

ARTÍCULO DE REVISIÓN

Análisis de perfiles transcripcionales: ventajas y dificultades de su ejecución

ARTÍCULO DE REVISIÓN

ANÁLISIS DE PERFILES TRANSCRIPCIONALES: VENTAJAS Y DIFICULTADES EN SU EJECUCIÓN

Federico Castro-Muñozledo* (1)

(1) Departamento de Biología Celular, Centro de Investigación y de Estudios
Avanzados del IPN. Ciudad de México, México.

*Autor de correspondencia correo E: federico.castro@cinvestav.mx

RESUMEN

La transcriptómica es la caracterización y cuantificación del conjunto de todos los transcritos, codificantes y no codificantes, que se expresan en una célula en un momento determinado. Su importancia estriba en que podemos analizar todos los transcritos de manera integral, a la par de que hace posible comparar patrones de expresión entre diferentes poblaciones celulares, estados de desarrollo o respuesta a diferentes tratamientos. En este trabajo se revisan las principales estrategias utilizadas para explorar y comparar perfiles transcripcionales, desde mediados de la década de 1980 hasta la fecha, haciendo hincapié en las técnicas más utilizadas en la actualidad: los microarreglos y la secuenciación de RNA de nueva generación (RNA-Seq). Actualmente, los microarreglos se han empleado con mayor frecuencia debido a su costo menor y a la relativa facilidad para procesar la información; sin embargo, la RNA-Seq tiene una mayor capacidad para identificar genes no caracterizados o de muy baja expresión. Además, con base en nuestra experiencia desarrollada al comparar transcriptomas de células troncales o precursoras de la superficie ocular y de linajes diferenciados de diferentes especies y mediante diferentes metodologías, discutimos las dificultades que puede enfrentar un usuario al realizar una transcriptómica comparativa, lo que podrá dar al lector mejores herramientas para llevarla a cabo.

PALABRAS CLAVE

transcriptomas,
microarreglos,
Secuenciación de
Nueva
Generación
(NGS),
RNA-Seq,
expresión génica
diferencial

ABSTRACT

Transcriptomics is the characterization and quantification of all the transcripts, coding and non-coding, expressed in a cell at a given time. Its importance is based on the fact that it enables us to compare the expression patterns between different cell populations, developmental stages, or response to different treatments. This paper reviews the major strategies used to explore and compare transcriptional profiles from the mid-1980s to the present, emphasizing the most commonly used techniques today: microarrays and RNA-Seq. Currently, microarrays have been used with more frequency due to their lower cost and the relative ease of information processing; however, RNA-Seq has a greater ability to identify genes that show very low expression. As a matter of fact, it is considered that both strategies are complementary. Finally, based on our experience when comparing transcriptomes of stem cells or precursors from the ocular surface obtained from different species and through different methodologies, we discuss the difficulties that a user may find when carrying out comparative transcriptomics.

KEYWORDS

transcriptomes,
microarrays,
Next Generation
Sequencing
(NGS),
RNA-Seq,
differential gene
expression

Introducción

La transcriptómica o el análisis de los transcriptomas completos, involucra a un conjunto de tecnologías orientadas a estudiar la suma de todos los transcritos que se expresan en una célula o tejido en un momento determinado. En otras palabras, es el análisis y caracterización de todas las moléculas de RNA, tanto codificantes como no codificantes, contenidas en una población celular o bien en una sola célula.

La idea de identificar a todos los RNAs contenidos en una célula, se basa en la necesidad de conocer y comparar los patrones de expresión entre poblaciones que i) responden a estímulos específicos, ii) se encuentran en diferentes etapas de desarrollo, o bien, iii) sufren alteraciones ante diversos estímulos. En cualquier caso, el propósito final consiste en entender los cambios que conllevan a la expresión de funciones, de fenotipos especializados, o de enfermedades desde una perspectiva global e integrativa.

Aunque los primeros intentos de estudiar transcriptomas completos se hicieron al inicio de los años 90's, los antecedentes directos a este abordaje se encuentran en experimentos de la década previa. En estas estrategias pioneras se buscaba distinguir a los genes que se expresan diferencialmente entre dos o más poblaciones celulares mediante una técnica conocida como hibridación sustractiva. Esta metodología consiste en la construcción de bibliotecas de cDNA a partir de cada una de las poblaciones celulares que se desean comparar, y posteriormente hibridarlas entre sí o contra RNA total aislado de cada una de las poblaciones estudiadas, para eliminar secuencias comunes compartidas (normalmente

aquellas que codifican proteínas constitutivas o de “housekeeping”); y posteriormente, seleccionar positivamente a los mensajeros característicos de cada población. Este abordaje permitió el aislamiento y caracterización de algunos genes de importancia biológica como el factor de transcripción MyoD que dirige el programa de diferenciación del músculo esquelético (1) o de genes que codifican proteínas con actividad antifúngica (2).

Otra metodología que se desarrolló con el mismo propósito fue la hibridación sustractiva-supresiva (SSH por sus siglas en inglés) basada en la reacción de la polimerasa en cadena (PCR) (3). Esta técnica consiste en la hibridación del cDNA aislado de dos poblaciones celulares (un control y una muestra experimental) para eliminar los elementos comunes (hibridación supresiva), y de manera subsecuente, los elementos no eliminados en esta fase son amplificados mediante PCR para finalmente obtener los cDNA que se expresan de manera diferencial en la población experimental (4).

Otro de los intentos para determinar el contenido de genes que se expresan en un organismo consistió en el uso de la secuenciación de bajo rendimiento basada en el método de Sanger para producir “*Expressed Sequence Tags*” (EST por sus siglas en inglés). Esta estrategia, empleada durante la década de 1990 fue un abordaje eficiente para generar transcritos de 200 a 500 nucleótidos de longitud, complementarios a los diferentes mRNAs representados en una biblioteca de expresión (5). Aunque *per se*, esta metodología no constituye un abordaje de tipo transcriptómico, fue extremadamente útil en la búsqueda e identificación de transcritos y de los genes corres-

pondientes, así como para la determinación de sus secuencias que fueron utilizadas para constituir las primeras bases de datos transcriptómicos.

A pesar de la existencia de estos abordajes, el procedimiento más utilizado en la década de 1990 para comparar perfiles de expresión entre células eucarióticas fue el *Differential Display* también conocido como *Differential display-PCR* (DDPCR) o DDRT-PCR (6). En este procedimiento, el mRNA se retro-transcribe para obtener cDNA, que a su vez se amplifica por PCR utilizando un juego de oligonucleótidos “anclados” al extremo de poli-A de los mRNAs (el anclado se logra porque los oligonucleótidos contienen una secuencia corta de desoxitimidinas que hibridan con la secuencia de poli-A del mRNA), y un juego de oligonucleótidos arbitrarios o aleatorios (para la amplificación del cDNA) (6, 7). Posteriormente, los numerosos productos de PCR obtenidos en esta fase se visualizan por electroforesis en gel y pueden identificarse mediante metodologías como el *Southern blot*, para detectar los amplicones que se expresan diferencialmente entre las poblaciones comparadas (6, 7). Entre las proteínas identificadas mediante esta técnica están la proteína ribosomal S19 de humanos que es reconocida como un marcador universal de tejidos que crecen rápidamente (8) o algunos marcadores de enfermedades cardiovasculares (7).

El desarrollo del DDPCR y la SSH fue simultáneo al diseño de otras tecnologías desarrolladas para descubrir genes cuya expresión es diferencial. Entre éstas, se puede mencionar la técnica conocida como SAGE (*Serial Analysis of Gene Expression*), que consiste en el aislamiento de secuencias nucleotídicas cortas (de 9 a 10 pares de bases, llamadas SAGE tags) correspondientes a los genes expresados en una población celular. Estas secuencias se concatenan serialmente para secuenciarse en un sólo evento y se comparan con las bases de datos del GenBank o de otros repositorios, lo que facilita el análisis rápido y detallado de miles de transcritos (9).

Estas estrategias condujeron al uso generalizado de los transcriptomas para el análisis global de la expresión genética. No obstante, la aplicación de estas primeras técnicas quedó rápidamente en desuso debido a su poca sensibilidad o a las dificultades asociadas a su ejecución. Como ejemplos de técnicas sujetas a esta problemática se tienen la hibridación sustractiva y el DDPCR. Debido a su baja sensibilidad, la primera se caracterizó por permitir únicamente la identificación de genes cuya expresión es abundante. La segunda tenía el gran inconveniente de generar falsos positivos con alta frecuencia (7).

