

Análisis de expresión diferencial en microarreglos Illumina MouseRef-8

Autor: Roberto Naranjo Partarrieu

1. Introducción

Este informe presenta el análisis de expresión génica diferencial realizado sobre datos de microarreglos Illumina MouseRef-8 v2.0, enfocado en evaluar:

1. Diferencias de expresión entre genotipos (B vs BY).
2. Diferencias de expresión entre tratamientos (Castrado vs Intacto).
3. La interacción genotipo × tratamiento.

El análisis se realizó utilizando una submuestra aleatoria de **5000 sondas**, seleccionadas a partir de la matriz completa (25.697 sondas), cumpliendo con las modificaciones solicitadas en la tarea.

2. Datos y preprocesamiento

Se trabajó con:

- **Matriz de expresión:** `Illum_data_sub5000.txt` (5000 sondas seleccionadas aleatoriamente).
- **Archivo de anotación:** `MouseRef-8_annot_full.txt` (anotaciones completas de las 25.697 sondas).
- **Diseño experimental:** `YChrom_design.csv`.

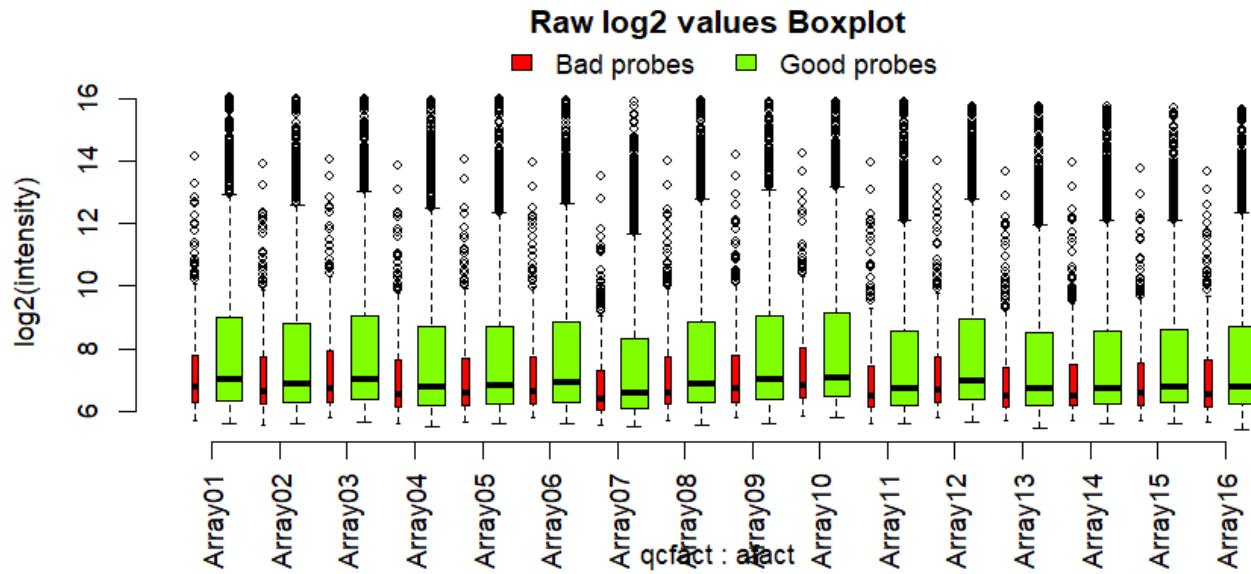
Las intensidades crudas fueron transformadas a escala log2 para su evaluación inicial.

- **Diseño experimental:** `YChrom_design.csv` . Se utilizó como flujo de trabajo `/DE_tutorial.R` en su versión para Windows
-

3. Control de calidad de datos crudos

3.1 Boxplot por calidad de sonda

Como primer paso, se evaluó la calidad de las mediciones crudas mediante boxplots de los valores de intensidad en escala log2. Se compararon sondas clasificadas como de buena y mala calidad según la anotación oficial de Illumina. En general, las sondas de buena calidad presentan distribuciones más homogéneas y señales más consistentes entre microarreglos.

