

Título: Análisis de expresión diferencial en poblaciones celulares metastáticas de pez cebra

Autor: Victor Ulises Plascencia Perez

Introducción

La metástasis es la dispersión de las células hacia sitios distantes del tumor primario y es considerada la principal causa de muerte en pacientes con cáncer¹. A pesar de su relevancia clínica, no existen tratamientos para evitar esta diseminación celular, esto principalmente es debido a que no se ha logrado caracterizar componentes moleculares clave de este proceso que sean comunes para varios tipos de cáncer². Uno de los principales obstáculos con los que se topan los investigadores al estudiar poblaciones de células metastáticas es la elección de un modelo biológico. Por años se ha utilizado el ratón (*Mus musculus*), sin embargo, la xenoimplantación de células tumorales humanas para inducir tumores en el ratón es complicada, ya que muchas de éstas pueden viajar en el torrente sanguíneo sin tener un fenotipo metastático, lo que complica la identificación de los marcadores moleculares antes mencionados¹. Por lo anterior, se ha comenzado a explorar la utilización de nuevos modelos, tales como el pez cebra (*Danio rerio*).

En embriones de pez cebra se puede inducir tumores humanos mediante la inyección de células tumorales en el saco vitelino y, dado que los embriones son transparentes, se puede identificar si la migración celular ocurrió tras la generación de un tumor primario, tal como en la metástasis, o si es un artefacto de la xenoimplantación, como anteriormente se mencionó que ocurre en ratón¹. Las células metastáticas pueden ser posteriormente caracterizadas y comparadas con las células del tumor primario³.

En el presente trabajo se realiza un análisis de expresión diferencial de células metastáticas de cáncer de mama triple negativo (TNBC) de un modelo de cáncer xenoimplantado en pez cebra, en donde la comparación se realiza con células parentales en el tumor primario.

Objetivos

1. General

Realizar un análisis de expresión diferencial de células metastáticas de TNBC en un modelo de cáncer xenoimplantado de pez cebra

Metodología

1. Descarga y procesamiento de archivos FASTQ para obtener el número de cuentas de transcritos

El script para el análisis y preprocesamiento de los datos está disponible en el presente repositorio en la carpeta *bash* con el nombre de *preprocesamiento.jdl*

Los archivos FASTQ fueron descargados de la base de datos GEO, utilizando el identificador SRR para cada población celular: SRR12082751, SRR12082752, SRR12082753 y SRR12082754. Los primeros dos corresponden a réplicas de la población celular del tumor primario (parentales) y los últimos dos a poblaciones metastáticas. Los archivos contenían lecturas de tipo *paired-end*. Se realizó la visualización de la calidad de los archivos FASTQ y, con base en ella, se determinó realizar un *trimming* para mejorar su calidad. Una vez con calidad óptima, los archivos se alinearon a un genoma de referencia (HG38) con la herramienta *hisat2*, con lo cual se obtuvieron archivos *sam*. Posteriormente, estos archivos fueron convertidos a archivos *bam*. En el último paso del procesamiento, se utilizó el lenguaje de programación *R* para obtener el archivo de cuentas de cada transcrito en cada población celular (el script utilizado para ello está disponible en este repositorio en la carpeta *R*, bajo el nombre *count_generation.R*).

2. Análisis de expresión diferencial

Una vez con los archivos de cuentas obtenido, se procedió a realizar un análisis de expresión diferencial. Para ello, se utilizó la librería *EdgeR*. Se generó un objeto *EdgeR* y una matriz de diseño adecuada. Genes con un bajo nivel de expresión fueron descartados para el análisis. Previo al análisis, se visualizaron los niveles de expresión de un gráfico de componentes principales (PCA).

Se realizó el análisis de expresión diferencial y se obtuvieron genes regulados a la alza y a la baja. El criterio para determinar si un gen estaba regulado a la alza fue si su expresión era 4 veces, o mayor, en la población metastática que en la parental, así como una tasa de descubrimiento de falsos negativos menor a $1e-5$. Se consideró que un gen era regulado a la baja si su expresión era 4 veces menor en la población metastática que en la parental. Finalmente se realizó una visualización de los genes diferencialmente expresados en un gráfico de volcán

En el último paso, se obtuvieron los TPM (transcritos por millón de kilobase) y se realizó un *heatmap*.

El script utilizado para el análisis de expresión diferencial puede consultarse en el presente repositorio con el nombre de `DEA_analysis.R`

3. Análisis de ontología

Con los identificadores de los genes regulados a la alza y a la baja se realizó un análisis de ontología utilizando la herramienta web *DAVID*.