Microarreglos de DNA

Aun así, gracias al progreso tecnológico la generación de perfiles transcripcionales evolucionó rápidamente. Uno de los primeros procedimientos que permitió la detección simultánea de miles de genes fue el arreglo bidimensional de colecciones pre-determinadas de fragmentos de DNA unidos a una superficie sólida: los microarreglos o chips de DNA (10). La técnica se basa en el uso de cDNA (arreglos de cDNA) o de oligonucleótidos adheridos a la matriz sólida (arreglos de oligonucleótidos) que funcionan como sondas específicas para hibridar con el mRNA de la población celular o tejido que se analiza y comparar con un tejido o muestra control. La hibridación entre los fragmentos del microarreglo y los ácidos nucleicos contenidos en la muestra se detecta por la excitación de fluoróforos distintos que marcan a cada una de las muestras y al control. En cualquiera de los ensayos, la relación entre la intensidad de excitación entre el fluoróforo del RNA control y el fluoróforo que marca al RNA de la muestra experimental, indica el nivel de expresión del gene (11) (Fig. 1). Esta metodología se usó por primera vez en 1995 para comparar los patrones de expresión entre plantas transgénicas y silvestres del género *Arabidopsis*, utilizando microarreglos compuestos por 45 genes de esta especie (12). Actualmente, los microarreglos permiten detectar de 10,000 a 40,000 genes diferentes, mediante el uso de fragmentos de DNA de 30pb a 2 kb (sondas o “probes”) dispuestos de manera ordenada en columnas y filas. Por mucho tiempo, los microarreglos fueron las principales herramientas para el análisis y la comparación de los cambios en la expresión genética ya sea ocasionados por la interacción entre poblaciones celulares, asociados al desarrollo embrionario (13, 14), o bien para estudiar los perfiles transcripcionales en ecología microbiana (15).

En la actualidad, la tecnología de microarreglos de DNA permite estudiar variantes generadas por *splicing*, polimorfismos de un solo nucleótido (SNP), así como también los patrones de metilación del DNA. Sin embargo, estas nuevas matrices solo están disponibles para unas pocas especies como el ratón, la rata, y el humano (16).

Aunque los microarreglos de DNA pueden emplearse para estudiar la respuesta transcripcional a cambios genéticos y ambientales, la tecnología tiene varias limitaciones, entre las que destacan:

- 1) Su falta de linealidad y proporcionalidad cuando las concentraciones utilizadas de mRNA/cDNA son altas, hecho que los convierte en un método fundamentalmente semicuantitativo.

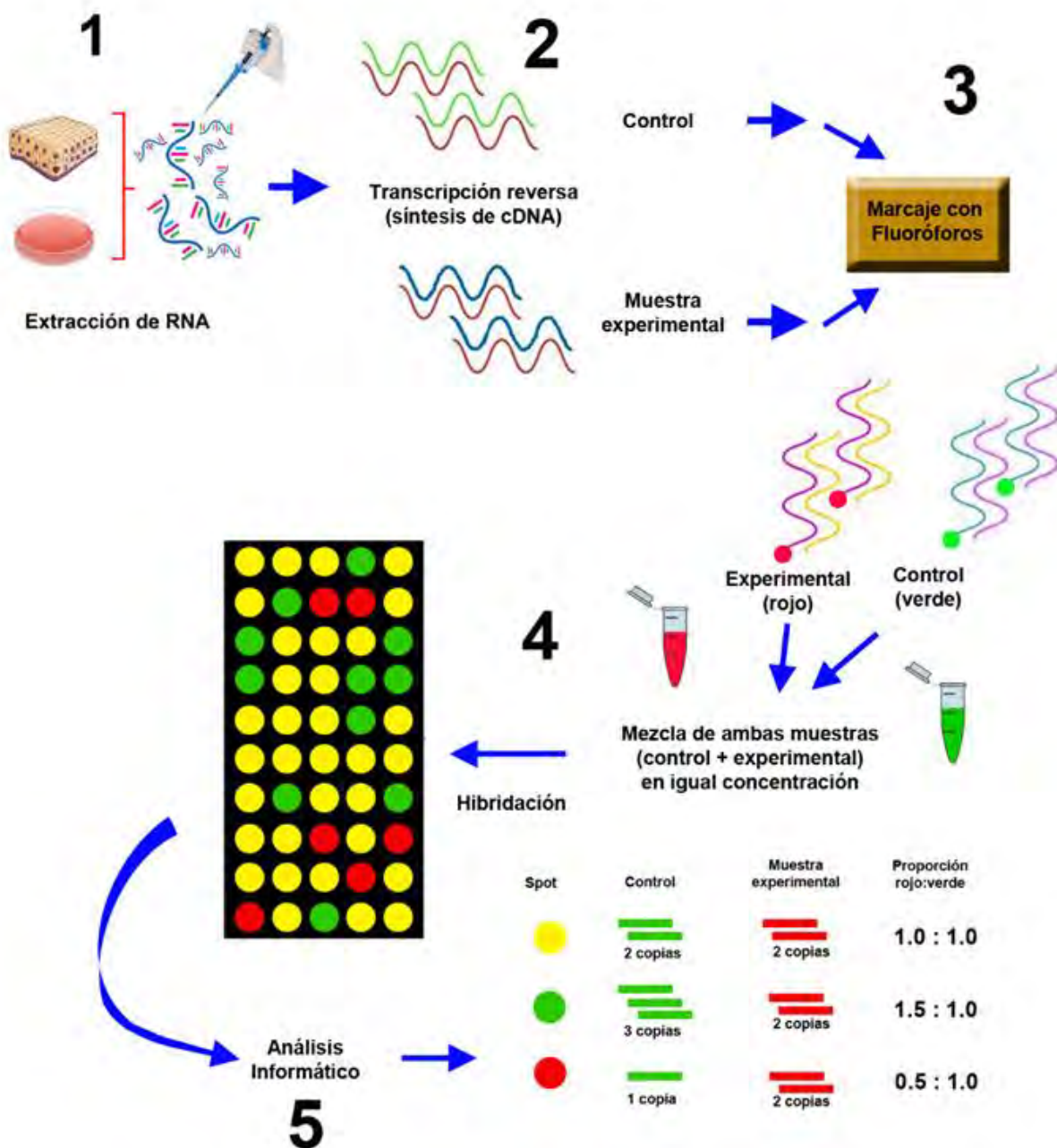


Figura 1. Representación esquemática del análisis de la expresión génica diferencial mediante microarreglos. El análisis del transcriptoma de una población celular o de un tejido requiere **1)** la extracción del RNA, **2)** la transcripción reversa para generar cDNA y su posterior **3)** marcaje con fluoróforos (tanto del control como de la muestra experimental) y **4)** su hibridación contra el microarreglo que contiene secuencias conocidas, para la posterior **5)** cuantificación mediante herramientas bioinformáticas.

2) Puesto que los microarreglos implican la hibridación entre secuencias, la metodología es susceptible de generar falsos positivos y/o negativos.

3) La hibridación inespecífica limita la precisión de las mediciones, particularmente si se desea detectar transcritos con baja abundancia.

4) Debido a su naturaleza específica, el diseño del microarreglo requiere que el genoma del organismo

del que se derivan las muestras a estudiar esté caracterizado o anotado. Si la anotación del genoma es incorrecta o incompleta, los resultados pueden afectarse o malinterpretarse, y por lo mismo no permite la identificación de transcritos cuya secuencia no esté anotada.

5) Considerando que las sondas que componen al microarreglo difieren en sus propiedades de hibrida-

ción, se puede decir que la técnica está limitada a ensayos en los que se examina la expresión de las secuencias para las que el microarreglo se diseñó. En otras palabras, una limitante es la disponibilidad en el chip de sondas para detectar genes conocidos.

6) Los transcriptomas resultantes del análisis no consideran eventos postranscripcionales que conducen a la expresión de variantes.

Métodos de secuenciación masiva de nueva generación

Las estrategias para caracterizar un perfil transcripcional utilizaron a los microarreglos como una de las herramientas esenciales en el estudio de casi cualquier problema biológico, incluyendo la identificación de genes que se expresan diferencialmente durante un proceso de desarrollo, entre tejidos sanos y aquellos con diversas patologías, o asociados con la respuesta a tratamientos farmacológicos, así como para efectuar diagnóstico a nivel clínico (17-20). Esta preferencia se debe a la relación costo-beneficio de su uso, pues considerando su resolución, la cantidad de muestra y el tiempo necesarios para ejecutarlos, y la facilidad del análisis de los resultados, se trata de una metodología asequible para la mayor parte de los laboratorios.

No obstante, las tecnologías de secuenciación de nueva generación (*Next Generation Sequencing*, NGS) se han extendido a casi todas las áreas de investigación debido a su capacidad de análisis masivo y, porque a diferencia de los microarreglos, no necesariamente requieren del conocimiento previo de un genoma de referencia (21). Estas metodologías permiten la secuenciación en poco tiempo de millones de secuencias, de miles de genes, o aún de un genoma completo. En general los procesos de NGS implican una serie de pasos fundamentales necesarios para obtener resultados:

- i) una vez aislados los ácidos nucleicos (en el caso de transcriptoma, RNA total), es indispensable eliminar los RNA ribosomales para, posteriormente, fragmentar al RNA en muchos segmentos cortos con una longitud que oscila entre 100 y 300 nucleótidos.
- ii) Posteriormente, se genera una biblioteca de cDNA. Para ello, los fragmentos de RNA se convierten en cDNA a los que se les añaden secuencias adaptadoras necesarias para el proceso de secuenciación.
- iii) Subsecuentemente, la biblioteca se somete a secuenciación masiva paralela, en cualquiera de las plataformas existentes (Roche 454, Illumina GA, ABI Solid, Ion Torrent, entre otras) (22, 23). Cabe señalar que existen dos diferentes estrategias básicas de secuenciación masiva: secuenciación por

un solo extremo (*Single-read sequencing*, SE) y secuenciación por ambos extremos (*Paired-end sequencing*, PE). La primera es la manera más simple e implica la secuenciación de los fragmentos de DNA por un solo extremo; en la segunda, los fragmentos se secuencian por ambos extremos y debido a ello, tiene la ventaja de generar secuencias alineables de alta calidad (Fig. 2), permite estimar el tamaño del fragmento original y situarlo con mayor precisión en el genoma mejorando así la cobertura de las zonas de interés. Asimismo, resulta útil para la detección de eventos estructurales.

