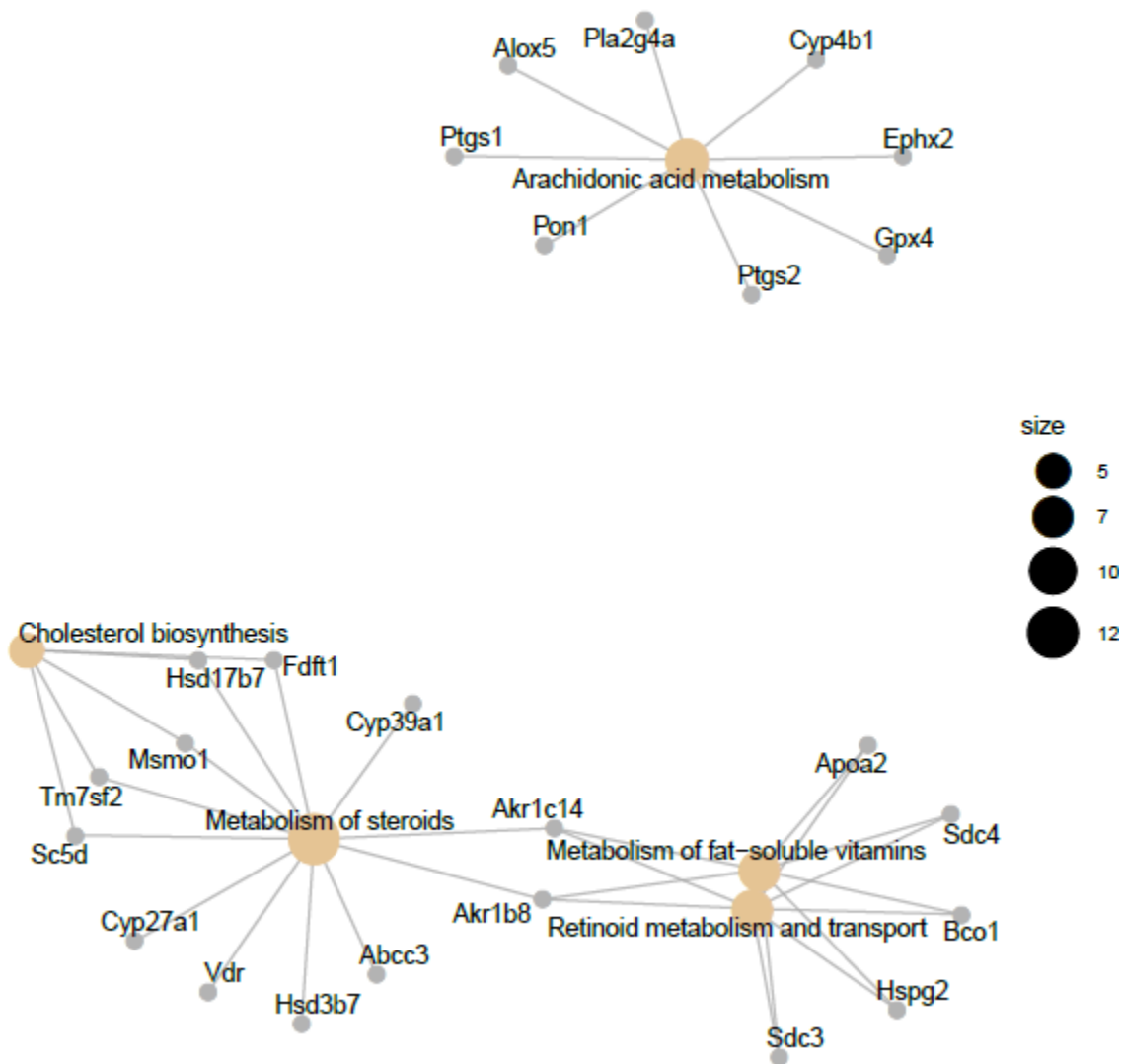


Figure 10: Gráfico de redes de la comparación CAV1.TGFvsWT.TGF

De esta manera podemos observar qué genes afectan a qué ruta metabólica significativamente.