Resultados

1. Calidad de los archivos FASTQ

Tal como se mencionó anteriormente, se visualizó la calidad de archivos FASTQ previo y tras un procesamiento de *trimming*. En la figura 1 se pueden observar las calidad del archivo SRR12082751 antes del *trimming* (a) y posterior a éste (b) a manera de ejemplo. Un comportamiento similar se apreció para todos los archivos tras el *trimming*.

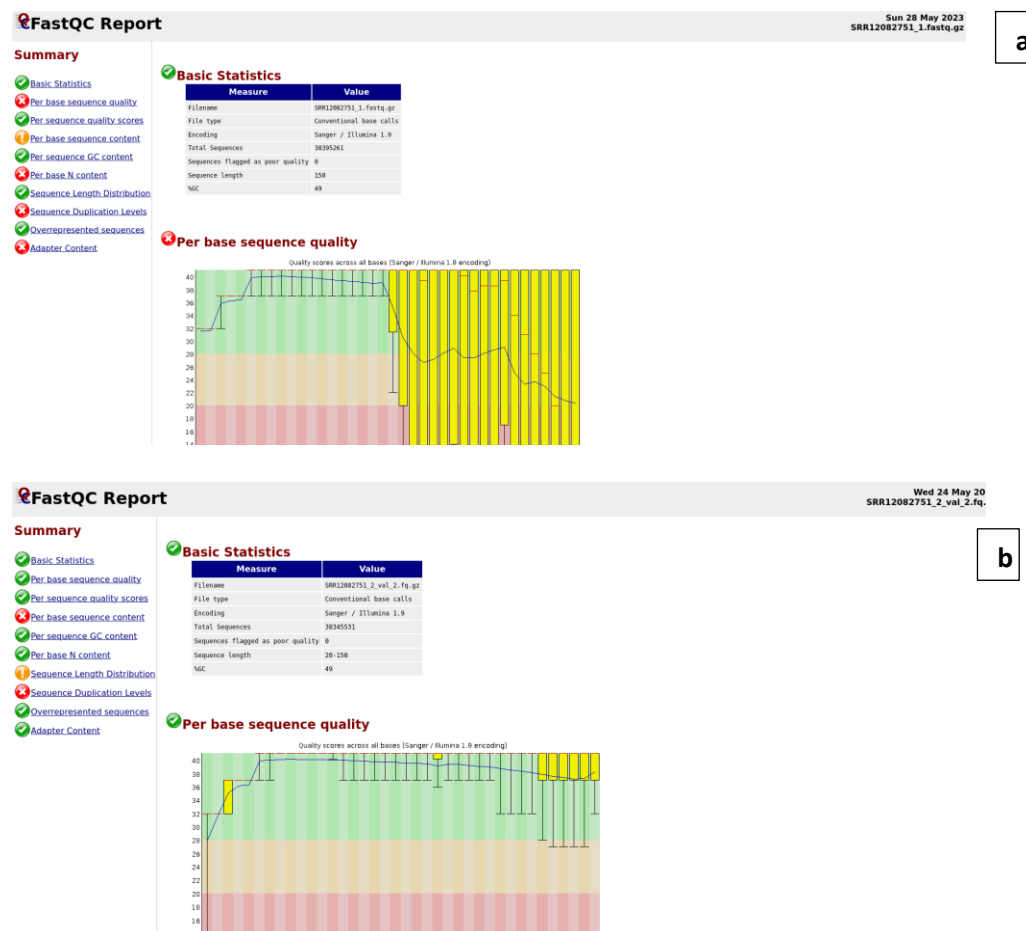


Figura 1. Análisis de calidad de los archivos FASTQ antes (a) y después (b) de la realización del *trimming*

2. Análisis de expresión diferencial

Previo al análisis de expresión diferencial se generó un PCA para ver una posible agrupación de los niveles de expresión en las muestras. En la figura 2. Se observa un PCA de los niveles de expresión; se puede ver claramente que las réplicas de las poblaciones metastáticas tienden a agruparse, así como las de las poblaciones parentales, aunque esto último en menor medida.

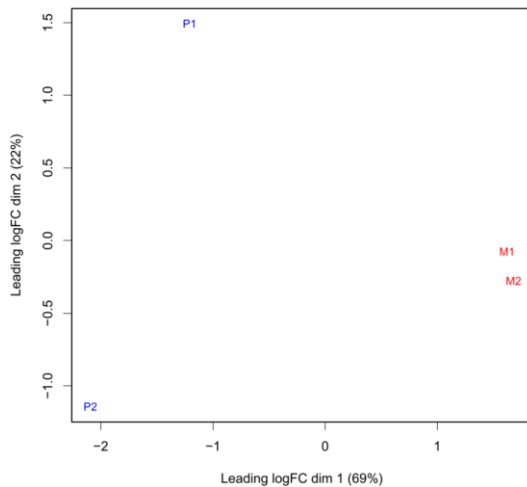


Figura 2. PCA de niveles de expresión de las poblaciones celulares metastáticas(M) y en el tumor primario (P).