iv) Posteriormente, se evalúan tanto el rendimiento como la calidad de los datos obtenidos del proceso de secuenciación, utilizando la herramienta FastQC que permite el establecimiento de la métrica, de los límites, y la generación de gráficos para este análisis (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Este primer análisis genera un archivo FASTQ que contiene tanto las secuencias biológicas (secuencias nucleotídicas) como su nivel de calidad.

v) Todo el análisis subsecuente de los resultados se realiza *in silico*. En primer término, se efectúa la “limpieza” de la biblioteca de expresión obtenida; para ello, las secuencias de los fragmentos se procesan eliminando los adaptadores.

vi) En siguiente término, las secuencias se alinean respecto a un genoma de referencia (que puede corresponder, incluso, al de una especie diferente) mediante las herramientas de alineamiento disponibles para la comunidad, se generan Mapas de Alineamiento Binario (archivos BAM, *Binary Alignment Map*) y se obtienen los resultados finales que corresponden a las secuencias completas de los RNAs provenientes de la población de interés. Los resultados de la secuenciación se procesan para normalizarlos, considerando la abundancia de transcritos y evitando el conteo repetido de secuencias correspondientes a un mismo fragmento (Fig. 3).

Finalmente, se determinan los niveles de expresión. La manera de hacerlo depende del método de normalización utilizado. Cuando la secuenciación se lleva a cabo por un extremo (*SE-sequencing*) el nivel de expresión se cuantifica en RPKM (*reads per kilobase per million mapped reads*), consistente en determinar el número de lecturas (*reads*) por millón con relación a la longitud del gene y es una forma adecuada para comparar los niveles de expresión dentro de una sola muestra. Alternativamente, en el caso de la secuenciación por ambos extremos (*PE-sequencing*), la normalización se efectúa mediante el método FPKM (*Fragments per kilobase per million mapped fragments*), basado en la longitud del gen y el número total de lecturas ma-

peadas. Por último, otro método utilizado para el *PE-sequencing* consiste en cuantificar la expresión génica en la cantidad de transcritos por millón (TPM) y se utiliza para comparar muestras diferentes.

Es importante señalar que para llevar a cabo el análisis de transcriptomas es esencial considerar la cobertura o profundidad del proceso de secuenciación. En este caso, la profundidad o cobertura de secuencia se refiere al número promedio de lecturas

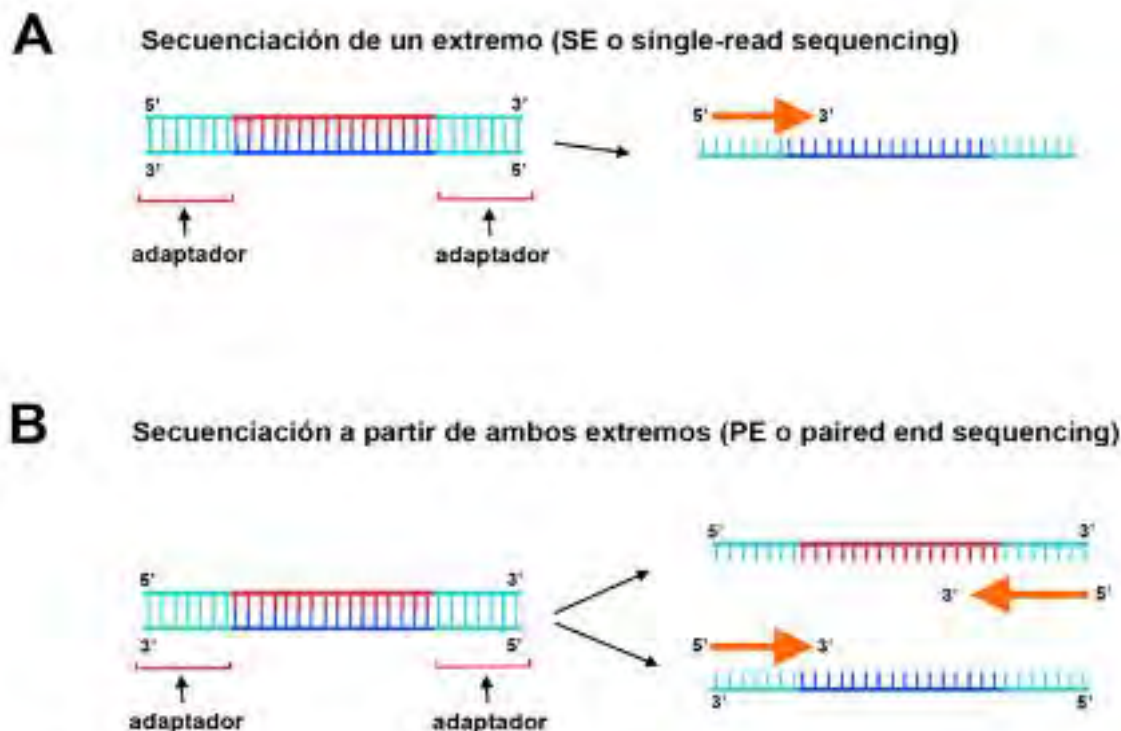


Figura 2. Esquematzación de las dos posibles estrategias de secuenciación masiva. (A) Secuenciación por un solo extremo (*Single-read sequencing*, SE) y **(B)** secuenciación por ambos extremos (*Paired-end sequencing*, PE).

que se alinean a secuencias conocidas. El objetivo de considerar este parámetro es el de tener el mayor número de lecturas únicas de cada región de una secuencia. En general se considera que para tener un transcriptoma confiable (o preciso), el número de lecturas de un transcriptoma oscila entre 20 y 50 millones. Esto disminuye significativamente los errores de secuenciación y permite detectar polimorfismos de un solo nucleótido.

De manera adicional, una vez que el investigador ha obtenido el transcriptoma de las muestras de interés, mediante el uso de herramientas informáticas puede proceder a realizar el estudio comparativo entre éstas, pudiendo analizar la expresión génica diferencial (DEGs, *differentially expressed genes*), así como identificar los procesos metabólicos, componentes celulares y funciones con herramientas como *Gene Ontology* (GO), IPA (*Ingenuity Pathway Analysis*) o WGCNA (*Weighted Gene Co-expression Network Analysis*). Estas herramientas, en conjunto, nos permiten explorar, caracterizar, dar seguimiento, y verificar los mecanismos molecu-

lares subyacentes a procesos específicos como el desarrollo, la diferenciación celular o mecanismos que subyacen en enfermedades como el Alzheimer o el cáncer.

RNA-Seq. La obtención de transcriptomas mediante secuenciación masiva o RNA-Seq es una metodología que permite la caracterización y cuantificación de transcritos sin que se requiera que el experimentador conozca un genoma en particular. La RNA-Seq permite al investigador buscar tanto variantes transcripcionales derivadas de *splicing* alternativo como detectar modificaciones postraduccionales, mutaciones, polimorfismos de un solo nucleótido (SNP), o cambios en la expresión génica entre diferentes poblaciones, estados de desarrollo, etc. (24). De manera importante, cuando la RNA-Seq se lleva a cabo con la cobertura o profundidad adecuada, permite la detección tanto de los transcritos producidos en proporciones muy bajas, como de aquellos que se expresan con niveles muy altos. Es indispensable señalar que el número de lecturas (o *reads*) requerido para detectar transcritos

poco abundantes oscila entre 100 y 200 millones de *reads*. Actualmente, la RNA-Seq se ha convertido en la opción preferida para obtener un perfil transcripcional en lugar de los microarreglos dadas las

ventajas que confiere para examinar a detalle la expresión genética, además de que se trata de una herramienta caracterizada por su alta eficiencia y precisión.