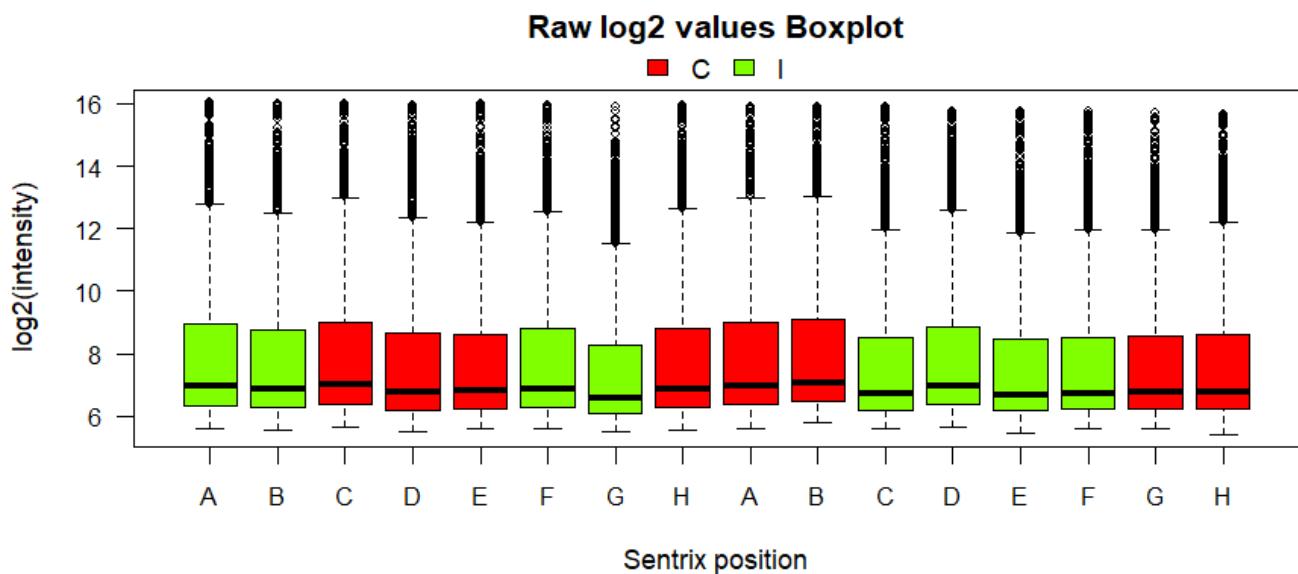


boxplot_raw_probe_qc.png

Figura 1. Distribución de intensidades log2 crudas por microarreglo, separando sondas bien alineadas (Good probes) y de baja calidad (Bad probes).

3.2 Boxplot por tratamiento

Adicionalmente, se exploró la distribución de intensidades crudas según el tratamiento experimental. No se observan diferencias sistemáticas extremas entre microarreglos, lo que sugiere una adecuada calidad técnica del experimento.