Tras la realización del análisis de expresión diferencial se obtuvieron 53 genes regulados a la alza y 105 regulados a la baja, se puede ver la representación de ellos en la figura 3a, en donde se muestra un *heatmap*. En la figura 3b podemos observar el gráfico de volcán, en donde las líneas punteadas presentan el corte de descubrimiento de falsos negativos (línea horizontal) y el logaritmo del cambio de duplicación de expresión génica (logFC; línea vertical) .

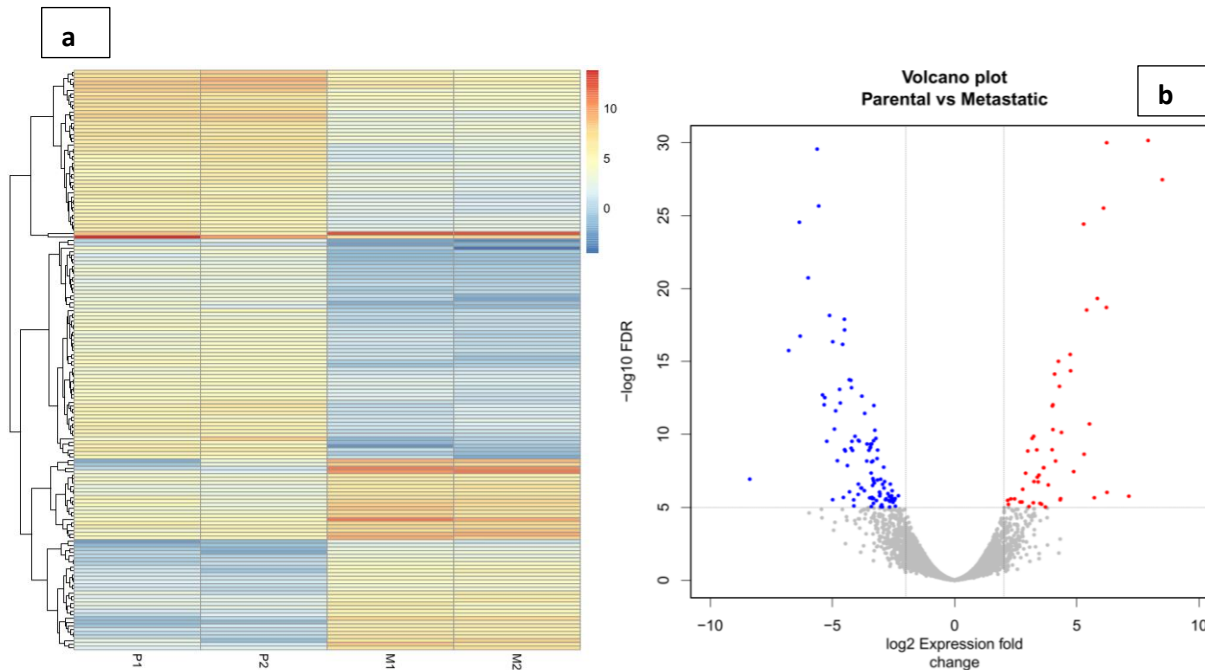


Figura 3. a) Heatmap con los 53 genes regulados al alza y los 105 regulados a la baja en poblaciones celulares parentales (P1 y P2) y metastáticas (M1 y M2). b) Gráfico de volcán donde se ven a los genes regulados al alza (rojos) y los regulados a la baja (azules).

3. Análisis de ontología

Finalmente, los análisis de ontología permitieron identificar los procesos biológicos en los cuales se encuentran involucrados los genes regulados al alza (Fig. 4) y regulados a la baja (Fig. 5). En el caso de los genes regulados al alza, los procesos que llaman la atención son aquellos relacionados con señalización, desarrollo, procesos del sistema inmune y muerte celular. En el caso de los genes regulados a la baja podemos encontrar que se encuentran involucrados en procesos similares a los de los genes regulados al alza; los que llaman la atención son los relacionados con la localización celular, la adhesión celular y localización.