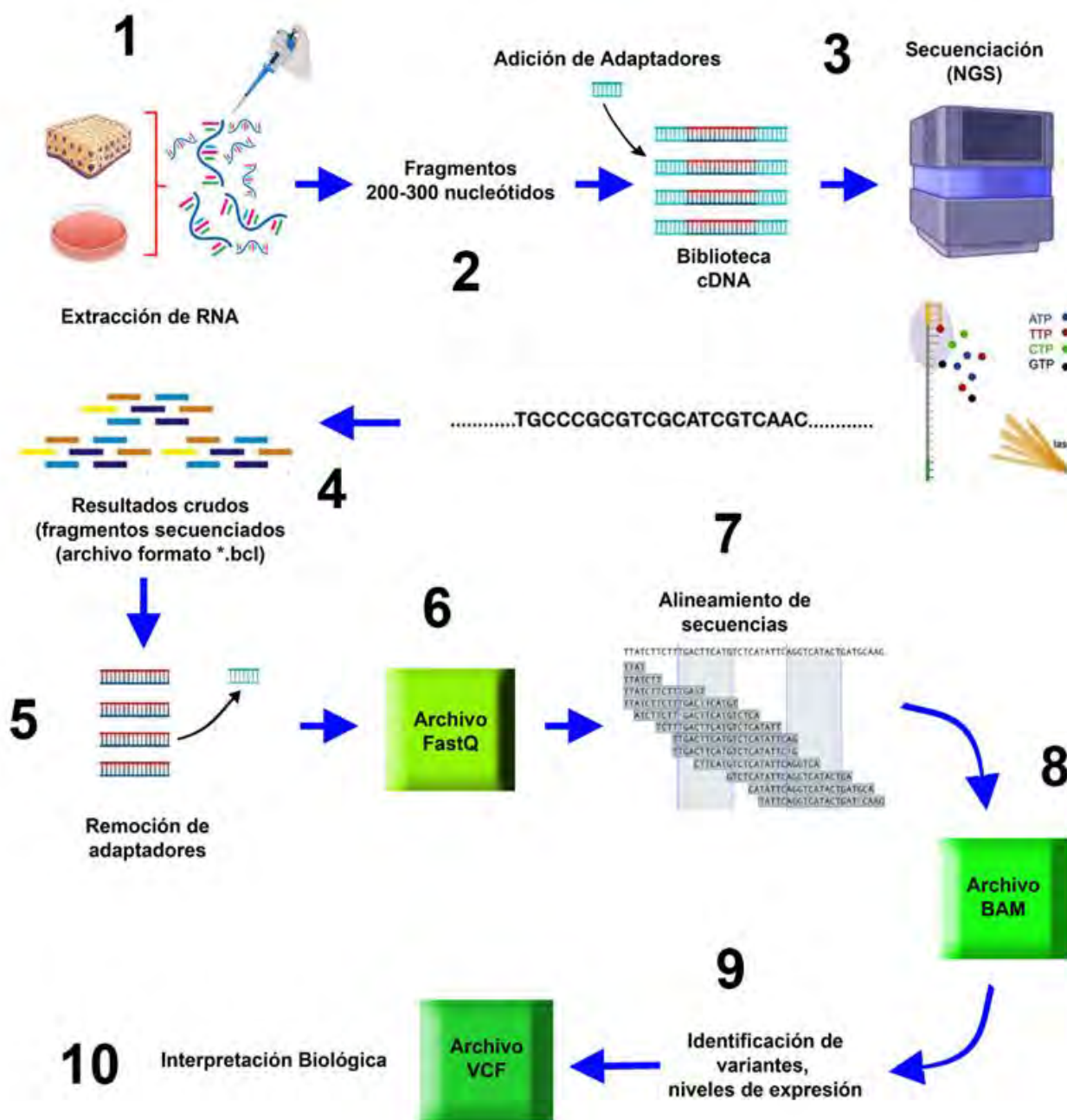


Figura 3. Esquema general que representa los pasos fundamentales necesarios para obtener resultados en la secuenciación de nueva generación (Next Generation Sequencing, NGS). El proceso se inicia con **1)** la purificación del RNA y la subsecuente eliminación de los RNA ribosomales. **2)** Posteriormente se fragmenta el RNA en segmentos cortos de 100 a 300 nucleótidos de longitud y se añaden lo adaptadores, cuya función es facilitar las reacciones de amplificación y secuenciación. **3)** Secuenciación y **4)** obtención de los resultados crudos en formato *.bcl, **5)** seguida por la remoción de los adaptadores y **6)** la obtención de un archivo FastQ que contiene las secuencias biológicas y es esencial para evaluar la calidad de las secuencias. El subsecuente análisis *in silico* permite **7)** el alineamiento para tener un archivo BAM que permite **8)** la identificación de niveles de expresión e identificación de variantes y finalmente un archivo VCF que se utiliza en toda la interpretación biológica.

scRNA-Seq y snRNA-Seq. Variantes más recientes de este enfoque son la RNA-Seq de una sola célula (scRNA-Seq) y la RNA-Seq de un solo núcleo (snRNA-Seq). Mientras la RNA-Seq considera que las células de un mismo tejido son homogéneas, las tecnologías de una célula o núcleo único (25, 26) consideran la heterogeneidad de la expresión entre células y permiten identificar subpoblaciones dentro de una muestra de tejido, y más aún, poblaciones muy raras o poco frecuentes. Esta característica ha permitido identificar relaciones entre linajes celulares durante etapas tempranas del desarrollo embrionario (27). No obstante, la scRNA-Seq está limitada por factores como la baja cantidad de material inicial, que tiene como consecuencia una mayor variabilidad de los datos obtenidos debido a la abundancia predominante de los mensajeros con mayor expresión, y la baja eficiencia para capturar mensajeros cuya expresión es baja (28). De manera adicional, dado que este tipo de análisis implica la disociación del tejido en sus componentes celulares, el daño causado por las enzimas utilizadas para la disgregación tisular puede afectar a un gran número de células, alterando la calidad de los ácidos nucleicos extraídos. Además, los tratamientos empleados durante la disgregación de las células pueden inducir cambios en la expresión genética ocasionados por la respuesta tisular al estrés.

Entre los diferentes intentos experimentales para evitar los inconvenientes asociados a la scRNA-Seq, algunos autores probaron aplicar métodos previamente utilizados para disociar tejidos difíciles de disgregar. Estas variantes metodológicas consistieron en el aislamiento de los núcleos celulares para su posterior extracción, aislamiento de los ácidos nucleicos y secuenciación. Esta estrategia permitió generar la secuenciación de núcleos aislados (*single nucleus* RNA-Seq, snRNA-Seq) usando una resolución equivalente a la empleada en la scRNA-Seq (29, 30).

En particular, la snRNA-Seq se ha convertido en una alternativa atractiva al scRNA-Seq debido a que disminuye el sesgo asociado a la abundancia de los subtipos celulares predominantes en una población. Una de sus ventajas es conferida por la posibilidad de analizar patrones de expresión genética en muestras procesadas mediante técnicas histológicas tradicionales, como la fijación con formalina y la inclusión en parafina (31). De manera adicional, es menos susceptible de alteraciones asociadas a perturbaciones en la expresión genética ocasionadas por la separación de las células durante la colecta de la muestra (32).

Microarreglos vs. scRNA-Seq y snRNA-Seq

Actualmente, el análisis del perfil transcripcional se basa en el uso de los microarreglos y de las tecnologías de secuenciación masiva (principalmente RNA-Seq). De éstas, los microarreglos se hicieron populares y son los más utilizados debido a la relación costo/beneficio y a la facilidad del análisis de sus resultados. A pesar de estas ventajas, los microarreglos adolecen de la disponibilidad de sondas para genes aún no caracterizados, y de la profundidad o sensibilidad que se alcanza con la RNA-Seq.

En contraste, los métodos de secuenciación masiva permiten detectar nuevos transcritos, y no generan falsos positivos asociados a la hibridación errónea o parcial observada con los microarreglos. Sin embargo, la RNA-Seq también presenta desventajas como las ocasionadas por el uso de muestras de RNA de mala calidad. De manera adicional, el análisis de los resultados obtenidos es arduo y consume mucho tiempo, ya que es necesario eliminar las secuencias sobrerrepresentadas o de baja calidad, y remover las secuencias adaptadoras añadidas para llevar a cabo la secuenciación. Más aún, diferentes autores consideran que el análisis de los resultados de los estudios de RNA-Seq tiene limitaciones, a pesar de las numerosas herramientas informáticas creadas para efectuar el ensamblaje y alineamiento de secuencias, la cuantificación de la expresión genética, el análisis de expresión diferencial, y el análisis de rutas reguladoras y grupos funcionales (33-35).

Considerando lo anterior, el lector puede cuestionar cuál podría ser la herramienta de elección para realizar el perfil transcripcional de un conjunto de poblaciones celulares o de un tejido. Para responder esta pregunta, algunos autores compararon las cualidades técnicas y el desempeño de ambas estrategias. Para ello, obtuvieron transcriptomas de un mismo organismo, órgano o tejido utilizando ambas estrategias de manera simultánea y bajo las mismas condiciones (36-38). En todos los casos, los investigadores encontraron que la eficiencia de detección de genes que se expresan diferencialmente por ambas plataformas es muy parecida y correlaciona bien tanto a nivel absoluto como a nivel relativo, rindiendo resultados similares en un 72-78% de los genes examinados. Sin embargo, el resto de los genes cuya expresión se identificó, sólo se detectó mediante RNA-Seq (36, 38). En este tipo de comparación, algunos autores han considerado que ambas metodologías se complementan; no obs-

tante, la RNA-Seq tiene la ventaja de tener mayor sensibilidad, lo que facilita la identificación de una mayor cantidad de genes cuya expresión es diferencial, así también, la RNA-Seq tiene la capacidad de identificar nuevos transcritos y transcritos no codificantes (36, 38).

Transcriptómica comparativa. Como hemos visto, las plataformas de secuenciación masiva son herramientas poderosas que nos capacitan para

analizar transcriptomas de genoma completo, pues permiten detectar respuestas transcripcionales bajo diversas condiciones y estímulos. Esto las hace invaluable para cuantificar el impacto ambiental sobre un tejido, o el efecto de las enfermedades. Sin embargo, diferentes factores pueden limitar el meta-análisis de los estudios de la RNA-Seq, tal es el caso de las variaciones en el diseño experimental, la tecnología de secuenciación y los métodos estadísticos.