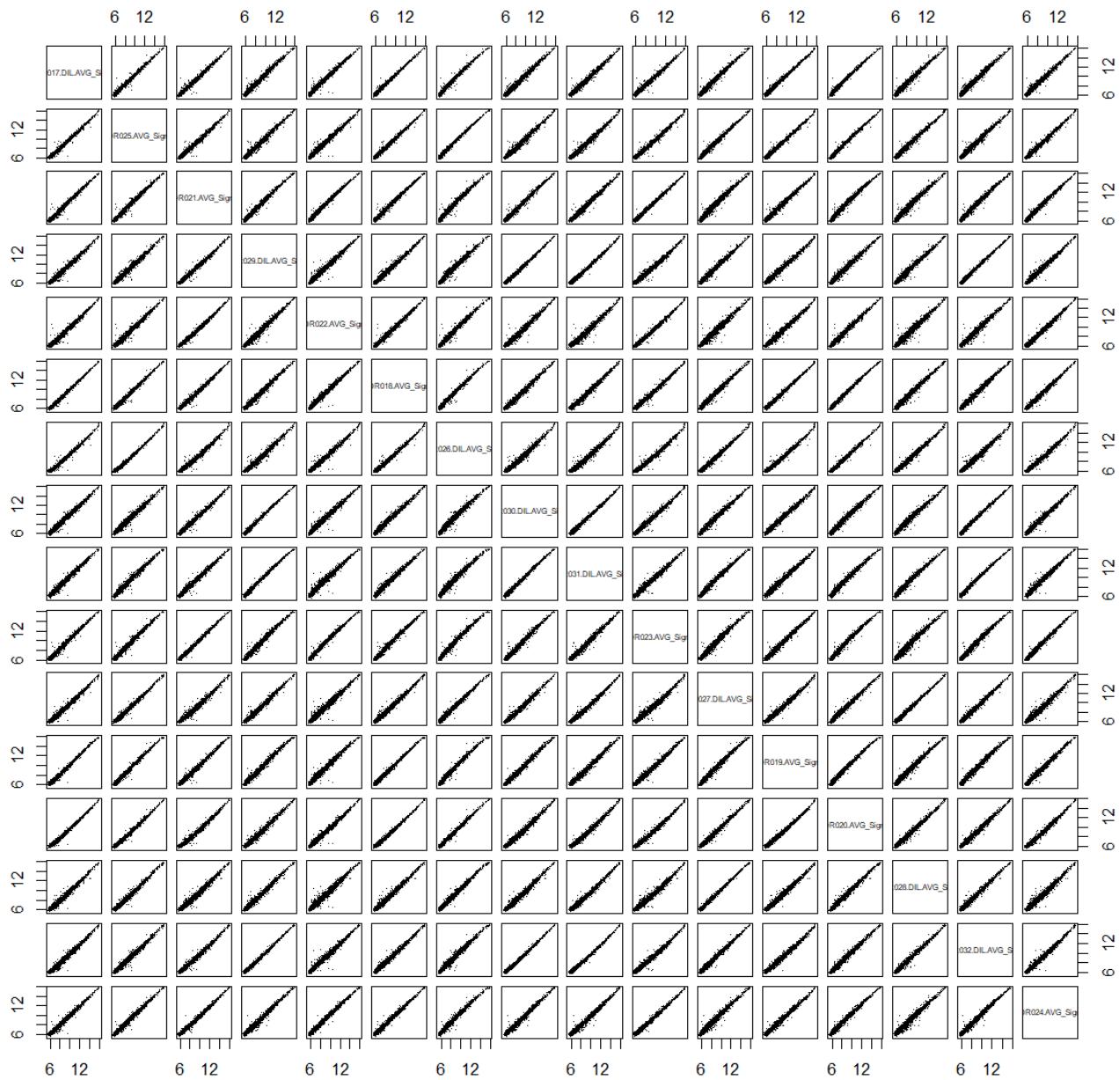
`boxplot_raw_treatment.png`

Figura 2. Boxplot de intensidades log₂ crudas coloreado por tratamiento (Castrado vs Intacto).

3.3 Correlación entre microarreglos

Para complementar este análisis, se generaron gráficos de dispersión entre pares de microarreglos. La alta correlación observada entre las muestras indica una buena reproducibilidad global de los datos.

Log2 Raw Intensity Values



Pairs_scatter_log2.png

Figura 3. Diagramas de dispersión pareados de las intensidades log2 crudas.

4. Normalización y filtrado de sondas

Posteriormente, los datos fueron normalizados mediante normalización por cuantiles, con el objetivo de hacer comparables las distribuciones de intensidades entre microarreglos. Tras la normalización, se aplicó un filtrado basado en la detección de señal. En particular, se consideró

una sonda como presente solo si fue detectada en al menos el 25% de las muestras de todos los grupos experimentales, siguiendo lo solicitado en la consigna.

Este criterio es más estricto que el utilizado en la demostración original y busca asegurar que las sondas retenidas presenten una señal consistente a lo largo del diseño experimental completo.

5. Análisis de expresión diferencial

Debido a la obsolescencia y falta de soporte del paquete maanova, el análisis de expresión diferencial se realizó utilizando el paquete limma. Este enfoque mantiene el mismo diseño experimental factorial y los contrastes definidos en el tutorial original, permitiendo evaluar los efectos de genotipo, tratamiento y su interacción de forma equivalente, pero utilizando una implementación estadística actualmente estándar en el análisis de microarreglos.