## Resumen de archivos obtenidos

Aquí ponemos la tabla resumen con los datos y los distintos gráficos y tablas obtenidos:

Table 3: Lista de archivos obtenidos en el análisis

Lista_de_archivos
ArrayQualityMetrics_Norm
ArrayQualityMetrics_rawData
data4Heatmap.csv
normalized_data.csv
normalized_filt_data.csv
ReactomePA.Results.CAV1.CTRLvsWT.CTRL.csv

---

#### Lista\_de\_archivos

---

ReactomePA.Results.CAV1.TGFvsWT.TGF.csv  
ReactomePABarplot.CAV1.CTRLvsWT.CTRL.pdf  
ReactomePABarplot.CAV1.TGFvsWT.TGF.pdf  
ReactomePAcnetplot.CAV1.CTRLvsWT.CTRL.pdf  
ReactomePAcnetplot.CAV1.TGFvsWT.TGF.pdf  
topAnnotated\_CAV1.TGFvsWT.TGF.csv  
topAnnotated\_INT.csv  
topAnnotated\_topTab\_CAV1.CTRLvsWT.CTRL.csv

---

## Discusión

Para obtener unos resultados con mayor fiabilidad sería recomendable utilizar una mayor cantidad de muestras, lo cual nos daría un mayor número de datos en los que confirmar las significaciones biológicas entre comparaciones.

## Conclusión

A partir de los datos del análisis de significación biológica hemos observado que solo hay diferencias significativas en una ruta metabólica entre los dos tipos distintos de células que no han sido tratadas con TGF-B. Aunque estamos hablando de una sola ruta metabólica, al ser concretamente la ruta de la fase de mitosis entre G1 y G1-S, un cambio significativo en esta fase puede afectar en gran medida al desarrollo celular y del organismo. Por ello, es una ruta metabólica a tener muy en cuenta y al que se podría investigar en más detalle qué efecto tiene la inhibición del gen CAV1 en esta ruta metabólica.

En la tabla de significación biológica de la comparación de los dos tipos distintos de células cuando se han tratado con TGF-B se observan concretamente doce rutas metabólicas que se ven significativamente afectadas, indicando que en una situación de presencia de TGF-B, las células se comportan de manera distinta cuando tienen (o no) el gen CAV1 silenciado. Debido a la gran cantidad de rutas metabólicas afectadas y las posibles relaciones o interacciones entre ellas, no me veo con conocimientos necesarios para dar unas conclusiones de los datos detalladas o certeras. Por ello, tan solo se puede concluir en este informe en que se han observado unas claras diferencias en la comparación CAV1.TGFvsWT.TGF las cuales merecen un estudio más detallado.

## Apéndice

Todo el código utilizado se encuentra en el archivo “rmd” en el repositorio github de la PEC1.

## References

- Gautier, Laurent, Leslie Cope, Benjamin M. Bolstad, and Rafael A. Irizarry. 2004. “Affy—Analysis of Affymetrix Genechip Data at the Probe Level.” *Bioinformatics* 20 (3): 307–15. <https://doi.org/10.1093/bioinformatics/btg405>.
- Gentleman, R., V. Carey, W. Huber, and F. Hahne. 2019. *Genefilter: Genefilter: Methods for Filtering Genes from High-Throughput Experiments*.
- Han, Mei, Zeribe Chike Nwosu, Weronika Piorońska, Matthias Philip Ebert, Steven Dooley, and Christoph Meyer. 2019. “Caveolin-1 Impacts on Tgf- $\beta$  Regulation of Metabolic Gene Signatures in Hepatocytes.” *Frontiers in Physiology* 10.
- Irizarry, Rafael A., Bridget Hobbs, Francois Collin, Yasmin D. Beazer-Barclay, Kristen J. Antonellis, Uwe Scherf, and Terence P. Speed. 2003. “Exploration, normalization, and summaries of high density

oligonucleotide array probe level data.” *Biostatistics* 4 (2): 249–64. <https://doi.org/10.1093/biostatistics/4.2.249>.

Kauffmann, Audrey, Robert Gentleman, and Wolfgang Huber. 2009. “ArrayQualityMetrics—a Bioconductor Package for Quality Assessment of Microarray Data.” *Bioinformatics* 25 (3): 415–6.

Pagès, Hervé, Marc Carlson, Seth Falcon, and Nianhua Li. 2019. *AnnotationDbi: Manipulation of Sqlite-Based Annotations in Bioconductor*.

Pla, Alex Sanchez, Ricardo, and Gonzalo. 2019. “Statistical Analysis of Microarray Data (Adapted for Teaching Purposes).”

Ritchie, Matthew E, Belinda Phipson, Di Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. 2015. “limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies.” *Nucleic Acids Research* 43 (7): e47. <https://doi.org/10.1093/nar/gkv007>.

Yu, Guangchuang, and Qing-Yu He. 2016. “ReactomePA: An R/Bioconductor Package for Reactome Pathway Analysis and Visualization.” *Molecular BioSystems* 12 (12): 477–79. <https://doi.org/10.1039/C5MB00663E>.