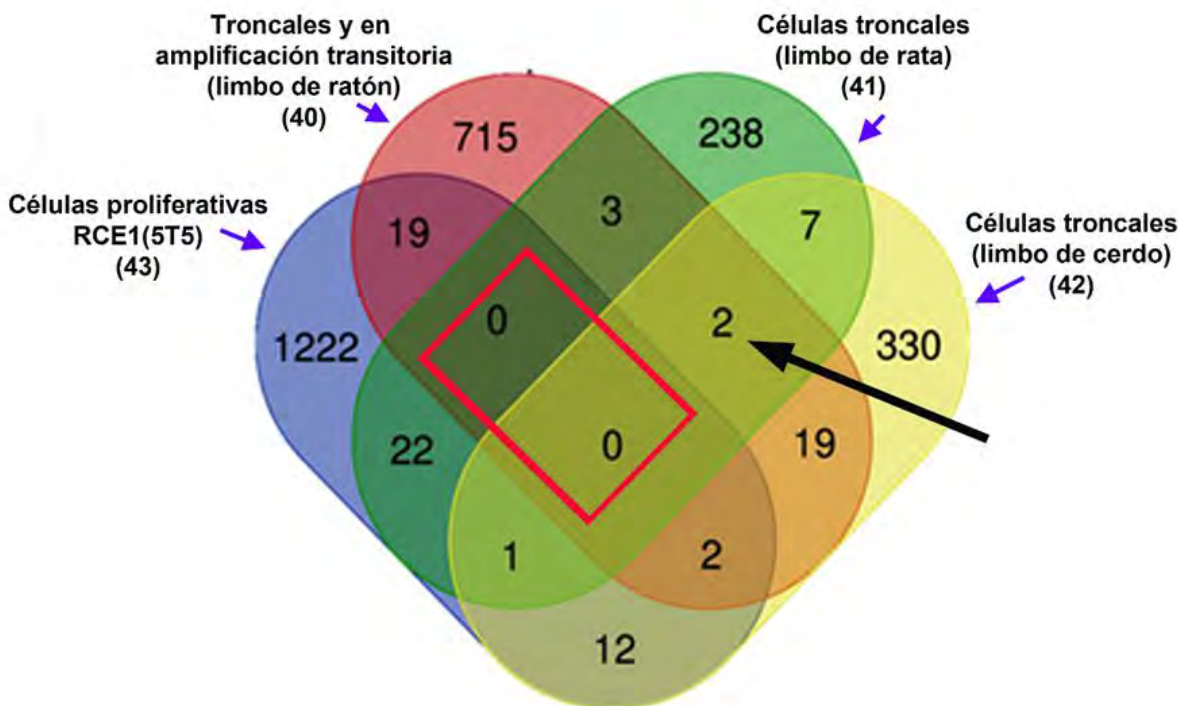


Figura 4. Diagrama de Venn que presenta el número de genes que se expresan diferencialmente (DEGs), y los genes comunes encontrados al comparar únicamente las firmas genéticas características de los transcriptomas de células limbales de ratón (40), de rata (41), de cerdo (42) y de células proliferativas de la línea de epitelio corneal de conejo RCE1(5T5) (43). En el análisis se compararon de 18,000 a cerca de 29,000 transcritos que constituyen a estos transcriptomas (39-42), disponibles en el *Gene Expression Omnibus* (GEO), *National Center for Biotechnology Information* (NCBI). Una vez que se eliminaron los transcritos constitutivos, se definió la firma genética de cada población celular y considerando como marco los transcritos ya identificados por diversos grupos como marcadores de células troncales del limbo esclero-corneal, se elaboró el diagrama de Venn. El gráfico es un resultado preliminar. Nótese que las células troncales/progenitoras de ratón, rata y de la línea celular de conejo no necesariamente expresan genes en común (recuadro en rojo), y en el caso de la rata, el ratón y el cerdo sólo encontramos 2 genes en común (flecha negra). Las flechas azules indican la especie a la que pertenece el epitelio limbal empleado para obtener los transcritos agrupados en cada uno de los conjuntos del diagrama. El diagrama se generó empleando la herramienta disponible en línea (<http://bioinformatics.psb.ugent.be/webtools/Venn/>).

La transcriptómica comparativa es un abordaje que se hizo más factible debido a la reducción de los costos de la secuenciación. Esta permite hacer comparaciones entre especies, y proporciona información sobre los mecanismos que impulsan los cambios fenotípicos. Al hacer comparaciones entre diferentes especies, las similitudes entre los patrones de transcripción identificados pueden indicar

mecanismos reguladores conservados; en contraste, las divergencias permiten resaltar el impacto que tiene la variación genética sobre los niveles de expresión de genes específicos.

A pesar de estas características, el equiparar respuestas transcripcionales entre diferentes especies puede ser un desafío debido que no todos los genomas disponibles están anotados por completo,

como ocurre en el caso del conejo; o bien, la disponibilidad o accesibilidad de aquellos genomas completamente secuenciados puede ser variable. Más aún, cuando se realiza un análisis comparativo entre diferentes especies, estas limitantes hacen factible que los investigadores encuentren inconsistencias entre los genomas de referencia, lo que origina sesgos en las anotaciones de los genes y en la identificación de vías. Otras limitaciones que dificultan este análisis están asociadas a las diferencias técnicas entre laboratorios, como pueden ser la manera en que se procesan los datos y los parámetros de ensamblaje, lo que a su vez puede afectar la cuantificación de los patrones de expresión génica.

Como ejemplo, en las siguientes líneas discutimos el análisis preliminar que realizamos en nuestro laboratorio para comparar diferentes perfiles transcripcionales obtenidos mediante el uso de diferentes plataformas y por diferentes grupos de investigación. En este caso, planteamos como meta encontrar moléculas que participan en la regulación transcripcional de las células troncales del epitelio que recubre la córnea de los mamíferos. Para alcanzar este objetivo, tomamos en cuenta que el sitio donde se alojan las células troncales de la superficie corneal es la región anatómica conocida como limbo esclero-corneal (39). Entonces, comparamos los transcriptomas ya publicados de: i) células del limbo esclero-corneal del epitelio corneal del ratón obtenido mediante scRNA-Seq (40); ii) células de epitelio limbal y corneal de rata determinado por SAGE (41), y iii) células de epitelio limbal del cerdo que se generó utilizando microarreglos (42). iv) A este análisis, se sumó el perfil transcripcional generado por RNA-Seq, de la línea celular RCE1(5T5) de epitelio corneal de conejo en fase de crecimiento exponencial (43). En particular, esta línea celular posee características de expresión genética que la asemejan a las etapas muy tempranas de la expresión del proceso de diferenciación de este tejido (43, 44).

Los perfiles transcripcionales analizados tuvieron variaciones en tamaño asociadas a la metodología empleada (microarreglos, RNA-Seq, scRNA-Seq), a la muestra utilizada como fuente de RNA, o la especie de origen. De esta manera comparamos la expresión de 18,000 a cerca de 29,000 transcritos, eliminando a los transcritos constitutivos, para obtener las firmas genéticas de cada una de las poblaciones estudiadas, cuya expresión diferencial permitiera la identificación de biomarcadores de interés. Este abordaje redujo significativamente el número de genes resultantes del análisis comparativo, y planteó la posibilidad de encontrar mar-

cadores compartidos entre las distintas poblaciones estudiadas que pudieran ser considerados como marcadores específicos de células troncales del epitelio corneal.

Al hacer esta comparación, encontramos que existe poca semejanza entre los transcriptomas de células limbales de cerdo, rata y ratón (40-42) (Fig. 4), a pesar de que numerosas publicaciones sugieren la existencia de elementos y mecanismos reguladores comunes entre las células progenitoras del epitelio corneal de estas especies. Algunas de estas diferencias también pueden deberse a disparidades entre las especies de las que se deriva el tejido (45,46), a diferencias entre etapas del desarrollo (47), a la edad de los individuos (48-49), a diferencias metodológicas (50) o a la cobertura del proceso de secuenciación.

Considerando estos factores, es posible explicar disparidades tan importantes como las encontradas al comparar perfiles transcripcionales obtenidos por diferentes grupos de investigación cuando comparamos 5 diferentes perfiles transcripcionales de células del limbo esclero-corneal de diferentes especies reportados recientemente, y obtenidos mediante RNA-Seq, scRNA-Seq y microarreglos. En la Fig. 5, mostramos el comparativo entre mapas de calor elaborados únicamente con los transcritos cuya expresión diferencial corresponde a los marcadores moleculares característicos del carácter troncal reportados como posibles reguladores de la diferenciación de células del limbo esclero-corneal y marcadores del estado terminalmente diferenciado. En este ejemplo se presenta únicamente la comparación entre los transcriptomas de humano, ratón y cerdo, incluyendo como controles a los perfiles de células de epitelio corneal de conejo, tanto proliferativas como diferenciadas (Fig. 5).

Como puede observarse, los perfiles comparados poseen muy pocos marcadores moleculares en común, y la presencia de elementos involucrados en la regulación de la expresión genética asociada al carácter troncal, es exigua. Todo esto, a pesar de que los transcriptomas se obtuvieron de poblaciones celulares que, teóricamente tienen las mismas características. El caso extremo lo ejemplifican los perfiles de células troncales del epitelio limbal de ratón y de humano, ambos obtenidos por scRNA-Seq (Fig. 5), que prácticamente difieren en su totalidad.