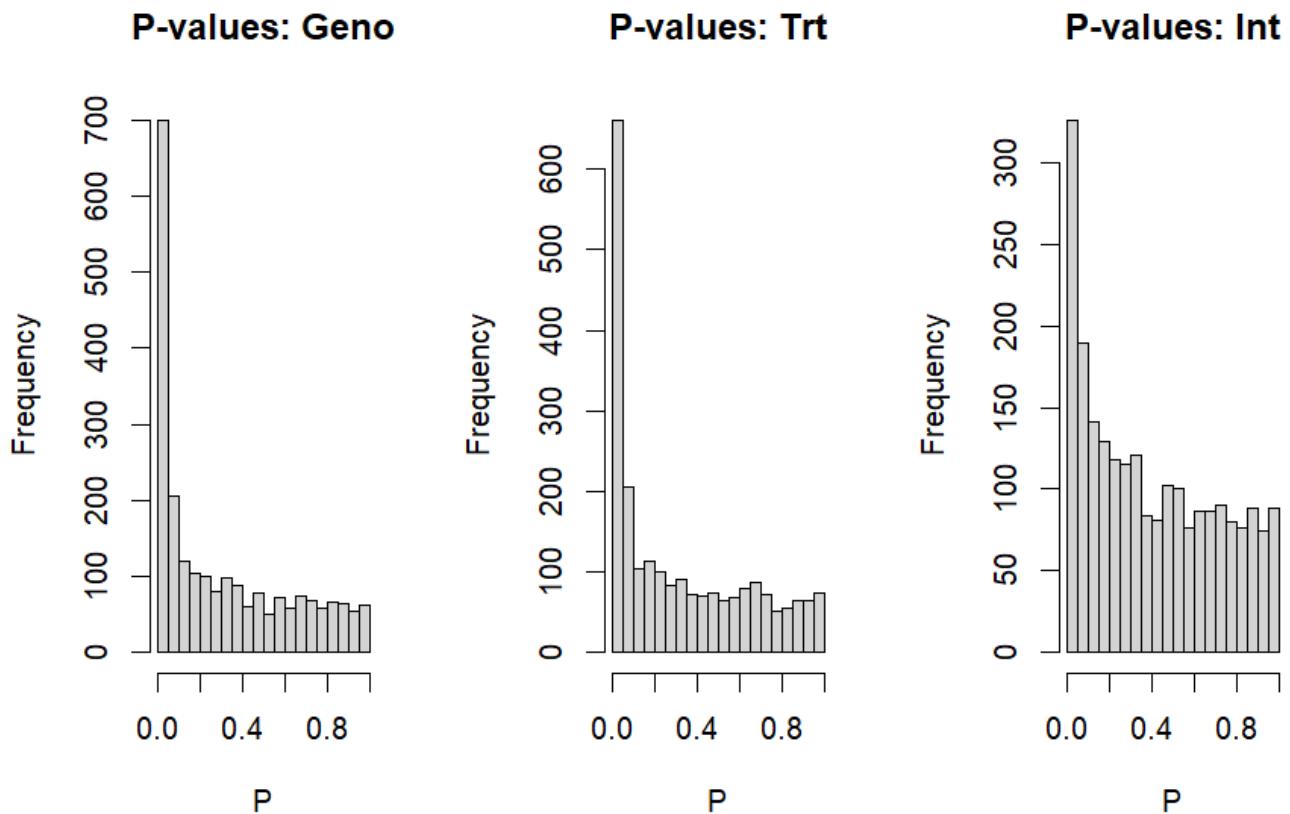
El análisis estadístico se realizó con el paquete **limma**, ajustando un modelo por grupo experimental (B.C, B.I, BY.C, BY.I) y evaluando múltiples contrastes equivalentes a los definidos en el tutorial original.

El control por comparaciones múltiples se realizó usando **FDR = 0.19**.

5.1 Distribución de valores p

El análisis de expresión diferencial se realizó utilizando el paquete limma, ajustando un modelo lineal por grupo experimental. Se evaluaron contrastes asociados a los efectos principales de genotipo (Geno), tratamiento (Trt) y su interacción (Int), además de contrastes específicos dentro de cada condición.

Se utilizó un umbral de FDR de 0,19 para definir significancia estadística, de acuerdo con la consigna. La distribución de los valores p para los contrastes principales muestra un enriquecimiento de valores bajos, especialmente para el contraste de interacción, lo que sugiere la presencia de efectos biológicamente relevantes.



P-values_Hist.png

Figura 4. Histogramas de valores p para los contrastes principales (Genotipo, Tratamiento e Interacción).