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	GOTERM_BP_1	immune system process	RT	<div></div>	24	47,1	1,5E-8	3,2E-7
<input type="checkbox"/>	GOTERM_BP_1	locomotion	RT	<div></div>	18	35,3	8,4E-7	8,8E-6
<input type="checkbox"/>	GOTERM_BP_1	developmental process	RT	<div></div>	31	60,8	2,8E-6	1,7E-5
<input type="checkbox"/>	GOTERM_BP_1	signaling	RT	<div></div>	31	60,8	3,3E-6	1,7E-5
<input type="checkbox"/>	GOTERM_BP_1	multi-organism process	RT	<div></div>	18	35,3	2,2E-5	9,2E-5
<input type="checkbox"/>	GOTERM_BP_1	response to stimulus	RT	<div></div>	34	66,7	2,1E-4	7,5E-4
<input type="checkbox"/>	GOTERM_BP_1	localization	RT	<div></div>	27	52,9	5,1E-4	1,4E-3
<input type="checkbox"/>	GOTERM_BP_1	multicellular organismal process	RT	<div></div>	30	58,8	5,3E-4	1,4E-3
<input type="checkbox"/>	GOTERM_BP_1	biological regulation	RT	<div></div>	40	78,4	8,2E-4	1,9E-3
<input type="checkbox"/>	GOTERM_BP_1	cell killing	RT	<div></div>	5	9,8	1,9E-3	4,0E-3
<input type="checkbox"/>	GOTERM_BP_1	single-organism process	RT	<div></div>	41	80,4	4,5E-3	8,5E-3
<input type="checkbox"/>	GOTERM_BP_1	growth	RT	<div></div>	8	15,7	4,8E-3	8,5E-3
<input type="checkbox"/>	GOTERM_BP_1	cellular component organization or biogenesis	RT	<div></div>	23	45,1	4,7E-2	7,7E-2
<input type="checkbox"/>	GOTERM_BP_1	reproductive process	RT	<div></div>	8	15,7	6,0E-2	7,9E-2
<input type="checkbox"/>	GOTERM_BP_1	biological adhesion	RT	<div></div>	8	15,7	6,0E-2	7,9E-2
<input type="checkbox"/>	GOTERM_BP_1	reproduction	RT	<div></div>	8	15,7	6,0E-2	7,9E-2

Figura 4. Procesos biológicos en los que se ven involucrados los genes regulados al alza en poblaciones celulares metastáticas.





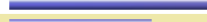






Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
<input type="checkbox"/>	GOTERM_BP_1	localization	RT		56	53,8	1,3E-6	2,8E-5
<input type="checkbox"/>	GOTERM_BP_1	locomotion	RT		23	22,1	2,7E-4	2,8E-3
<input type="checkbox"/>	GOTERM_BP_1	multicellular organismal process	RT		55	52,9	9,4E-4	6,5E-3
<input type="checkbox"/>	GOTERM_BP_1	biological adhesion	RT		18	17,3	1,4E-3	7,5E-3
<input type="checkbox"/>	GOTERM_BP_1	single-organism process	RT		82	78,8	8,0E-3	3,4E-2
<input type="checkbox"/>	GOTERM_BP_1	response to stimulus	RT		58	55,8	1,3E-2	4,1E-2
<input type="checkbox"/>	GOTERM_BP_1	developmental process	RT		44	42,3	1,4E-2	4,1E-2
<input type="checkbox"/>	GOTERM_BP_1	signaling	RT		43	41,3	2,6E-2	6,8E-2
<input type="checkbox"/>	GOTERM_BP_1	immune system process	RT		23	22,1	3,3E-2	7,8E-2
<input type="checkbox"/>	GOTERM_BP_1	biological regulation	RT		73	70,2	5,2E-2	1,1E-1
<input type="checkbox"/>	GOTERM_BP_1	cellular component organization or biogenesis	RT		43	41,3	8,0E-2	1,5E-1

Figura 4. Procesos biológicos en los que se ven involucrados los genes regulados a la baja en poblaciones celulares metastáticas.

Discusión y conclusión

En el presente trabajo se presentan los resultados de un análisis de expresión diferencial comparando poblaciones celulares metastáticas contra poblaciones parentales en el tumor primario de un modelo de cáncer xenoinplantado en pez cebra. Se lograron identificar 53 y 105 genes regulados al alza y a la baja, respectivamente, que, según los resultados de ontología, están involucrados en procesos celulares clave, que pudieran estar involucrados, a su vez, en la generación de fenotipos metastáticos. De particular interés son aquellos genes involucrados en procesos de señalización celular en células metastáticas, puesto que resulta relevante conocer cómo estas células responden a estímulos en su camino hacia nuevos tejidos y si esas señales pudieran estar contribuyendo a su supervivencia. Otro proceso relevante, en el cual genes regulados al alza estuvieron involucrados es la muerte celular; las células metastáticas se ven sujetas a estrés al alejarse de su tumor primario lo que, eventualmente, puede conducir a su muerte⁴. Determinar si la expresión de ciertos genes puede favorecer la resistencia a muerte celular en células metastáticas tendría gran relevancia clínica.

Los resultados de este análisis de expresión diferencial representan apenas un punto de partida para nuevos estudios que permitan dilucidar el papel de múltiples genes en el proceso de metástasis que, en última instancia, desemboquen en el desarrollo de nuevos fármacos que permitan bloquear este proceso en procesos de cáncer.