A partir de los resultados anteriores se hace evidente que las diferencias encontradas son consecuencia de algunas particularidades técnicas. Una de éstas es la dificultad para aislar a las células troncales del epitelio corneal debido a la carencia de metodologías y marcadores adecuados que permitan selec-

cionar a esta población celular (51). Hasta ahora, los datos provenientes de diferentes laboratorios han ligado el carácter troncal en el epitelio corneal a la expresión de un grupo de factores de transcripción entre los que destacan Pax6 y $\Delta Np63\alpha$, así como de

moléculas que forman parte de los complejos de adhesión como la integrina $\alpha6\beta4$, y de proteínas involucradas en las cascadas de señalización como NOTCH1, entre otros (revisado en 39). No obstante, la evidencia sugiere que en el epitelio limbal coexis-

EPITELIO LIMBAL/CELULAS TRONCALES DEL EPITELIO CORNEAL

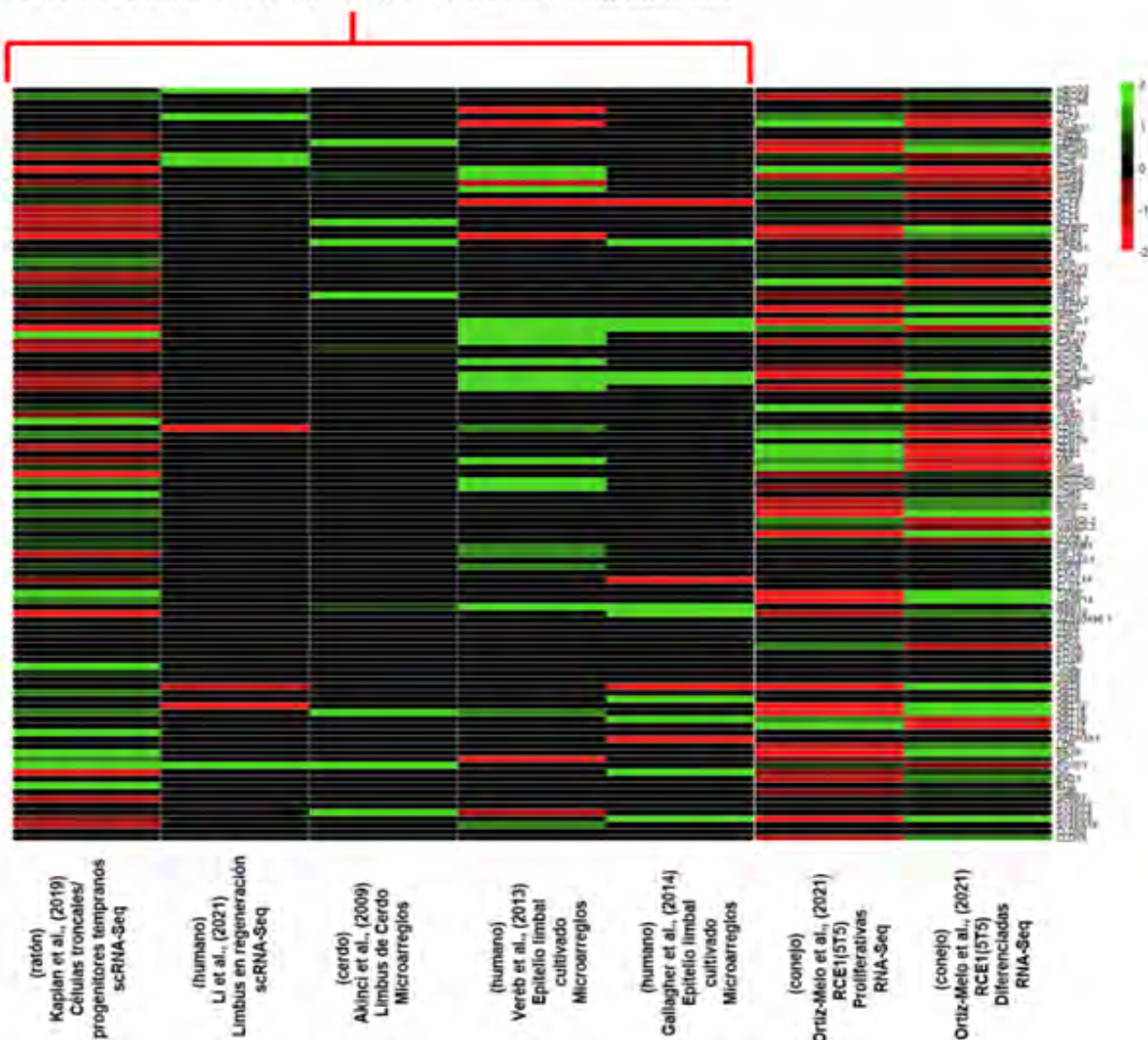


Figura 5. Mapa de calor elaborado con los genes que se expresan diferencialmente (DEGs) entre las firmas genéticas que forman parte de los transcriptomas comparados obtenidos de células troncales/progenitoras del limbo esclero-corneal de humano, ratón, cerdo. Se ilustran los genes cuya expresión aumenta (verde) o disminuye (rojo) significativamente, o bien no cambian (negro); véase escala de referencia en el margen superior derecho. Se comparan únicamente los genes que se expresan diferencialmente y que forman parte de los perfiles transcriptómicos completos obtenidos para esta región de la superficie ocular mediante diferentes estrategias metodológicas, en particular: microarreglos (42,56,57), scRNA-Seq (40,58), y como referencia RNA-Seq de la línea celular RCE1(5T5) (43). Es notable que los DEGs seleccionados para esta comparación provienen de células cultivadas (56,57), tejidos en regeneración (58), y el tejido normal (40). Asimismo, la ilustración nos muestra que es factible equiparar entre especies: humano (56-58), ratón (40), y cerdo (42). Los niveles de abundancia para los genes que se expresan diferencialmente (DEGs) se utilizaron para generar el mapa de calor, observado en esta figura, con la herramienta XLSTAT2017 (Addinsoft, Paris, France 2017). Para esta comparación se emplearon los transcriptomas reportados (40,42,43,56-58), y disponibles en el *Gene Expression Omnibus* (GEO), *National Center for Biotechnology Information* (NCBI).

ten las células troncales del epitelio y las células que se encuentran en etapas muy tempranas del proceso de diferenciación, y dado que comparten varios marcadores moleculares, constituyen una población celular heterogénea (52,53).

También se deben considerar las diferencias técnicas tanto en el aislamiento del material biológico como las diferencias en el procesamiento, análisis y generación de resultados. Aunque existen estudios que sugieren que los resultados obtenidos mediante microarreglos y métodos de secuenciación masiva como el RNA-Seq son complementarios y presentan una correlación alta (36-38), esta conclusión se alcanzó partiendo de una misma muestra, sometida a los mismos métodos, para finalmente ser examinada en las diferentes plataformas de análisis de expresión (36-38).

Considerando las diferencias observadas, el investigador debe estar consciente de que las comparaciones entre conjuntos de transcriptomas independientes alojados en un repositorio público como el NCBI *Gene Expression Omnibus* (GEO), requieren que las bases de datos estén curadas de manera exhaustiva para poder reutilizar la información y ejecutar el análisis secundario que permita obtener conclusiones útiles. Actualmente, este tipo de estudios representa un reto que demanda el uso de herramientas informáticas y estadísticas que permitan el reprocesado de la información para ordenarla y normalizarla y hacer comparables las diferentes bases de datos disponibles, eliminando los artefactos técnicos existentes entre los transcriptomas comparados. Esto se aplica tanto a las comparaciones entre especies pertenecientes a diferentes linajes filogenéticos como a los experimentos con tejidos y tipos celulares de un mismo organismo.

Considerando lo anteriormente expuesto, la transcriptómica y la transcriptómica comparativa son herramientas extremadamente poderosas para identificar genes que participan en la regulación de la expresión genética asociada a programas de desarrollo y procesos de diferenciación, así como posibles blancos terapéuticos y farmacológicos, o bien para establecer relaciones filogenéticas entre individuos o poblaciones. Entre las diferentes estrategias existentes para el análisis de los perfiles transcripcionales, nos hemos enfocado principalmente en las plataformas predominantes, tanto microarreglos como las variaciones más frecuentes de secuenciación masiva (RNA-Seq).

Debo aclarar que existen otros tipos de abordaje, como los métodos basados en el uso de *tags*. Tal es el caso del análisis de expresión genética cap (CAGE, *Cap Analysis of Gene Expression*), que

monitorea con precisión a los sitios de iniciación de la transcripción y a los promotores (54); o la secuenciación paralela masiva de firmas (*Massively Parallel Signature Sequencing*, MPSS) que determina los niveles de expresión de RNA mensajero al cuantificar el número de copias de cada mRNA individual (55). Sin embargo, la mayoría de estas técnicas utilizan como base el método de secuenciación de Sanger, lo que las hace muy costosas en relación con los microarreglos y a las técnicas de secuenciación masiva.

