

Efecto del alcohol en células madre neurales derivadas de células madre de embriones humanos

Marina Ballesteros

4/9/2020

Contents

Lectura de los datos	1
1. Datos del estudio	

Los datos del análisis se han cargado desde la base de datos Gene Expression Omnibus (GEO). El conjunto de datos seleccionado se identifica con el número de adhesión: **GSE56906**.

El estudio que generó dichos datos investiga el efecto del alcohol en el desarrollo de células madre neurales derivadas de células madre de embriones humanos. Existen indicios de que el alcohol interviene negativamente en el desarrollo de las células madre neurales. Para corroborar estos indicios, se han cultivado células madre embrionarias durante cinco días en un medio de inducción neural (NIM). A continuación, los agregados neuronales fueron sembrados en placas recubiertas con poli-L-ornitina/laminina y cultivados con NIM durante siete días adicionales para desarrollar la estructura de roseta neuronal. Después de siete días, las rosetas neurales fueron desalojadas y luego replataadas para la expansión de las células precursoras neurales (NPC) durante 3-5 días.

Los microarrays utilizados para este experimento fueron del siguiente tipo: Affymetrix Human Genome U133 Plus 2.0 Array.

#Materiales y métodos

##Preparación del ambiente de trabajo

Antes de empezar con el análisis y a manejar la enorme cantidad de datos y ficheros que ello conlleva, crearé tres carpetas para la organización del mismo:

- La carpeta principal del análisis será “Effect of alcohol in ESC differentiation”, la cual también será mi directorio de trabajo.
- Una carpeta llamada **data** para almacenar todo tipo de datos del experimento y en los cuales basaré mi análisis. Es decir, guardaré los archivos *.CEL* y el archivo *targets*, en el cual se describirán los factores de estudio y sus niveles.
- En la carpeta **results** guardaré todos los resultados obtenidos en el análisis.
- La carpeta **figures** servirá para almacenar todo tipo de imágenes y figuras generadas durante el análisis.

Lectura de los datos

En el experimento se han analizado 10 muestras, las cuales se analizan en torno a dos factores: el tipo celular y el tipo de tratamiento durante la diferenciación celular. Los niveles de dichos factores son los siguientes:

Para el factor **tipo celular** existen tres niveles:

- Células madre embrionarias (ESC) indiferenciadas *H1 p40*, dos repeticiones.
- Células neurales en forma de roseta *H1 Rosett*
- Células progenitoras neurales (NPC) *NPC H1*

Para el factor **tipo de tratamiento** existen dos niveles:

- Células diferenciadas sin tratamiento con alcohol *EtOH 0*
- Células diferenciadas bajo tratamiento con alcohol *EtOH 20*

Una tabla que recoge todas las muestras sería la siguiente:

Referencia	Tipo de muestra	Características
GSM1371025	H1 EtOH0 Rosett1	Roseta sin alcohol
GSM1371026	H1 EtOH0 Rosett2	Roseta sin alcohol
GSM1371027	H1 EtOH20 Rosett1	Roseta con alcohol
GSM1371028	H1 EtOH20 Rosett2	Roseta con alcohol
GSM1371029	H1 p40 Und-1	ESC indiferenciada
GSM1371030	H1 p40 Und-2	ESC indiferenciada
GSM1371031	NPC H1 EtOH 0-1	NPC sin alcohol
GSM1371032	NPC H1 EtOH 0-2	NPC sin alcohol
GSM1371033	NPC H1 EtOH 20-1	NPC con alcohol
GSM1371034	NPC H1 EtOH 20-2	NPC con alcohol

Para la lectura de los datos, primero he descargado los 10 archivos *.CEL* y he creado un archivo llamado *targets.csv* donde recopilé las características de cada muestra así como su referencia en la base de datos GEO. Tras haber instalado Bioconductor y el paquete “GEOquery”, procedo a la lectura de los datos.

```
library(oligo)
celFiles <- list.celfiles("./data", full.names = TRUE)
library(Biobase)
#Asociación de los archivos CEL con el archivo targets
my.targets <- readAnnotatedDataFrame(file.path("./data", "targets.csv"),
                                     header = TRUE, row.names = 1,
                                     sep=";")
rawData <- read.celfiles(celFiles, phenoData = my.targets)
print(pData(rawData))
```

2. Control de calidad de los datos crudos

El control de calidad de los datos en crudo nos permite conocer la calidad de los datos recogidos durante el experimento a través de boxplot de intensidad o un estudio de los componentes principales (PCA). Este paso es muy importante ya que es donde se evalúa si los datos tienen suficiente calidad para la normalización de los mismos. Si alguno de los arrays no tiene la calidad suficiente, este será marcado y expuesto a revisión para decidir si mantenemos o no dicho array en el análisis. El paquete que desarrolla el control de calidad se llama **ArrayQualityMetrics**.

```
library(arrayQualityMetrics)
#Guardar los resultados del control calidad en el directorio results
arrayQualityMetrics(rawData, outdir = "./results/rawData_quality", force = TRUE)
```

La siguiente tabla resumen nos ofrece el resultado del control de calidad.

<input type="checkbox"/>	array	sampleNames	*1	*2	*3	SampleTitle	Classification
<input type="checkbox"/>	1	GSM1371025				H1 EtOH 0 Rosett-1	Roseta diferenciada sin alcohol
<input type="checkbox"/>	2	GSM1371026				H1 EtOH 0 Rosett-2	Roseta diferenciada sin alcohol
<input type="checkbox"/>	3	GSM1371027				H1 EtOH 20 Rosett-1	Roseta diferenciada con alcohol
<input type="checkbox"/>	4	GSM1371028				H1 EtOH 20 Rosett-2	Roseta diferenciada con alcohol
<input type="checkbox"/>	5	GSM1371029				H1 p40 Und-1	ESC indiferenciada
<input type="checkbox"/>	6	GSM1371030				H1 p40 Und-2	ESC indiferenciada
<input checked="" type="checkbox"/>	7	GSM1371031			x	NPC H1 EtOH 0-1	NPC diferenciada sin alcohol
<input type="checkbox"/>	8	GSM1371032				NPC H1 EtOH 0-2	NPC diferenciada sin alcohol
<input checked="" type="checkbox"/>	9	GSM1371033			x	NPC H1 EtOH 20-1	NPC diferenciada con alcohol
<input type="checkbox"/>	10	GSM1371034				NPC H1 EtOH 20-2	NPC diferenciada con alcohol

Figure 1: Aspect of the summary table, in the index.html file, produced by the arrayQualityMetrics package on the raw data