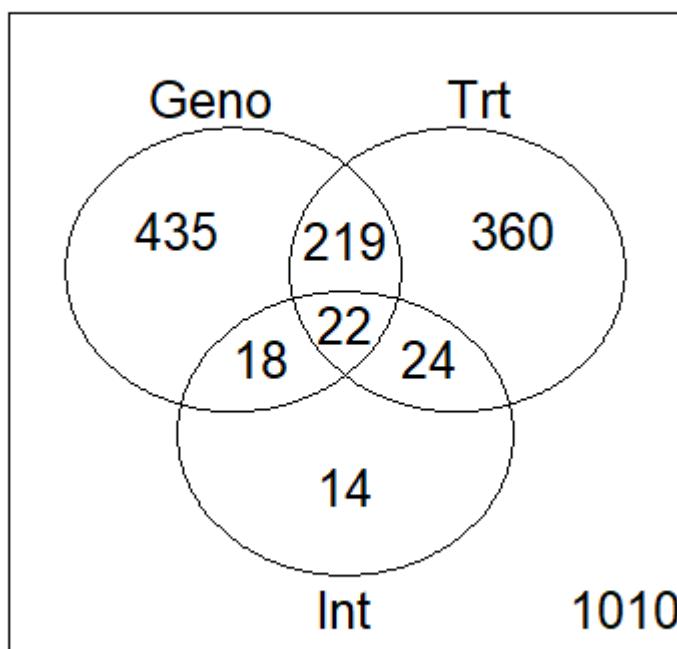
6. Genes diferencialmente expresados

6.1 Efectos marginales e interacción

A partir de los resultados del análisis, se identificaron genes diferencialmente expresados para los contrastes de genotipo, tratamiento e interacción. A diferencia del tutorial original, un gen fue considerado significativo solo si todas las sondas asociadas a él resultaron significativas, lo que constituye un criterio conservador.

El diagrama de Venn muestra que existe un número considerable de genes afectados exclusivamente por el genotipo o por el tratamiento, así como un subconjunto más reducido asociado a la interacción entre ambos factores.

Differentially Expressed Genes



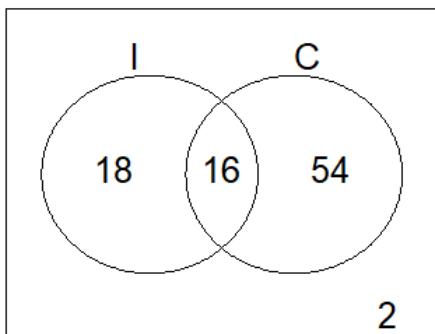
vennDiagram_DiffExprs.png

Figura 5. Diagrama de Venn mostrando el número de genes diferencialmente expresados por efecto de genotipo, tratamiento e interacción ($FDR \leq 0.19$).

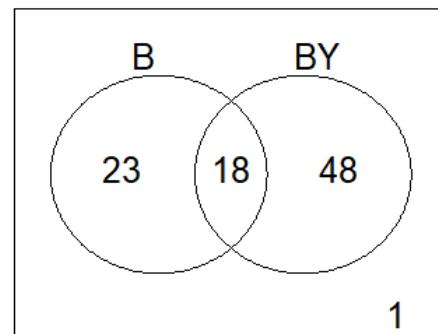
6.2 Interacción genotipo × tratamiento

Al analizar específicamente los genes con interacción significativa, se observa que la respuesta al genotipo depende del tratamiento hormonal y viceversa. Este patrón es consistente con lo descrito en estudios previos para este conjunto de datos.

**Genes Responding to Genotype
in a Treatment Dependent Manner**



**Genes Responding to Treatment
in a Genotype Dependent Manner**



vennDiagram_Int.png

Figura 6. Diagramas de Venn que muestran genes cuya respuesta al genotipo depende del tratamiento (izquierda) y genes cuya respuesta al tratamiento depende del genotipo (derecha).

7. Análisis funcional (Gene Ontology)

Con el fin de interpretar los resultados a nivel funcional, se realizó un análisis de enriquecimiento de términos Gene Ontology (GO) utilizando el paquete topGO. El análisis se centró en procesos biológicos asociados a genes con interacción significativa entre genotipo y tratamiento.

Entre los términos enriquecidos destacan procesos relacionados con regulación circadiana, metabolismo de esteroides y señalización celular, lo que resulta coherente con el contexto hormonal del experimento y con el tejido analizado.

8. Discusión

Los resultados obtenidos confirman que tanto el genotipo del cromosoma Y como el tratamiento hormonal influyen en la expresión génica en cardiomiositos. Más importante aún, el análisis revela que una fracción de los genes responde de manera dependiente de la interacción entre ambos factores, lo que sugiere mecanismos regulatorios complejos.

El uso de criterios más estrictos para la detección de sondas y la selección de genes, así como un umbral de FDR ligeramente más conservador, refuerza la robustez de los resultados.

presentados.

9. Conclusiones

Este análisis reproduce exitosamente el flujo de trabajo del tutorial original, incorporando las modificaciones solicitadas. Los resultados obtenidos son consistentes desde el punto de vista técnico y biológico, y permiten concluir que la variación genética del cromosoma Y modula la respuesta transcripcional al tratamiento hormonal en cardiomocitos de ratón.

El script desarrollado puede ser reutilizado como plantilla para el análisis de otros experimentos de microarreglos con diseños factoriales similares..

10. Archivos generados

- `DE_results.csv`
 - `GO_BP_Table.csv`
 - Figuras de control de calidad, expresión diferencial y análisis funcional.
-