En cualquiera de las metodologías planteadas, es esencial una consideración cuidadosa de la consistencia metodológica y los factores técnicos para obtener información confiable y significativa. Es importante resumir las principales dificultades a las que se enfrenta el investigador involucrado en transcriptómica comparada:

- 1) La comparación de transcriptomas de diferentes tejidos o especies puede ser limitada como consecuencia de la inconsistencia en la anotación de los genomas disponibles en las bases de datos. Diferentes especies pueden tener diferentes niveles de completitud en la anotación del genoma. Mientras los genomas mejor anotados corresponden (en animales) al humano, rata y ratón, los genomas de otras especies no están caracterizados por completo. Alternativamente, pueden existir varios genomas anotados para una especie, lo que complica las comparaciones.
- 2) Variabilidad técnica. El diseño experimental puede variar en factores como la preparación de la muestra, las plataformas de secuenciación y los protocolos de procesamiento.
- 3) Otras variaciones que alteran la confiabilidad de las comparaciones radican en los métodos de procesamiento de datos, incluidos el control de calidad, la alineación de lectura y la normalización.
- 4) Variabilidad biológica. La expresión génica puede cambiar significativamente entre diferentes tejidos (expresión tejido-específica) de un mismo organismo. Por otra parte, la abundancia relativa de los tipos celulares en un tejido puede influir en los perfiles de expresión génica.
- 5) La divergencia evolutiva puede originar diferencias en los mecanismos que controlan la expresión genética entre diferentes especies. En este caso puede existir divergencia en los patrones de transcripción.
- 6) Finalmente, existen diferencias asociadas a la normalización e integración de los datos empleados en el análisis comparativo.

Todos estos factores participan en la integración de un buen análisis de los perfiles transcripcionales. En

este sentido, diferentes autores consideran que es necesario desarrollar mejores herramientas informáticas para mejorar la calidad de las comparaciones.

Conclusión

En el futuro, es indudable que con el abatimiento de los costos, la RNA-Seq se convertirá en el método de selección para la elaboración de transcriptomas. Asimismo, el desarrollo y perfeccionamiento de la metodología y de las plataformas de análisis permi-

tirán evitar los problemas asociados a las variaciones que hemos citado a lo largo de este trabajo.



Agradecimientos

Este trabajo fue financiado en parte por el proyecto de CONAHCYT 320450.

El autor agradece a la M. en C. Erika Sánchez-Guzmán y al Sr. José Juan Prado Barajas por su apoyo, que permite el funcionamiento diario del grupo de trabajo.

REFERENCIAS

1. Davis RL, Weintraub H, Lassar AB. Expression of a Single Transfected cDNA Converts Fibroblasts to Myoblasts. *Cell* 1987; 51:987-1000. DOI: 10.1016/0092-8674(87)90585-x.
2. Ng TB. Antifungal proteins and peptides of leguminous and non-leguminous origins. *Peptides* 2004; 25:1215-22. DOI: 10.1016/j.peptides.2004.03.012.
3. Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam, F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Nat Acad Sci USA* 1996; 93:6025-30. DOI: 10.1073/pnas.93.12.6025
4. Diatchenko L, Lukyanov S, Lau YF, Siebert PD. Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes. *Methods Enzymol.* 1999; 303:349-80. DOI: 10.1016/s0076-6879(99)03022-0.
5. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, W. McCombie R, Venter JC. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 1991; 252:1651-6. DOI: 10.1126/science.2047873.
6. Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 1992; 257:967-71. DOI: 10.1126/science.1354393.
7. Wang X, Feuerstein GZ. The use of mRNA differential display for discovery of novel therapeutic targets in cardiovascular disease. *Cardiovasc Res* 1997; 35:414-21. DOI: 10.1016/s0008-6363(97)00169-7.
8. Li B, Sun M, He B, Yu J, Zhang YD, Zhang YL. Identification of differentially expressed genes in human uterine leiomyomas using differential display. *Cell Res* 2002; 12:39-45. DOI: 10.1038/sj.cr.7290108.
9. Tuteja R, Tuteja N. Serial Analysis of Gene Expression: Applications in Human Studies. *BioMed Res Int.* 2004; 2004:113-20. DOI: 10.1155/S1110724304308119.
10. Bumgarner R. Overview of DNA microarrays: types, applications, and their future. *Curr Protoc Mol Biol.* 2013; Chapter 22:Unit 22.1. DOI: 10.1002/0471142727.mb2201s101.
11. Tarca AL, Romero R, Draghici S. Analysis of microarray experiments of gene expression profiling. *Am J Obstet Gynecol* 2006; 195:373-88. DOI: 10.1016/j.ajog.2006.07.001.
12. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995; 270:467-70. DOI: 10.1126/science.270.5235.467.
13. Lee KF, Yao YQ, Kwok KL, Xu JS, Yeung WS. Early developing embryos affect the gene expression patterns in the mouse oviduct. *Biochem Biophys Res Commun* 2002; 292:564-70. DOI: 10.1006/bbrc.2002.6676.

14. Niemann H, Carnwath JW, Kues W. Application of DNA array technology to mammalian embryos. *Theriogenology* 2007; 68:S165-77. DOI: 10.1016/j.theriogenology.2007.05.041.
15. Miller MB, Tang YW. Basic concepts of microarrays and potential applications in clinical microbiology. *Clin Microbiol Rev* 2009; 22:611-33. DOI: 10.1128/CMR.00019-09.
16. Sadikovic B, Yoshimoto M, Al-Romaih K, Maire G, Zielenska M, Squire JA. In vitro analysis of integrated global high resolution DNA methylation profiling with genomic imbalance and gene expression in osteosarcoma. *PLoS ONE* 2008; 3:e2834. DOI: 10.1371/journal.pone.0002834.
17. Almiñana C, Fazeli A. Exploring the application of high-throughput genomics technologies in the field of maternal-embryo communication. *Theriogenology* 2012; 77:717-37. DOI: 10.1016/j.theriogenology.2011.11.009.
18. Okamura H, Yamano H, Tsuda T, Morihiro J, Hirayama K, Nagano H. Development of a clinical microarray system for genetic analysis screening. *Pract Lab Med* 2022; 33:e00306. DOI: 10.1016/j.plabm.2022.e00306.
19. Joseph SM, Sathidevi PS. An automated cDNA microarray image analysis for the determination of gene expression ratios. *IEEE/ACM Trans Comput Biol Bioinform* 2023; 20:136-50. DOI: 10.1109/TCBB.2021.3135650.
20. Sani A, Idrees Khan M, Shah S, Tian Y, Zha G, Fan L, Zhang Q, Cao C. Diagnosis and screening of abnormal hemoglobins. *Clin Chim Acta* 2024; 552:117685. DOI: 10.1016/j.cca.2023.117685.
21. Raz T, Kapranov P, Lipson D, Letovsky S, Milos PM, Thompson JF. Protocol dependence of sequencing-based gene expression measurements. *PLoS One* 2011; 6:e19287. DOI: 10.1371/journal.pone.0019287.
22. Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, Frazer KA. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol* 2009; 10(3):R32. DOI: 10.1186/gb-2009-10-3-r32.
23. Mandlik JS, Patil AS, Singh S. Next-Generation Sequencing (NGS): Platforms and Applications. *J Pharm Bioallied Sci.* 2024; 16(Suppl 1):S41-S45. DOI: 10.4103/jpbs.jpbs_838_23.
24. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics.* 2009; 10:57-63. DOI: 10.1038/nrg2484.
25. Eberwine J, Yeh H, Miyashiro K, Cao Y, Nair S, Finnell R, Zettel M, Coleman P. Analysis of gene expression in single live neurons. *Proc Natl Acad Sci USA.* 1992; 89:3010-14. DOI: 10.1073/pnas.89.7.3010.
26. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 2009; 6:377-82. DOI: 10.1038/nmeth.1315.
27. Petropoulos S, Edsgård D, Reinius B, Deng Q, Panula SP, Codeluppi S, Plaza Reyes A, Linnarsson S, Sandberg R, Lanner F. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell* 2016; 165:1012-26. DOI: 10.1016/j.cell.2016.03.023.
28. Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* 2018; 50:1-14. DOI: 10.1038/s12276-018-0071-8.
29. Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, Hofree M, Choudhury SR, Aguet F, Gelfand E, Ardlie K, Weitz DA, Rozenblatt-Rosen O, Zhang F, Regev A. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods.* 2017; 14:955-58. DOI: 10.1038/nmeth.4407.
30. Lake BB, Ai R, Kaeser GE, Salathia NS, Yung YC, Liu R, Wildberg A, Gao D, Fung HL, Chen S, Vijayaraghavan R, Wong J, Chen A, Sheng X, Kaper F, Shen R, Ronaghi M, Fan JB, Wang W, Chun J, Zhang K. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science.* 2016; 352:1586-90. DOI: 10.1126/science.aaf1204.
31. Guo Y, Wang W, Ye K, He L, Ge Q, Huang Y, Zhao X. Single-Nucleus RNA-Seq: Open the Era of Great Navigation for FFPE Tissue. *Int J Mol Sci.* 2023; 24:13744. DOI: 10.3390/ijms241813744.
32. Lacar B, Linker SB, Jaeger BN, Krishnaswami SR, Barron JJ, Kelder MJ, Parylak SL, Paquola AC, Venepally P, Novotny M, O'Connor C, Fitzpatrick C, Erwin JA, Hsu JY, Husband D, McConnell MJ, Lasken R, Gage FH. Nuclear RNA-seq of single neurons reveals molecular

- signatures of activation. *Nat Commun.* 2016; 7:11022. DOI: 10.1038/ncomms11022.
33. Zhao, S, Fung-Leung, W-P, Bittner, A, Ngo, K, Liu, X. Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. *PLoS ONE.* 2014; 9:e78644. DOI: 10.1371/journal.pone.0078644.
34. Garber M, Grabherr M, Guttman M, Trapnell C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Methods* 2011; 8:469-77. DOI: 10.1038/nmeth.1613.
35. Rosati D, Palmieri M, Brunelli G, Morrione A, Iannelli F, Frullanti E, Giordano A. Differential gene expression analysis pipelines and bioinformatic tools for the identification of specific biomarkers: A review. *Comput Struct Biotechnol J.* 2024; 23:1154-68. DOI: 10.1016/j.csbj.2024.02.018.
36. Rao MS, Van Vleet TR, Ciurlionis R, Buck WR, Mittelstadt SW, Blomme EAG, Liguori MJ. Comparison of RNA-Seq and Microarray Gene Expression Platforms for the Toxicogenomic Evaluation of Liver From Short-Term Rat Toxicity Studies. *Front. Genet.* 2019; 9:636. DOI: 10.3389/fgene.2018.00636.
37. Rai MF, Tycksen ED, Sandell LJ, Brophy RH. Advantages of RNA-seq compared to RNA microarrays for transcriptome profiling of anterior cruciate ligament tears. *J Orthop Res* 2018; 36:484-97. DOI: 10.1002/jor.23661.
38. Kogenaru S, Qing Y, Guo Y, Wang N. RNA-seq and microarray complement each other in transcriptome profiling. *BMC Genomics.* 2012; 13:629. DOI: 10.1186/1471-2164-13-629.
39. Castro-Muñozledo F. The Mammalian Limbal Stem Cell Niche: A Complex Interaction between Cells, Growth Factors and Extracellular Matrix. In: *Biology in Stem Cell Niche, Chapter 2. Stem Cell Biology and Regenerative Medicine Book Series.* (Turksen, K., editor). Springer International Publishing Switzerland. 2015; 23-56. DOI: 10.1007/978-3-319-2.
40. Kaplan N, Wang J, Wray B, Patel P, Yang W, Peng H, Lavker RM. Single-Cell RNA Transcriptome Helps Define the Limbal/Corneal Epithelial Stem/Early Transit Amplifying Cells and How Autophagy Affects This Population. *Invest Ophthalmol Vis Sci* 2019; 60:3570-83. DOI: 10.1167/iovs.19-27656.
41. Adachi W, Ulanovsky H, Li Y, Norman B, Davis J, Piatigorsky J. Serial analysis of gene expression (SAGE) in the rat limbal and central corneal epithelium. *Invest Ophthalmol Vis Sci.* 2006; 47:3801-10. DOI: 10.1167/iovs.06-0216.
42. Akinci MA, Turner H, Taveras M, Wolosin JM. Differential gene expression in the pig limbal side population: implications for stem cell cycling, replication, and survival. *Invest Ophthalmol Vis Sci.* 2009; 50:5630-8. DOI: 10.1167/iovs.09-3791.
43. Ortiz-Melo MT, Garcia-Murillo MJ, Salazar-Rojas VM, Campos JE, Castro-Muñozledo F. Transcriptional profiles along cell programming into corneal epithelial differentiation. *Exp Eye Res.* 2021; 202:108302. DOI: 10.1016/j.exer.2020.108302.
44. Castro-Muñozledo F. Development of a spontaneous permanent cell line of rabbit corneal epithelial cells that undergoes sequential stages of differentiation in cell culture. *J Cell Sci.* 1994; 107:2343-51. DOI: 10.1242/jcs.107.8.2343.
45. Okyere J, Oppon E, Dzidzienyo D, Sharma L, Ball G. Cross-species gene expression analysis of species specific differences in the preclinical assessment of pharmaceutical compounds. *PLoS One.* 2014; 9:e96853. DOI: 10.1371/journal.pone.0096853.
46. Wang X, Yang L, Wang YC, Xu ZR, Feng Y, Zhang J, Wang Y, Xu CR. Comparative analysis of cell lineage differentiation during hepatogenesis in humans and mice at the single-cell transcriptome level. *Cell Res* 2020; 30:1109-26. DOI: 10.1038/s41422-020-0378-6.
47. Sonam S, Bangru S, Perry KJ, Chembazhi UV, Kalsotra A, Henry JJ. Cellular and molecular profiles of larval and adult *Xenopus* corneal epithelia resolved at the single-cell level. *Dev Biol.* 2022; 491:13-30. DOI: 10.1016/j.ydbio.2022.08.007.
48. Notara M, Shortt AJ, O'Callaghan AR, Daniels JT. The impact of age on the physical and cellular properties of the human limbal stem cell niche. *Age* 2013; 35:289-300. DOI: 10.1007/s11357-011-9359-5.
49. Douvaras P, Webb S, Whitaker DA, Dorà N, Hill RE, Dorin JR, West JD. Rare corneal clones in mice suggest an age-related decrease of stem cell activity and support the limbal epithelial stem cell hypothesis. *Stem Cell Res* 2012; 8:109-19. DOI: 10.1016/j.scr.2011.08.007.

50. López-Paniagua M, Nieto-Miguel T, de la Mata A, Dziasko M, Galindo S, Rey E, Herreras JM, Corrales RM, Daniels JT, Calonge M. Comparison of functional limbal epithelial stem cell isolation methods. *Exp. Eye Res* 2016; 146:83-94. DOI: 10.1016/j.exer.2015.12.002.
51. Guo ZH, Zeng YM, Lin JS. Dynamic spatiotemporal expression pattern of limbal stem cell putative biomarkers during mouse development. *Exp. Eye Res* 2020; 192:107915. DOI: 10.1016/j.exer.2020.107915.
52. Wu YF, Chang NW, Chu LA, Liu HY, Zhou YX, Pai YL, Yu YS, Kuan CH, Wu YC, Lin SJ, Tan HY. Single-cell transcriptomics reveals cellular heterogeneity and complex cell-cell communication networks in the mouse cornea. *Invest Ophthalmol Vis Sci*. 2023; 64:5. DOI: 10.1167/iovs.64.13.5.
53. Schiffman JS, D'Avino AR, Prieto T, Pang Y, Fan Y, Rajagopalan S, Potenski C, Hara T, Suvà ML, Gawad C, Landau DA. Defining heritability, plasticity, and transition dynamics of cellular phenotypes in somatic evolution. *Nat Genet*. 2024; 56:2174-84. DOI: 10.1038/s41588-024-01920-6.
54. Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, Kodzius R, Watahiki A, Nakamura M, Arakawa T, Fukuda S, Sasaki D, Podhajski A, Harbers M, Kawai J, Carninci P, Hayashizaki Y. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci USA*. 2003; 100:15776-81. DOI: 10.1073/pnas.2136655100.
55. Zhou D, Rao MS, Walker R, Khrebtukova I, Haudenschild CD, Miura T, Decola S, Vermaas E, Moon K, Vasicek TJ. Massively parallel signature sequencing. *Methods Mol Biol*. 2006; 331:285-311. DOI: 10.1385/1-59745-046-4:285.
56. Veréb Z, Albert R, Póliska S, Olstad OK, Akhtar S, Moe MC, Petrovski G. Comparison of upstream regulators in human *ex vivo* cultured cornea limbal epithelial stem cells and differentiated corneal epithelial cells. *BMC Genomics*. 2013; 14:900. DOI: 10.1186/1471-2164-14-900.
57. Gallagher C, Clarke C, Aherne ST, Katikireddy KR, Doolan P, Lynch V, Shaw S, Bobart-Hone A, Murphy C, Clynes M, Power W, O'Sullivan F. Comparative transcriptomic analysis of cultivated limbal epithelium and donor corneal tissue reveals altered wound healing gene expression. *Invest Ophthalmol Vis Sci*. 2014; 55:5795-805. DOI: 10.1167/iovs.14-14664.
58. Li DQ, Kim S, Li JM, Gao Q, Choi J, Bian F, Hu J, Zhang Y, Li J, Lu R, Li Y, Pflugfelder SC, Miao H, Chen R. Single-cell transcriptomics identifies limbal stem cell population and cell types mapping its differentiation trajectory in limbal basal epithelium of human cornea. *Ocul Surf*. 2021; 20:20-32. DOI: 10.1016/j.jtos.2020.12.004.

LECTURAS RECOMENDADAS

1. Various Authors. Transcriptome Analysis (M. Blumenberg editor). IntechOpen. eBook (PDF) ISBN978-1-78985-263-9. 2019;108 pages. DOI: 10.5772/intechopen.77860.
2. Various Authors. Transcriptome Data Analysis. Methods and Protocols. (Yejun Wang & Ming-a Sun, editors). From the series: Methods in Molecular Biology Vol 1751. 2018: Humana Press NY. DOI: 10.1007/978-1-4939-7710-9.
3. Anamika K, Verma S, Jere A, Desai, A. Transcriptomic Profiling Using Next Generation Sequencing - Advances, Advantages, and Challenges. Published in: Next Generation Sequencing - Advances, Applications and Challenges. (Kulski, JK editor). InTech. 2016 DOI: 10.5772/61789.